

REGARDS DIALECTOMÉTRIQUES SUR LES DONNÉES  
DE L'ATLAS LINGUISTIQUE DE LA FRANCE (ALF): RELATIONS  
QUANTITATIVES ET STRUCTURES DE PROFONDEUR

Hans GOEBL  
Universität Salzburg

1. UN ASPECT LONGTEMPS NÉGLIGÉ DE LA GÉOGRAPHIE LINGUISTIQUE:  
L'AMÉNAGEMENT LINGUISTIQUE GLOBAL DE L'ESPACE GÉOGRAPHIQUE  
PAR L'HOMME

Par «géographie linguistique» nous entendons l'étude atlantographique de la variation linguistique de l'espace. C'est donc une discipline qui est aux prises avec tous les problèmes relatifs à la conception, à l'établissement et surtout au dépouillement final d'atlas linguistiques de toute sorte. Or, l'histoire et le développement des études romanes en général et de la géographie linguistique romane en particulier nous enseignent que les atlas linguistiques ont été élaborés d'abord et dépouillés ensuite le plus souvent dans la seule perspective d'approfondir et d'améliorer nos connaissances phonétiques, morphologiques, syntaxiques et lexicales relatives à une région donnée si bien que la géographie linguistique est devenue avant tout une sorte de complément à des outils de recherche déjà existants comme, par exemple, les grammaires ou les vocabulaires régionaux ou locaux.

Certes, les atlas linguistiques, pris comme mines inépuisables de formes phonétiques ou lexicales rares ou intéressantes, ont très bien servi la curiosité de leurs utilisateurs et continuent de le faire. Ceci vaut d'ailleurs non seulement pour les atlas nationaux ou (macro)régionaux, mais aussi pour ceux de dimensions plus réduites.

Mais cette finalité qui emboîte le pas au courant pluriséculaire de la curiosité philologique particularisante —c'est-à-dire visant à déceler les secrets de la genèse et de la fonction du signifiant et du signifié d'un nombre inouï d'unités linguistiques (plus ou moins détachées de leurs contextes fonctionnels)— était dès le début bien loin d'être la seule à être desservie par la géographie linguistique.

C'est que les atlas linguistiques offrent, à côté de la considération particularisante de leurs données, aussi une autre perspective, globale et généralisante, qui consiste à étudier, par delà l'analyse de données d'atlas isolées, les mécanismes et les régularités avec lesquels les différentes unités linguistiques occupent l'espace considéré en tant que tel.

Qu'il y ait là d'étonnantes découvertes à faire, c'était l'une des expériences, inattendues et spontanées, faites par les pionniers de n'importe quelle atlantographie linguistique, que ce soit Jules Gilliéron (1854-1926) en France ou, peu avant lui, Georg Wenker (1852-1911) en

Allemagne. Le théâtre principal de ces découvertes était —au moins pour les géolinguistes romans— la carte muette qui, une fois remplie d'un jeu différentiel de couleurs répandues sur le papier suivant des critères préalablement déterminés, révélait la bigarrure prodigieuse du dynamisme avec lequel l'activité basilectale de l'homme affrontait les contraintes tant naturelles que culturelles inhérentes à l'espace géographique.

Parmi les leçons centrales de la géographie linguistique classique (représentée par des maîtres incontestés tels que Jules Gilliéron, Albert Dauzat, Karl Jaberg, Jakob Jud et bien d'autres: cf. la synthèse magistrale de Sever Pop, 1950) figurait, entre autres, la constatation plutôt décevante que, sur les cartes muettes dûment remplies —et même sur celles d'orientation linguistique apparentée— ni les aires dont elles consistaient, ni, à plus forte raison, les lignes entourant les aires en question, ne tendaient à coïncider si bien que l'écrasante majorité des chercheurs se résignaient à accepter, devant le caractère apparemment flou, voire primesautier des analyses cartographiques des cartes d'atlas, une variante de l'adage —forgé d'ailleurs par Jakob Grimm (1785-1863) au début du XIX<sup>e</sup> siècle (cfr. Christmann, 1971: 122)— suivant lequel «chaque mot devait avoir son histoire particulière».

Cette sagesse, traduite en termes de géographie linguistique, ne signifiait rien d'autre que chaque attestation dialectale devait disposer, sur la carte géographique, de son aire individuelle. Mais ceci signifiait également que pour l'écrasante majorité des chercheurs compétents en la matière il semblait être vain, voire inutile de se mettre à la recherche de régularités majeures dans ce désordre apparent.

Or, la dialectométrie doit sa genèse aux doutes que certains géolinguistes ont fini par nourrir face au dogme universel du caractère chaotique et irrégulier de l'aménagement linguistique de l'espace par l'HOMO LOQUENS, et aussi à leur courage de passer outre la considération particularisante des données d'atlas pour se lancer dans l'aventure de la synthèse de ces dernières par voie numérique. Ce sont là les mérites de Jean Séguy (1914-1974), d'Henri Guiter (1909-1994) et de tous ceux qui ont contribué à élargir la panoplie méthodique et la richesse épistémologique de la géographie linguistique en jetant les bases de la pensée dialectométrique.

La dialectométrie telle qu'elle se présente de nos jours, c'est-à-dire quelque trente ans après la définition officielle de son nom (Séguy, 1973: 1), dispose d'un grand nombre de prérequis empiriques, théoriques et méthodiques qui appartiennent aussi à d'autres sciences tant humaines que naturelles à orientation quantitative, parmi lesquelles figurent, entre autres, la psycho-, socio-, anthropo- ou économétrie et aussi la géographie quantitative. C'est avec cette dernière discipline qu'elle partage avant tout l'intérêt et la curiosité pour tous les aspects quantitatifs de l'aménagement de l'espace par l'homme.<sup>1</sup> Il est donc d'autant plus regrettable de voir le peu de cas que font les géographes, surtout ceux qui pratiquent la géographie dite «humaine» ou «régionale», de la géographie linguistique: voir à titre d'exemple le mutisme total à cet égard dans la synthèse historique du géographe humain Paul Claval (1998).

## 2. LA DIALECTOMÉTRIE: UN OUTIL QUANTITATIF POUR L'ÉTUDE GLOBALE DE L'AMÉNAGEMENT LINGUISTIQUE DE L'ESPACE GÉOGRAPHIQUE PAR L'HOMME

Suivant une définition établie en 1981 (Goebel, 1981: 349), la dialectométrie (DM) représente un alliage méthodique entre la géographie linguistique et la taxonomie (ou classification) numérique. Cette dernière —appelée aussi «taxométrie, analyse des données, analyse ty-

1. Voir à ce sujet notre étude interdisciplinaire de 1996.

pologique, automatische Klassifikation, Cluster-Analyse, numerical classification», etc.— est une discipline mathématique dont les origines remontent aux années 60 et 70 du xx<sup>e</sup> siècle et dont la finalité consiste à déceler, dans des données empiriques de masse, des groupements et des structures cachées à l'observation directe de la part du chercheur, mais qui offrent néanmoins un intérêt gnoséologique majeur pour la science respective.

La classification numérique représente, à cause de son utilité générale, une méthode interdisciplinaire par excellence et dispose, de nos jours, d'un grand nombre de manuels, de revues et aussi d'associations scientifiques tant internationales que nationales.

Le caractère épistémologique de la DM est double:

C'est d'abord une discipline *inductive* et *généralisante*, étant donné que sa démarche part du niveau du particulier pour viser celui du général.

C'est une discipline *exploratoire* dont le but est de découvrir, par le biais de méthodes quantitatives (mathématiques ou numériques), des structures de profondeur qui échappent à l'introspection directe du chercheur et dont le rang épistémologique et gnoséologique dépasse considérablement celui des structures de superficie telles qu'elles se révèlent, entre autres, sur les mosaïques polychromes de cartes muettes dûment remplies.

De par sa démarche quantifiante, la DM est surtout appelée à produire un amas de chiffres qui pourraient cependant indisposer les géolinguistes non avertis et en outre peu familiers avec le quantitatif. C'est pourquoi nous avons développé, dès le début de nos travaux dialectométriques dans les années 70 du xx<sup>e</sup> siècle, une riche gamme de méthodes visualisatrices qui permettent, soit sous la forme de cartes isarithmiques ou choroplèthes, soit sous celle d'arborescences multicolores, de présenter les résultats taxométriques dans une forme iconique voire heuristique aisément intelligible et partant d'un accès facile pour le grand public.

Inutile de préciser que le dialogue interdisciplinaire auquel la DM est fortement attachée, dépend dans une large mesure de la bonne intelligibilité des visualisations de calculs numériques souvent fort complexes, réalisées selon les acquis plus récents de la cartographie numérisée (ou assistée par l'ordinateur).

L'éclat et la précision des figures et des cartes polychromes de cet article sont dus avant tout à l'engagement et la compétence de deux collaborateurs fidèles, Edgar Haimerl et Slawomir Sobota, dont le premier est l'auteur d'un logiciel taxométrique et cartographique ultra-puissant (VDM - Visual Dialectometry) et le second le réalisateur des 5 figures et 24 cartes. Nous leur exprimons déjà ici notre profonde reconnaissance quitte à la réitérer dans le paragraphe 10 de cette contribution.

### 3. LES DIFFÉRENTES ÉTAPES DE LA DIALECTOMÉTRISATION DE L'ALF: EXIGENCES TECHNIQUES ET TRAITEMENT INFORMATIQUE

L'idée de dialectométrer une partie aussi grande que possible de l'ALF est née vers le milieu des années 90 du siècle passé. Notre propos initial était d'y appliquer non seulement les principes de la taxation lexicale telle qu'elle avait été réalisée dans nos travaux dialectométriques antérieurs sur l'ALS (voir surtout Goebel, 1981, 1982, 1983a, 1984, 1992a et 1993a), mais aussi ceux de la taxation phonétique<sup>2</sup> et, si possible, morpho-syntaxique.

2. La taxation phonétique avait été éprouvée sur les données des «Tableaux phonétiques des patois de la Suisse romande» (TPPSR) en 1985 et 1987.

D'entrée de jeu, il fallait acheter un exemplaire vierge de la réimpression de l'*ALF* (disponible depuis 1968 auprès de l'éditeur Forni à Bologne, Italie) et le faire décomposer en feuilles (ou planches) détachées. Ce travail quelque peu inhabituel a été réalisé, au moins en ce qui concerne les cartes 1-1421 de la série A de l'*ALF*, par la main experte d'une relieuse salzbourgeoise. Ensuite, les planches détachées de l'*ALF* ont été dotées d'un réseau polychrome de 18 «parcours d'épreuve» dont la logique serrée assurait la consultation et la lecture contrôlées de toutes les données répertoriées sur les planches ainsi que le transfert exact des données taxées (ou codées) des planches de l'original de l'*ALF* sur des listes appropriées. C'est à partir de ces listes (remplies à la main) que les données ont fini par être enfournées dans l'ordinateur.

Il fallait en outre former une équipe de jeunes collaborateurs à qui confier le travail à la fois délicat et harassant de la taxation des planches de l'*ALF*. Il va de soi que notre tâche consistait aussi à encourager, animer et contrôler les activités de nos jeunes collaborateurs et à leur assurer une rétribution financière décente.

Ont fait partie, entre 1997 et 2000, de l'équipe salzbourgeoise pour la taxation de l'*ALF*: Barbara Aigner, Irmgard Dautermann, Hildegund Eder, Susanne Oleinek et Annette Ida Schatzmann.

Précisons enfin que la dialectométrisation de l'*ALF* bénéficiait de deux subventions de recherche (no. 12414 et 13349) accordées par l'organisme de recherche autrichien «Fonds zur Förderung der wissenschaftlichen Forschung in Österreich» (FWF), siégeant à Vienne.

Que nos jeunes collaboratrices salzbourgeoises et les responsables du FWF reçoivent ici l'expression de ma sincère reconnaissance!

Quant à la mise en place d'un logiciel capable d'assurer tant la saisie rationnelle des données codées que leur traitement taxométrique et cartographique consécutif, nous avons déjà précisé qu'il avait été conçu et rodé par Edgar Haimerl au cours d'un stage de recherche de trois années à Salzbourg. Pour d'ultérieures informations nous renvoyons nos lecteurs à notre site internet bilingue (allemand-anglais): <http://ald.sbg.ac.at/dm> où ils pourront trouver une description détaillée de notre démarche dialectométrique en général et de toutes les étapes de la dialectométrisation de l'*ALF* en particulier.

#### 4. STRUCTURE ET FONCTIONNEMENT DE LA CHAÎNE DIALECTOMÉTRIQUE UTILISÉE

Nous tenons à préciser d'entrée de jeu que la dialectométrisation de l'*ALF* et le logiciel élaboré par E. Haimerl ont été conçus et réalisés exactement suivant les méthodes et les normes dialectométriques établis par nous il y a maintenant plus de vingt ans: voir à ce sujet nos contributions respectives publiées depuis 1981. L'ensemble des méthodes dialectométriques actuellement disponibles est organisé sous la forme d'une «chaîne» dont la figure 1 montre les différents «maillons».

Le traitement dialectométrique de l'*ALF* commence donc avec les planches de l'*ALF* prises en considération qui, de ce fait, représentent la «source» originale. Par la suite, l'on obtient, à l'aide d'un processus de métabolisme informationnel bien défini (appelé aussi «taxation» ou «codage» des données de départ), la «matrice de données» qui constitue un schéma logique à double entrée avec  $N$  éléments (ici:  $N = 638 + 3$  points d'enquête) et  $p$  attributs linguistiques (ici:  $p = 1687$  cartes «de travail», toutes catégories linguistiques confondues). La matrice de données deviendra ensuite l'objet de différentes mensurations de la similarité (pondérée ou non) entre les vecteurs d'attributs des points-*ALF* qui, en l'occurrence, sont toujours considérés par paires.

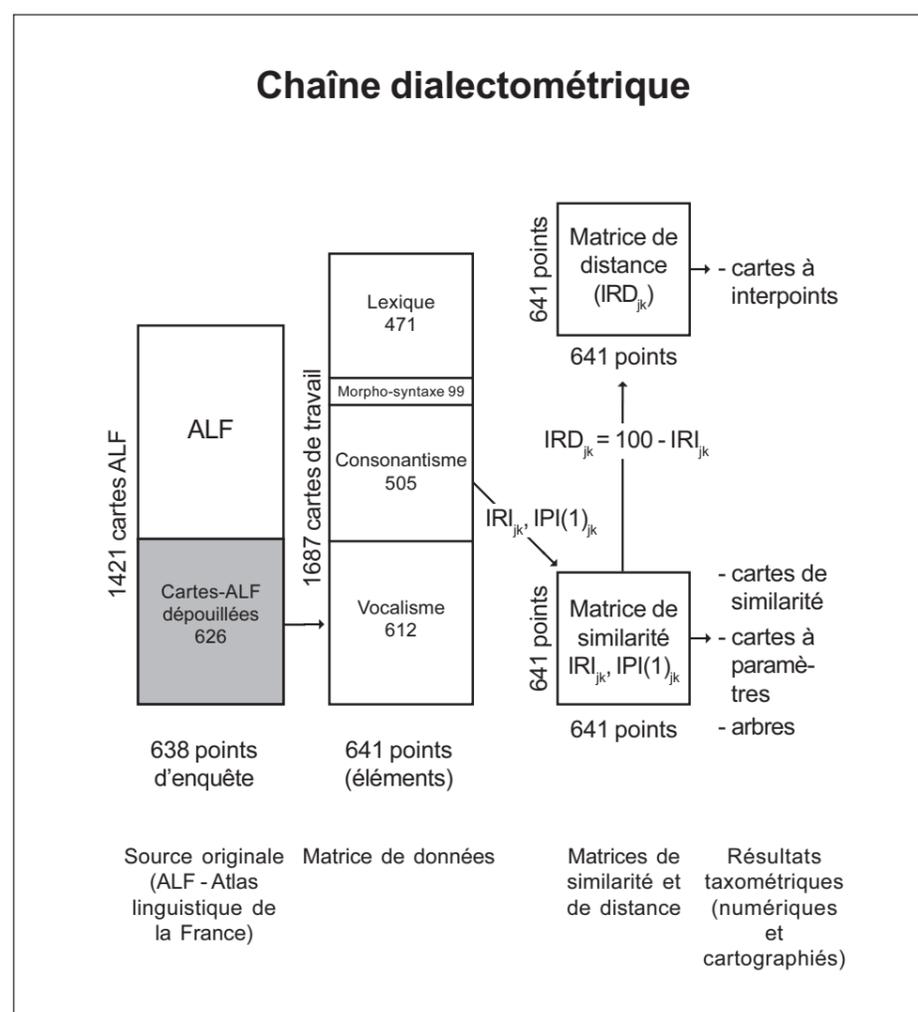


FIGURE 1: Diagramme des méthodes dialectométriques utilisées dans cet article.

Il en résulte un tableau symétrique avec les dimensions N fois N appelé «matrice de similarité» d'un côté ou —après une transformation numérique très simple par le biais de la formule  $IRD_{jk} = 100 - IRI_{jk}$ — «matrice de distance» de l'autre. Les deux matrices qui contiennent l'ensemble de toutes les informations taxométriques disponibles, deviendront ensuite le pivot central pour un grand nombre de dépouillements numériques (et aussi graphiques) tant uni- que multivariés.

Précisons que le schéma sériel de la figure 1 est valable pour n'importe quelle démarche (taxo)métrique et que, de ce fait, il revêt un caractère franchement pluridisciplinaire.

#### 4.1. DES DONNÉES ORIGINALES DE L'ALF À LA MATRICE DE DONNÉES

##### 4.1.1. Préparation empirique et cartographique du réseau de l'ALF

Ont été pris en considération tous les 638<sup>3</sup> points d'enquête de l'ALF auxquels nous avons ajouté trois points d'enquête artificiels qui correspondent à trois langues standards susceptibles d'avoir influé, tant hier que de nos jours, sur les dialectes galloromans:

P. 999: français standard (tel qu'il figure dans les entêtes des cartes de l'ALF), localisé en Île-de-France, à proximité des PP. 226 et 227.

P. 998: italien standard (suivant les indications d'un dictionnaire français-italien), localisé à la limite orientale de la Provence, entre les PP. 990 et 899.

P. 997: catalan standard (suivant les indications d'un dictionnaire français-catalan), localisé aux confins méridionaux du Roussillon, entre les PP. 796 et 798.

Cependant, il s'est avéré après coup que, de ces trois langues standards, seul le français constitue, pour les données de l'ALF, un agent géolinguistique sérieux.

Comme, pour des raisons d'espace, il est impossible de présenter les numéros des points d'enquête de l'ALF sur le fond des cartes, et que, par contre, il est facile de les identifier par le biais d'une comparaison visuelle directe avec une carte polygonale muette dotée de la numérotation complète de l'ALF, nous en avons mis un fichier pdf téléchargeable à la disposition de nos lecteurs sous l'adresse internet suivante: [http://ald.sbg.ac.at/dm/germ/Allgemein/ALF\\_Netz.pdf](http://ald.sbg.ac.at/dm/germ/Allgemein/ALF_Netz.pdf).

La polygonisation du réseau géographique naturel de l'ALF constitue la dernière étape de la préparation cartographique et s'applique donc à tous les 641 points d'atlas. Elle a été réalisée selon les principes bien connus de la géométrie de Delaunay-Voronoi (cfr. Goebel, 1981: 363-364; 1984, I: 90-92 et Okabe/Boots/Sugihara, 1992: passim) dont l'usage, pour les besoins de géographie linguistique, avait été déjà recommandé en 1898 par le germaniste Carl Haag.

Précisons finalement que pour garantir une meilleure lisibilité du fond de la carte polygonale les points-ALF correspondant aux îles situées le long de la côte atlantique ont été fusionnés avec la masse continentale de la France.

##### 4.1.2. Taxation (codage ou typisation) des données originales de l'ALF

Par «taxation» (appelée aussi «codage» ou «typisation») l'on entend un processus de mensuration au niveau de l'échelle (métrologique) nominale qui consiste à représenter l'essentiel de la variation des données originales de l'ALF par la variation analogue d'un certain nombre d'unités taxatoires (appelés «taxats») qui ne se distinguent entre elles que par leur seule différence nominale. Toute taxation revient donc, en dernière analyse, à une simplification

3. Etant donné qu' E. Edmont a effectué deux enquêtes à Saint-Pol-sur-Ternoise (= P.-ALF 284), sa ville natale, l'ALF dispose de 639 enquêtes faites dans 638 communes. Comme nous n'avons pris en compte que la première des deux réponses enregistrées au P.-ALF 284, notre réseau-ALF n'a que 638 points d'enquête, mis à part les trois points-ALF factices (PP. 997, 998 et 999).

des données de l'atlas considéré qui —ce qu'il ne faut jamais oublier— est issu, lui aussi, d'un processus de simplification de la réalité basilectale telle qu'elle s'est présentée aux oreilles (et au crayon!) de l'explorateur.

Nous n'avons taxé que des planches-*ALF* aussi complètes que possible tout en écartant celles qui, sur le total des 638 réponses théoriquement possibles, disposent de moins de 600 entrées transcrites et partant utilisables: ceci pour minimiser le danger de distorsions numériques qui résulteraient inévitablement de la prise en compte taxométrique de cartes d'atlas par trop lacunaires.

Le travail taxatoire contient, en ligne générale, deux volets: le volet *phonétique* d'un côté, et le volet *lexical* et *morpho-syntaxique* de l'autre.

Quant à la taxation phonétique, il est bien évident qu'elle ne pouvait se faire qu'à partir de cartes-*ALF* étymologiquement homogènes (mononymes) auxquelles nous avons appliqué une grille analytique aussi précise que possible. Cette grille d'analyse comprenait les catégories classiques de la phonétique historique, en partant des positions prétoniques, toniques et posttoniques relatives aux différents nexus vocaliques et/ou consonantiques latins, pour aboutir à la position (libre ou entravée) des syllabes situées sous et en dehors de l'accent tonique de la langue originale (i. e. du latin classique et vulgaire), ainsi qu'un certain nombre de nexus latins d'usage fréquent (comme, par exemple, -ĒRE, -ĪRE, -ĀRE, -ĀTU, -ĒLLU, Ē + Palatale, Ō + Palatale, etc.). Voir à ce sujet aussi notre site internet: <http://ald.sbg.ac.at/dm>.

C'est ainsi que, du point de vue *phonétique*, les cartes mononymes de l'*ALF* ont pu être analysées plus d'une fois si bien que nous avons tiré, de 247 cartes originales mononymes de l'*ALF*, 1117 cartes dites «de travail»: voir le tableau récapitulatif suivant.

			Corpus: TOT: 641 PP.	Corpus: OC: 220 PP.	Corpus: NORD: 421 PP.		
1	2		3	4	5	6	
Catégorie linguistique	Cartes - <i>ALF</i>		Cartes de travail	Taxats	Taxats	Taxats	
	Décompte <i>partiel</i>	Décompte intégral	Décompte intégral	Décompte intégral	Décompte intégral	Décompte intégral	
1	Phonétique	247	—	1.117	12.023	7.507	8.948
2	<i>Vocalisme</i>	242	—	612	8.657	5.218	6.518
3	<i>Consonantisme</i>	227	—	505	3.366	2.289	2.430
4	Morpho-syntaxe	84	—	99	989	634	625
5	Lexique	463	—	471	7.031	3.680	4.691
	Total	—	<b>626</b>	<b>1.687</b>	<b>20.043</b>	<b>11.821</b>	<b>14.264</b>

Remarque: Comme, d'une carte originale de l'*ALF*, il était possible de tirer plusieurs cartes de travail appartenant en outre à plus d'une catégorie linguistique, il est impossible de calculer le total de la colonne 2 du tableau précédent en additionnant les effectifs partiels des lignes 1-5.

En ce qui concerne la taxation *lexicale* et *morpho-syntaxique* par contre, il était le plus souvent impossible de tirer plus d'une carte de travail d'une seule planche originale de l'*ALF*. C'est pourquoi le rendement taxatoire du lexique et de la morpho-syntaxe est de loin inférieur à celui de la phonétique. Pour le lexique nous avons tiré, de 463 cartes originales de l'*ALF*, 471 cartes de travail et, pour la morpho-syntaxe, 99 cartes de travail à partir de 84 cartes-*ALF* originales.

Les taxats lexicaux équivalaient *grosso modo* aux étymons des formes-*ALF* respectives. L'adjonction de suffixes, préfixes et infixes à la forme simple issue directement du latin (etc.) entraînait toujours l'établissement d'un nouveau taxat. En règle générale, notre propos était de capter ainsi le dynamisme de la genèse et de la fragmentation lexicales de l'espace galloroman tout en nous inspirant vaguement des comportements de recherche utilisés en matière d'onomasiologie («Wörter und Sachen») et aussi de l'esprit analytique dans lequel les différentes entrées du *FEW* ont été rédigées.

Pour ce qui est de la sélectivité empirique de notre travail taxatoire précisons que, pour la phonétique, les différents taxats correspondent le plus souvent aux caractères simples de la transcription de l'*ALF*. Quant aux caractères combinés (diphthongues, triphthongues, caractères en position élevée, etc.) certains parmi eux ont été réunis en groupes pour enrayer le morcellement interne du bilan taxatoire. Ce procédé correspond d'ailleurs au principe du «lumping», opposé à celui du «splitting», fréquemment utilisé en matière de taxonomie biologique.

Comme le travail taxatoire et la saisie électronique de données taxées sont très onéreux et demandent beaucoup de temps et d'énergie, nous n'avons pu considérer que 626 —ou quelque 44% —du total des 1421 cartes originales de la série A de l'*ALF*. La figure 2 montre la répartition des 626 cartes analysées sur l'abscisse de l'ensemble des 1421 cartes-*ALF* originales.

L'ordonnée de la figure 2 illustre en outre le fait de la taxation multiple des cartes originales de l'*ALF*: l'on y voit clairement que le nombre des cartes de travail tirées d'une seule carte-*ALF* est très inégale et peut varier entre 1 et 10.

L'examen combiné du nombre des cartes-*ALF* taxées, des cartes de travail établies et des taxats analysés nous permet de mieux saisir la variabilité interne de la matrice de données: c'est ainsi que nous avons tiré en moyenne 2,69 cartes de travail (=  $1687 : 626$ ) de chaque planche originale de l'*ALF* et que chaque carte de travail dispose, en général, de 11,88 taxats (=  $20\,043 : 1687$ ) dont l'aire respective dispose théoriquement de 53,95 points du réseau-*ALF* (=  $641 : 11,88$ ).

L'examen analogue pour le nord (sous-corpus NORD) et le sud (sous-corpus OC) de notre réseau nous montre que le morcellement taxatoire (et partant la fragmentation géolinguistique) du sud est supérieur à celui du nord. Des 14 264 taxats répertoriés dans le sous-corpus NORD l'on obtient 8,45 taxats/carte de travail (=  $14\,264 : 1687$ ) dont l'aire respective moyenne embrasse 49,82 points-*ALF* (=  $421 : 8,45$ ). Pour le sud, le rendement taxatoire moyen d'une carte de travail est de seulement 7 taxats/carte de travail (=  $11\,821 : 1687$ ) dont l'aire moyenne occupe respectivement 31,41 points-*ALF* (=  $220 : 7$ ). L'on peut donc dire qu'en général les aires des différents taxats sont plus réduites dans le sud (sous-corpus OC) que dans le nord (sous-corpus NORD).

Le même décalage numérique entre le sud et le nord résulte d'une comparaison de la richesse taxatoire des vecteurs d'attributs des N points d'atlas des deux sous-corpus:

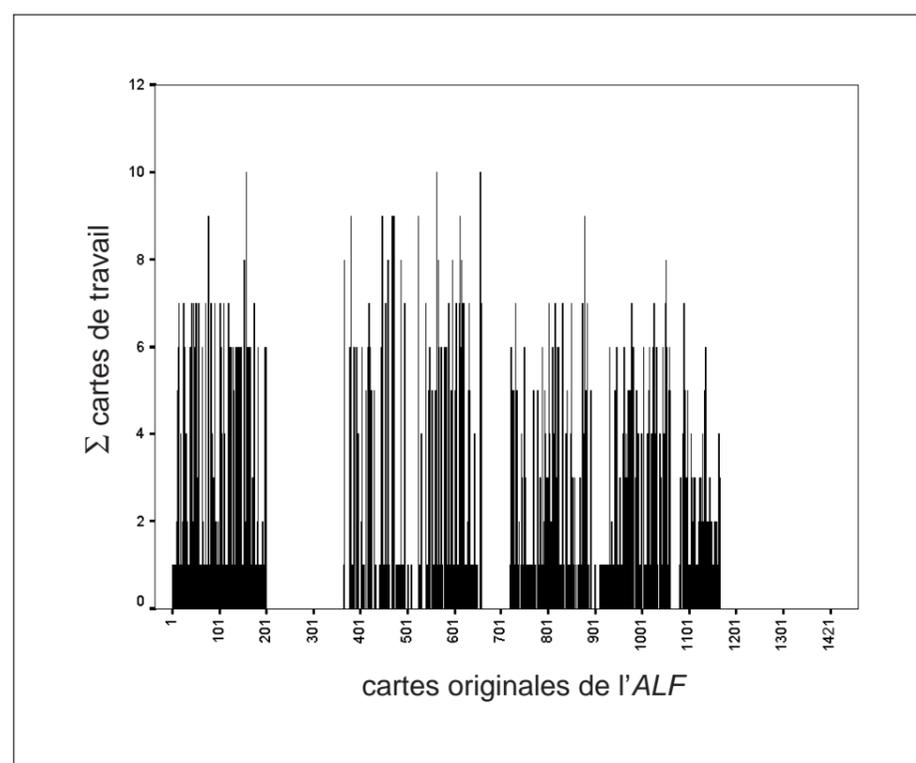


FIGURE 2: Diagramme du dépouillement taxatoire de l'ALF effectué à Salzbourg entre 1997 et 2000.

OC: 53,75 taxats/vecteur d'attributs (= 11 821 : 220)  
 NORD: 33,88 taxats/vecteur d'attributs (= 14 264 : 421).

La considération concomitante du nombre des taxats par carte de travail (en abscisse) et du nombre des cartes de travail (en ordonnée) est du plus haut intérêt: voir les figures 3-5. L'on obtient ainsi des courbes dont la déclivité est bien accusée et d'une allure souvent très régulière.

Pour le corpus total (voir la figure 3, cliché supérieur) la déclivité continue s'inscrit entre 221 cartes de travail 3-nyms (c'est-à-dire comprenant *trois* taxats) et la seule carte 91-nyme du corpus alors que, pour le seul corpus consonantique (voir la figure 4, cliché intermédiaire), la déclivité respective s'inscrit entre 87 cartes de travail 2-nyms et l'unique carte de travail 27-nyme.

Par contre, la silhouette de la courbe vocalique (voir la figure 4, cliché supérieur) est plus accidentée. C'est qu'il s'agit d'une distribution bimodale, c'est-à-dire disposant de deux sommets numériques.

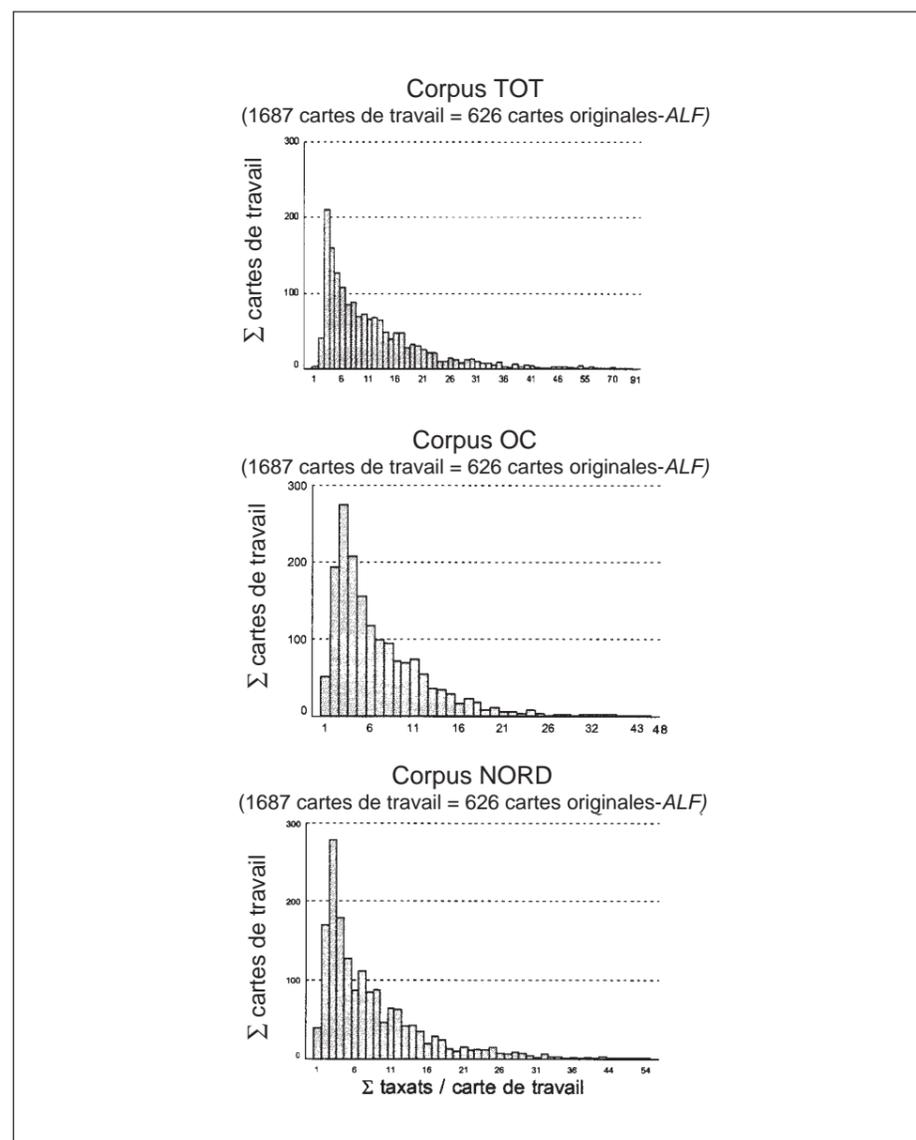


FIGURE 3: Histogrammes relatifs à la relation entre le nombre des cartes de travail et leur morcellement taxatoire interne.

Attributs: corpus intégral (1687 cartes de travail = 626 cartes originales de l'ALF).  
Points-ALF: corpus TOT (641), corpus OC (220), corpus NORD (421).

L'étalement des valeurs relatives au lexique (voir la figure 4, cliché inférieur) est moins régulier que celui des scores relatifs au corpus intégral: en ordonnée, les valeurs maximales (36 et 35 cartes de travail) correspondent aux cartes de travail 5- et 10-nymes, et la valeur minimale à la seule carte de travail 91-nyme.

Le morcellement taxatoire du lexique (de 1 à 91) dépasse donc largement celui du vocalisme (de 2 à 76 taxats/cartes de travail) et du consonantisme (de 2 à 27 taxats/cartes de travail).

Précisons encore que les régularités visualisées sur les figures 3-5 reflètent certaines particularités du changement linguistique pluriséculaire tel qu'il s'est déroulé dans l'espace gallo-roman. Elles peuvent être modélisées du point de vue mathématique comme résultantes d'une interaction complexe entre différents processus de genèse et de dégénérescence d'unités linguistiques diverses (cfr. Altmann, 1985 avec la déduction stochastique d'une «loi de Goebel»).

#### 4.2. De la matrice de données aux matrices de similarité et de distance

Le passage de la matrice de *données* ou bien à la matrice de *similarité* ou bien à la matrice de *distance* —voir la figure 1— se fait moyennant l'application d'un «indice de similarité» ou «de distance». Or, la taxométrie moderne (cfr. p. ex. Sneath/Sokal, 1973; Bock, 1974; Chandon/Pinson, 1981; Bailey, 1994 et Gordon, 1999) offre à ce propos une gamme richissime d'indices utilisables qui répondent cependant tous à d'autres nécessités et finalités exploratoires. Il est donc indispensable que le dialectométricien règle son choix sur les besoins de sa propre démarche taxométrique. Nous utilisons dans le cadre de cet article deux indices de similarité —l'«Indice Relatif d'Identité» ( $IRI_{jk}$ ) et l'«Indice Pondéré (avec le poids 1) d'Identité» ( $IPI(1)_{jk}$ )— qui ont été mis au point par nous il y a belle lurette et dont l'utilité taxométrique a été éprouvée à d'innombrables reprises.

Pour l' $IRI_{jk}$ : cfr. Goebel, 1982: 25 s., 1984, I: 47 s. et 2001: 858 s. (en allemand) ainsi que 1981: 357 s., 1992a: 434 s. (en français) et 1993a: 283 s. (en anglais); pour l' $IPI(1)_{jk}$ : cfr. Goebel, 1983b: 12 s., 1984, I: 83 s. et 1998a: 981 s. (en allemand) ainsi que 1987a: 70 s. (en français).

Alors que l' $IRI_{jk}$  est un indice isocratique (ou «adansonien», baptisé d'après le naturaliste français Michel Adanson, 1727-1806) et traite ainsi tous les taxats de la matrice de données numériquement sur un pied d'égalité, l' $IPI(1)_{jk}$  confère, comme indice anisocratique, un poids numérique plus grand aux taxats géographiquement moins répandus tout en défavorisant les taxats de grande diffusion spatiale. L'action pondératrice de cet indice traduit en termes taxométriques une opinion très habituelle parmi les linguistes selon laquelle les traits linguistiques universellement répandus disposeraient d'une importance classificatoire moindre que les traits linguistiques moins fréquents auxquels on attribue souvent la prérogative d'être ou bien des archaïsmes ou bien des innovations dignes d'un intérêt majeur.

La formule de l' $IPI(x)_{jk}$  contient un facteur de pondération  $x$ . L'importance numérique (ou le «poids») des taxats de moindre diffusion augmente avec la diminution du facteur  $x$ : pour les détails mathématiques cfr. Goebel, 1983b: 12 s. (en allemand) et 1987a: 70 s. (en français). Les exemples iconiques respectifs présentés dans cet article, y ont été calculés avec le facteur de pondération maximal, à savoir avec  $x = 1$ . L'utilité taxométrique de l'indice pondéré  $IPI(x)_{jk}$  réside avant tout dans la détection de paysages dialectaux moins saillants et dans la saisie privilégiée des enchevêtrements des taxats du vecteur d'attributs du point de référence, repérables dans son entourage immédiat et aussi aux antipodes typologiques de la carte de similarité respective.

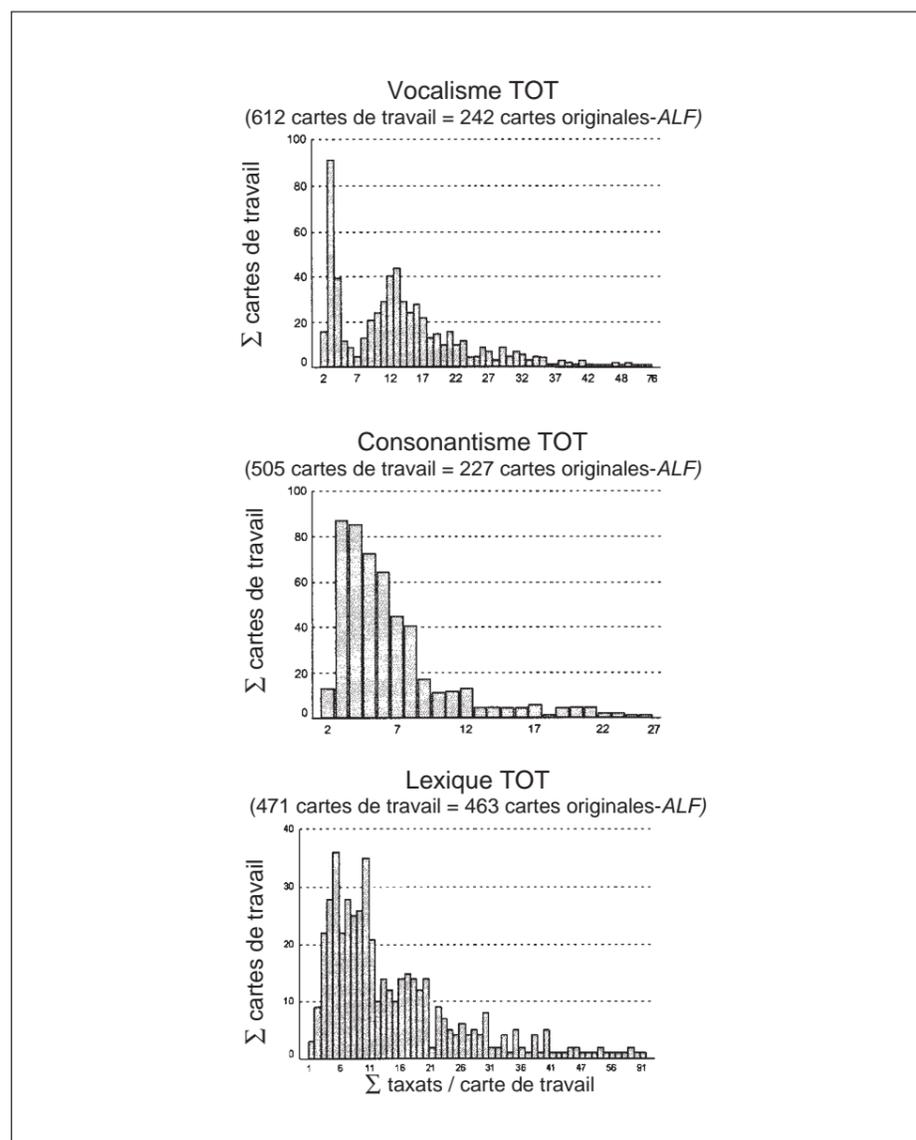


FIGURE 4: Histogrammes relatifs à la relation entre le nombre des cartes de travail et leur morcellement taxatoire interne.

Attributs: vocalisme, consonantisme et lexique.

Points-ALF: corpus TOT (641).

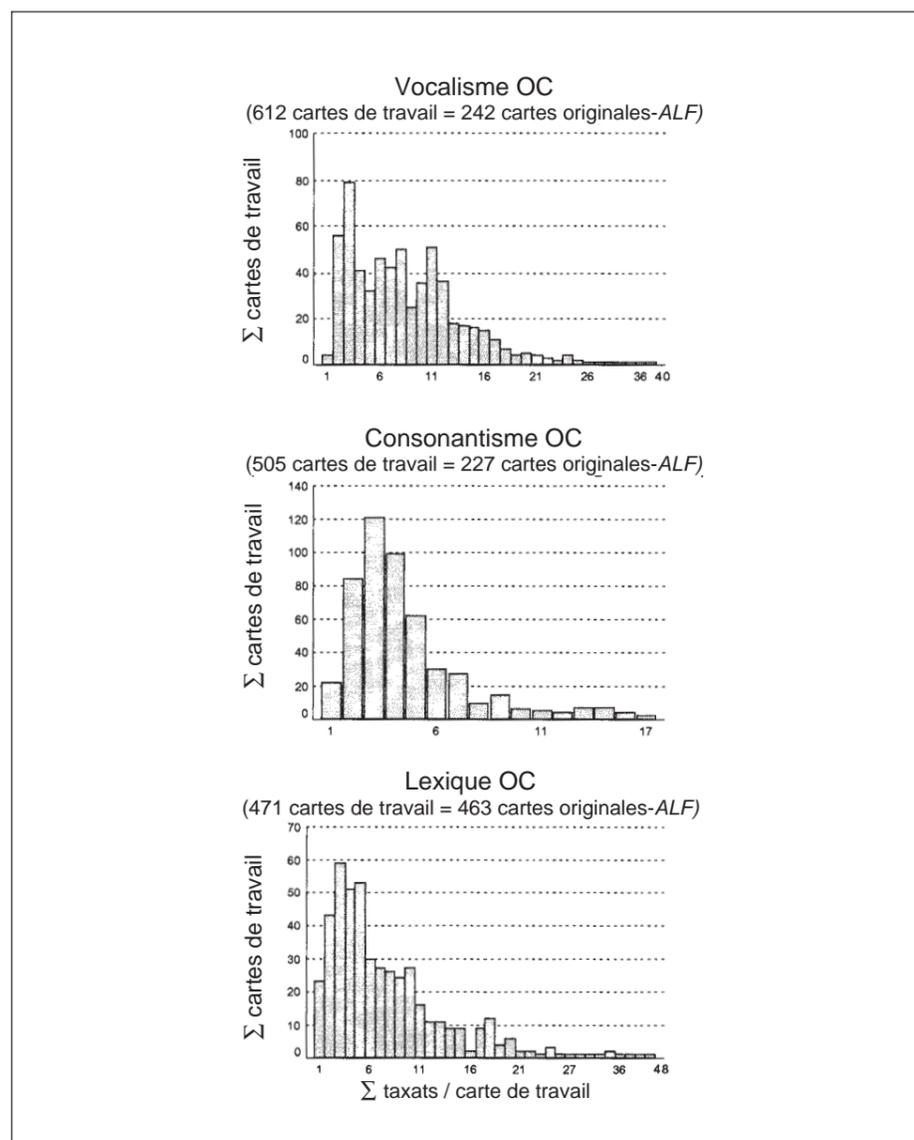


FIGURE 5: Histogrammes relatifs à la relation entre le nombre des cartes de travail et leur morcellement taxatoire interne.

Attributs: vocalisme, consonantisme, lexique.

Points-ALF: corpus OC (220).

#### 4.3. Les prérequis statistiques et cartographiques de la mise en carte des résultats numériques

Les 24 cartes présentées dans cet article obéissent à des conventions statistiques et cartographiques précises qui ont été élaborées il y a plus de vingt ans. Le but central des cartes est de transformer la variabilité numérique des différentes distributions dialectométriques (c'est-à-dire de similarité, de distance, etc.) en une variabilité analogue d'un certain nombre de paliers chromatiques ordonnés selon la logique du spectre solaire (ou de l'arc-en-ciel). Or, les acquis de la psychologie visuelle nous disent que la sensibilité de l'oeil humain n'est bien desservie que jusqu'à un contraste différentiel de 6 à 8 paliers chromatiques différents. Il est donc indiqué de prévoir la possibilité de changer librement le nombre des paliers chromatiques utilisables pour satisfaire ainsi tant les besoins physiologiques de l'oeil que la curiosité exploratoire du dialectométricien. Le logiciel VDM répond parfaitement à toutes ces nécessités par la mise à disposition de 10 paliers chromatiques comprenant, au choix, entre 2 et 20 intervalles numériques.

La définition des paliers chromatiques s'opère à l'aide d'algorithmes d'intervallisation bien définis, dont deux —MINMWMAX et MEDMW— sont utilisés pour la confection des 24 cartes de cette contribution. Pour une description détaillée cfr. Goebel, 1981: 365 s. et 1987a: 79 s. (en français) ainsi que 1982: 29 s. et 1984, I: 93-97 (en allemand).

Quant aux histogrammes présentés le plus souvent au-dessous du réseau polygonal des cartes isarithmiques et choroplèthes, ils servent avant tout, moyennant 2n colonnes verticales, à la visualisation des caractères statistiques de la distribution de fréquence mise en carte (à l'aide de n paliers chromatiques et/ou intervalles numériques). Leur construction est également assurée par un algorithme précis qui génère toujours 2n colonnes (équivalentes à autant d'intervalles numériques): cfr. à ce sujet Goebel, 1981: 363 (en français) et 1984, I: 97-98 (en allemand).

La projection de la courbe gaussienne (ou normale) théorique (calculée à partir de la *moyenne arithmétique* et de l'*écart-type* [all. *Standardabweichung*; angl. *standard deviation*] de la distribution de fréquence respective sur l'histogramme en question, offre la possibilité d'en contrôler rapidement certaines particularités statistiques.

### 5. LA MENSURATION DE LA DISTANCE LINGUISTIQUE ENTRE LES POINTS CONTIGUS DE L'ALF: LA DM INTERPONCTUELLE: VOIR LES CARTES 1 ET 2

#### 5.1. CONVENTIONS TAXOMÉTRIQUES ET CARTOGRAPHIQUES

Nous savons, depuis les explications théoriques (accompagnées d'excellentes démonstrations cartographiques multicolores) du germaniste Carl Haag publiées dès 1898, que la géographie de Delaunay-Voronoi est le seul moyen graphique apte à permettre un tracé rationnel d'isoglosses combinées. De nos jours, la cartographie assistée par l'ordinateur offre la possibilité d'en réaliser facilement tous les prérequis tant formels que pratiques. La polygonisation du réseau de l'ALF a abouti à la définition de 1791 côtés de polygone dont le maillage constitue le support graphique de la syntaxe iconique des cartes 1 et 2. Le tracé des tronçons d'isoglosses<sup>4</sup> qui en-

4. Précisons que le confinement linéaire de chacune des 20 043 aires taxatoires de notre corpus-ALF intégral ne s'effectue pas, dans l'optique de géométrie de Delaunay-Voronoi, moyennant une ligne ondulée continue (= isoglosse dans le sens traditionnel), mais bien plutôt à l'aide d'un certain nombre de côtés de polygone (ou tronçons isoglottiques) combinés dont chacun représente un segment rectiligne relativement court.

tourent les 20 043 aires de notre corpus intégral, s'effectue donc le long des 1791 côtés (ou segments) de polygone mentionnés ci-dessus.

Dans ce contexte, il semble indiqué de rappeler encore trois points:

1) que chacun des 1791 segments de polygone (ou tronçons d'isoglosse) est porteur d'un certain score de *distance* (de *dissemblance*, *dissimilarité*, etc.) entre les vecteurs d'attributs de deux points d'atlas contigus (j et k); cfr. Goebel, 1983a: 387.

2) que nous disposons, depuis 1953, du terme heureux d'«interpoint», proposé par l'abbé Théodore Lalanne, pour désigner la position *intermédiaire*, et partant *relationnelle*, de chaque côté de polygone entre deux points d'atlas contigus.

3) que la contiguïté des points de n'importe quel atlas linguistique est définie au cours de la triangulation du réseau d'atlas respectif qui, elle, constitue donc la première étape du processus de polygonisation. Elle s'appelle d'ailleurs «géométrie de Delaunay» (cfr. Okabe/Boots/Sugihara, 1992: 72 s.).

Précisons encore, avant d'aborder l'interprétation des cartes 1 et 2, qu'elles reposent, toutes les deux, sur le même stock de données interponctuelles. Il s'agit donc de deux *visualisations* différentes qui obéissent, chacune, à d'autres conventions cartographiques.

La carte 1 a été réalisée à l'aide de l'algorithme d'intervallisation MINMWMAX et de 4 intervalles numériques (correspondant à autant de paliers chromatiques différents). Les scores interponctuels équivalents aux couleurs «chaudes» (rouge et ocre) se situent entre le minimum (= 14,63) et la moyenne arithmétique (= 40,82) de la distribution de fréquence respective, alors que les scores interponctuels visualisés en turquoise et bleu foncé occupent l'espace numérique entre la moyenne arithmétique (= 40,82) et le maximum (= 78,83). Le nombre réduit des paliers chromatiques fait mieux ressortir les zones d'implantation des grandes distances interponctuelles: c'est là que le faciès de la surface polygonale change le plus rapidement.

Précisons en outre que les maxima des scores interponctuels de distance sont visualisés par l'apparition parallèle de deux facteurs iconiques: 1) de la *plus grande épaisseur* des côtés de polygone et 2) de la couleur *bleue* (évoquant symboliquement la «froideur» de relations interindividuelles). Le contraire est vrai pour les minima: apparition concomitante de l'*épaisseur minimale* des côtés de polygone et du coloriage en *rouge* (évoquant symboliquement la «chaleur» des relations d'amitié entre deux individus voisins).

Les bases numériques des cartes 1 et 2 ont été calculées avec l'«Indice Pondéré (avec le poids 1) de Distance»  $IPD(1)_{jk}$  (formule:  $IPD(1)_{jk} = 100 - IPI(1)_{jk}$ ) qui, par rapport à l'«Indice Relatif de Distance»  $IRD_{jk}$  (formule:  $IRD_{jk} = 100 - IRI_{jk}$ ), privilégie les scores interponctuels maximaux et confère de ce fait à la syntaxe iconique de la carte interponctuelle une allure plus accidentée, voire escarpée. L'effet-frontière entre le sud et le nord de la Galloromania devient ainsi plus explicite.

## 5.2. Interprétation linguistique des cartes 1 et 2

Malgré leur habillage cartographique différent, les deux cartes font apparaître nettement qu'entre le nord et le sud de la France, il existe une large zone de transition sillonnée d'innombrables cloisons plus ou moins saillantes. La logique iconique de la carte 1 laisse entrevoir, par le biais des côtés de polygone en bleu foncé, un effet séparateur qui tout en évoquant de loin l'impression d'une ligne, se creuse d'une part entre les domaines d'Oïl, francoprovençal et d'Oc et de l'autre entre les domaines d'Oïl et francoprovençal. Sur les deux cartes, des

effets séparateurs moins marqués se manifestent en outre dans le nord, l'est et le sud-ouest du domaine d'Oïl, dans les zones montagnardes du domaine d'Oc (Massif Central et Alpes provençales) et entre le languedocien et le catalan du Roussillon.

Bien que la carte 2 dispose d'une articulation cartographique (et surtout chromatique!) plus fine, son utilité heuristique semble être inférieure par rapport à celle de la carte 1, surtout en ce qui concerne la présentation des maxima du cloisonnement interponctuel de notre réseau d'enquête. La raison en est due à l'incapacité de l'œil humain de discerner, avec la même aisance, plusieurs teintes de couleurs froides d'une part et la variation concomitante de l'épaisseur des côtés de polygone de l'autre.

Les cloisonnements minimaux (symbolisés par la couleur rouge et l'épaisseur minimale des côtés de polygone) renvoient à des paysages dialectaux plus ou moins homogènes (Centre, Languedoc, Provence maritime). Remarquons en outre la position iconique détachée de la Petite Gavacherie (= P. 635, Andraut, Gironde) qui, comme chacun sait, est un îlot linguistique saintongeais (et partant de type oïlique) situé en territoire d'Oc, à l'est de Bordeaux, et les positions non moins périphériques du catalan et de l'italien standard adossés respectivement au sud du Roussillon et au sud-est de la Provence maritime.

Voir aussi les cartes interponctuelles numérisées dans Goebel, 2000a: 236 et 2002: 40 qui, basées sur d'autres corpus-ALF et réalisées avec des algorithmes d'intervallisation différents, se distinguent légèrement des cartes 1 et 2 de cet article.

Une comparaison des cartes 1 et 2 avec des synthèses isoglottiques antérieures, faites à la main et établies avec des données-ALF beaucoup plus réduites, s'impose: voir à ce sujet Rosenqvist, 1919<sup>5</sup> et Etmayer, 1924.<sup>6</sup> Il en appert la dimension du progrès méthodique et technique réalisé en quelque quatre-vingts ans de recherches. Mais il en résulte également la portée de l'erreur qu'avait commise Gaston Paris en 1888 quand il affirmait, en Sorbonne et devant l'assemblée annuelle des Sociétés savantes de France: «Et comment, je le demande, s'expliquerait cette étrange frontière qui de l'ouest à l'est couperait la France en deux en passant par des points absolument fortuits? Cette muraille imaginaire, la science, aujourd'hui mieux armée, la renverse, et nous apprend qu'il n'y a pas deux Frances, qu'aucune limite réelle ne sépare les Français du nord de ceux du midi [...].» (Paris, 1888: 435).

Le verdict de Gaston Paris qui, dans le même discours, réclamait la mise en chantier d'un atlas dialectal de la France, est l'émanation d'une pensée anti-classificatoire d'inspiration franchement nominaliste (ou «typophobe»; cfr. Goebel, 1984, I: 14) qui, outre qu'elle a été fomentée par une certaine exaspération politique due aux conséquences de la Guerre franco-prussienne de 1870/71, a longtemps paralysé le dépouillement classificatoire globalisant des données des atlas linguistiques, y compris de celles de l'ALF.<sup>7</sup>

La syntaxe iconique des cartes 1 et 2 nous montre en outre —et que Gaston Paris s'en rasure après coup!— qu'il n'y a aucune coupure linéaire nette au sein de la Galloromania et que les phénomènes différenciateurs s'échelonnent au sein de larges zones (ou bandes) de transition. Il en résulte aussi que la vieille métaphore de la «frontière linguistique» —qui inévitablement suggère l'idée (et partant l'existence réelle!) d'une rupture linguistique linéaire dans l'espace— est carrément inutilisable et qu'il est grand temps de la proscrire et de la remplacer définitivement par celle de la «zone (ou bande) de transition».

5. Les deux cartes de Rosenqvist ont été republiées, sous une forme légèrement remaniée, dans Berschin/Felixberger/Goebel, 1978: 261 et 266.

6. Voir surtout les sept planches en grand format annexées à ce livre *magistral* qui mérite ce qualificatif aussi du point de vue de la dialectologie théorique.

7. Voir à ce sujet les contributions bien documentées et riches en observations pertinentes de D. Baggioni, 1996 et G. Bergounioux (1984, 1994: 38 s. et 1997).

Revenons encore au fait quelque peu paradoxal que la syntaxe interponctuelle des cartes 1 et 2 nous cache un certain nombre de paysages dialectaux bien connus (tel le wallon, le limousin, l'auvergnat, etc.) qui semblent disparaître derrière le chassé-croisé des cloisonnements épais et bleus des deux cartes, quitte d'ailleurs à réapparaître de plus belle au cours de mensurations de *similarité* ou d'analyses *dendrographiques* appliquées aux mêmes données. Cette contradiction apparente tient à deux raisons:

- 1) aux divergences classificatoires qui existent entre les différentes méthodes dialectométriques,
- 2) au caractère *ambigu* du concept de «frontière».

Comme toute frontière marque le ralentissement ou la stagnation de certaines relations (ou communications) interponctuelles alors que d'autres continuent à agir bel et bien, l'analyse dialectométrique de l'effet-frontière doit recourir à différentes méthodes taxométriques pour déceler cette ambivalence. L'effet *ralentisseur* est saisi par la DM *interponctuelle* (cfr. Goebel, 2000b) alors que l'effet de *continuité* est l'objet de la DM de *similarité*. Pour un approfondissement interdisciplinaire de ce problème nous recommandons à nos lecteurs de consulter les livres stimulants de Nordman 1998 et Bromberger/Morel 2002.

Bien que la carte interponctuelle semble constituer, à cause de sa parenté méthodique avec les synthèses isoglottiques traditionnelles, un outil de recherche d'un accès apparemment facile, il est bon de ne pas en surestimer la portée classificatoire. Des 205 120 scores de distance (selon la formule:  $N/2 (N - 1)$ ; pour  $N = 641$ ) répertoriés dans la matrice de distance, seul un pourcentage minime (i. e.  $0,87\% = 1791 : 205\ 120$ ) a été utilisé pour la confection des cartes 1 et 2 alors que les cartes présentées dans les chapitres 7 et 8 de cet article reposent sur la totalité des scores numériques stockés dans la même matrice de distance (ou de similarité). L'importance *taxométrique* (et partant *classificatoire*) de la DM interponctuelle est donc fort limitée.

## 6. LA MENSURATION DE LA SIMILARITÉ LINGUISTIQUE ENTRE LES 641 POINTS-ALF: LA DM RELATIONNELLE (RÉFÉRENTIELLE) OU DE SIMILARITÉ

### 6.1. CONVENTIONS TAXOMÉTRIQUES ET CARTOGRAPHIQUES

Comme, sur les cartes isarithmiques et choroplèthes de cet article, la numérotation originale de l'ALF a dû être omise pour des raisons de lisibilité, nous mettons un exemplaire vierge de la carte polygonale dotée des numéros de tous les points-ALF à la disposition de nos lecteurs: consulter l'adresse internet suivante: [http://ald.sbg.ac.at/dm/germ/Allgemein/ALF\\_Netz.pdf](http://ald.sbg.ac.at/dm/germ/Allgemein/ALF_Netz.pdf).

La visualisation des cartes choroplèthes 3-12 repose sur l'algorithme d'intervallisation MINMWMAX avec 4 et 6 intervalles, alors que celle des cartes 13, 15 et 17 a été réalisée à l'aide de l'algorithme MEDMW avec 8 intervalles.

La mensuration de la similarité se fait surtout à travers  $IRI_{jk}$ , mais aussi à l'aide de l' $IPI(1)_{jk}$  (cartes 13 et 15).

Précisons d'emblée que la variation de l'étendue des corpus utilisés suivant le nombre des points d'enquête (voir les cartes 6, 9-10 et 17-18) et celle de la nature des attributs mis en synthèse suivant les différentes catégories linguistiques (voir les cartes 10 et 11), sont tout particulièrement intéressantes.

Voir aussi —à titre de comparaison— les cartes coloriées dans Goebel, 2000a (figures 1-12) et 2002 (cartes 1-20).

#### 6.2. DEUX PROFILS DE SIMILARITÉ «SUDISTES»: PRÉSENTATION DES CARTES 3 ET 4

Les cartes 3 et 4 sont deux représentants typiques de l'outil dialectométrique de la carte de similarité. Elles disposent, comme toutes les cartes de similarité, d'un point de référence (ou de repère) (à choisir librement parmi les N points d'enquête du réseau étudié) et fournissent, par le biais du jeu interférentiel de 6 paliers de couleur, des profils iconiques bien ordonnés, c'est-à-dire pourvus de structurations spatiales très claires. L'étalement des 6 plages coloriées dans l'espace montre fort bien le degré d'insertion du vecteur d'attributs du point de référence dans l'ensemble du réseau, et illustre de ce fait la manière dans laquelle le potentiel linguistique du point de référence influe sur le reste du réseau ou, à l'inverse, est lui-même influencé par son entourage plus ou moins éloigné.

Pour les multiples usages de la carte de similarité voir les explications détaillées dans Goebel, 1981: 368 s. (en français), 1984, I: 100 s. (en allemand) et 1993a: 284 s. (en anglais).

La carte 3 montre un profil iconique typiquement catalan: ceci signifie qu'à partir des autres points-*ALF* du Roussillon les profils iconiques respectifs sont plus ou moins les mêmes. Les polygones rouges qui symbolisent les plus grandes similarités relatives au point de référence 796 (Arles-sur-Tech, Pyrénées Orientales), couvrent la totalité du Roussillon alors que les polygones orange se projettent, à l'instar d'un «glacis de transition», à l'intérieur du domaine languedocien. Le pourtour septentrional de la zone jaune marque la fin de la latinité de type occitano-catalan. Il coïncide presque parfaitement avec son pendant de la carte 4. Rappelons que les couleurs *chaudes* correspondent à des scores de similarité situés *au-dessus* de la moyenne arithmétique et signalent ainsi une parenté linguistique *extraordinaire* face au point de repère respectif. Le contraire est vrai pour les couleurs *froides* qui, équivalentes aux scores d'identité situés *au-dessous* de la moyenne arithmétique, marquent la présence d'affinités linguistiques *peu accusées*.

Le profil de la carte 4 (point de référence: 664, Luxey, département des Landes) est typiquement «gascon». L'ensemble des polygones rouges et orange coïncide presque entièrement avec la silhouette traditionnelle de la Gascogne. Remarquons en outre que, pour les deux cartes, les antipodes (c'est-à-dire les points d'enquête linguistiquement plus distants face au point de référence) occupent le nord et l'est du réseau et s'alignent donc entre la Wallonie, la Lorraine et la Suisse romande. La carte 4 offre la particularité très rare que l'intervalle 1 (en bleu foncé) ne dispose que d'un seul point-*ALF* (= P. 998, italien standard).

Ajoutons que notre logiciel VDM permet de visualiser rapidement, par un simple clic sur la souris, tous les autres profils de similarité stockés dans la matrice de similarité respective, d'où la possibilité d'en étudier avec aisance la richesse des structurations choroplèthes y ayant trait.

#### 6.3. MENSURATION DE SIMILARITÉ ET VARIATION DE CORPUS: PRÉSENTATION DES CARTES 5 ET 6

Etant donné la nature relationnelle (ou référentielle) de la mensuration de similarité, il est bien évident que la position d'un point de référence donné et l'agencement du profil choroplèthe y ayant trait doivent varier en fonction de l'étendue du réseau au sein duquel le point de référence respectif se trouve. Le contraste iconique entre les cartes 5 et 6 en offre un exemple très éloquent. La carte 5 repose sur notre corpus-*ALF* intégral (avec 641 points-*ALF*) alors que la carte 6 est basée sur un corpus réduit qui, sur le total de ses 221 points-*ALF*, comprend

218 points-ALF relatifs au domaine occitano-catalan et les trois points artificiels habituels, à savoir le français (P. 999), l'italien (P. 998) et le catalan (P. 997) standards.

Le point de référence 743 (Gaillac, département du Tarn) se situe en plein domaine languedocien. Les fondements généraux du profil iconique de la carte 5 —définis par la distribution spatiale des couleurs chaudes et froides— sont les mêmes que sur les cartes 3 et 4.

Quant au profil iconique de la carte 6, il est bien évident qu'il diffère de celui de la carte 5. Toujours est-il que l'ensemble des plages rouges, orange et jaunes de la carte 6 recouvre presque entièrement la somme des polygones rouges et orange de la carte 5. Sur la carte 6, les antipodes typologiques se trouvent en Provence (P. 990), en Auvergne (PP. 702, 703, 804, 805, 809) et à l'emplacement du catalan, de l'italien et du français standards.

#### 6.4. L'INCIDENCE TAXOMÉTRIQUE ET CARTOGRAPHIQUE DE LA MENSURATION PONDÉRÉE DE LA SIMILARITÉ: PRÉSENTATION DES CARTES 7 ET 8

Malgré le fait que les cartes 7 et 8 se partagent le même point de référence (886, Aups, département du Var), leurs profils choroplèthes diffèrent considérablement. Ceci s'explique par l'application de deux indices de similarité différents:  $IRI_{jk}$  (carte 7) et  $IPI(1)_{jk}$  (carte 8). Les plus grandes différences se manifestent moins dans l'étalement des couleurs chaudes, mais surtout dans celui des couleurs froides. C'est qu'à l'intérieur de la zone froide se détachent — sous la forme d'un hémicycle en bleu foncé, adossé à la frontière linguistique vers le néerlandais et l'allemand— des buttes-témoins équivalent aux similarités linguistiques minimales. Ajoutons encore que, sur les deux cartes, l'étendue de la zone rouge —qui coïncide *grosso modo* avec le domaine linguistique provençal— est sensiblement la même. Cette particularité vaut également pour les autres points-ALF de la Provence au cas où ils figureraient comme points de référence d'une carte de similarité. Signalons encore que sur la carte 8, réalisée avec l'indice anisocratique  $IPI(1)_{jk}$ , la zone chaude est plus cohérente que celle de la carte 7 et ne comporte aucun polygone en vert, bleu ou bleu foncé, mis à part les deux points-ALF factices du catalan et de l'italien standards ainsi que l'îlot linguistique de la Petite Gavacherie (P. 635, Andraut, département de la Gironde) répertorié en vert (et situé à proximité de Bordeaux).<sup>8</sup>

#### 6.5. LA FRANCISATION DU DOMAINE D'OC ET L'INCIDENCE TAXOMÉTRIQUE DE DIFFÉRENTES CATÉGORIES LINGUISTIQUES: PRÉSENTATION DES CARTES 9 ET 10

Le corpus réduit occitan qui comprend également le point-ALF artificiel du français standard (P. 999) permet d'examiner, à l'aide d'une mensuration de la similarité faite à partir de ce dernier, la logique spatiale de la diffusion, voire de l'intrusion de la «langue du Roi et de la République Une et Indivisible» dans les terres d'Oc. Il est en outre très intéressant de poser la même question sur la base de corpus partiels relatifs au *vocalisme* (612 cartes de travail), au *consonantisme* (505 cartes de travail) et au *lexique* (471 cartes de travail). L'ampleur numérique de ces sous-corpus est suffisamment grande pour permettre une comparaison mutuelle fiable.

Comme il s'agit également de déceler aussi bien que possible les zones de la plus grande distance linguistique face au français standard, il est indiqué de recourir, pour la mensuration de la similarité par rapport au point 999, à l'indice pondéré  $IPI(1)_{jk}$ .

8. Pour les origines historiques et linguistiques de la Petite Gavacherie cfr. Tourtoulon-Bringuier, 1876: 567 s., Suchier, 1888: 598 et Ronjat, 1930: 24-25.

La carte 9 montre clairement que la progression du français dans le domaine d'Oc dépend avant tout de la distance kilométrique face au cœur politique, social et culturel de la France. Les zones plus francisées coiffent le domaine d'Oc vers le nord, à l'exception notable du couloir rhodanien et de la côte provençale qui très souvent ont fait fonction de zones d'irruption et de propagation de formes centrales (parachutées ou non): lire à ce sujet les pages lumineuses que l'historien Fernand Braudel (1902-1985) a dédiées à ce problème (p. ex. 1986, I: 267 s., sur l'«isthme français»).

Signalons encore que le domaine languedocien et quelques zones pyrénéennes et catalanes semblent résister le mieux à l'infiltration de la latinité centrale d'Oïl.

Les trois profils choroplèthes de la carte 10 confèrent encore plus de précision au message de la carte 9. C'est que l'on y perçoit avec netteté que la diffusion spatiale du *vocalisme* et du *consonantisme* français obéit à des logiques diatopiques *différentes*: la plus grande avancée du *vocalisme* se déploie le long de la *côte atlantique* tout en rejoignant le cœur des Pyrénées alors que le *consonantisme* se lance vers la Méditerranée à travers le *couloir rhodanien*. L'implantation du *lexique* français (ou «francisant») en territoire d'Oc, par contre, paraît s'opérer tant par *progression linéaire* sur le terrain que par *parachutages ponctuels*.

Les antipodes typologiques (ou «noyaux de résistance») face au français se trouvent pour le *vocalisme* du côté du languedocien, pour le *consonantisme* dans la Gascogne méridionale et pyrénéenne, et pour le *lexique* dans le Roussillon.

Soulignons encore le fait tout autre qu'insignifiant que le point 635 (Andraut, Petite Gavaucherie) qui représente un îlot linguistique d'Oïl, n'a conservé sa parenté linguistique originelle avec le français que pour le *vocalisme* et le *consonantisme* alors que, pour le *lexique*, il semble avoir encouru une certaine méridionalisation, surtout de la part du limousin et du gascon septentrional: voir à ce sujet aussi la carte 13 dans Goebel, 2002: 52.

#### 6.6. LA POSITION HYBRIDE DE LA LATINITÉ CRESCENTINE: PRÉSENTATION DE LA CARTE 11

Le latinisme «crescentain» constitue l'adjectif relatif au terme technique «Croissant» forgé, comme chacun sait, en 1913 par Jules Ronjat. Celui-ci voulait caractériser par cette métaphore en dernière analyse (g)astronomique, l'alignement curviligne des isoglosses délimitant le domaine d'Oc vers le nord tel qu'il se manifestait sur les premières cartes muettes tracées d'après les données de l'*ALF*, déjà disponible alors avec la totalité de ses cartes. Les discussions relatives au Croissant ou —en d'autres termes— à la zone de transition entre le domaine d'Oïl et d'Oc sont d'ailleurs beaucoup plus anciennes. Guylaine Brun-Trigaud en a tiré, en 1990, la substance pour une description historique détaillée et très intéressante sous la forme d'une monographie brillante.

Dans l'optique de la DM, le Croissant entendu comme zone de transition entre Oc et Oïl, constitue une bande courbe relativement large qui s'étend de la Gironde jusqu'au domaine franco-provençal et peut revêtir, suivant les analyses dialectométriques appliquées, des articulations spatiales diversifiées: voir à ce sujet les cartes 11 dans Goebel, 2000a: 235 et 14 dans Goebel, 2002: 53.

La carte 11 montre trois profils d'identité relatifs au point-*ALF* 609 (Seilhac, Corrèze) dont chacun a été établi à l'aide d'un sous-corpus différent: *vocalisme* (612 cartes de travail), *consonantisme* (505 cartes de travail) et *lexique* (471 cartes de travail). Or, il s'avère que les profils choroplèthes de ces trois mensurations de similarité divergent considérablement entre eux. Alors que le profil *vocalique* est franchement «sudiste» et ressemble de ce fait beaucoup à nos cartes 3-5, le profil *consonantique* tire plutôt vers le nord tout en dégageant, par l'agrégation cohérente de plusieurs polygones rouges, une agglomération «crescentiforme». Le profil *lexical*, quant à lui, revêt une allure nettement hybride. De tels phénomènes d'hybridation

géotypologique peuvent être observés aussi à partir d'autres points-ALF situés à l'intérieur du Croissant. Ceci vaut également pour la dialectométrie de sous-corpus relatifs au vocalisme et au consonantisme qui, pour les points-ALF du Croissant, montrent des affiliations géotypologiques très diversifiées, allant tantôt vers le nord, tantôt vers le sud.

Dans la perspective de la DM, l'on peut donc dire qu'une des propriétés géotypologiques saillantes du Croissant consiste dans la composition mixte des latinités respectives en fonction des différentes (sous)catégories linguistiques.

#### 6.7. EN PASSANT DU DOMAINE D'OÏL AU DOMAINE D'OC: PRÉSENTATION DE LA CARTE 12

La structuration très éloquente des différents profils choroplèthes de similarité et leur stabilité géotypologique relativement grande nous avaient suggéré, dès le début de nos recherches dialectométriques, l'expérience visuelle suivante: de présenter un certain nombre de cartes choroplèthes de similarité dont les points de référence sont alignés le long d'un parcours d'épreuve cohérent, en fondu enchaîné et de profiter ainsi des avantages (typo)diagnostiques qu'offrent les effets de la cinématographie: cfr. à ce sujet Goebel, 1981: 374-375 et 1984, I: 104-105.

Les trois clichés de la carte 12 en illustrent —bien que d'une manière très approximative— le principe. Les profils de similarité de la carte 12 reposent sur trois points de repère dont le premier (504, Dun-le-Palleteau, Creuse) se trouve en domaine d'Oïl, le deuxième (609, Seilhac, Corrèze) en plein Croissant et le troisième (628, Villefranche-de-Belvès, Dordogne) en domaine d'Oc. Les points-ALF 504, 609 et 628 sont donc alignés le long d'un parcours imaginaire qui du domaine d'Oïl se dirige dans le domaine d'Oc tout en traversant le Croissant.

Or, la carte 12 montre très nettement que le point 504 dispose d'un profil d'identité franchement oïlique, que celui du P.-ALF 609 est tiraillé entre le nord et le sud avec une certaine prépondérance vers le sud, et que, enfin, le profil d'identité du P. 628 revêt une allure occitane à tous les égards.

Soit dit entre parenthèses, le cliché central de la carte 12 doit être comparé avec ceux de la carte 11 dont il constitue en quelque sorte la somme iconique.

A l'aide du programme VDM (ou du CD-ROM y ayant trait) il serait d'ailleurs facile d'étudier de plus près, l'une après l'autre, toutes les cartes de similarité calculables entre les points 505 et 628, à savoir celles qui se réfèrent aux points-ALF 603, 617, 616, 618 et 619.

Signalons encore un détail intéressant qui ressort du cliché de gauche de la carte 12 (profil d'identité relatif au point-ALF 504). L'on y discerne un provignement linéaire en ocre qui, au sud-est de la zone composée des couleurs rouge et ocre, se dégage de la superficie oïlique cohérente et s'insinue entre les domaines d'Oc et francoprovençal comme un espèce de cordon sanitaire. Cet alignement est constitué des PP.-ALF 808, 819, 818, 828, 920, 838, 940 et 849 et s'étend *grosso modo* entre Thiers et Grenoble.

Il s'agit là d'un effet de francisation rampante dû à une incertitude (méta)linguistique de la part des locuteurs autochtones qui se traduit avant tout par des changements (ou substitutions) phonétiques au niveau du consonantisme prétonique. Comme l'explication dialectométrique détaillée de ce phénomène, déjà signalé et discuté par A. Dauzat (1928), P. Gardette (1970) et d'autres,<sup>9</sup> serait fort complexe, nous devons renoncer à la présenter ici-même, quitte à la publier ailleurs.

9. «On s'étonne parfois du chemin que certaines aires de francisation dessinent sur la carte de notre pays, spécialement dans le domaine francoprovençal, qu'il s'agisse du lexique, de la morphologie ou de la phonétique. C'est ainsi qu'Albert Dauzat a été intrigué par le long et large chemin qu'a pris le *ch* français de Roanne à Grenoble. Je voudrais montrer que ce chemin n'est pas seulement celui du *ch* français, mais qu'il est aussi celui d'autres traits de

## 7. AU-DELÀ DE LA MENSURATION DE LA SIMILARITÉ: LA SYNOPSIS DE DEUX PARAMÈTRES CARACTÉRISTIQUES DES DISTRIBUTIONS DE SIMILARITÉ

Chacune des cartes de similarité présentées dans le paragraphe 6 de cet article repose sur une distribution de fréquence qui, elle, comprend, pour le réseau intégral de l'ALF, 640 (= 641 – 1 point de référence) scores numériques. Or, toute distribution de fréquence est passible d'être caractérisée selon les préceptes de la statistique descriptive. Dans cette perspective, il est de mise de relever, entre autres, les paramètres dits «caractéristiques» tels que le minimum, le maximum, la moyenne arithmétique, l'écart-type et différentes mesures de symétrie parmi lesquelles compte aussi le «coefficient d'asymétrie de Fisher» (CAF; all. *Schiefe*, angl. *skewness*).

Comme une des finalités majeures de la DM consiste à déterminer la position relationnelle de chaque vecteur d'attributs au sein de la matrice de données et que ceci peut être réalisé surtout à l'aide de l'analyse statistique des N distributions de similarité stockées dans la matrice de similarité, il est indiqué de se pencher sur l'utilité taxométrique des différents paramètres caractéristiques tout en les réunissant en synthèse d'abord et en les visualisant ensuite: cfr. à ce propos nos explications détaillées dans Goebel, 1981: 381 s. (en français) et 1984, I: 136 s. (en allemand).

### 7.1. À LA DÉCOUVERTE DE NOYAUX DIALECTAUX: LA SYNOPSIS DES MAXIMA DES DISTRIBUTIONS DE SIMILARITÉ: PRÉSENTATION DES CARTES 13 ET 14

Par «noyau dialectal» l'on entend, en matière de géolinguistique, une section d'un tissu dialectal donné où, à travers la concomitance territoriale d'un certain nombre de traits linguistiques, se produit une certaine homogénéisation géotypologique. Il arrive très souvent que cet effet de compactage est ressenti aussi de la part des locuteurs des dialectes respectifs.

Le concept de noyau dialectal présuppose encore deux choses:

- 1) qu'un tissu dialectal donné dispose de plusieurs noyaux dialectaux,
- 2) qu'entre les différents noyaux dialectaux s'étend un réseau de failles ou de zones de passage (amphizones) où la combinaison respective des différents traits linguistiques se désagrège progressivement pour se recomposer d'une autre manière par la suite, quitte à former, à plus ou moins longue distance, un autre noyau dialectal.

En ligne générale, les maxima se trouvent, sur les cartes de similarité, dans le voisinage immédiat du point de référence. Pour chacune des N distributions de similarité stockées dans la matrice de similarité, le maximum sera donc transféré à l'endroit du point de référence respectif. Ce faisant, l'on obtient une distribution de fréquence composée de N valeurs maximales différentes qui, elle, finira par être soumise au traitement cartographique habituel.

Pour obtenir des résultats plus finement articulés, nous avons choisi l'indice  $IPI(1)_{jk}$  qui privilégie numériquement les taxats linguistiques peu répandus (et partant retenus comme étant plus «caractéristiques» pour une région donnée) tout en circonscrivant le poids numérique des taxats universellement connus et passant, de ce fait, pour être «insignifiants».

francisation: qu'il est moins un chemin qu'une zone francisante, allongée le long de la frontière qui sépare, à l'ouest et au sud-ouest, le francoprovençal du provençal; qu'il est le résultat d'une sorte d'effondrement de la région francoprovençale entière, effondrement produit par le climat d'incertitude régnant au sujet des mots et des formes francoprovençales et occitanes.» (Gardette, 1970: 291). Voir aussi Dauzat, 1928: p. 72 s. ainsi que Wüest, 1981: 14 s.

Les résultats de cette analyse se présentent sous la forme d'un paysage ondulé où crêtes et vallées, plateaux et pentes alternent harmonieusement.

Les cartes 13 et 14 se basent en même temps sur les mêmes données *taxométriques* et sur deux philosophies *visualisatrices* différentes. Toujours est-il qu'elles fournissent des perspectives très détaillées de l'organisation «nucléaire» de l'espace dialectal de la Galloromania. Le stéréogramme de la carte 14 a été réalisé à l'aide du logiciel *MapInfo*. Pour d'autres visualisations stéréographiques (relatives à l'Italie septentrionale et aux données de l'AIS) voir Goebel, 1981: 383 (en français) et 1984, I: 115 s. (en allemand).

Ajoutons encore que les résultats de cette analyse «nucléaire» dépendent dans une large mesure de la granulation du réseau de l'atlas considéré. Plus le maillage respectif sera fin et équidistant, plus l'allure «paysagiste» de la synthèse des maxima sera accentuée. Précisons en outre qu'à cet égard le maillage de l'ALF laisse un peu à désirer, surtout en ce qui concerne l'exploration du centre de la Galloromania.

Ceci n'empêche que les profils des cartes 13 et 14 sont très éloquents. Sont à remarquer:

1) *pour l'ensemble du réseau*: la grande faille située entre Oïl et Oc et les fossés plus ou moins profonds (en bleu foncé, bleu marin et vert) qui s'inscrivent entre un certain nombre de noyaux de taille inégale situés entre les régions suivantes: Picardie/Wallonie, Normandie occidentale/Normandie centrale, Maine/Touraine, Touraine/Saintonge, Champagne/Bourgogne, Lorraine/Franche-Comté.

2) *pour le domaine d'Oïl*: des noyaux solides (en rouge et ocre) dans l'ouest, le sud-est, dans le Centre, en Picardie, en Franche-Comté et en Bourgogne; des noyaux moins marqués (en orange et jaune) en Normandie, en Wallonie, en Champagne et en Franche-Comté.

3) *pour le domaine d'Oc*: des noyaux solides (en rouge et orange) dans le Languedoc, en Auvergne, dans la Gascogne orientale et méridionale ainsi qu'en Provence.

4) *pour le domaine francoprovençal*: une ultime butte-témoin (en jaune) en Haute-Savoie (PP. 933, 945 et 944).

Remarquons en outre le trajet ondulé de la grande faille (en bleu foncé) qui sépare les domaines d'Oïl et d'Oc, et le fait que les valeurs minimales de la synthèse des maxima se trouvent le long du pourtour méridional du francoprovençal.

## 7.2. L'ANALYSE DU COMPROMIS (OU BRASSAGE) LINGUISTIQUE (ALL. «SPRACHAUSGLEICH») À L'AIDE DE LA SYNOPSIS DES «COEFFICIENTS D'ASYMÉTRIE DE FISHER» (CAF)

Une des propriétés mathématiques les plus importantes de toutes les distributions de similarité réside dans leur plus ou moins grande (a)symétrie. Est symétrique toute distribution de fréquence dont les scores se répartissent à part égale des deux côtés de la moyenne arithmétique. Or, les distributions de similarité ont ceci de particulier qu'il est possible d'associer à un surplus de scores de similarité situé *au-dessus* de la moyenne arithmétique, l'idée de capacités communicatives *extraordinaires* avec le reste des données examinées. Ceci vaut *mutatis mutandis* aussi pour la situation inverse —c'est-à-dire pour un excédent de scores de similarité situé *au-dessous* de la moyenne arithmétique— d'où résulte une *mauvaise* insertion communicative de la distribution de similarité (et du vecteur d'attributs du point de référence respectif) dans l'ensemble du réseau de l'atlas examiné.

Pour se convaincre de la variation du degré de symétrie des différentes distributions de similarité, il suffit de comparer entre eux les histogrammes des cartes 3, 4, 5, 7 et 8. Il en appert

que la considération serrée du degré de symétrie (ou d'asymétrie) des différentes distributions de similarité est du plus haut intérêt géolinguistique (ou dialectologique).

Parmi les nombreuses possibilités qu'offre la statistique descriptive pour la mensuration de l'asymétrie d'une distribution de fréquence, le «coefficient d'asymétrie de Fisher» (CAF), proposé en 1925 par le statisticien anglais Sir Ronald A. Fisher (1890-1962),<sup>10</sup> occupe une place de choix. Pour le cas d'une symétrie parfaite de la distribution de fréquence respective, le CAF prend la valeur 0, des valeurs négatives quand la distribution de fréquence est asymétrique vers la droite tout en accumulant la majorité des scores au-dessus de la moyenne arithmétique, et des valeurs positives quand la symétrie de la distribution respective tend vers la gauche et que, par conséquent, la plupart des scores numériques se concentrent entre la moyenne arithmétique et le minimum.

Le CAF est une mesure très importante en matière de DM. Ceci s'explique surtout par le fait qu'il permet de saisir, par voie quantitative et partant dialectométrique, un des problèmes majeurs de n'importe quel domaine linguistique, à savoir celui de sa *genèse diachronique*. Quant au développement du faciès géolinguistique de la Galloromania actuelle, il est bien connu depuis longtemps qu'il est la résultante d'une multiplicité de processus d'irradiation intenses à partir de l'Île-de-France qui provoquaient des mouvements de retrait et de résistance de différentes latinités submergées ou minorées, parmi lesquelles la latinité de Lugdunum, vieille capitale des Gaules avant Lutetia/Paris, figure au premier rang.

La logique algébrique du CAF<sup>11</sup> permet de capter, à travers les maxima des scores négatifs (et visualisés ensuite en bleu foncé, etc.), les endroits du réseau d'atlas examiné où s'opère la plus forte interaction géolinguistique à échelle macro-régionale, alors que la considération des scores positifs (et visualisés par la suite en rouge, etc.) aboutit à la détection de zones conservatrices ou de retrait. La mise en carte de N scores-CAF permet donc de broser un tableau synthétique et comparatif à la fois, des différentes *avancées* (ou *percées*) entropiques du compromis (ou brassage) linguistique à grande échelle, et aussi des actions de *défense* et de conservation qui s'y opposaient.

#### 7.2.1. *L'incidence taxométrique et cartographique de la pondération au-delà de la mensuration de la similarité: présentation des cartes 15 et 16*

Passons maintenant à la considération des cartes 15 (cartes choroplèthes) et 16 (surfaces statistiques lissées). Pour mieux faire ressortir les accidents des types iconiques respectifs, les deux profils choroplèthes de la carte 15 ont été établis à l'aide de huit (au lieu de quatre ou six) paliers chromatiques et de deux indices de similarité différents ( $IRI_{jk}$  et  $IPI(1)_{jk}$ ). De cette manière, il est possible de juger rapidement des conséquences taxométriques de l'effet pondérateur de l' $IPI(1)_{jk}$  au niveau du CAF.

Les deux stéréogrammes de la carte 16 visualisent le même bilan numérique d'une façon iconiquement plus évidente encore.

Le message classificatoire central des deux profils choroplèthes de la carte 15 peut être résumé en trois constatations élémentaires:

1) implantation *circulaire* d'un compromis linguistique intense dans le nord (voir l'agencement spatial des polygones en bleu foncé, bleu marin et bleu clair). Le caractère circulaire

10. Cfr. Brentari, 1990: 40-41.

11. Pour la formule et une description mathématique précise cfr. Goebel, 1981: 394-401 et 1984, I: 1250-153 ainsi que Kreyszig, 1968: 94 et Brentari, 1990: 40-41.

du rayonnement du compromis linguistique ressort mieux encore à la lumière de l'indice de similarité pondéré  $IPI(1)_{jk}$ : voir le profil supérieur de la carte 15.

2) même remarque pour le domaine francoprovençal qui est entouré de phénomènes de brassage linguistique de façon *plus* (cliché *inférieur* de la carte 15) ou *moins* (cliché *supérieur* de la carte 15) marquée.

3) persistance de plusieurs poches de résistance linguistique (coloriées surtout en rouge et orange) dans les domaines d'Oc et francoprovençal. L'indice pondéré  $IPI(1)_{jk}$  rehausse, sur la carte choroplèthe respective (voir le cliché *inférieur* de la carte 15), la position excentrique du limousin, de l'auvergnat et du francoprovençal.

Quant au domaine d'Oïl, les plages vertes (rangées dans l'intervalle 4) peuvent être associées à des zones «pacifiées» où le dynamisme expansif du brassage linguistique à grande échelle s'est arrêté après avoir généré une certaine saturation sur le plan des enchevêtrements de traits linguistiques à grande diffusion géographique.

Constatons encore pour le domaine d'Oïl qu'il y existe, mise à part la région centrale, aussi trois aires latérales qui autrefois ont dû être le théâtre d'un brassage linguistique plus important: la Saintonge, la Picardie et la Wallonie: pour les deux dernières provinces voir surtout le cliché supérieur de la carte 15.

Encore une ultime remarque relative à une particularité diatopique oïlique de la carte 15: l'alignement de polygones soit en *bleu foncé* (cliché *supérieur* de la carte 15) soit en *vert* (cliché *inférieur* de la carte 15) qui clôture le francoprovençal vers le sud, correspond exactement à celui que nous avons déjà analysé dans le paragraphe 6.7. (voir la carte 12, profil de gauche). L'origine en est due à une forte immission de traits consonantiques de type d'Oïl le long d'une ligne de force qui pénètre entre le francoprovençal et le domaine d'Oc.

Passons maintenant à l'analyse de la structuration choroplèthe du domaine d'Oc. L'on y constate que les plages rouges et orange sont entourées de couloirs jaunes et ocres qui représentent également, bien que dans une moindre mesure, des zones d'interaction et d'échange. Comme aucun des grands sous-dialectes occitans (gascon, languedocien, etc.) n'a réussi à s'imposer dans le sud aux dépens des autres, il s'est établi, parmi eux, à la différence du domaine d'Oïl, une sorte de coexistence diatopique équilibrée à l'instar de celle qui s'est instaurée entre les grands domaines dialectaux de l'Italie septentrionale (cfr. Goebel, 1981: 394 s. et 1984, I. 150 s.). Ceci n'empêche, bien sûr, les échanges et contacts inter-régionaux capillaires de «tous les jours».

Mais pour mieux étudier cette interaction intra-occitane du point de vue dialectométrique, il faut évidemment enlever des points-ALF d'Oc de notre matrice de données le poids numérique (et relationnel!) de leurs équivalents d'Oïl, et réitérer ensuite les analyses dialectométriques à l'aide de ce corpus occitan réduit. Pour les résultats de cette démarche voir le paragraphe suivant.

Soulignons encore, en guise de conclusion de ce chapitre, un fait méthodique et gnoséologique très important: notre matrice de données dispose, pour l'ensemble de notre réseau de recherche, de 20 043 taxats qui correspondent, dans la logique diatopique de nos données, à autant d'aires géolinguistiques. Or, chacune de ces 20 043 aires, dont les étendues respectives peuvent varier considérablement, constitue, en tant que *dépositaire* d'une *géographie* et aussi d'une *histoire particulières* (cfr. Christmann, 1971), un agent coopérateur dans l'élaboration bimillénaire du faciès géolinguistique de la Galloromania. La saisie, par voie dialectométrique, des modalités relationnelles de cette coopération aux facettes multiples, permet donc de brosser un *portrait global* de la genèse de l'*aménagement géolinguistique* de la Galloromania telle qu'elle s'est présentée à la fin du XIX<sup>e</sup> siècle.

7.2.2. *L'incidence taxométrique de l'ampleur du réseau: la France du sud face à la France du nord: présentation des cartes 17 et 18*

La bipartition de notre corpus-ALF permet d'étudier de plus près, dans chacune des deux moitiés, deux phénomènes mal observables dans la perspective intégrale:

- 1) pour le sud: l'implantation linéaire du compromis (ou brassage) linguistique entre les différents sous-domaines d'Oc,
- 2) pour le nord: l'implantation ponctuelle de plusieurs poches de résistance d'Oïl.

Ad 1: cartes 17 et 18, clichés inférieurs

L'ensemble des polygones en bleu (foncé, marin et clair; intervalles 1-3) qui signalent le brassage linguistique maximal, sillonnent le domaine d'Oc d'une façon cohérente tout en se creusant entre les grands sous-domaines d'Oc: gascon, limousin, languedocien, auvergnat, provençal. Les positions excentriques du gascon, du catalan et du provençal face au reste du réseau sont encore plus marquées que sur la carte 15. Soulignons que parmi tous les sous-domaines d'Oc, le languedocien est le plus tributaire du compromis linguistique pan-occitan, c'est-à-dire rayonnant à l'intérieur du seul tiers sud de la France. Paradoxalement ceci vaut aussi pour le français standard (rangé dans l'intervalle 6: colorié en ocre) dont le «droit de cité» en Occitanie en tant que «coopérateur linguistique» date de quelque 600 ans.

Ad 2: cartes 17 et 18, clichés supérieurs

En effet, l'on y discerne mieux que sur les cartes 15 et 16 plusieurs buttes-témoins de la résistance anti-française. Il s'agit, en ordre décroissant de l'altérité linguistique, des domaines dialectaux suivants: francoprovençal (en rouge et orange), franc-comtois (en orange et ocre), lorrain (en orange et ocre), wallon (en orange et ocre), picard (en orange et ocre), poitevin et saintongeais (en jaune). Sur les deux clichés en question, l'action de l'effet circulaire est d'ailleurs plus accusée dans l'est du domaine d'Oïl que sur la carte 15 (voir surtout le cliché inférieur). Pour encore mieux le faire ressortir, il faudrait établir la même carte à l'aide de l'indice pondéré  $IPI(1)_{jk}$  et du même corpus, c'est-à-dire réduit à la seule moitié septentrionale de la Galloromania.

## 8. LA DIALECTOMÉTRIE DENDROGRAPHIQUE: VOIR LES CARTES 19-24

Ce type de DM se sert, en termes mathématiques, de la «classification ascendante hiérarchique» (CAH). C'est un ensemble de processus taxométriques couramment utilisés depuis plusieurs décennies, lesquels ordonnent, par voie statistique, des objets scientifiques divers dans des hiérarchies arborescentes de classes disjonctives. Le résultat iconique en est l'arbre généalogique: voir les cartes 20, 22 et 24.

Comme, depuis le XIX<sup>e</sup> siècle, la métaphore de l'arbre généalogique a été très souvent utilisée en matière de linguistique (tant générale qu'indo-européenne, germanique et romane; cfr. Schleicher 1863) pour décrire les relations de parenté ou de filiation diachroniques entre les langues (ou dialectes), il est évident que l'utilité et la valeur de telles analyses aussi pour la géolinguistique ne sont plus à démontrer. Rappelons en outre que des classifica-

tions dendrographiques ont été employées également en matière de *lexicostatistique*<sup>12</sup> et de *glottochronologie*: voir à ce sujet les ouvrages synthétiques de Hoenigswald/Wiener 1987 et de Dyan/Kruskal/Black 1992 ainsi que nos remarques respectives publiées en 1992b et 1993b.

Il va de soi que, pour une bonne intelligence des cartes 19-24, il faut se familiariser —au moins dans les grandes lignes— avec le principe taxométrique de l'algorithme utilisé (c'est-à-dire avec la méthode proposée en 1963 par J. H. Ward jr.). En général, l'on en trouve des descriptions détaillées dans tous les manuels modernes de taxométrie. Pour la méthode de Ward voir en particulier Sneath/Sokal, 1973: 204 s. (en anglais), Bock, 1974: 407 s. (en allemand) et Chandon/Pinson, 1981: 123-124 (en français).

L'algorithme proposé par Ward régit —comme d'ailleurs toutes les autres procédures agglomératives— un processus de fusions (ou d'agglomérations) des objets étudiés qui, lui, comprend  $N - 1$  étapes différentes. Il résulte, pour  $N = 641$  points-ALF, 640 fusions (agglomérations ou agglutinations) auxquelles correspondent autant de bifurcations (ramifications ou embranchements) binaires au sein de la hiérarchie arborescente. Le processus agglomératif commence par la considération des *maxima* des  $N^2$  scores de *similarité* (ou des *minima* des scores de *distance*) emmagasinés dans la matrice de similarité (ou distance) respective, et par la fusion consécutive de ceux qui, suivant les critères spéciaux de l'algorithme appliqué, se différencient le moins. Après chaque fusion, l'algorithme arborescent génère une nouvelle matrice de similarité  $[(N - x)(N - x)]$  avec des dimensions toujours décroissantes, pour reprendre ensuite son action agglomératrice.

En procédant ainsi, les algorithmes agglomératifs prennent en compte d'un côté les hétérogénéités «intra-group» (i. e. à l'intérieur des différents groupes ou classes d'objets) et les hétérogénéités «inter-group» (i. e. entre les différents groupes ou classes d'objets) de l'autre, dont l'importance numérique augmente au fur et à mesure que le processus d'agglomération avance tout en s'approchant de la racine de l'arborescence en train de naître: voir à ce propos, en bas des cartes 19, 21 et 23, la ligne fléchée avec la mention «distance linguistique (hétérogénéité intra-group)».

La taxométrie moderne dispose d'un grand nombre de tels algorithmes agglomératifs dont la valeur classificatoire et l'utilité heuristique dépendent cependant dans une large mesure des présuppositions théoriques et empiriques de celui qui les utilise. En dernière analyse, la tâche du classificateur (i. e. en l'occurrence du dialectométricien) consiste à arranger un mariage heureux entre les nécessités et les contraintes *statistiques* du monde taxométrique et celles de l'univers *empirique* considéré. Des algorithmes «passe-partout» ou «bons à tout faire» n'existent pas, de même qu'il n'existe pas de solution-CAH «vraie» ou «unique».

C'est pourquoi nous avons dû procéder, en tant que dialectométricien, comme tous les autres taxométriciens empiriques: expérimenter un certain nombre d'analyses arborescentes d'abord et en apprécier l'utilité géolinguistique par la suite. Cette tâche nous a été rendue tout particulièrement facile par l'admirable logiciel VDM qui permet non seulement de calculer et de construire des arborescences selon différents algorithmes du type CAH, mais aussi d'en spatialiser rapidement les résultats arborescents et de colorier par la suite tant les «dendrèmes» de l'arbre que les «chorèmes»<sup>13</sup> de la spatialisation y ayant trait.

Les cartes 19, 21 et 23 contiennent des représentations simplifiées des arbres originaux qui, eux, disposent respectivement de 641 (carte 19), de 421 (carte 21) et de 221 (carte 23)

12. Pour une discussion mathématique très équilibrée des problèmes de la lexicostatistique cfr. Sankoff, 1973: passim.

13. Les deux termes se rapportent à des groupements significatifs de l'arbre (dendrème) d'un côté et de la spatialisation de l'arbre (chorème) de l'autre. Nous avons emprunté le terme de «chorème» à R. Brunet (1987: 190 et 211) et forgé nous-même, sur son modèle, le terme de «dendrème» (dérivé du grec ancien *déndron* «arbre») en 1998b (296).

feuilles terminales. Aux neuf branches classificatoires (ou dendrèmes) des trois arbres correspondent autant d'agrégations spatiales cohérentes (ou chorèmes) sur les cartes 20, 22 et 24.

A la différence des graphiques isarithmiques et choroplèthes précédents (cartes 1-18), le coloriage des cartes 19-24 ne sert qu'à des fins qualitatives et n'a donc aucune importance quantitative. À souligner l'étonnante cohésion spatiale de tous les chorèmes des cartes 20, 22 et 24 qui ne souffre que peu d'exceptions.

L'interprétation linguistique des arbres et de leurs spatialisations peut se faire tant du point de vue *diachronique* que du point de vue *synchronique*.

#### 8.1. LE VERSANT DIACHRONIQUE DE LA DM DENDROGRAPHIQUE

Quant à l'interprétation diachronique des arbres calculés, la DM dendrographique rejoint la pensée de la lexicostatistique selon laquelle la fragmentation diatopique d'une région donnée s'est développée, au cours du temps, à partir d'une phase initiale une et non diversifiée. Pour problématique que soit une telle vision des choses pour la description et l'explication de la genèse de la richesse phénoménologique actuelle de la Romania, elle connaît néanmoins, bien qu'avec force nuances, de nombreux adhérents parmi les linguistes romans. La même constatation vaut aussi pour les indo-européanistes et les partisans de la classificabilité du total des langues de ce monde dans la lignée de Joseph H. Greenberg et Merritt Ruhlen (cfr. Ruhlen, 1994).

L'interprétation diachronique des cartes 19, 21 et 23 consiste dans une suite de N scénarios de fragmentations successives qui commencent près de la racine (ou du tronc) de l'arbre respectif.

Considérons à titre d'exemple l'arbre de la carte 19. La progression des bifurcations s'opère dans le sens de la flèche «distance diachronique» (en haut sur la carte 19).

*Scénario 1* (voir la ligne verticale bleue):

branche 1: dendrèmes/chorèmes F-I (domaine d'Oc, etc.)

branche 2: dendrèmes/chorèmes A-E (domaine d'Oïl)

*Scénario 2* (voir la ligne verticale verte):

branche 3: dendrèmes/chorèmes F-I (domaine d'Oc)

branche 4: dendrème/chorème E (francoprovençal)

branche 5: dendrèmes/chorèmes A-D (domaine d'Oïl)

*Scénario 3*: (voir la ligne verticale rouge):

branche 6: dendrèmes/chorèmes G-I (domaine occitan extra-gascon)

branche 7: dendrème/chorème F (gascon)

branche 8: dendrème/chorème E (francoprovençal)

branche 9: dendrèmes/chorèmes C-D (domaine d'Oïl du nord et de l'ouest)

branche 10: dendrèmes/chorèmes A-B (domaine d'Oïl du Centre et de l'ouest).

De la logique ramifiée (binaire) de l'arbre de la carte 19 il résulte donc que la première scission interne de la Galloromania originaire aurait abouti à la genèse des domaines d'Oïl et d'Oc, et qu'à la deuxième scission serait due la constitution du francoprovençal. Ce n'est qu'ensuite que le gascon et les dialectes d'Oïl du nord et de l'est se seraient séparés de leurs branches-mères respectives.

Alors que la *glottochronologie* tente de conférer des datations *absolues* aux différents points de bifurcation, la *lexicostatistique* se contente de proposer des datations *relatives*. En

matière de DM dendrographique, il est également préférable de s'en tenir aux datations relatives et de ne surtout pas en exagérer la portée heuristique. Malgré l'apparente plausibilité de la majorité des bifurcations des arbres dialectométriques présentés dans cet article, il faut néanmoins rester prudent et ne pas oublier le caractère foncièrement fictif des structures arborescentes et aussi l'âge réel des données brutes (= ALF: exploré entre 1897 et 1901!). Toujours est-il qu'à la richesse mathématique souvent déroutante des méthodes dendrographiques modernes correspond une abondance non moins troublante relative à l'utilité et aux facultés heuristiques des arbres généalogiques qui en dérivent.

A côté des arbres des cartes 19, 21 et 23, nous avons expérimenté, au cours de nos analyses dialectométriques, un grand nombre d'autres algorithmes arborescents. Fort de ces expériences, nous pouvons donc affirmer —et ceci aussi en vue de rassurer les lecteurs non avertis— que les structures dendrographiques que nous avons étudiées, se ressemblent beaucoup si bien qu'il est possible d'entrevoir une structuration ramifiée «moyenne».

#### 8.2. LE VERSANT SYNCHRONIQUE DE LA DM DENDROGRAPHIQUE

L'utilité synchronique de la DM dendrographique est double: il s'agit d'un côté d'établir une classification régionale des composantes (points d'enquête, etc.) de l'espace (appelée aussi «régionalisation»: cfr. p. ex. Claval, 1993: 45 s.) et des liens de similarité (ou de proximité, etc.) linguistique qui existent entre les groupements de la régionalisation respective de l'autre. Une fois de plus, la logique ramifiée de l'arbre calculé nous permet de résoudre rapidement ce problème. C'est ainsi que la carte 19 nous enseigne que le francoprovençal (dendème E) est —toujours dans la logique algébrique de l'arborescence calculée et selon celle de l'indice de similarité appliqué (ici:  $IPI(1)_{jk}$ )— plus proche du français de l'ouest (dendème B) que du gascon (dendème F), et que le provençal (dendème H) est plus proche des latinités du Limousin, de l'Auvergne et du Croissant (dendème G) que de la latinité du languedocien (dendème I).

#### 8.3. LA CLASSIFICATION DENDROGRAPHIQUE DU RÉSEAU INTÉGRAL DE L'ALF: PRÉSENTATION DES CARTES 19 ET 20

L'arbre respectif a été établi —tout comme ceux des cartes 21 et 23— à l'aide de l'algorithme de Ward et des données de la matrice de similarité calculée avec l'indice de similarité  $IPI(1)_{jk}$ . Pour une comparaison avec des analyses dendrographiques analogues voir les cartes 21 (et 22: méthode-CAH: «Complete Linkage»; indice de similarité:  $IRI_{jk}$ ) et 23 (et 24: méthode-CAH: algorithme de Ward; indice de similarité:  $IRI_{jk}$ ) dans Goebel, 2002: 60 et 63.

Quant à la carte 20, nous renvoyons nos lecteurs encore aux particularités suivantes: 1) appartenance de l'îlot linguistique de la Petite Gavacherie (P. 635) au chorème B; agrégation des points artificiels du catalan et de l'italien au chorème «de transition» G qui, lui, revêt l'aspect d'un «Super-croissant» étiré vers l'est qui finit par s'insinuer entre le francoprovençal (chorème E) et le provençal (chorème H).

Evidemment, le programme VDM permet d'affiner encore la «dendémisation» de l'arbre de la carte 19 et d'étudier ainsi la structuration interne des dendèmes A-I, chose évidemment irréalisable dans le cadre de cet article mais à la portée de tous ceux qui sont prêts à s'initier au maniement de VDM. L'on constatera ainsi que la cohésion spatiale de chorèmes résultant de l'analyse même de dendèmes très petits, reste étonnamment stable.

Il ne faut cependant jamais oublier que le statut mathématique et partant taxométrique de

l'analyse *dendrographique* et de ses résultats diffère considérablement de celui de la DM des *interpoints* et de la *similarité*, et ceci malgré la conformité souvent parfaite avec les résultats diatopiques respectifs.

#### 8.4. LA CLASSIFICATION DENDROGRAPHIQUE DE LA FRANCE DU NORD: PRÉSENTATION DES CARTES 21 ET 22

La comparaison de résultats dialectométriques calculés à partir de corpus de taille divergente sert surtout à tester la stabilité spatiale des arborescences respectives. Quant à la dialectométrisation de l'*ALF*, il en résulte que celle-ci est très grande. En confrontant les cartes 20 et 22 l'on constate que les chorèmes respectifs concordent presque toujours: voir le tableau synoptique suivant:

Correspondance spatiale de chorèmes entre les cartes 20 et 22	
Chorèmes de la carte 20	Chorèmes de la carte 22
A	B + A
D	H + G
B	C
E	I

Dans la perspective *diachronique*, la première bifurcation (voir la ligne verticale bleue) dégage le francoprovençal (branche 1: dendrème/chorème I), la deuxième sépare l'est et le nord (branche 4) du domaine d'Oïl du centre et de l'ouest (branche 5), et la troisième (voir la ligne verticale rouge) détache le bloc picardisant (branche 7) du reste de la moitié orientale du domaine d'Oïl (branche 8).

#### 8.5. LA CLASSIFICATION DENDROGRAPHIQUE DE LA FRANCE DU SUD: PRÉSENTATION DES CARTES 23 ET 24

Ici encore l'on constate d'étonnantes convergences avec l'analyse analogue faite à l'aide du corpus intégral. En voici le tableau synoptique des concordances:

Correspondance spatiale de chorèmes entre les cartes 20 et 24	
Chorèmes de la carte 20	Chorèmes de la carte 24
F	A + B
I	I + G + H
G	C + D + E

Examinons encore les affiliations particulières des points artificiels: le P. 999 (français) passe au chorème D, les PP. 998 (italien) et 997 (catalan) fusionnent avec le chorème E. Alors que les affiliations des PP. 999 et 998 semblent être plausibles, celle du P. 997 (catalan) qui, en bonne logique, aurait dû s'agréger au chorème (roussillonnais) I, surprend. Il faut cependant savoir que les vecteurs d'attributs des points artificiels 997 (catalan) et 998 (italien) sont très lacunaires du point de vue quantitatif et que, du point de vue qualitatif, les matériaux linguistiques respectifs —tirés de dictionnaires bilingues— n'ont pas la même validité (sémasiologique et aussi onomasiologique) que le vocabulaire français «officiel» du point artificiel 999 qui, tout en figurant dans le questionnaire de l'*ALF*, avait été couramment utilisé par Edmond Edmont dans ses enquêtes conduites entre 1897 et 1901.

L'îlot linguistique de la Petite Gavacherie (P. 635) a été incorporé au chorème D qui, parmi tous les chorèmes de ce sous-corpus occitan, semble contenir le plus d'attributs linguistiques oïliques.

Dans la perspective *diachronique*, la première bifurcation (voir la ligne verticale bleue) a détaché le gascon (branche 2: dendrèmes/chorèmes A-B), alors que la deuxième ramification (voir la ligne verticale verte) a séparé le bloc catalano-languedocien (branche 3: dendrèmes/chorèmes G-I) du bloc limousino-provençal (branche 4: dendrèmes/chorèmes C-F). La troisième bifurcation (voir la ligne verticale rouge) oppose, enfin, la Provence (branche 7: dendrème/chorème F) à la couronne septentrionale du domaine d'Oc (branche 8: dendrèmes/chorèmes C-E) qui constitue —comme nous l'avons déjà dit plus haut (paragraphe 8.3.)— une sorte de Croissant démesuré.

## 9. EPILOGUE

Nous concluons avec deux remarques d'inspiration épistémologique (9.1. et 9.2.) et une brève remarque d'ordre purement pratique (9.3.).

9.1. Malgré son indéniable utilité pour beaucoup de philologies modernes (des études africaines à celles des langues baltiques, en passant par les philologies germanique, anglaise, italienne et française),<sup>14</sup> la DM continue d'être regardée avec un certain scepticisme de la part de certains linguistes, surtout de ceux qui voient dans le recours au traitement quantitatif de données foncièrement qualitatives au premier abord, une détérioration de l'essence intrinsèque de ces dernières. Certains de ces scepticismes sont franchement incurables,<sup>15</sup> voire intraitables et évoquent les apories méthodiques de la querelle médiévale dite «des Universaux».

A part cela, la DM est —à cause de sa nature quantifiante— exposée aux mêmes malentendus et incompréhensions dont souffrent aussi d'autres secteurs de la linguistique quantitative. Il s'agit là d'un épiphénomène du conflit impérissable «entre les deux cultures» scientifiques, à savoir entre les sciences dites «humaines» et celles dites «pures» ou «naturelles».<sup>16</sup>

9.2. Le verdict de la «non-existence des dialectes» —prononcé, il y a belle lurette, par Paul Meyer, Gaston Paris, Jules Gilliéron et d'autres linguistes, d'abord contre Graziadio Isaia Ascoli (au cours de la fameuse querelle sur le «franco-provenzale» éclatée, en 1875, entre P. Meyer et G. I. Ascoli)<sup>17</sup> et ensuite contre tous ceux qui osaient grouper leurs données dialectales en entités spatiales majeures tout en dépassant le niveau de l'innocente et simpliste «géographie des traits linguistiques isolés»<sup>18</sup> a pesé lourdement sur le développement méthodique ultérieur du traitement de données géolinguistiques brutes en matière de dialectologie française. Les séquelles de cette consigne fallacieuse étaient encore sensibles du temps de Jean Séguy.

14. Voir les indications bibliographiques dans Goebel, 1992a: 433-434 et 1993a: 277-278.

15. Lire à ce sujet les remarques imprégnées d'une profonde incompréhension face à toute recherche non-particularisante de C. Grassi (2001: 226 s.).

16. Lire à sujet le classique de Sir Charles Percy Snow (original: 1959; dernière édition: 1993).

17. Voir surtout Meyer (1875), Ascoli (1876, réplique à Meyer, 1875) et Paris, 1888.

18. Pour l'historique de cette querelle lire les bilans critiques contemporains d'A. Horning (1893) et de L. Gauchat (1903) ainsi que nos analyses méthodiques rétrospectives de 1986, 1990 (version allemande) et 1995 (version italienne). La paternité de la consigne délétère de «faire la géographie non des dialectes, mais des traits linguistiques» revient à G. Paris (1881: 606).

De nos jours, cette querelle entre le parti des «typophobes» —comme P. Meyer et G. Paris— et celui des «typophiles» se trouve résolue en faveur des derniers. Ceci concerne non seulement la dialectologie française (ou romane), mais aussi beaucoup d'autres sciences tant humaines que naturelles, qui toutes sont aux prises avec l'observation, la mensuration et le traitement classificatoire consécutif d'un grand nombre de phénomènes empiriques apparemment changeants et primesautiers.

Tant la classification elle-même<sup>19</sup> que les techniques de la classification numérique de grandes quantités de données empiriques sont de nos jours universellement reconnues et ont fourni d'innombrables exemples de leur grande utilité et fécondité scientifiques. Quant au développement de la pensée classificatoire dans les seules sciences humaines, l'on peut d'ailleurs constater que pratiquement toutes ont été hantées de tentatives ou incertitudes typophobes de toute sorte<sup>20</sup>. En dernière analyse, il s'agit là d'un épiphénomène de la fameuse «Querelle des Universaux» du moyen âge (cfr. Stegmüller, 1956/57).

L'exemple des sciences naturelles, méthodiquement plus mûres en la matière que les sciences humaines (y compris aussi les différentes branches de la linguistique), et l'essor de l'informatique accompagné de la diffusion rapide d'un jeu bien assorti de logiciels puissants pour la classification numérique et l'analyse des données, ont fini par saper, dans le dernier quart du xx<sup>e</sup> siècle, les ultimes bastions de la pensée anticlassificatoire et typophobe. L'«existence», au sens platonicien du mot, est désormais reconnue non seulement aux dialectes mais à tous les agrégats cognitifs (cfr. Rosch/Lloyd, 1978), qu'ils soient le résultat ou bien de processus cognitifs humains (ou naturels) ou bien de procédés taxométriques artificiels, destinés à la «reconnaissance de structures» (all. «Mustererkennung», angl. «pattern recognition»), tels qu'ils ont été présentés dans cet article.

9.3. Les 24 cartes que nous venons de présenter et d'analyser n'ont fait qu'effleurer la richesse des résultats et méthodes dialectométriques actuellement disponibles sur le système informatique VDM à Salzbourg. Les lecteurs désireux de recevoir un CD-ROM avec le logiciel VDM et la totalité des calculs dialectométriques effectués jusqu'aujourd'hui et à partir de la matrice de données relative à l'ALF, pourront l'obtenir sur simple demande.

## 10. REMERCIEMENTS

### *Taxation du corpus de l'ALF:*

Barbara Aigner, Irmgard Dautermann, Hildegund Eder, Susanne Oleinek, Annette Ida Schatzmann (toutes de Salzbourg)

### *Programmation, rodage et entretien courant du logiciel VDM (Visual Dialectometry):*

Edgar Haimerl (Blaustein, Bade-Wurtemberg, Allemagne)

19. Citons, à titre d'exemple, les ouvrages synthétiques de Broadfield, 1946, Centre National [...], 1963, Rosch/Lloyd, 1978, Parrochia, 1991, Bailey, 1994, Gordon, 1999 et Moriconi, 2000.

20. Voir à ce sujet les contributions d'Altmann/Lehfeldt, 1973 (pour la linguistique), Needham, 1975 (pour l'anthropologie), Mahmood/Armstrong, 1992 (pour l'ethnologie) et Dos Santos, 2001 (pour la géographie humaine), Sériot, 1999 (pour le structuralisme «eurasiste» de N. S. Troubetzkoy, etc.) et de Hempfer, 1973 (pour la critique littéraire et la théorie des genres littéraires).

*Cartographie assistée par l'ordinateur:*  
Slawomir Sobota (Salzbourg)

*Soutien financier:*  
Fonds zur Förderung der wissenschaftlichen Forschung (FWF) in Österreich (Vienne):  
projets de recherche 12414 et 13349

*Supervision technique de l'impression électronique de cet article:*  
Antoni Escolà i Banqué, Servei de Producció Editorial de l'Institut d'Estudis Catalans  
(IEC) (Barcelone)

*Toilette stylistique de notre texte français:*  
Lily Ditz-Fuhrich (Salzbourg)

Que toutes les personnes mentionnées ci-dessus ainsi que les organismes de recherche  
FWF et IEC reçoivent ici l'expression de ma profonde gratitude.

## 11. ABRÉVIATIONS SOUVENT UTILISÉES

CAF	«coefficient d'asymétrie de Fisher»: voir le paragraphe 7.2.
CAH	«classification ascendante hiérarchique»: voir le paragraphe 8.
DM	dialectométrie
IPD(1) <sub>jk</sub>	Indice Pondéré (avec le poids 1) de Distance (selon la formule: IPD(1) <sub>jk</sub> + IRI(1) <sub>jk</sub> = 100): voir le paragraphe 5.1. ainsi que la figure 1.
IPI(1) <sub>jk</sub>	Indice Pondéré (avec le poids 1) d'Identité: voir le paragraphe 4.2.
IRD	Indice Relatif de Distance
IRI <sub>jk</sub>	Indice Relatif d'Identité: voir le paragraphe 4.
MEDMW	algorithme d'intervallisation: voir le paragraphe 4.3.
MINMWMAX	algorithme d'intervallisation: voir le paragraphe 4.3.
VDM	Visual Dialectometry: voir le paragraphe 2.

## 12. RÉFÉRENCES BIBLIOGRAPHIQUES

- AIS Jaberg, Karl/Jud, Jakob (eds.) (1928-1940): *Sprach- und Sachatlas Italiens und der Südschweiz*. Zofingen: Ringier, 8 vol. (réimpression: Nendeln: Kraus, 1971).
- ALF Gilliéron, Jules/Edmont, Edmond (eds.) (1902-1910): *Atlas linguistique de la France*. Paris: Champion, 10 vol. (réimpression: Bologne: Forni, 1968).
- Altmann 1985 ALTMANN, Gabriel (1985): «Die Entstehung diatopischer Varianten. Ein stochastisches Modell». *Zeitschrift für Sprachwissenschaft* 4, p. 139-155.
- Altmann/Lehfeldt 1973 ALTMANN, Gabriel/LEHFELDT, Werner (1973): *Allgemeine Sprachtypologie. Prinzipien und Meßverfahren*. München: Fink.

- Ascoli 1878 ASCOLI, Graziadio Isaia: «Schizzi franco-provenzali [1874]». *AGI* 3, p. 61-120.
- Ascoli 1876 ASCOLI, Graziadio Isaia: «P. Meyer e il franco-provenzale». *AGI* 2, p. 385-395.
- Baggioni 1996 BAGGIONI, Daniel (1996): «De Coquebert de Montbret et Raynouard au duo G. Paris/P. Meyer: Aux sources de la linguistique et dialectologie romanes et françaises». *RLaR*, p. 135-162.
- Bailey 1994 BAILEY, Kenneth D. (1994): *Typologies and Taxonomies. An Introduction to Classification Techniques*. Thousand Oaks, Londres, New Delhi: Sage.
- Bergounioux 1984 BERGOUNIOUX, Gabriel (1984): «La science du langage en France de 1870 à 1885: du marché civil au marché étatique». *Langue française* 63, p. 7-14.
- Bergounioux 1994 BERGOUNIOUX, Gabriel (1994): *Aux origines de la linguistique française*. Paris: Pocket.
- Bergounioux 1997 BERGOUNIOUX, Gabriel (1997): «L'Université et les patois». *Lengas* 42, p. 135-170.
- Berschin/Felixberger/Goebel 1978 BERSCHIN, Helmut/FELIXBERGER, Josef/GOEBL, Hans (1978): *Französische Sprachgeschichte. Lateinische Basis. Interne und externe Geschichte. Sprachliche Gliederung Frankreichs*. Munich: Hueber.
- Bock 1974 BOCK, Hans Hermann (1974): *Automatische Klassifikation. Theoretische und praktische Methoden zur Gruppierung und Strukturierung von Daten (Cluster-Analyse)*. Göttingen: Vandenhoeck & Ruprecht.
- Braudel 1986 BRAUDEL, Fernand (1986): *L'identité de la France. Vol. I: Espace et histoire*. Paris: Flammarion.
- Brentari 1990 BRENTARI, Eugenio (1990): *Asimmetria e misure di asimmetria*. Turin: Giapichelli.
- Broadfield 1946 BROADFIELD, A. (1946): *The Philosophy of Classification*. Londres: Grafton.
- Bromberger/Morel 2001 BROMBERGER, Christian/MOREL, Alain (eds.) (2001): *Limites floues, frontières vives. Des variations culturelles en France et en Europe*. Paris: Editions de la Maison des sciences de l'homme.
- Brun-Trigaud 1990 BRUN-TRIGAUD, Guylaine (1990): *Le Croissant: le concept et le mot. Contributions à l'histoire de la dialectologie française au XIX<sup>ème</sup> siècle*. Lyon: Centre d'Etudes linguistiques Jacques Goudet.
- Brunet 1987 BRUNET, Roger (1987): *La carte. Mode d'emploi*. Paris: Fayard/Reclus.
- Centre National [...] 1963 CENTRE NATIONAL DE RECHERCHES DE LOGIQUE/SOCIÉTÉ BELGE DE LOGIQUE ET PHILOSOPHIE DES SCIENCES (eds.) (1963): *La classification dans les sciences*. Gembloux: Duculot.
- Chandon/Pinson 1981 CHANDON, Jean-Louis/PINSON, Suzanne (1981): *Analyse typologique. Théories et applications*. Paris, New York, Barcelone, Milan: Masson.
- Christmann 1971 CHRISTMANN, Hans Helmut (1971): «Lautgesetze und Wortgeschichte. Zu dem Satz "Jedes Wort hat seine eigene Geschichte"». Stempel, Wolf-Dieter/Coseriu, Eugenio (eds.): *Sprache und Geschichte. Festschrift für Harri Meier zum 65. Geburtstag*. Munich: Fink, p. 111-124.
- Claval 1993 CLAVAL, Paul (1993): *Initiation à la géographie régionale*. Paris: Nathan.
- Claval 1998 CLAVAL, Paul (1998): *Histoire de la géographie française de 1870 à nos jours*. Paris: Nathan.
- Dauzat 1928 DAUZAT, Albert (1928): *Essais de géographie linguistique. Deuxième série: Problèmes phonétiques*. Paris: Champion.
- Dos Santos 2001 DOS SANTOS, José Rodrigues (2001): «La constitution historique des "Cévennes"». *Réflexions sur le statut des entités culturelles*. Bromberger/Morel, p. 183-217.
- Dyen/Kruskal/Black 1991 DYEN, Isidore/KRUSKAL, Joseph. B./BLACK, Paul (1991): *An In-*

- doeuropean Classification: A Lexicostatistical Experiment*. Philadelphia: The American Philosophical Society.
- Ettmayer 1924 ETTMAYER, Karl von (1924): «Über das Wesen der Dialektbildung, erläutert an den Dialekten Frankreichs». *Denkschriften [Mémoires] der Akademie der Wissenschaften in Wien, philosophisch-historische Klasse*, vol. 66, Vienne/Leipzig: Hölder-Pichler-Tempsky, p. 1-56, 7 planches.
- FEW WARTBURG, Walther von (1922 ss.): *Französisches etymologisches Wörterbuch*. Bonn: Bouvier.
- Gardette 1970 GARDETTE, Pierre (1970): «Rencontre de synonymes et pénétration du français dans les aires marginales». *RLiR* 34, p. 280-305 (réimpression: GARDETTE, Pierre (1983): *Etudes de géographie linguistique*. Strasbourg: Société de linguistique romane, p. 521-546).
- Gauchat 1903 GAUCHAT, Louis (1903): «Gibt es Mundartgrenzen?». *ASNSL* 111, p. 365-403.
- Goebel 1981 GOEBL, Hans (1981): «Eléments d'analyse dialectométrique (avec application à l'AIS)». *RLiR* 45, p. 349-420.
- Goebel 1982 GOEBL, Hans (1982): «Dialektometrie. Prinzipien und Methoden des Einsatzes der Numerischen Taxonomie im Bereich der Dialektgeographie». *Denkschriften [Mémoires] der Österreichischen Akademie der Wissenschaften, philosophisch-historische Klasse*, vol. 157, Vienne: Verlag der Österreichischen Akademie der Wissenschaften, p. 1-123, 41 figures.
- Goebel 1983a GOEBL, Hans (1983): «Parquet polygonal et treillis triangulaire: les deux versants de la dialectométrie interponctuelle». *RLiR* 47, p. 353-412.
- Goebel 1983b GOEBL, Hans (1983): «"Stammbaum", und "Welle". Vergleichende Betrachtungen aus numerisch-taxonomischer Sicht». *Zeitschrift für Sprachwissenschaft* 2,1, p. 3-44.
- Goebel 1984 GOEBL, Hans (1984): *Dialektometrische Studien. Anhand italoromanischer, rätoromanischer und galloromanischer Sprachmaterialien aus AIS und ALF*. Tübingen: Niemeyer, 3 vol.
- Goebel 1985 GOEBL, Hans (1985): «Coup d'oeil dialectométrique sur les *Tableaux phonétiques des patois suisses romands (TPPSR)*». *VR* 44, p. 189-233.
- Goebel 1986 GOEBL, Hans (1986): «Typophilie und Typophobie. Zu zwei problembeladenen Argumentationstraditionen innerhalb der *Questione ladina*». Holtus Günter/Ringger, Kurt (eds.): *Raetia antiqua et moderna. Wilhelm Theodor Elwert zum 80. Geburtstag*. Tübingen: Niemeyer, p. 513-536.
- Goebel 1987a GOEBL, Hans (1987): «Points chauds de l'analyse dialectométrique: pondération et visualisation». *RLiR* 51, p. 63-118.
- Goebel 1987b GOEBL, Hans (1987): «Encore un coup d'oeil dialectométrique sur les *Tableaux phonétiques des patois suisses romands (TPPSR)*. Deux analyses interponctuelles: parquet polygonal et treillis triangulaire». *VR* 46, p. 91-125.
- Goebel 1990 GOEBL, Hans (1990): «"Ma il distintivo necessario del determinato tipo sta appunto nella simultanea presenza o nella particolare combinazione di quei caratteri". Methodische und wissenschaftsgeschichtliche Bemerkungen zum Diskussionskomplex "unità ladina"». *Ladinia* 14, p. 219-257.
- Goebel 1992a GOEBL, Hans (1992): «Problèmes et méthodes de la dialectométrie actuelle (avec application à l'AIS)». Euskaltzaindia/Académie de la langue basque (ed.): *Nazioarteko dialektologia biltzarra. Agiriak/ Actes du Congrès international de dialectologie (Bilbo/Bilbao 1991)*. Bilbo/Bilbao: Euskaltzaindia, p. 429-475.
- Goebel 1992b GOEBL, Hans (1992): «Dendrogramme im Dienst der Dialektometrie. Zwei hierarchisch-agglomerative Klassifikationen von Daten des Sprachatlases AIS». Klenk, Ur-

- sula (ed.): *Computatio linguae. Aufsätze zur algorithmischen und quantitativen Analyse der Sprache*. Stuttgart: Steiner, p. 54-73.
- Goebel 1993a GOEBL, Hans (1993): «Dialectometry. A Short Overview of the Principles and Practice of Quantitative Classification of Linguistic Atlas Data». Köhler, Reinhard/Rieger, Burghard (eds.): *Contributions of Quantitative Linguistics*. Dordrecht, Boston, Londres: Kluwer, p. 277-315.
- Goebel 1993b GOEBL, Hans (1993): «Eine hierarchisch-agglomerative Klassifikation von Daten des Sprachatlases AIS». Schmidt-Radefeldt, Jürgen/Harder, Andreas (eds.): *Sprachwandel und Sprachgeschichte. Festschrift für Helmut Lüdtke zum 65. Geburtstag*. Tübingen: Narr, p. 63-78.
- Goebel 1995 GOEBL, Hans (1995): «Che cos'è un geotipo? Il problema dell'unità ladina in chiave ascoliana». Banfi, Emanuele/Bonfadini, Giovanni/Cordin, Patrizia/Iliescu, Maria (eds.): *Italia settentrionale: crocevia di idiomi romanzi. Atti del Convegno internazionale di studi (Trento 1993)*. Tübingen: Niemeyer, p. 103-131.
- Goebel 1996 GOEBL, Hans (1996): «La convergence entre les fragmentations géo-linguistique et géo-génétique de l'Italie du Nord». *RLiR* 60, p. 25-49.
- Goebel 1998a GOEBL, Hans (1998): «Dialektometrische Beschreibung der Romania». *LRL* VII, p. 977-1003.
- Goebel 1998b GOEBL, Hans (1998): «Zu einer dialektometrischen Analyse der Daten des Dees-Atlas von 1980». Werner, Edeltraud/Liver, Ricarda/Storck, Yvonne/Nicklaus, Werner (eds.): *Et multum et multa. Festschrift für Peter Wunderli zum 60. Geburtstag*. Tübingen: Narr, p. 293-309.
- Goebel 2000a GOEBL, Hans (2000): «La dialectométrisation de l'ALF: présentation des premiers résultats». *Linguistica* 40, p. 209-236.
- Goebel 2000b GOEBL, Hans (2000): «Langues standards et dialectes locaux dans la France du sud-est et l'Italie septentrionale sous le coup de l'effet-frontière: une approche dialectométrique». *International Journal of the Sociology of Language* 145, p. 181-215.
- Goebel 2001 GOEBL, Hans (2001): «Dialektometrie». *LRL* I, 2, p. 856-874.
- Goebel 2002 GOEBL, Hans (2002): «Analyse dialectométrique des structures de profondeur de l'ALF». *RLiR* 66, p. 5-63.
- Gordon 1999 GORDON, A. D. (1999): *Classification*. Boca Raton/Londres/New York/Washington D. C.: Chapman & Hall.
- Grassi 2001 GRASSI, Corrado (2001): «La geografia linguistica». *LRL* I, 1, p. 207-235.
- Haag 1898 HAAG, Carl (1898): *Die Mundarten des oberen Neckar- und Donautales (Schwäbisch-alemannisches Grenzgebiet: Baarmundarten)*. Reutlingen: Hutzler.
- Hempfer 1973 HEMPFER, Klaus W. (1973): *Gattungstheorie*. München: Fink.
- Hoenigswald/Wiener 1987 HOENIGSWALD, Henry M./WIENER, Linda F. (eds.) (1987): *Biological Metaphor and Cladistic Classification. An Interdisciplinary Perspective*. Philadelphia: University of Pennsylvania Press.
- Horning 1903 HORNING, Adolf (1893): «Über Dialektgrenzen im Romanischen». *ZrPh* 17, p. 160-187 (réimpression: Spitzer, Leo (ed.) (1930): *Meisterwerke der romanischen Sprachwissenschaft*. Munich: Hueber, vol. II, p. 264-298).
- Kreyszig 1968 KREYSZIG, Erwin (1968): *Statistische Methoden und ihre Anwendungen*. Göttingen.
- Lalanne 1953 LALANNE, Théodore (1953): «Indice de polyonymie, indice de polyphonie». *FM* 21, p. 263-274.
- LRL HOLTUS, Günter/METZELTIN, Michael/SCHMITT, Christian (eds.) (1988 ss.): *Lexikon der Romanistischen Linguistik*. Tübingen: Niemeyer.

- Mahmood/Armstrong 1992 MAHMOOD, Cynthia K./ARMSTRONG, Sharon L. (1992): «Do Ethnic Groups exist? A Cognitive Perspective on the Concepts of Cultures». *Ethnology* 31, p. 1-14.
- Meyer 1875 MEYER, Paul (1875): «compte rendu de: Ascoli 1874». *Romania* 4, p. 293-296.
- Moriconi 2000 MORICONI, Anna (2000): *La classificazione nelle scienze sociali*. Formello: Seam.
- Needham 1975 NEEDHAM, Rodney (1975): «Polythetic Classifications: Convergence and Consequences». *Man* 10, p. 349-369.
- Nordman 1998 NORDMAN, Daniel (1998): *Frontières de France. De l'espace au territoire. XVI<sup>e</sup> - XIX<sup>e</sup> siècle*. Paris: Gallimard.
- Okabe/Boots/Sugihara 1992 OKABE, Atsuyuki/BOOTS, Barry/SUGIHARA, Kokichi (1992): *Spatial Tessellations. Concepts and Applications of Voronoi Diagrams*. Chichester/New York/Brisbane/Toronto/Singapour: Wiley & Sons.
- Paris 1881 PARIS, Gaston (1881): «compte rendu de: Adam, Lucien: Les patois lorrains, Paris 1881». *Romania* 10, p. 601-609.
- Paris 1888 PARIS Gaston (1888): «Les parlers de France». PARIS, Gaston (1909): *Mélanges linguistiques*. Paris, p. 432-448.
- Parrochia 1991 PARROCHIA, Daniel (ed.) (1991): *Mathématiques & Existence. Ordres. Fragments. Empiètements*. Seyssel: Champ Vallon.
- Pop 1950 POP, Sever (1950): *La dialectologie. Aperçu historique et méthodes d'enquêtes linguistiques*. Louvain: Chez l'auteur /Gembloux: Duculot, 2 vol.
- Ronjat 1913 RONJAT, Jules (1913): *Essai de syntaxe des parlers provençaux modernes*. Mâcon: Protat Frères.
- Ronjat 1930 RONJAT, Jules (1930/41): *Grammaire Istorique [sic] des Parlers Provençaux modernes*. Montpellier: Société des langues romanes, 4 vol. (réimpression: Genève: Slatkine/Marseille: Laffitte, 1980).
- Rosch/Lloyd (1978) ROSCH, Eleanor/LLOYD, Barbara B. (eds.) (1978): *Cognition and Categorization*. Hillsdale N. J.: Lawrence Erlbaum.
- Rosenqvist 1919 ROSENQVIST, Arvid (1919): «Limites administratives et division dialectale de la France». *NphM* 20, p. 87-119.
- Ruhlen 1994 RUHLEN, Merritt (1994): *On the Origin of Languages. Studies in Linguistic Taxonomy*. Stanford, CA: Stanford University Press.
- Sankoff 1973 SANKOFF, David (1973): «Mathematical Developments in Lexicostatistic Theory». *Current Trends in Linguistics* 11, p. 93-113.
- Schleicher 1863 SCHLEICHER, August (1863): *Die Darwinsche Theorie und die Sprachwissenschaft. Offenes Sendschreiben an Herrn Dr. Ernst Hückel, a. o. Professor der Zoologie und Director des zoologischen Museums an der Universität Jena*. Weimar: Böhlau.
- Séguy 1973 SÉGUY, Jean (1973): «La dialectométrie dans l'Atlas linguistique de la Gascogne». *RLiR* 37, p. 1-24.
- Sériot 1999 SÉRIOT, Patrick (1999): *Structure et totalité. Les origines intellectuelles du structuralisme en Europe centrale et orientale*. Paris: Presses Universitaires de France.
- Sneath/Sokal 1973 SNEATH, Peter A./SOKAL, Robert R. (1973): *Numerical Taxonomy. The Principles and Practice of Numerical Classification*. San Francisco: Freeman.
- Snow 1993 SNOW, Charles Percy (1993): *The Two Cultures: with an introduction by Stefan Collini*. Cambridge: Cambridge University Press (édition originale: ibidem, 1959).
- Stegmüller 1956/57 STEGMÜLLER, Wolfgang (1956/57): «Das Universalienproblem einst und jetzt». *Archiv für Philosophie* 6, p. 192-225 et 7, p. 45-81 (réimpression: Darmstadt: Wissenschaftliche Buchgesellschaft, 1965).
- Suchier 1888 SUCHIER, Hermann (1888): «Die französische und provençalische Sprache und

- ihre Mundarten». Gröber, Gustav (ed.): *Grundriß der romanischen Philologie*. Strasbourg: Trübner, vol. I, p. 561-668.
- Tourtoulon/Bringuier 1876 TOURTOULON, Charles de/BRINGUIER, Octavien (1876): «Rapport sur la limite géographique de la langue d'oc et de la langue d'oïl [sic]». *Archives des Missions scientifiques et littéraires* III/3, p. 544-605, 1 carte.
- Wüest 1981 WÜEST, Jakob (1981): «Absence d'«Intercourse» ou «Esprit de clocher». À propos des zones frontalières du domaine francoprovençal». *VR* 40, p. 13-21.

## RÉSUMÉ

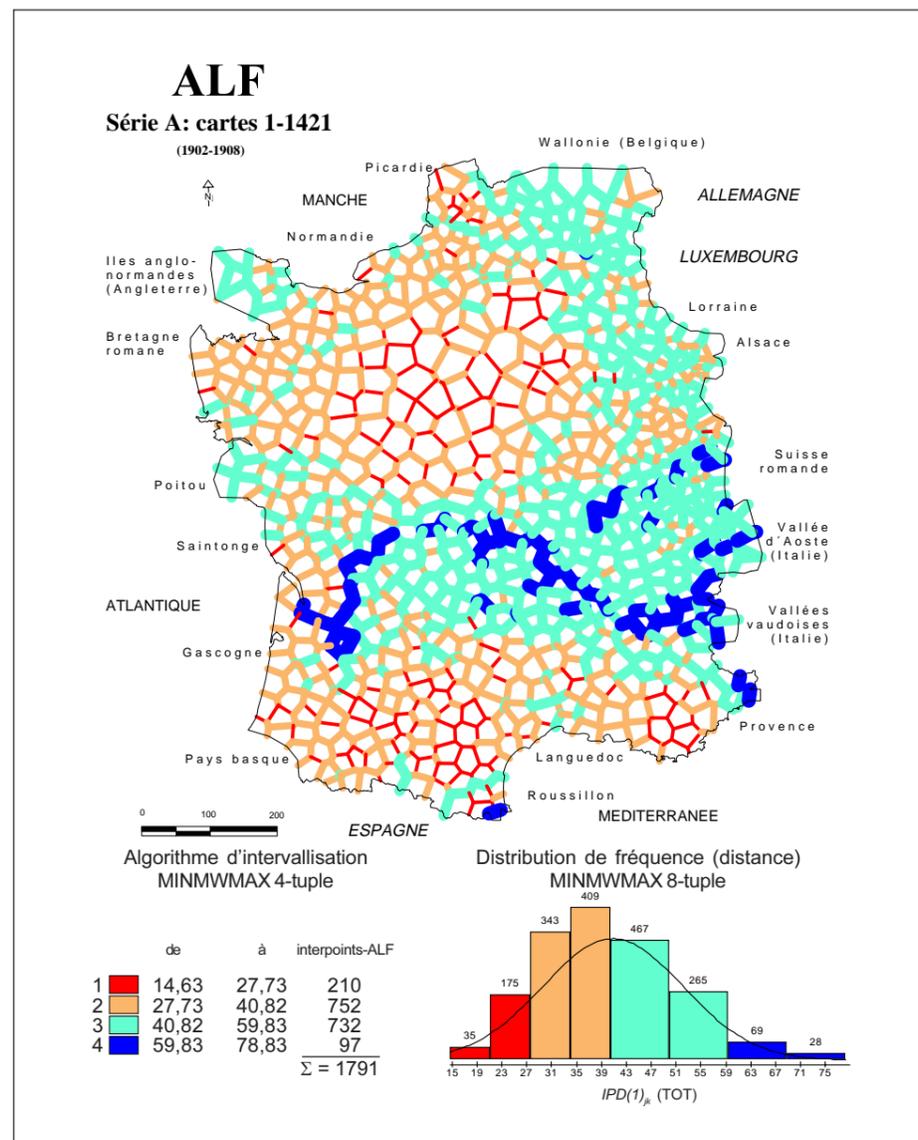
L'article décrit, à l'aide de données géolinguistiques tirées de l'Atlas linguistique de la France de Jules Gilliéron et Edmond Edmont (*ALF*), la démarche théorique et méthodique ainsi que les possibilités typodiagnostiques de la dialectométrie (DM) moderne dont le but central est la découverte de structures diatopiques de profondeur qui normalement restent cachées à l'observation directe de la part du géolinguiste. Le texte, pourvu de 5 figures en noir-et-blanc et de 24 cartes en couleurs (dont chacune occupe une page entière), fournit une introduction détaillée aux procédés dialectométriques suivants: DM *interponctuelle* (mensuration des distances interponctuelles), DM de *similarité* (mensuration des similarités entre tous les points-*ALF* considérés), DM des *paramètres caractéristiques* et DM *dendrographique*. Il en résulte une vision globale très détaillée de la genèse et de la structure actuelle de l'espace gallo-roman tel qu'il est représenté dans le maillage des 638 points d'enquête et les transcriptions répertoriées sur les 626 planches de l'*ALF* prises en considération.

MOTS CLÉS: dialectologie galloromane, géolinguistique, dialectométrie, classification numérique, Atlas linguistique de la France (*ALF*), géographie linguistique quantitative, géotypologie.

## ABSTRACT

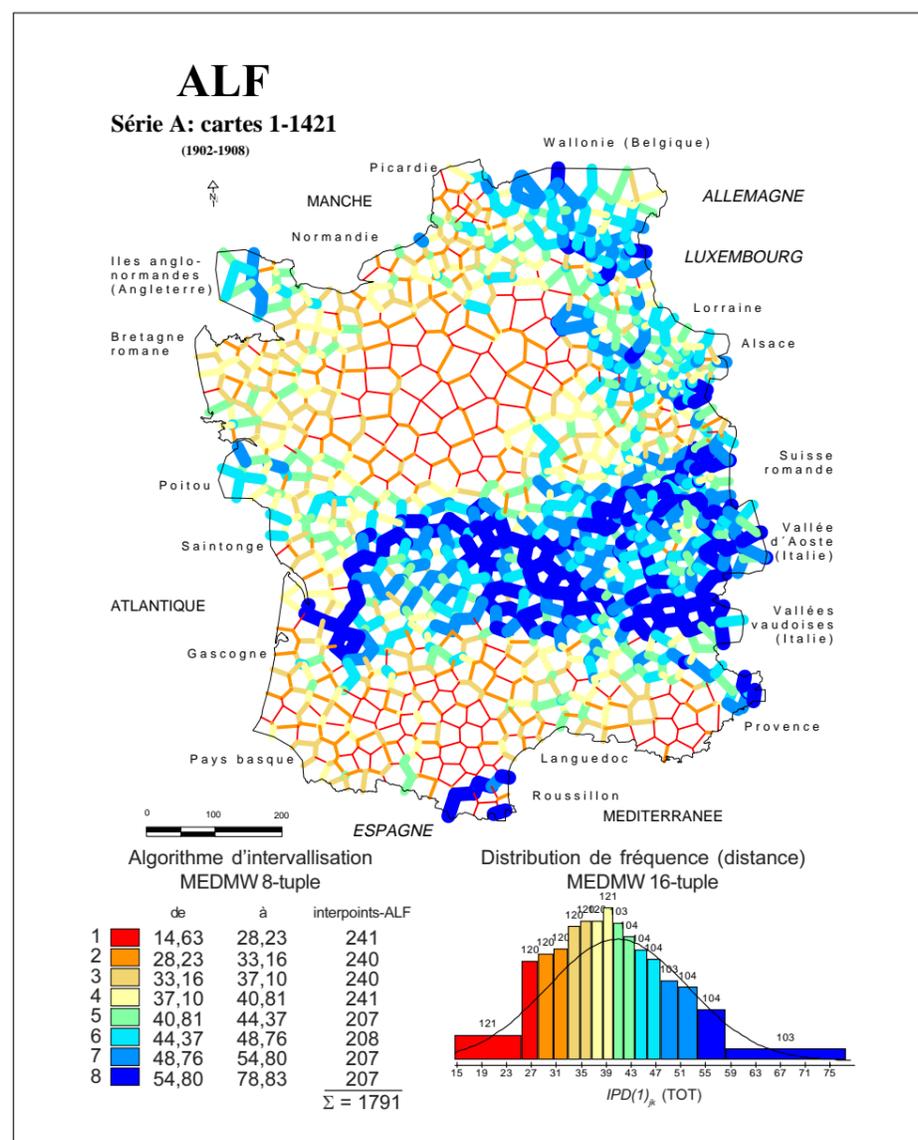
This paper describes, with the help of geolinguistic data drawn from the linguistic Atlas of France by Jules Gilliéron and Edmond Edmont (*ALF*), the theoretical and methodological approach and the typodiagnostic possibilities of modern dialectometry (DM), whose central aim is to discover the underlying diatopic structures that geolinguists are normally not able to observe directly. The paper has 5 black-and-white figures and 24 full-page colour maps, and it provides a detailed introduction to the following procedures in dialectometry: *interpoint* DM (measuring distances between points), *similarity* DM (measuring the similarities between the all the *ALF* points taken into account), *characteristic-parameter* DM, and *dendrographic* DM. The result is a very detailed overview of the genesis and current structure of the Gallo-Romance area as it appears from cross-linking the 638 survey points and the transcriptions identified in the 626 plates of the *ALF* taken into account.

KEY WORDS: Gallo-Romance dialectology, geolinguistics, dialectometry, numerical classification, Atlas linguistique de la France (*ALF*), quantitative linguistic geography, geotypology.



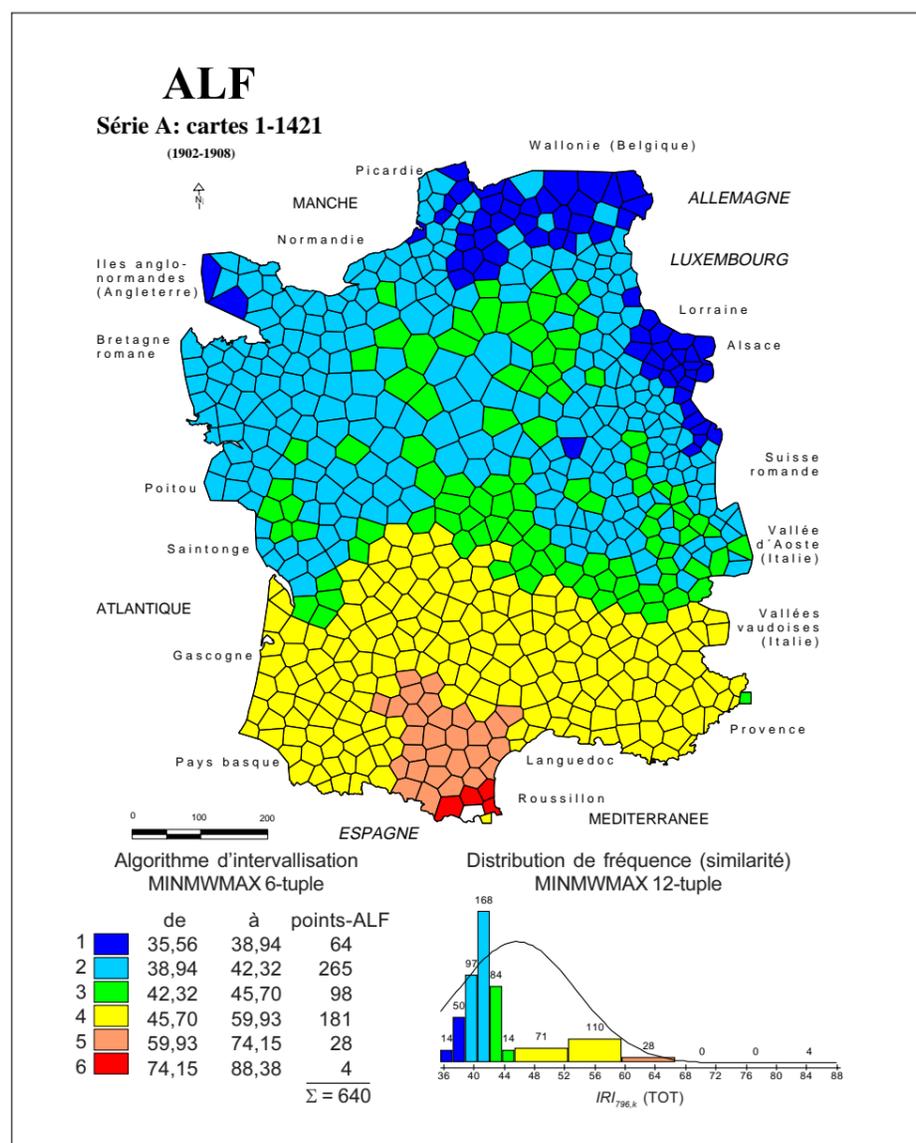
CARTE 1: Carte à interpoints en fonction discriminatoire (isoglosses quantitatives).

Corpus: 1687 cartes de travail (données intégrales). Indice de distance:  $IPD(1)_k$ . Algorithme d'intervallisation: MINMWMAX 4-tuple. Nombre des paliers d'épaisseur variable des segments de polygone: 4. Voir aussi le paragraphe 5.2.



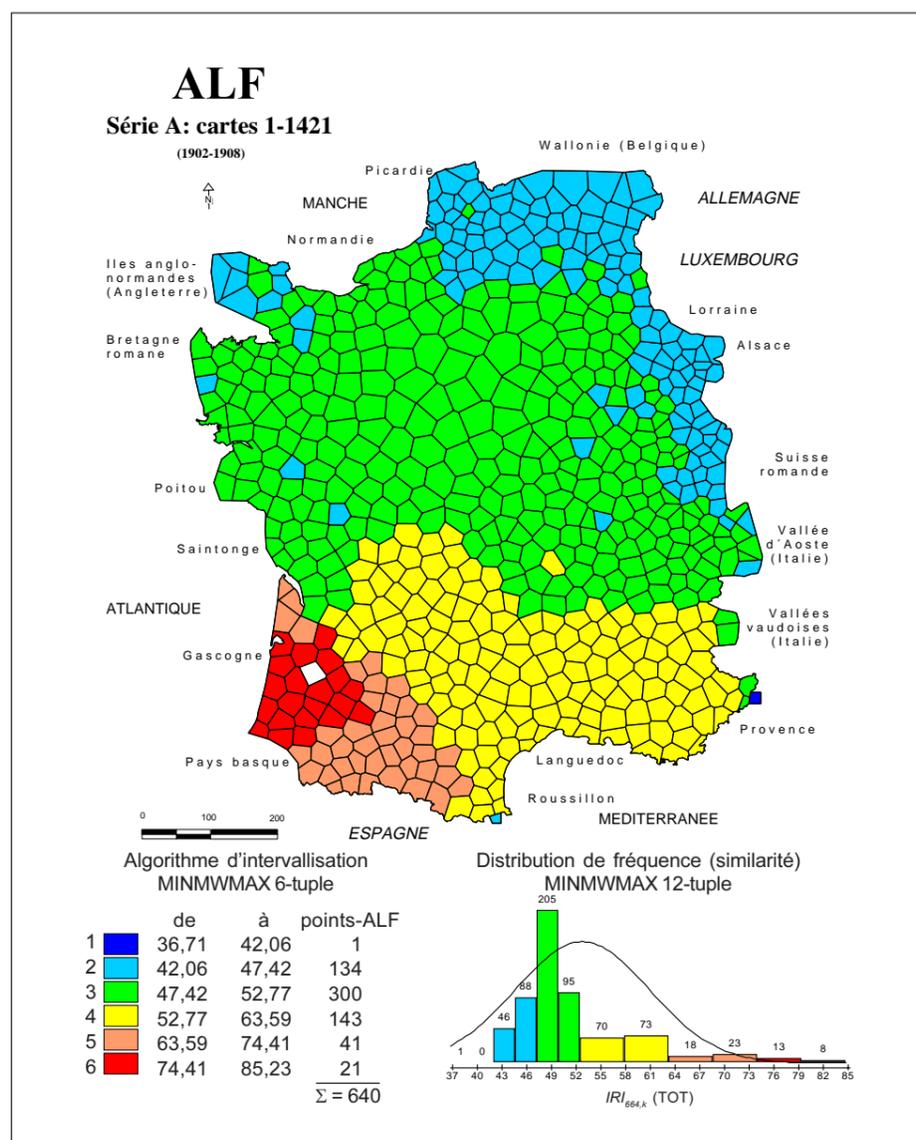
CARTE 2: Carte à interpoints en fonction discriminatoire (isoglosses quantitatives).

Corpus: 1687 cartes de travail (données intégrales). Indice de distance:  $IPD(1)_k$ . Algorithme d'intervallisation: MEDMW 8-tuple. Nombre des paliers d'épaisseur variable des segments de polygone: 8. Voir aussi le paragraphe 5.2.



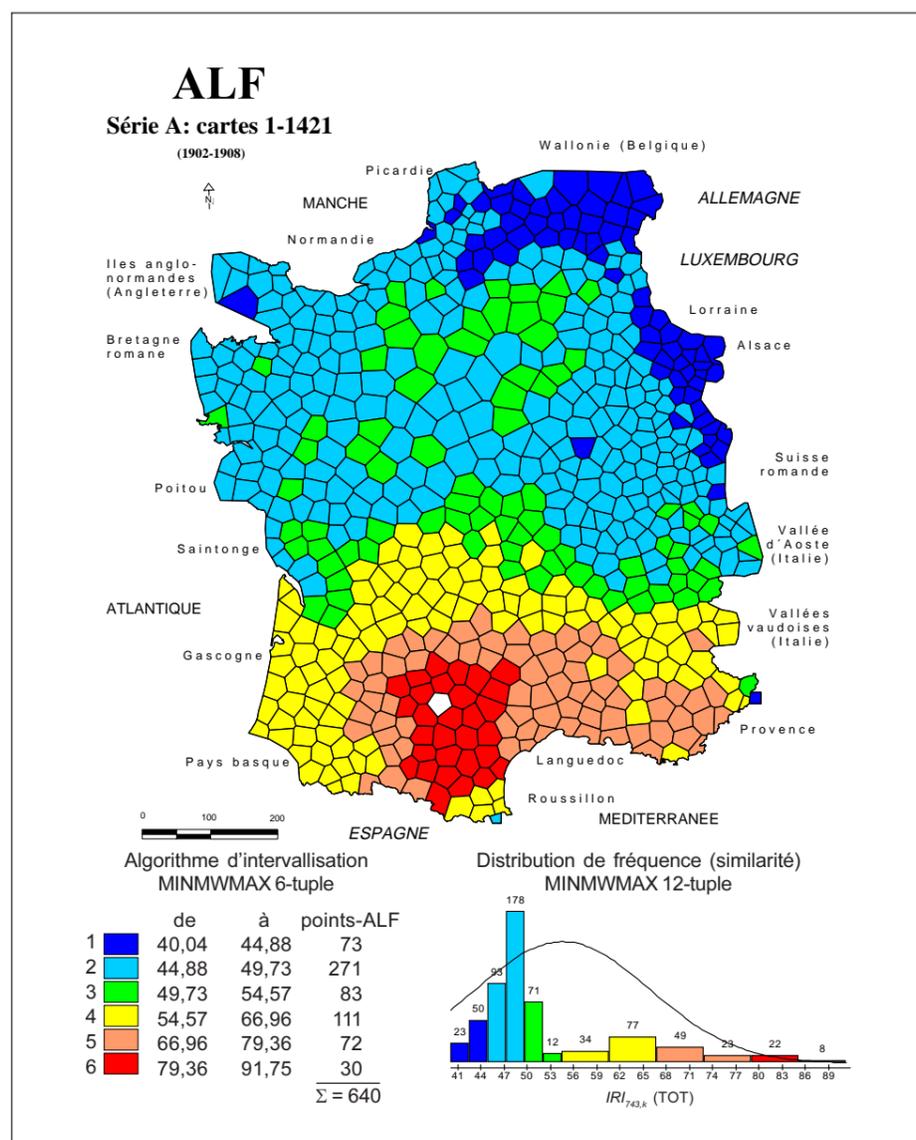
CARTE 3: Carte choroplèthe de la distribution de similarité ( $IRI_{796,k}$ ) relative au point de référence 796, Arles-sur-Tech (Pyrénées Orientales).

Corpus: 1687 cartes de travail (données intégrales). Indice de similarité:  $IRI_{jk}$ . Algorithme d'intervallisation: MINMWMAX 6-tuple. Nombre des paliers chromatiques: 6. Voir aussi le paragraphe 6.2.



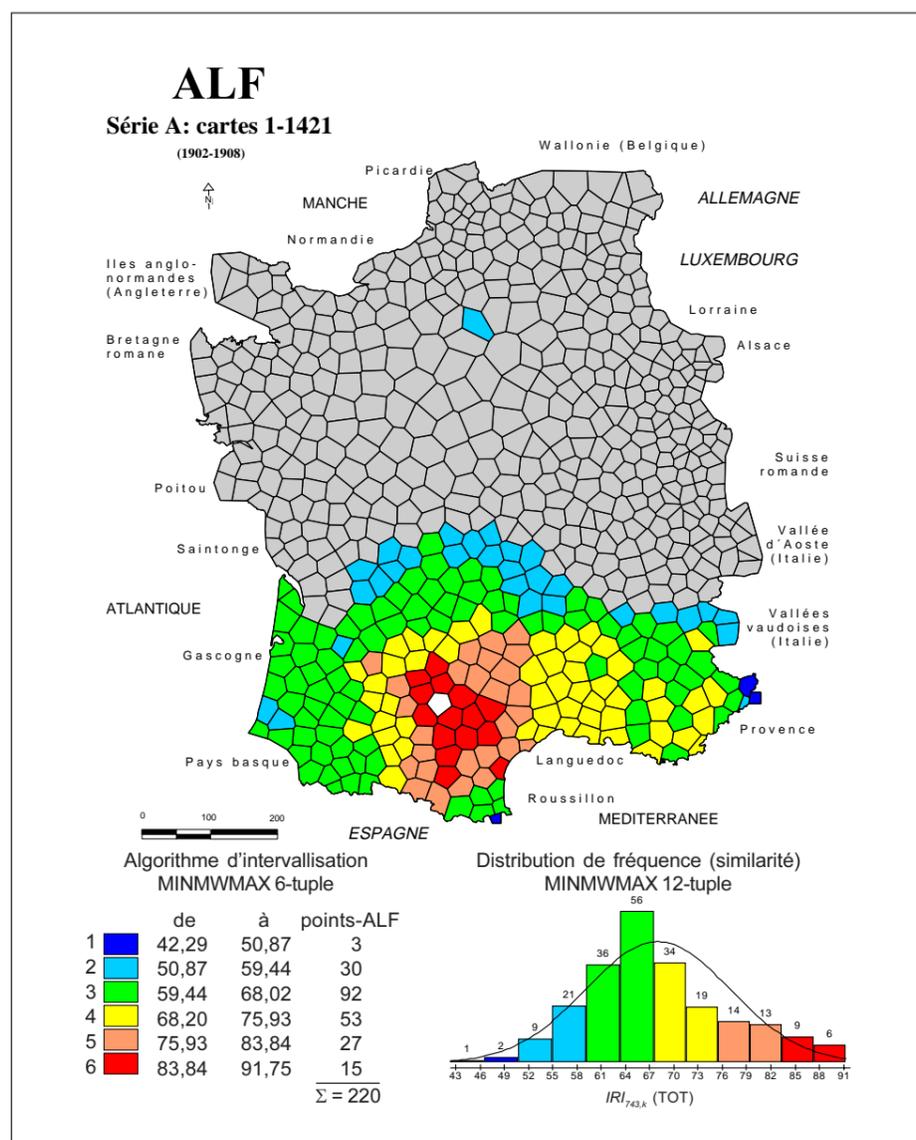
CARTE 4: Carte choroplèthe de la distribution de similarité ( $IRI_{664,k}$ ) relative au point de référence 664, Luxey (Landes).

Corpus: 1687 cartes de travail (données intégrales). Indice de similarité:  $IRI_{jk}$ . Algorithme d'intervallisation: MINMWMAX 6-tuple. Nombre des paliers chromatiques: 6. Voir aussi le paragraphe 6.2.



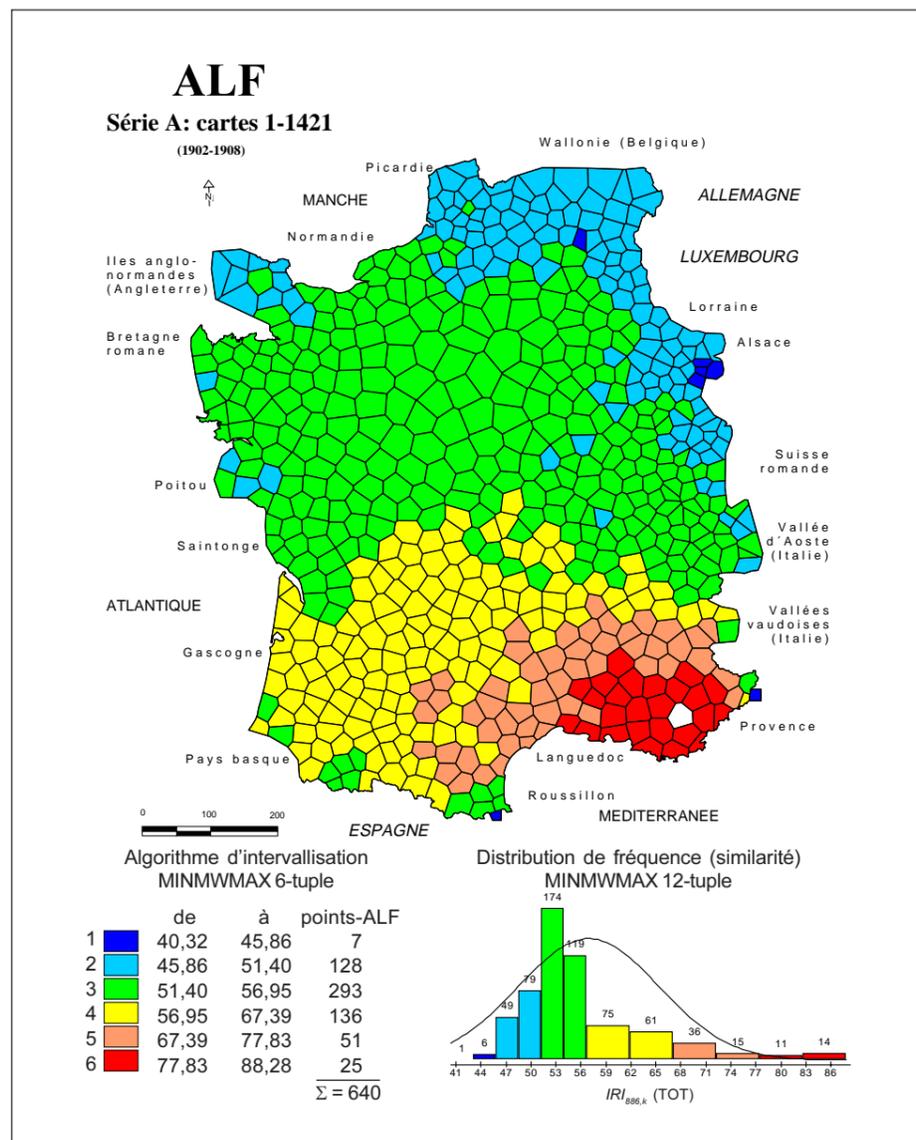
CARTE 5: Carte choroplèthe de la distribution de similarité ( $IRI_{743,k}$ ) relative au point de référence 743, Gaillac (Tarn).

Corpus: 1687 cartes de travail (données intégrales). Indice de similarité:  $IRI_{jk}$ . Algorithme d'intervallisation: MINMWMAX 6-tuple. Nombre des paliers chromatiques: 6. Voir aussi le paragraphe 6.3.

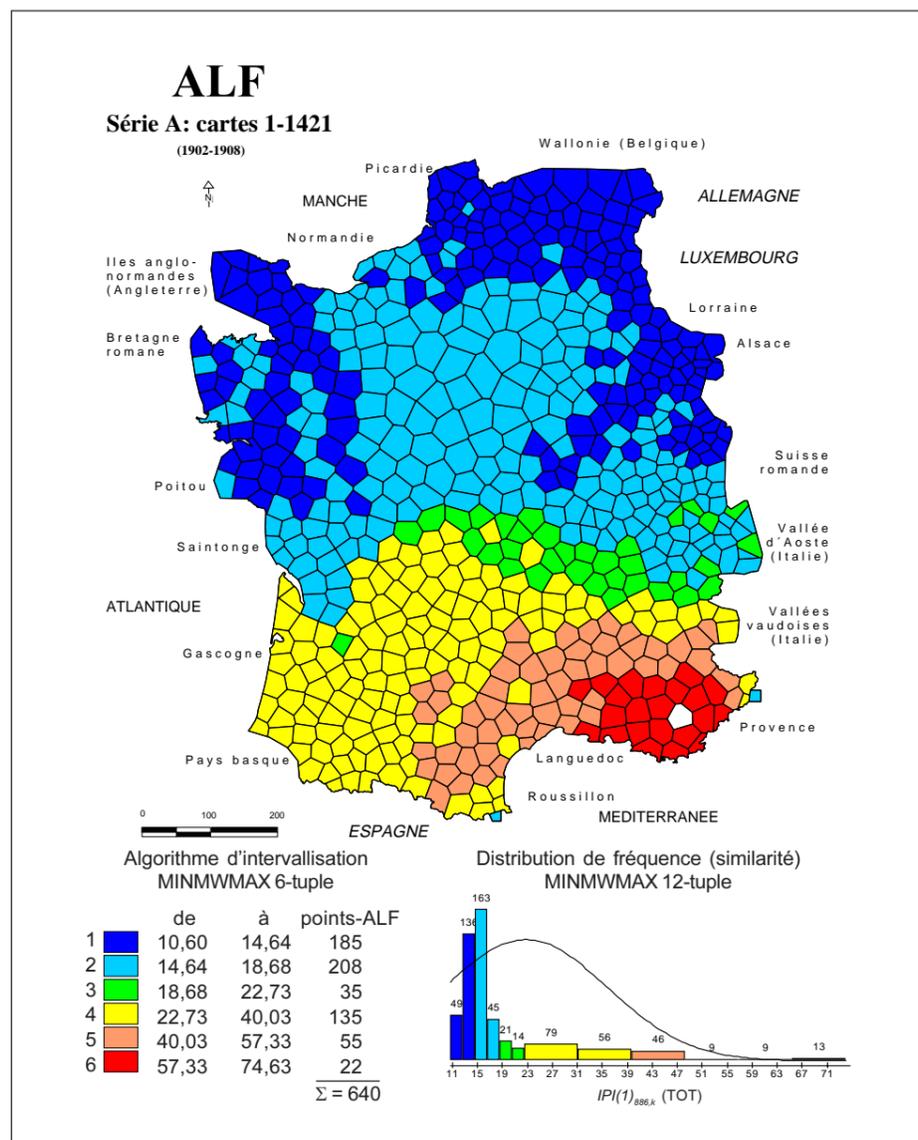


CARTE 6: Carte choroplèthe de la distribution de similarité ( $IRI_{743,k}$ ) relative au point de référence 743, Gaillac (Tarn).

Corpus: 1687 cartes de travail (données intégrales); moitié sud du réseau-ALF. Indice de similarité:  $IRI_{jk}$ . Algorithme d'intervallisation: MINMWMAX 6-tuple. Nombre des paliers chromatiques: 6. Voir aussi le paragraphe 6.3.



CARTE 7: Carte choroplèthe de la distribution de similarité ( $IRI_{886,k}$ ) relative au point de référence 886, Aups, (Var).  
Corpus: 1687 cartes de travail (données intégrales). *Indice de similarité*:  $IRI_{j,k}$ . *Algorithme d'intervallisation*:  
MINMWMAX 6-tuple. *Nombre des paliers chromatiques*: 6. Voir aussi le paragraphe 6.4.



CARTE 8: Carte choroplèthe de la distribution de similarité ( $IRI_{886,k}$ ) relative au point de référence 886, Aups, (Var).  
Corpus: 1687 cartes de travail (données intégrales). Indice de similarité:  $IPI(1)_k$ . Algorithme d'intervallisation:  
MINMWMAX 6-tuple. Nombre des paliers chromatiques: 6. Voir aussi le paragraphe 6.4.

# ALF

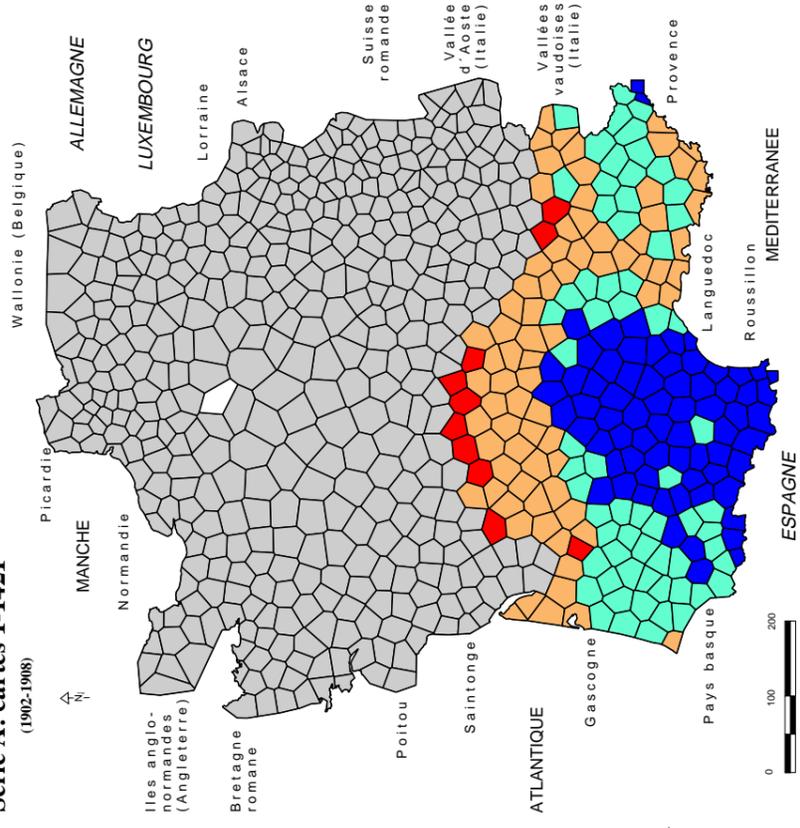
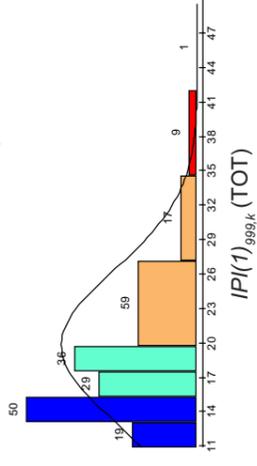
Série A : cartes 1-1421

(1902-1908)

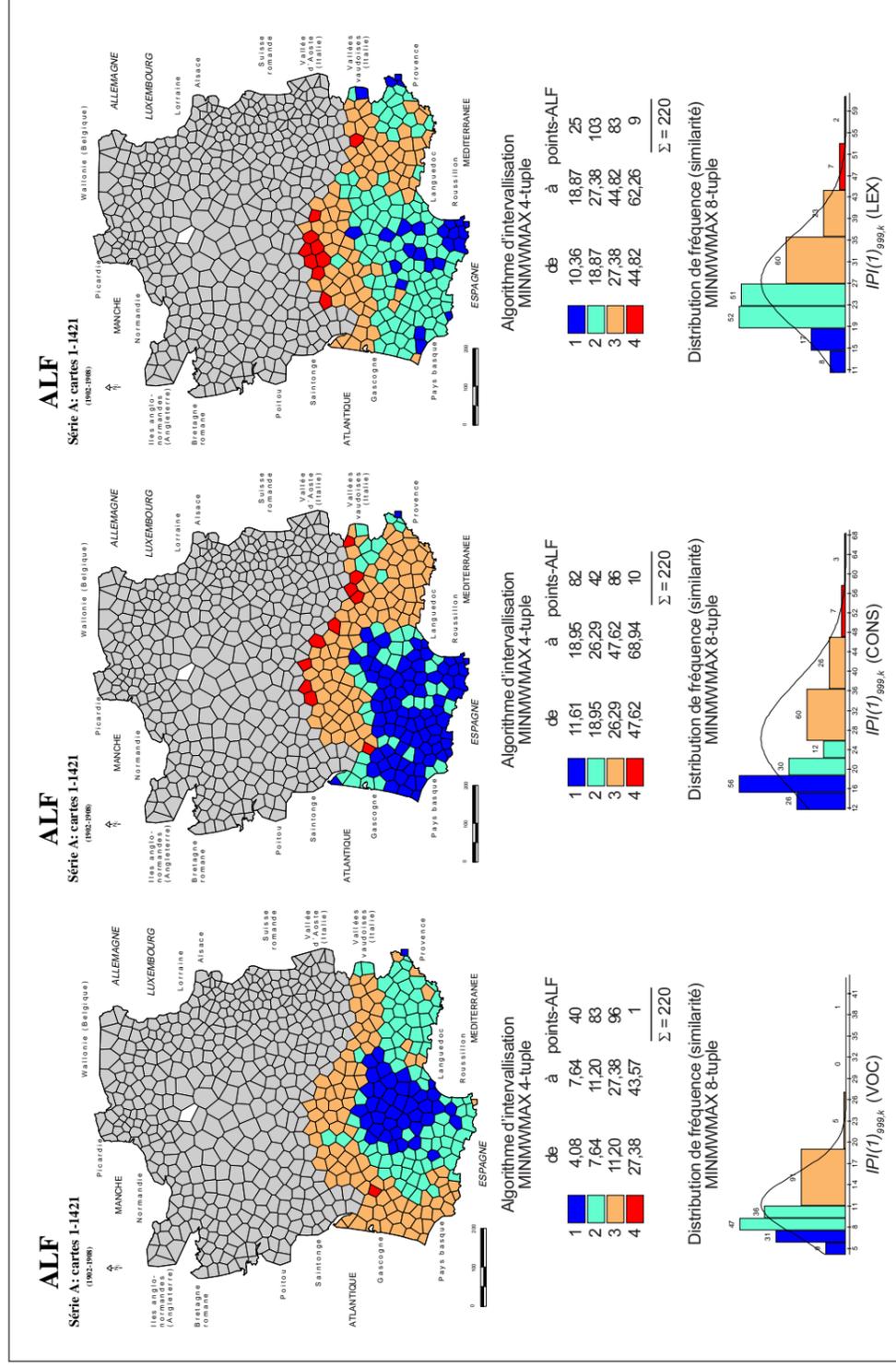
Algorithme d'intervallisation  
MINMWMAX 4-tuple

de	à	points-ALF	
1	10,87	15,32	69
2	15,32	19,77	65
3	19,77	34,82	76
4	34,82	49,88	10
		$\Sigma = 220$	

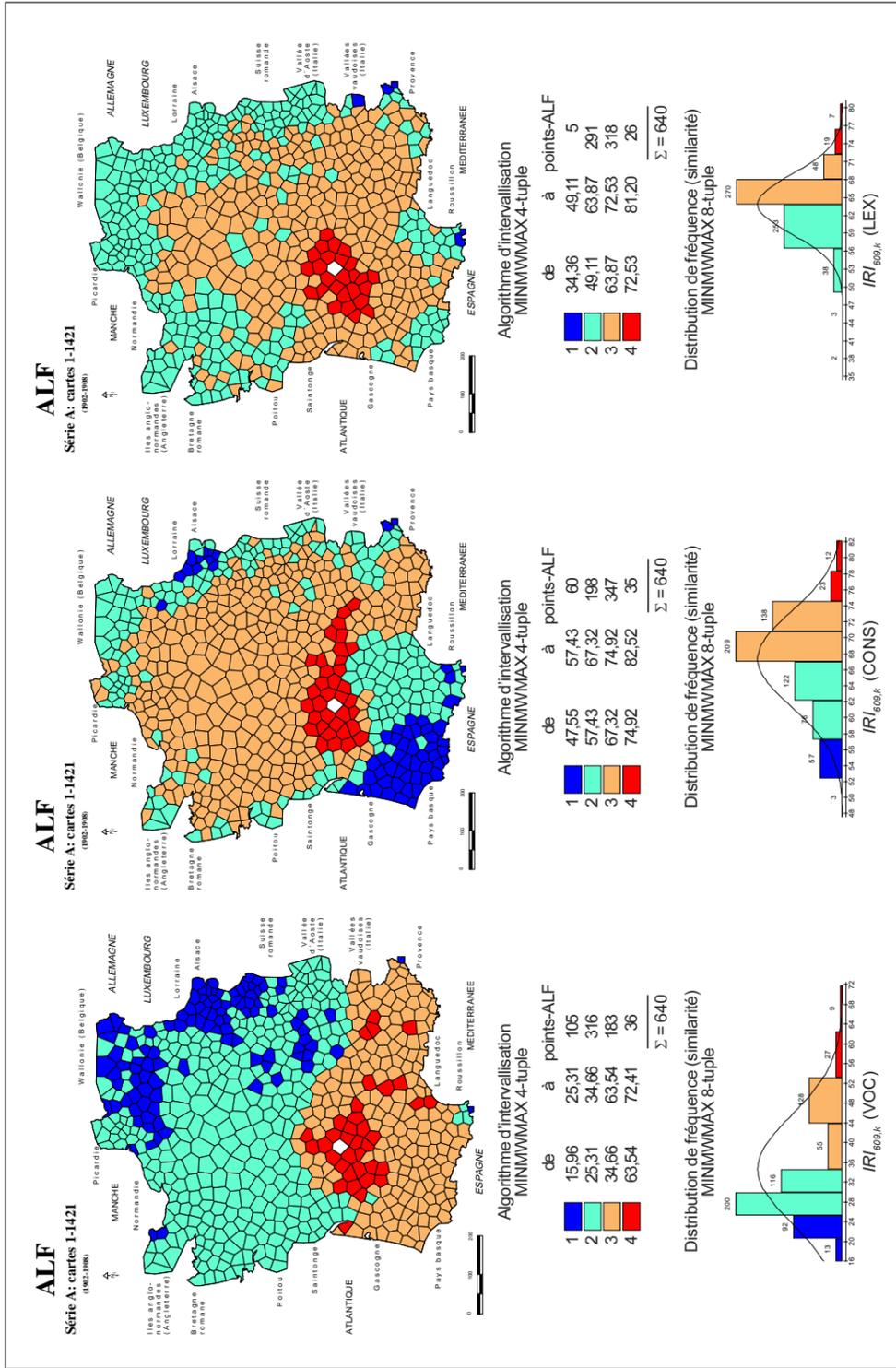
Distribution de fréquence (similarité)  
MINMWMAX 8-tuple



CARTE 9: Carte choroplèthe de la distribution de similarité ( $IPI(1)_{999,k}$ ) relative au point de référence 999, français standard [en blanc, au milieu de la zone grise].  
Corpus: 1687 cartes de travail (données intégrales); moitié sud du réseau-ALF. *Indice de similarité*:  $IPI(1)_{jk}$ . *Algorithme d'intervallisation*: MINMWMAX 4-tuple. *Nombre des paliers chromatiques*: 4. Voir aussi le paragraphe 6.5.

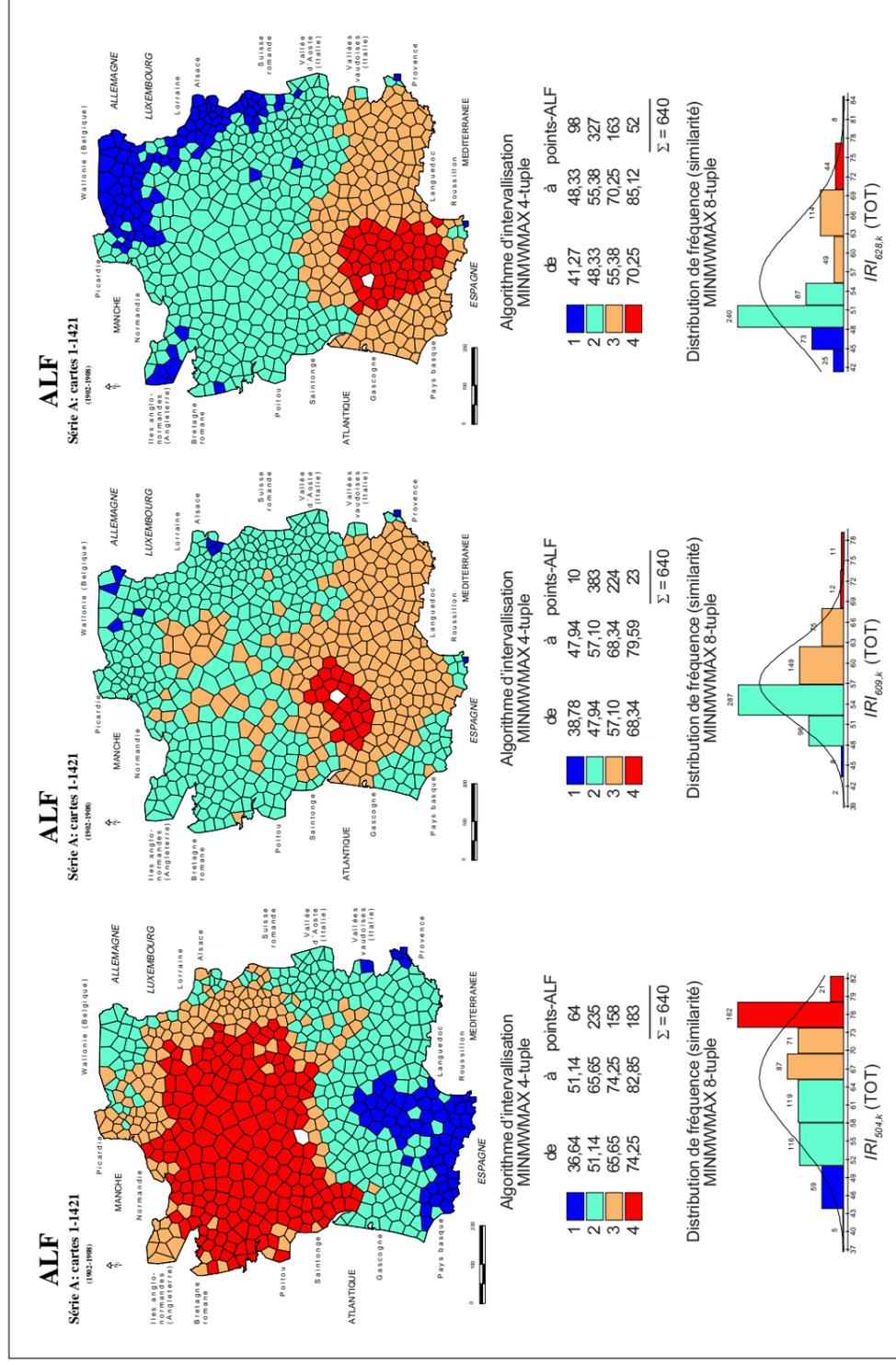


CARTE 10: Cartes choroplèthes de trois distributions de similarité ( $IP(1)_{999,k}$ ) relatives au point de référence 999, français standard [en blanc, au milieu de la zone grise].  
Corpora: *Vocalisme*: 612, *consonantisme*: 505, *lexique*: 471 cartes de travail; moitié sud du réseau-ALF. *Indice de similarité*:  $IP(1)_{999,k}$ . *Algorithme d'intervallisation*: MINMWMAX 4-tuple.  
Nombre des paliers chromatiques: 4. Voir aussi le paragraphe 6.5.



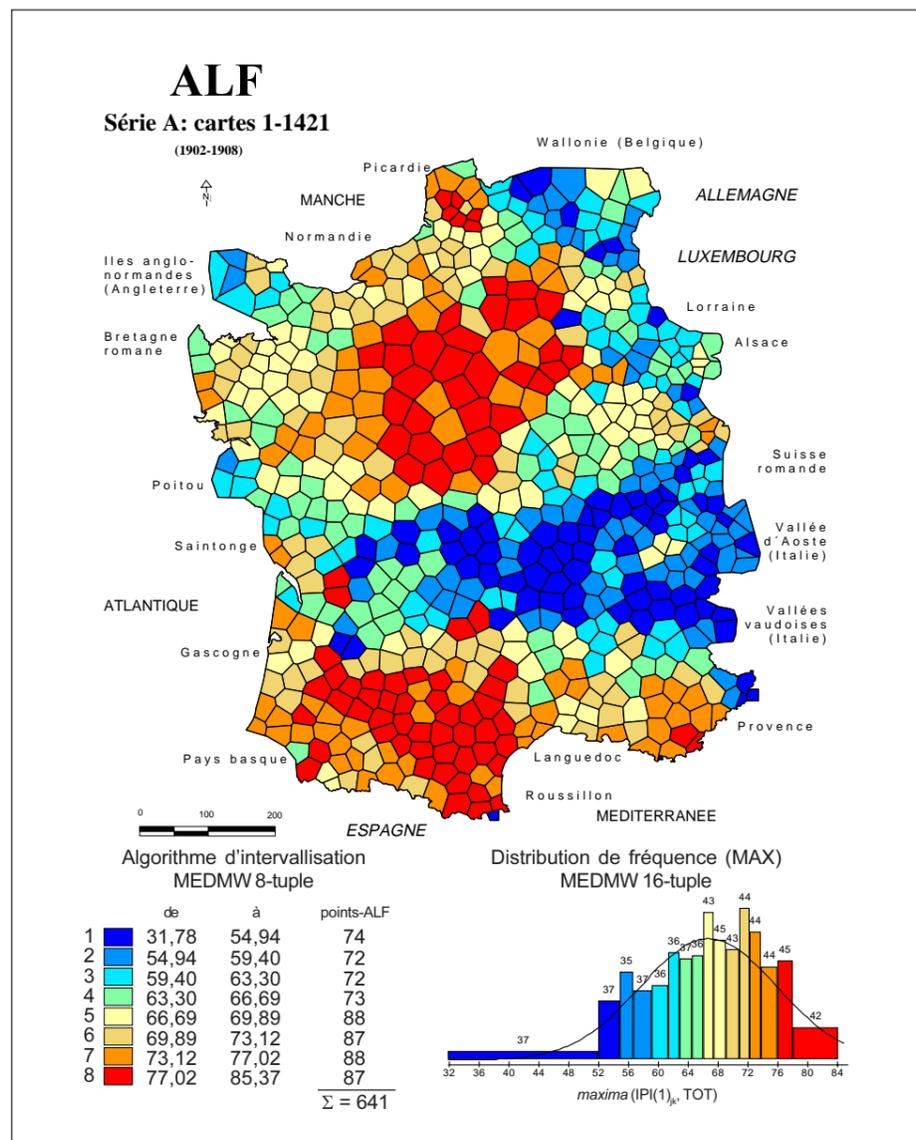
CARTE 11: Cartes choroplèthes de trois distributions de similarité (IRI<sub>609,k</sub>) relatives au point de référence 609, Seilhac (Corrèze). *Corpora: Vocalisme: 612, consonantisme: 505, lexique: 471* cartes de travail.

Indice de similarité: IRI<sub>k</sub>. *Algorithme d'intervallisation: MINMMWMAX 4-tuple. Nombre des paliers chromatiques: 4. Voir aussi le paragraphe 6.6.*

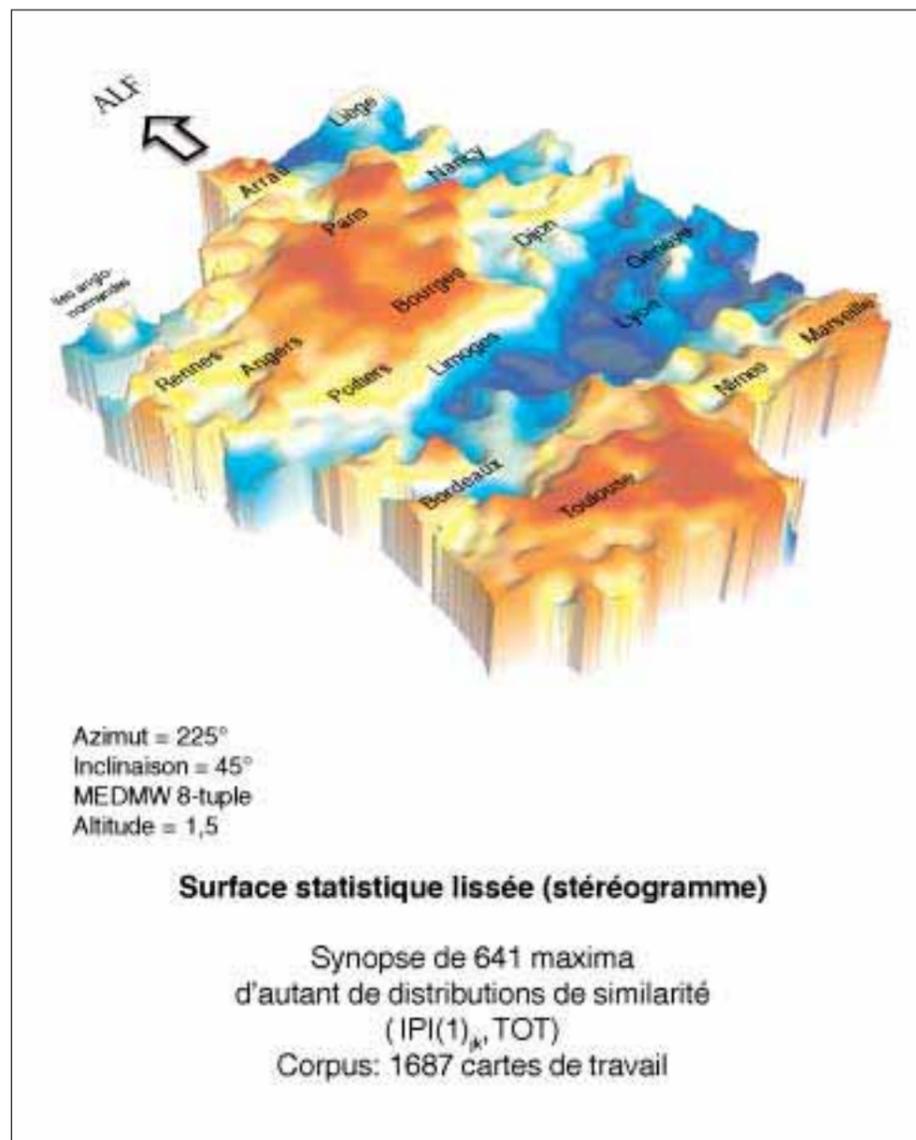


CARTE 12: Cartes choroplèthes de trois distributions de similarité relatives au point de référence 504, Dun-le-Palleteau (Creuse), 609, Seilhac (Corrèze) et 628, Villefranche-de-Belvès (Dordogne).

Corpus: 1687 cartes de travail (données intégrales). *Indice de similarité*: IRI<sub>k</sub>. *Algorithme d'intervallisation*: MINMMWMAX 4-tuple. *Nombre des patiers chromatiques*: 4. Voir aussi le paragraphe 6.7.

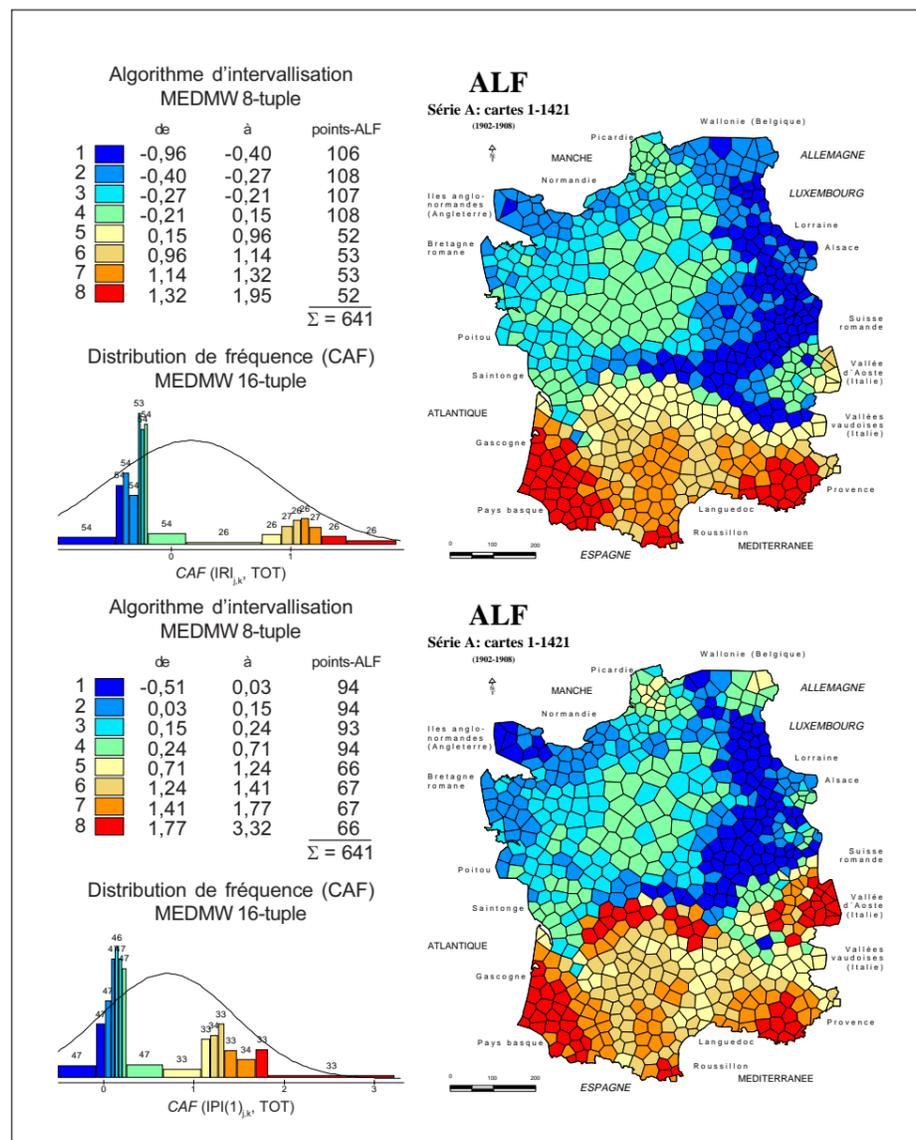
CARTE 13: Carte choroplèthe de la synopse des maxima de 641 distributions de similarité (IPI(1)<sub>jk</sub>).

Corpus: 1687 cartes de travail (données intégrales). Indice de similarité: (IPI(1)<sub>jk</sub>). Algorithme d'intervallisation: MEDMW 8-tuple. Nombre des paliers chromatiques: 8. Voir aussi le paragraphe 7.1.



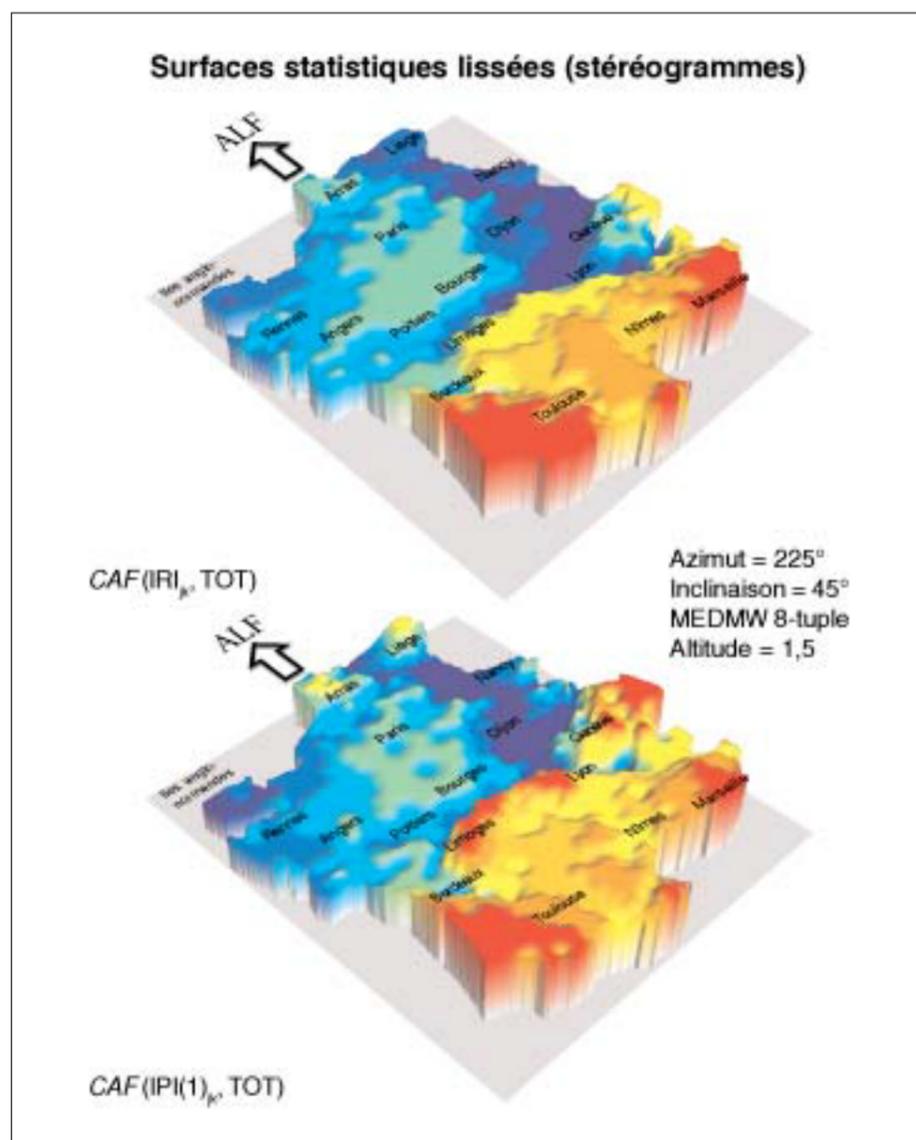
CARTE 14: Surface statistique lissée (stéréogramme) de la synopse des maxima de 641 distributions de similarité (IPI(1)<sub>μ</sub>).

Mêmes données que dans la carte 13. *Point de vue de l'observateur*: à partir du sud-ouest (225°). *Angle d'élévation de l'observateur*: 45°. Voir aussi le paragraphe 7.1.



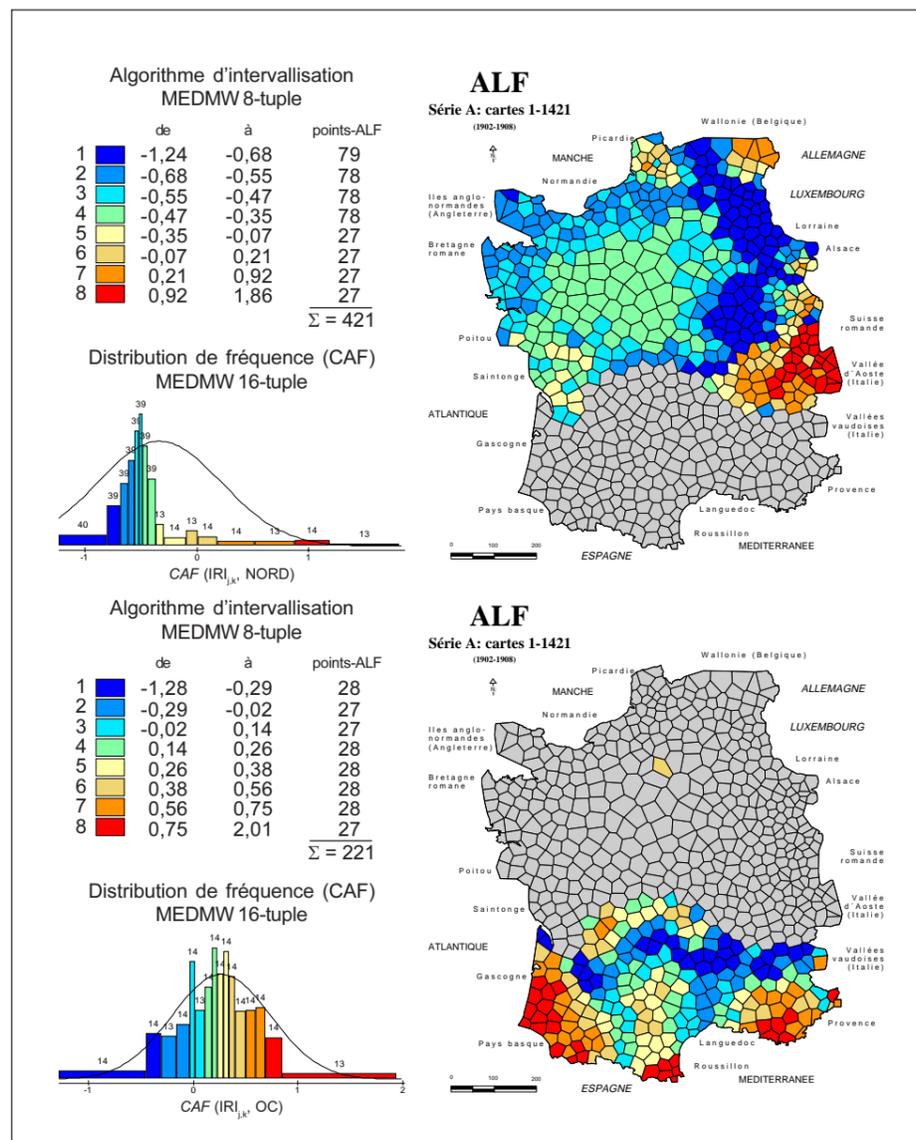
CARTE 15: Cartes choroplèthes de deux synopses de 641 coefficients d'asymétrie de Fisher (CAF).

Corpus: 1687 cartes de travail (données intégrales). Indices de similarité: en haut:  $IRI_{jk}$ ; en bas:  $IPI(1)_{jk}$ . Algorithme d'intervallisation: MEDMW 8-tuple. Nombre des paliers chromatiques: 8. Voir aussi le paragraphe 7.2.1.



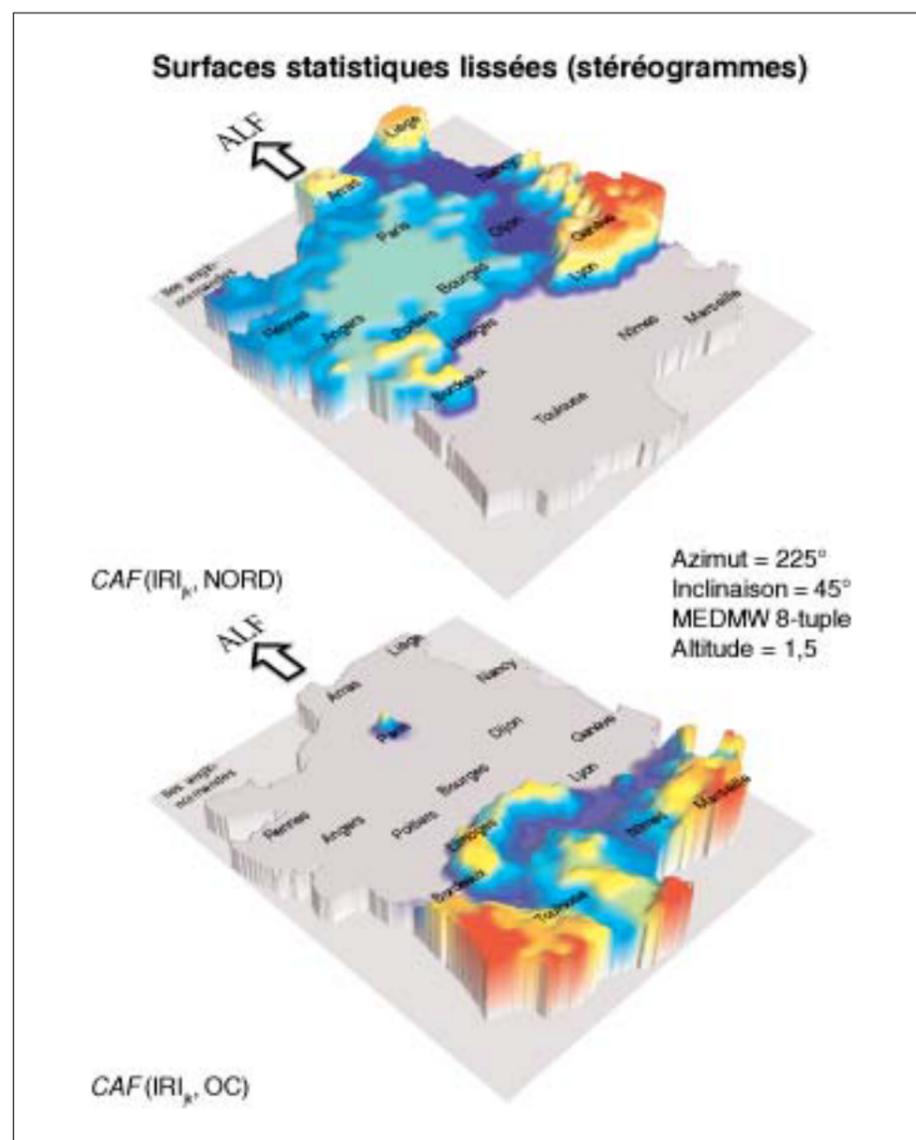
CARTE 16: Surfaces statistiques lissées (stéréogrammes) de deux synopses de 641 coefficients d'asymétrie de Fisher (CAF).

Mêmes données que dans la carte 15. *Point de vue de l'observateur*: à partir du sud-ouest (225°). *Angle d'élévation de l'observateur*: 45°. Voir aussi le paragraphe 7.2.1.



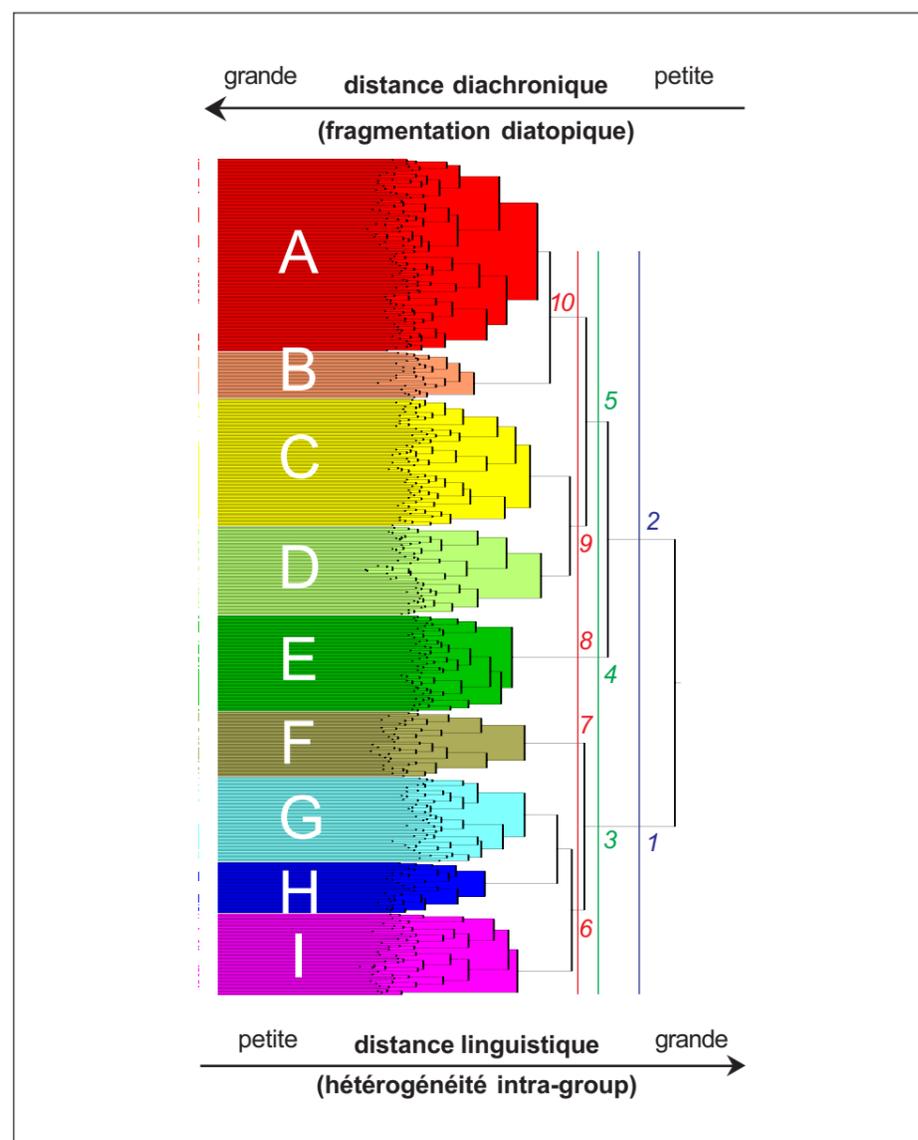
CARTE 17: Cartes choroplèthes de deux synopses de 641 coefficients d'asymétrie de Fisher (CAF).

Corpora: 1687 cartes de travail (données intégrales); moitiés nord (en haut) et sud (en bas) du réseau-ALF. Indice de similarité: IRI<sub>jk</sub>. Algorithme d'intervallisation: MEDMW 8-tuple. Nombre des paliers chromatiques: 8. Voir aussi le paragraphe 7.2.2.



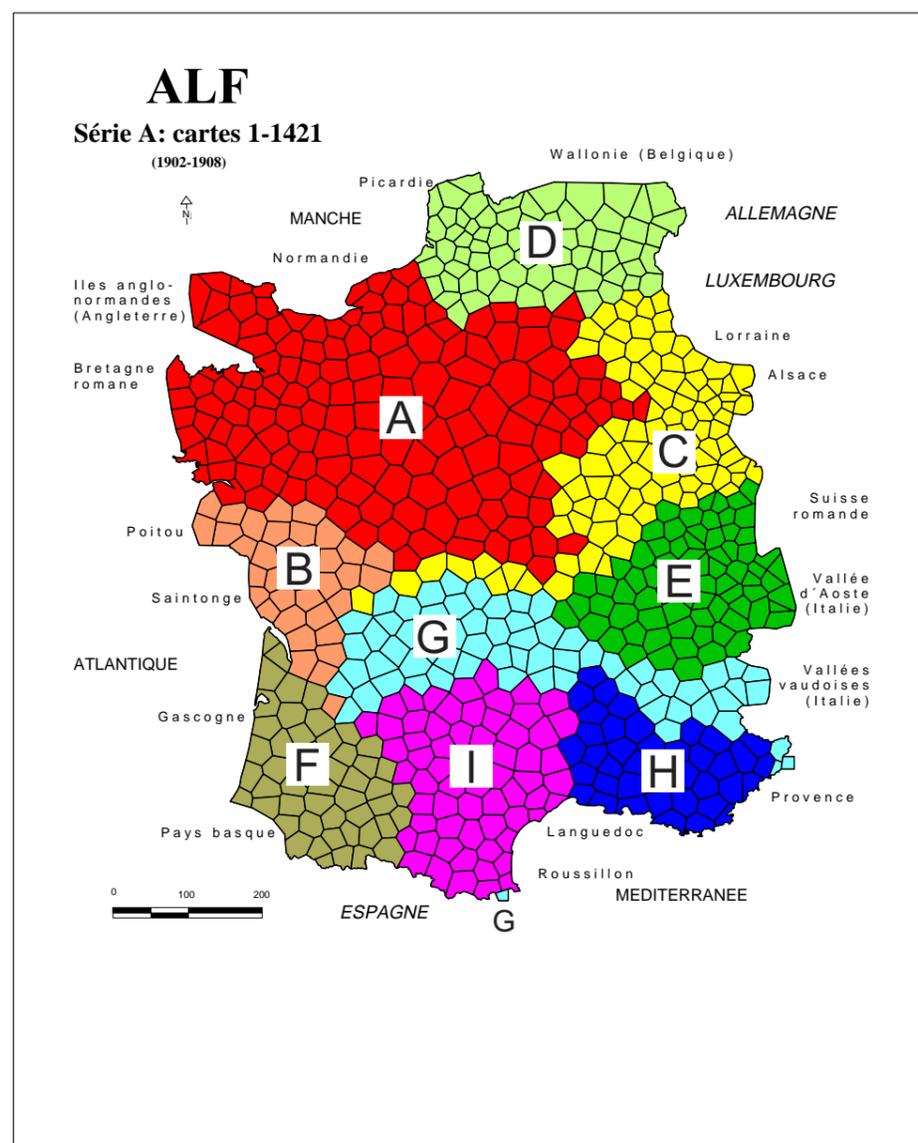
CARTE 18: Surfaces statistiques lissées (stéréogrammes) de deux synopses de 641 coefficients d'asymétrie de Fisher (CAF).

Mêmes données que dans la carte 17. *Point de vue de l'observateur*: à partir du sud-ouest (225°). *Angle d'élévation de l'observateur*: 45°. Voir aussi le paragraphe 7.2.2.



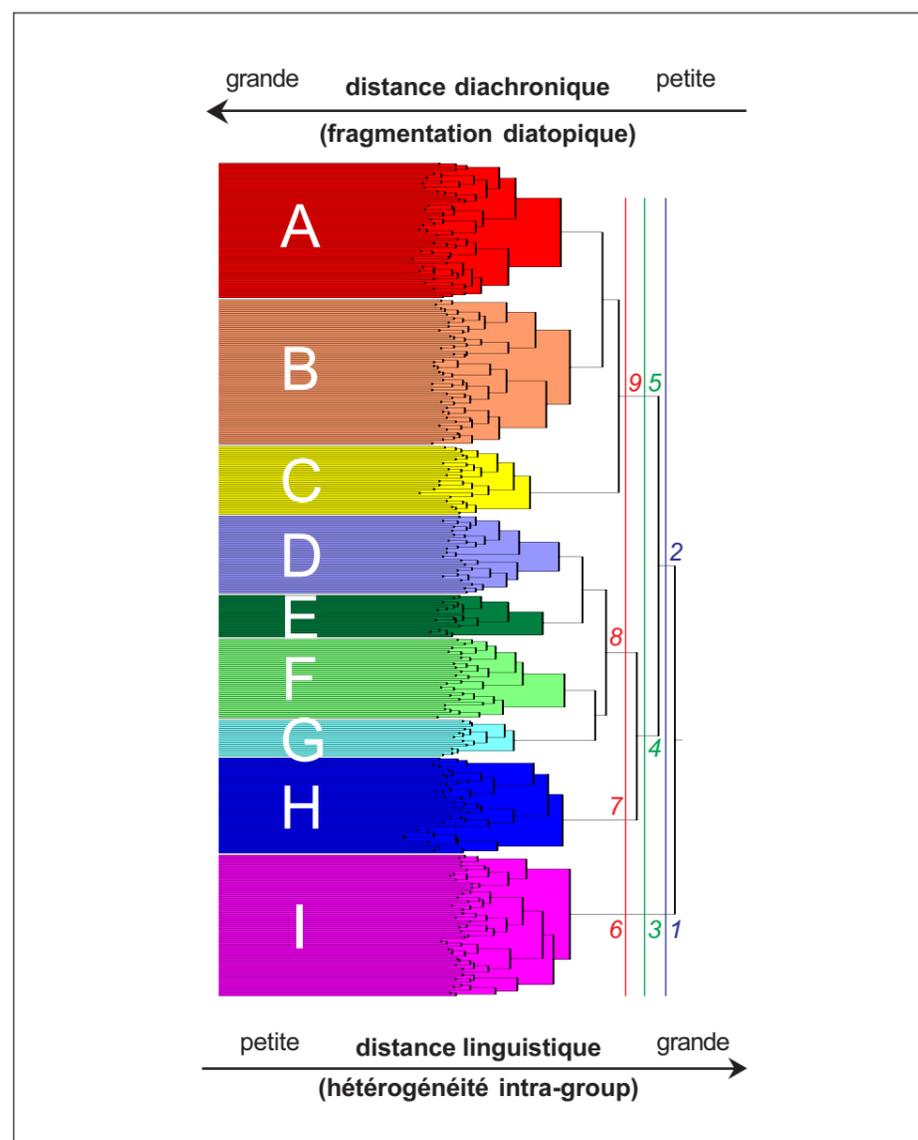
CARTE 19: Classification dendrographique des vecteurs d'attributs de 641 points-ALF.

Algorithme agglomératif: méthode de Ward. Corpus: 1687 cartes de travail (données intégrales). Indice de similarité:  $IPI(1)_k$ . Nombre des dendrèmes coloriés: 9. Voir aussi le paragraphe 8.3. et la carte 20 (spatialisation de cet arbre).



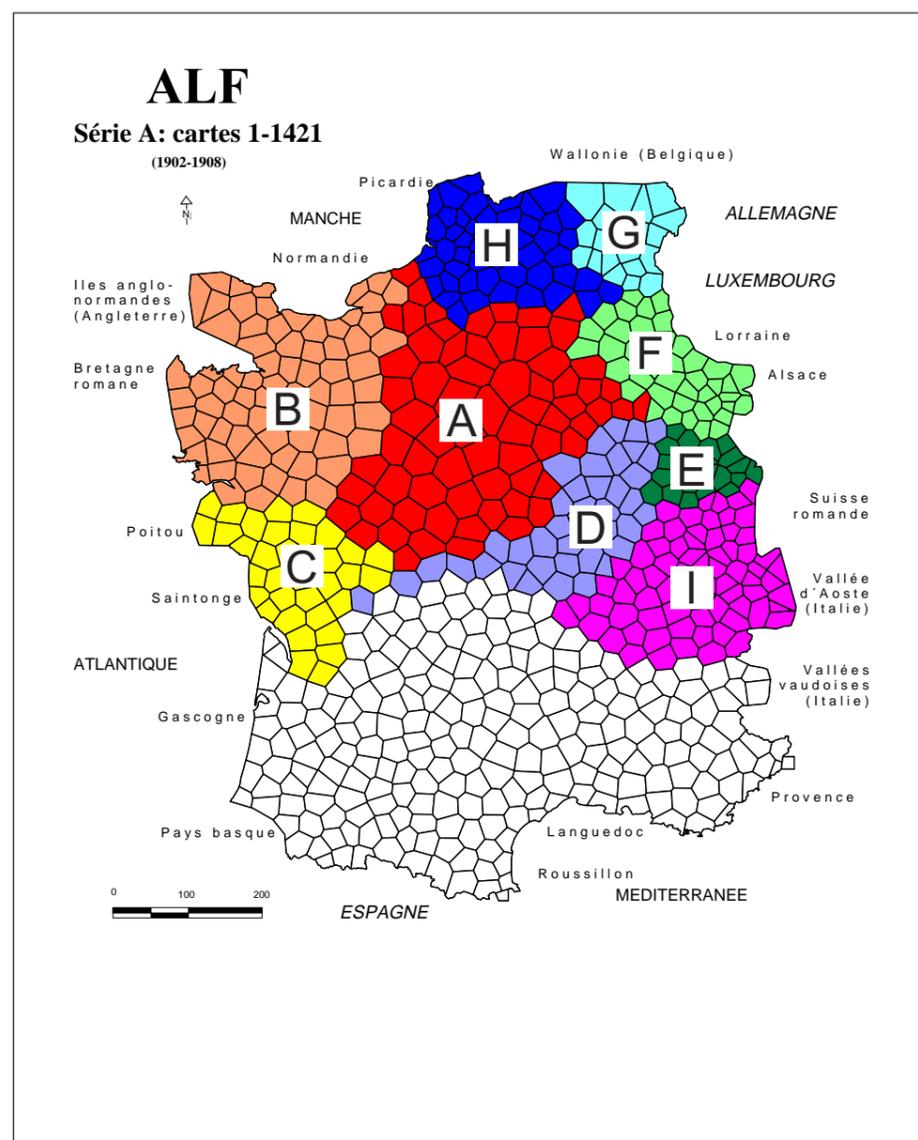
CARTE 20: Spatialisation des données de la carte 19.

Nombre des chorèmes coloriés: 9. Voir aussi le paragraphe 8.3. et la carte 19 (arbre original relatif à cette spatialisation).



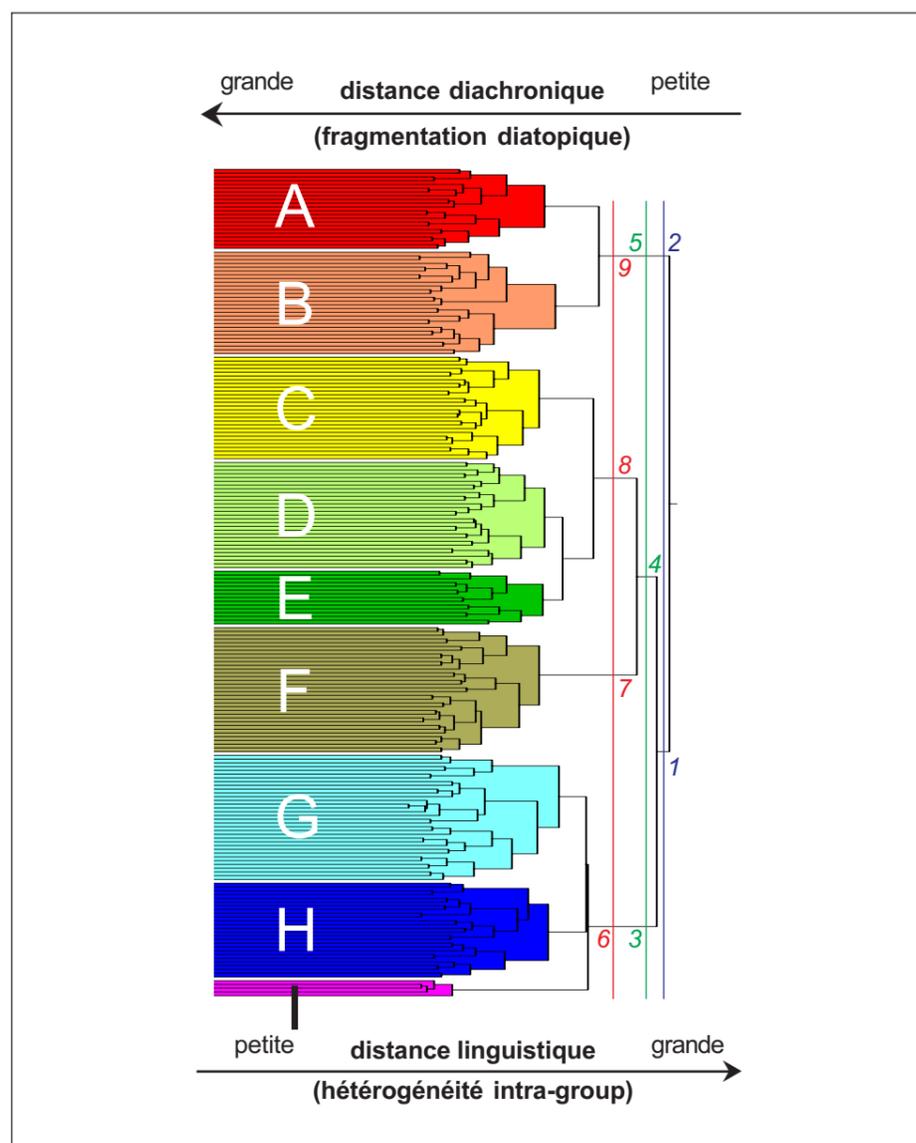
CARTE 21: Classification dendrographique des vecteurs d'attributs de 421 points-ALF.

Algorithme agglomératif: méthode de Ward. Corpus: 1687 cartes de travail; moitié nord du réseau-ALF. Indice de similarité:  $IPI(1)_k$ . Nombre des dendrèmes coloriés: 9. Voir aussi le paragraphe 8.4. et la carte 22 (spatialisation de cet arbre).



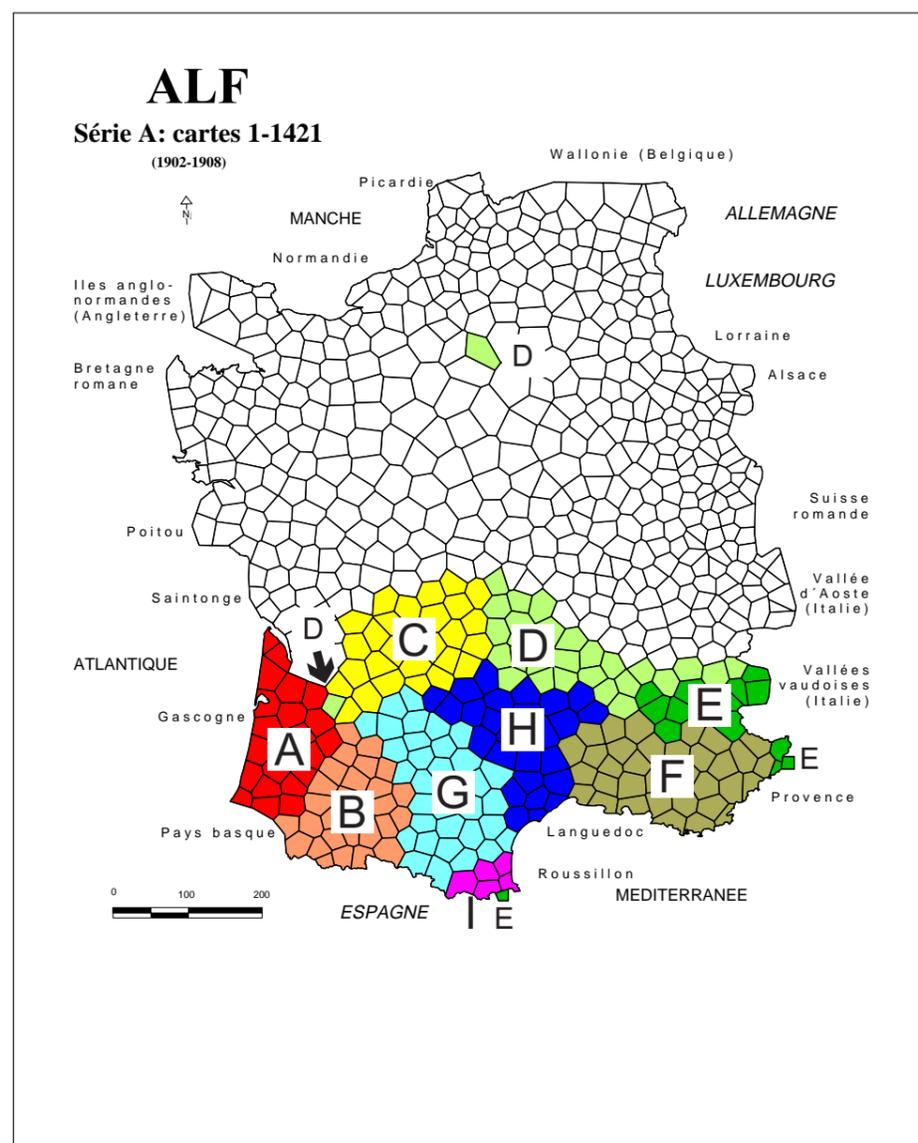
CARTE 22: Spatialisation des données de la carte 21.

Nombre des chorèmes coloriés: 9. Voir aussi le paragraphe 8.4. et la carte 21 (arbre original relatif à cette spatialisation).



CARTE 23: Classification dendrographique des vecteurs d'attributs de 221 points-ALF.

Algorithme agglomératif: méthode de Ward. Corpus: 1687 cartes de travail; moitié sud du réseau-ALF. Indice de similarité:  $IPI(1)_R$ . Nombre des dendrèmes colorés: 9. Voir aussi le paragraphe 8.5. et la carte 24 (spatialisation de cet arbre).



CARTE 24: Spatialisation des données de la carte 23.

Nombre des chorèmes coloriés: 9. Voir aussi le paragraphe 8.5. et la carte 23 (arbre original relatif à cette spatialisation).