Zeitschrift: Revue de linguistique romane
Herausgeber: Société de Linguistique Romane
Band: 66 (2002)
Heft: 261-262

Artikel: Analyse dialectométrique des structures de profondeur de l'ALF
Autor: Goebl, Hans
DOI: https://doi.org/10.5169/seals-400039

Nutzungsbedingungen
Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Mehr erfahren

Conditions d'utilisation
L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. En savoir plus

Terms of use
The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. Find out more

Download PDF: 13.07.2025

ETH-Bibliothek Zürich, E-Periodica, https://www.e-periodica.ch

ANALYSE DIALECTOMÉTRIQUE DES STRUCTURES DE PROFONDEUR DE L'ALF

1. La dialectométrie (DM): buts, méthodes, pratique et intérêt épistémologique

La dialectométrie – définie selon une formule établie il y a quelque vingt ans («géographie linguistique + taxonomie numérique»: Goebl 1981, 349) – est issue d'un mariage heureux entre la géographie linguistique et la taxonomie (ou classification) numérique. Il s'agit donc d'une méthode quantitative qui sert à l'analyse et, partant, à la découverte de structures profondes dans les données d'un atlas linguistique, qui, autrement, resteraient cachées à l'œil de l'observateur humain. Par cette finalité exploratoire, elle se range parmi d'autres disciplines taxométriques telles la psycho-, socio-, anthropo- ou économétrie – pour ne citer que quelques exemples situés dans le halo des sciences du langage –, avec lesquelles elle partage un grand nombre de prérequis empiriques, théoriques et méthodiques. Elle entretient en outre des liens privilégiés avec la géographie quantitative et d'autres disciplines engagées dans l'étude scientifique de l'aménagement de l'espace par l'homme.

Il en résulte qu'en dernière analyse l'intérêt épistémologique de la DM consiste avant tout à élargir nos connaissances sur les mécanismes de la gestion linguistique de l'espace géographique par l'HOMO LOQUENS.

La valeur et l'importance de la DM pour les études romanes et aussi pour la géographie linguistique, l'histoire, l'ethnographie, l'anthropologie, la démographie, etc., ne sont plus à démontrer. Ceci n'empêche que la pratique courante de la DM présuppose – du côté du chercheur intéressé – une bonne maîtrise d'un certain nombre de méthodes taxométriques et cartographiques, et aussi la disponibilité d'outils informatiques bien rodés.

Il ne faut cependant pas passer sous silence le fait tout autre qu'insignifiant que la DM – 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[1] –

(1) Voir les indications bibliographiques dans Goebl 1992a, 433-434 et 1993b, 277-278.

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. À part cela, la DM est en butte aux mêmes malentendus et incompréhensions dont souffrent aussi d'autres secteurs de la linguistique quantitative.

2. L'ALF: mine inépuisable de structures diatopiques de toute sorte

L'ALF – que malgré son âge je continue de considérer comme l'astre central de la richissime nébuleuse des atlas linguistiques romans – permet le dégagement de deux sortes de structures diatopiques:

1) de structures de *superficie* résultant d'analyses typisantes de cartes-ALF isolées: la fameuse typisation de la carte 1 *l'abeille* de l'ALF, dressée par Jules Gilliéron en 1918, en serait un exemple éloquent et universellement connu;

2) de structures de *profondeur* qui résultent de la superposition d'un certain nombre d'analyses particularisantes. Citons, comme cas exemplaire de ce genre d'analyses, la carte des frontières dialectales de la France, établie par Arvid Rosenqvist en 1919 (et réimprimée en 1978 dans Berschin/Felixberger/Goebl, 261-266) et les planches établies par K. v. Ettmayer (1924).

Quant à la recherche des structures de profondeur, la DM permet de pousser plus avant l'expérience globalisante et d'explorer ainsi un nombre pratiquement illimité de structurations diatopiques latentes. Comme l'ALF constitue, par l'ampleur de son réseau d'enquête (638 points) et la richesse linguistique de son questionnaire (1421 questions pour la série A), une source empirique de toute première importance, sa dialectométrisation se révélait pratiquement «incontournable». Nous avons donc décidé, dès avant 1997, de relever ce défi et de nous atteler à la tâche «bénédictine» d'une dialectométrisation de l'ALF entier(2).

3. La dialectométrisation de l'ALF: prérequis techniques et informatiques

Tout d'abord, nous nous sommes procuré un exemplaire de l'ALF dont il existe, depuis 1968, une bonne réimpression de l'édition originale publiée entre 1902 et 1910. Il fallait ensuite le décomposer carte par carte

(2) Les travaux décrits dans cet article se sont déroulés dans le cadre de deux projets de recherche (no. 12414 et 13349) financés par l'organisme de recherche autrichien «Fonds zur Förderung der wissenschaftlichen Forschung in Österreich» (FWF) à Vienne.

parce que le travail de typisation (de taxation, codage, etc.) exige la mise à disposition des planches de l'ALF en tant que feuilles détachées. Ces préparatifs n'ont d'ailleurs été appliqués qu'aux 1421 cartes de la série A de l'ALF. Par la suite, nous avons fait imprimer, sur les 1421 planches dégagées de l'ALF, un réseau multicolore de 18 «parcours d'épreuve» pour assurer ainsi le transfert contrôlé des données de l'ALF sur des listes appropriées.

Le travail de transfert – qui comprend la taxation (typisation, codage, etc.) à proprement parler et l'enfournement consécutif des données taxées dans l'ordinateur – a été confié à une équipe de jeunes romanistes préalablement formée. Il va de soi que le travail de taxation s'est déroulé suivant des consignes aussi précises que possible pour assurer ainsi l'homogénéité des travaux taxatoires de nos jeunes collaborateurs(3).

Du côté informatique enfin, il fallait trouver un collaborateur ultra-compétent à qui confier la confection d'un logiciel capable d'assurer la saisie, le contrôle et le stockage des données-ALF taxées d'un côté, et d'effectuer tous les calculs dialectométriques ainsi que leur visualisation consécutive selon les standards cartographiques les plus récents de l'autre. Heureusement ce collaborateur idéal s'est présenté en la personne de M. Edgar Haimerl qui, entre 1990 et 1997, avait déjà réalisé l'informatisation complète de notre atlas linguistique ladin ALD-I. E. Haimerl a par la suite initié un jeune romaniste salzbourgeois d'origine polonaise – M. Slawomir Sobota – aux secrets de la cartographie assistée par ordinateur. Le programme mis au point par E. Haimerl et à l'aide duquel toutes les cartes de cet article ont été préparées, s'appelle «Visual Dialectometry» (VDM) alors que le logiciel cartographique utilisé par S. Sobota pour la confection définitive des cartes est un produit industriel dénommé «Map Info». Que MM. Haimerl et Sobota reçoivent ici l'expression de ma reconnaissance sincère et profonde!

Les lecteurs intéressés pourront d'ailleurs consulter notre site internet bilingue (allemand-anglais) http://ald.sbg.ac.at/dm où ils trouveront une description détaillée de la démarche dialectométrique en général et des différentes étapes de la dialectométrisation de l'ALF en particulier.

---

(3) À cet égard nous tenons à remercier chaleureusement Barbara Aigner, Irmgard Dautermann, Hildegund Eder, Susanne Oleinek et Annette Schatzmann (toutes de Salzbourg) pour leur collaboration compétente, efficace et tenace. Nous incluons dans notre gratitude Mme Lily Ditz-Fuhrich (Salzbourg) qui, une fois de plus, a bien voulu assurer le contrôle stylistique de notre texte français.

4. Brève description de la chaîne dialectométrique mise en œuvre

Précisons d'entrée de jeu que la dialectométrisation de l'ALF a été faite exactement suivant les normes et méthodes dialectométriques établies par nous il y a plus de vingt ans. C'est pourquoi les lecteurs de la «Revue de linguistique romane» se reporteront utilement à nos contributions respectives de 1981, 1983 et 1987. L'ensemble des méthodes dialectométriques actuellement disponibles est agencé selon le modèle d'une chaîne dont la figure 1 montre les différents maillons.

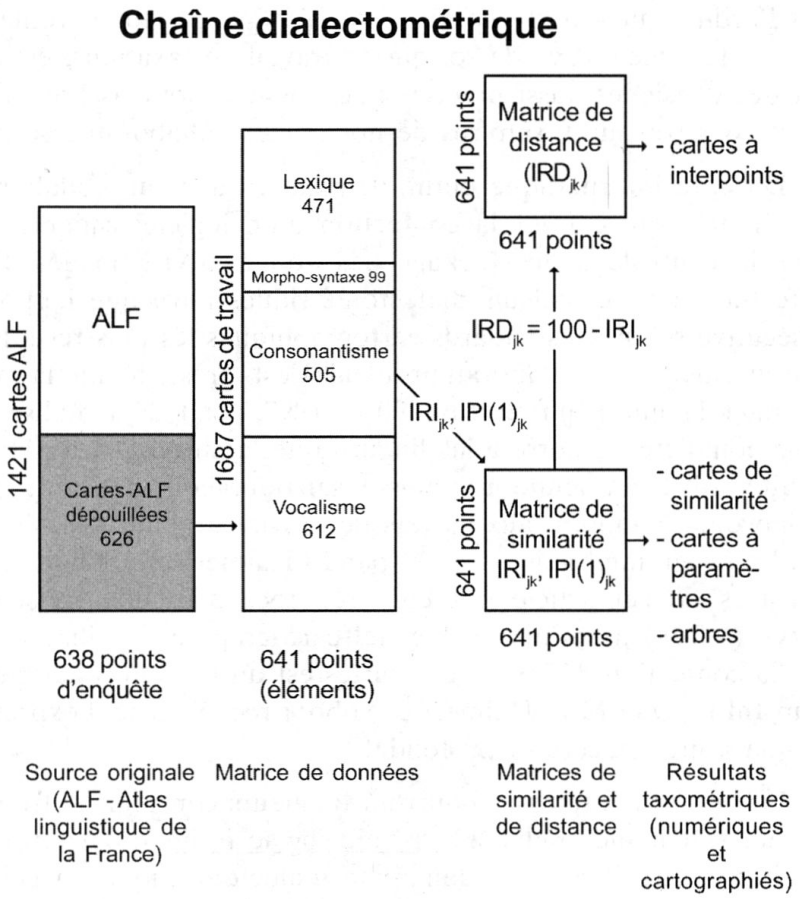


Figure 1: Diagramme des méthodes dialectométriques utilisées dans cet article.

De la source originale – c'est-à-dire de l'atlas linguistique à dialectométriser – l'on passe, par le biais d'un métabolisme informationnel bien défini (i. e. par la taxation ou codage des données originales), à la matrice de données (comprenant N points d'enquête et p cartes de travail) qui, elle, représente donc une sorte d'image, pour ne pas dire un modèle, des

données originales. Rappelons le fait bien connu (mais souvent occulté dans la pratique courante de la recherche) que les données originales d'un atlas linguistique sont également le résultat d'un transfert d'informations et ne doivent en aucune manière être confondues avec la réalité dialectale à proprement parler.

Par la suite, la *matrice de données* doit être soumise à un traitement exploratoire particulier. À cet effet il faut choisir un indice de *similarité* ou de *distance* approprié, et calculer ensuite, suivant les besoins et les finalités de l'analyse dialectométrique en cours, une *matrice* ou bien de *similarité* ou bien de *distance*. Par définition les deux matrices sont toujours symétriques: *N* points d'enquête fois *N* points d'enquête. Par le dépouillement taxométrique des deux matrices mentionnées l'on finit par dégager les structures de profondeur recherchées, d'abord de manière numérique, et ensuite, par le biais d'une visualisation appropriée, aussi de manière graphique.

4.1. Des données de l'ALF à la matrice de données

4.1.1. Traitement empirique et cartographique du réseau de l'ALF

Nous avons pris en considération l'ensemble des 638 points d'enquête de l'ALF(4) tout en leur ajoutant trois points artificiels qui correspondent à autant de langues standards qui, d'une façon d'ailleurs très inégale, ont influé sur les dialectes galloromans:

P. 999: français standard; localisé en Ile-de-France, entre les points-ALF 226 et 227;

P. 998: italien standard; localisé aux confins orientaux de la Provence, entre les points-ALF 990 et 899;

P. 997: catalan standard; localisé au sud du Roussillon, entre les points-ALF 796 et 798.

La dialectométrisation de l'ALF une fois accomplie, les résultats en démontrent cependant que seul le français constitue, pour les données de l'ALF, un agent géolinguistique important.

La polygonisation, comme dernière étape de la préparation cartographique du réseau de l'ALF, portait donc sur 641 points d'enquête. Elle a

(4) Étant donné que E. Edmont a fait 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.

été faite suivant les principes dorénavant bien connus de la géométrie de Delaunay-Voronoi (cf. Goebl 1981, 363-364; 1984 I, 90-92 et Okabe/Boots/Sugihara 1992, passim). Ajoutons encore que, pour des raisons d'une meilleure lisibilité cartographique du fond de la carte polygonale, les îles situées le long de la côte atlantique et prises en considération dans le réseau de l'ALF, ont été accolées directement à la masse continentale de la France.

4.1.2. Taxation (codage, typisation, etc.) des cartes de l'ALF

Dans le souci d'obtenir une matrice de données aussi complète que possible, nous avons écarté de l'analyse taxatoire des cartes-ALF trop lacunaires, à savoir celles qui disposent de moins de 600 (sur 638) réponses transcrites (et partant utilisables). La typisation à proprement parler se faisait suivant des critères phonétiques, morpho-syntaxiques et lexicaux. Évidemment, la taxation *phonétique* ne pouvait se faire qu'à partir de cartes-ALF étymologiquement homogènes et à l'aide d'une grille analytique précise. Ont été insérées dans cette grille non seulement les catégories classiques de la phonétique historique (comme les positions prétonique, tonique et posttonique des différents nexus vocaliques et consonantiques, les syllabes libres et entravées, etc.) mais aussi un grand nombre de nexus étymologiques d'usage fréquent (comme p. ex. -ĔRE, -ĂRE, -ĂTU, -ĔLLU, Ĕ + Palatale, -ĂRIU, etc.). Pour plus ample information voir notre site internet http://ald.sbg.ac.at/dm.

Les unités taxatoires émanant de l'analyse taxatoire s'appellent «taxats». En matière de phonétique, les taxats correspondent le plus souvent aux différents symboles de la transcription de l'ALF dont certains ont d'ailleurs été réunis en groupe pour éviter ainsi un débordement de la fragmentation interne du bilan taxatoire. Ce faisant, nous avons obéi au principe du «lumping», opposé à celui du «splitting», fréquemment observé en matière de taxonomie biologique.

Quant au *lexique*, la taxation visait à dégager des unités lexicales abstraites équivalant grosso modo aux étymons des formes-ALF respectives. L'adjonction de suffixes, préfixes et infixes impliquait toujours l'établissement d'un nouveau taxat. Notre propos était de saisir ainsi la dynamique de l'évolution lexicale de l'espace galloroman, tout en nous inspirant vaguement de la tradition de recherche de l'onomasiologie («Wörter und Sachen») et aussi de l'esprit analytique avec lequel les différents chapitres du FEW ont été conçus.

Il va de soi que le bilan taxatoire de chaque carte originale de l'ALF est plus grand pour la phonétique que pour le lexique ou la morpho-syntaxe. Ceci s'explique par le fait que les données d'une carte-ALF étymologiquement homogène peuvent être soumises à plusieurs taxations, portant, p. ex., sur les multiples succédanés des nexus initiaux, intermédiaires et finals de l'étymon respectif. C'est ainsi que 1117 cartes de travail d'inspiration phonétique ont été tirées de 247 cartes originales de l'ALF alors que le même bilan s'avère plus équilibré pour le lexique: de 463 cartes originales de l'ALF nous avons dégagé seulement 471 cartes de travail.

Vu la complexité du travail taxatoire et de la saisie électronique des données taxées, nous n'avons pu prendre en considération que 626 cartes originales de l'ALF, soit quelque 44% du total des 1421 planches de la série A. Les 626 cartes-ALF analysées se répartissent d'une façon plus ou moins régulière sur l'abscisse des 1421 cartes de l'ALF: voir la figure 2 qui montre en outre que de chaque carte originale de l'ALF il a été tiré un nombre inégal de cartes de travail, allant de 1 à 10.

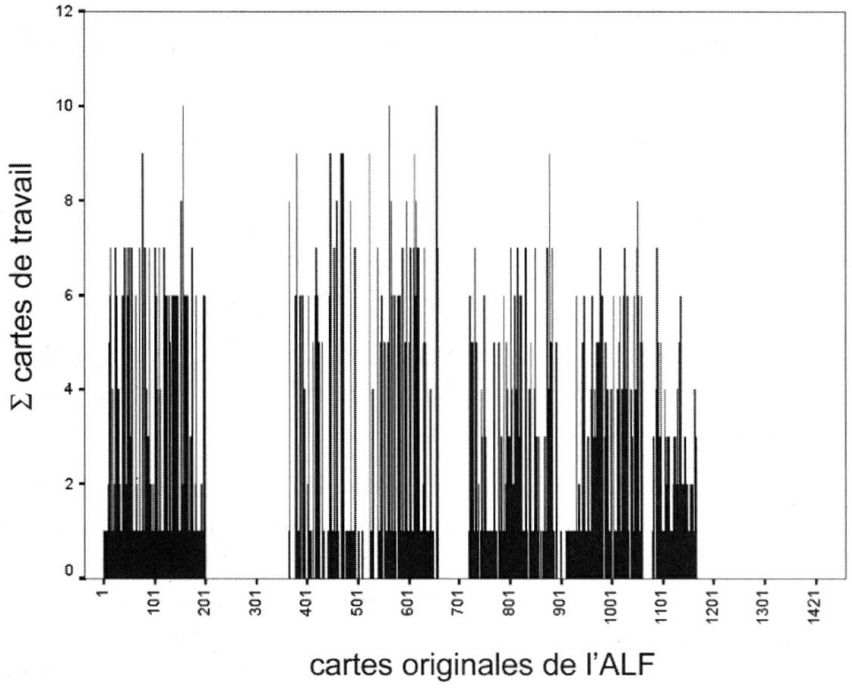


Figure 2: Diagramme du dépouillement taxatoire de l'ALF effectué à Salzbourg entre 1997 et 2000.

Comme l'échantillonnage des 626 cartes-ALF dépouillées s'est fait en fonction de l'ordre alphabétique des titres des 1421 cartes de l'ALF, l'on

peut dire que notre choix a été opéré d'une façon quasiment aléatoire. Le tableau synoptique suivant renseigne sur les effectifs des différentes catégories linguistiques considérées dans cet article.

1		2		3	4
Catégorie linguistique		cartes - ALF		cartes de travail	taxats
		décompte *partiel*	décompte intégral	décompte intégral	décompte intégral
1	Phonétique	247	------	1.117	12.023
2	*vocalisme*	*242*	------	*612*	*8.657*
3	*consonantisme*	*227*	------	*505*	*3.366*
4	Morpho-syntaxe	84	------	99	989
5	Lexique	463	------	471	7.031
	Total	------	**626**	**1.687**	**20.043**

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 moyenne, nous avons tiré, de chaque carte originale de l'ALF, 2,69 cartes de travail (= 1687 : 626). Comme les 1687 cartes de travail fournissent un effectif total de 20 043 taxats, chaque carte de travail dispose en moyenne de 11,88 taxats (= 20 043 : 1687) dont les aires respectives occupent en général 53,95 points du réseau-ALF (= 641 : 11,88).

Entre le nombre des cartes de travail et celui des taxats analysés sur chaque carte de travail il existe d'ailleurs des relations mathématiques très intéressantes: voir à ce sujet les figures 3-5.

Le nombre des cartes de travail décroît d'une manière plus ou moins continue avec l'augmentation de la fragmentation interne des cartes de travail. La figure 3 en offre l'exemple le plus éloquent: la courbe décline harmonieusement entre le score maximal de 211 cartes de travail 3-nymes (c'est-à-dire disposant de trois taxats) et le score minimal d'une seule carte de travail 91-nyme. Pour la phonétique, la situation est très similaire: voir la figure 4. La déclivité de la courbe s'inscrit entre 178 cartes de travail 3-nymes et une carte de travail 76-nyme, alors que les scores relatifs au lexique révèlent une allure moins régulière: les valeurs maximales (36 et 35 cartes de travail) correspondent aux cartes 5- et 10-nymes et la valeur minimale (1) appartient à une carte de travail 91-nyme.

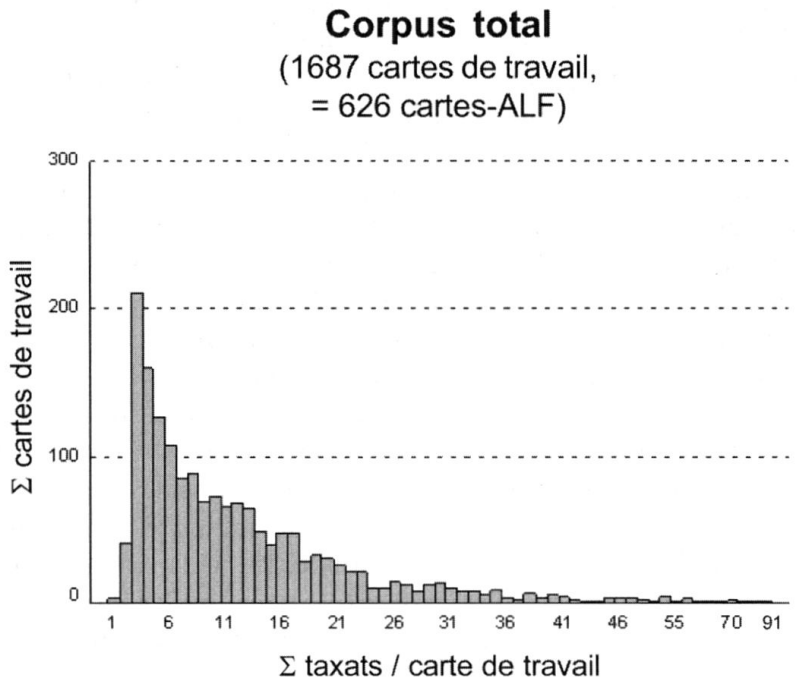


Figure 3: Histogramme relatif à la relation entre le nombre des cartes de travail et leur morcellement taxatoire interne.
Corpus total (1687 cartes de travail = 626 cartes originales de l'ALF).

Le morcellement taxatoire du lexique (de 1 à 91) dépasse largement celui de la phonétique: de 2 à 76 en général; vocalisme: de 2 à 76 taxats/cartes de travail; consonantisme: de 2 à 27 taxats/cartes de travail, en particulier. Les régularités visualisées sur les figures 3-5 reflètent d'ailleurs certaines particularités du changement linguistique tel qu'il s'est déroulé dans l'espace galloroman. 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 (cf. Altmann 1985 avec la déduction stochastique d'une «loi de Goebl»).

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

C'est par l'application d'un indice ou bien de *similarité* ou bien de *distance* que l'on passe – le long de la chaîne dialectométrique (voir la figure 1) – de la matrice de données à la matrice de similarité ou de distance. Comme la taxométrie moderne (cf. p. ex. Sneath/Sokal 1974, Bock 1974, Chandon/Pinson 1981, Bailey 1994 ou Gordon 1999) offre un grand

nombre d'indices utilisables et que chacun d'entre eux répond à d'autres besoins et finalités exploratoires, il est de toute première importance que le dialectométricien règle son choix sur les nécessités de sa propre démarche dialectométrique. Les deux indices de similarité utilisés dans le cadre de cet article – l'«Indice Relatif d'Identité» (IRI_{jk}) et l'«Indice Pondéré (avec le poids 1) d'Identité» ($IPI(1)_{jk}$) – ont prouvé leur utilité dialectologique plus d'une fois (cf. Goebl 1981, 357 s. et 1987, 702 s.). Alors que l'IRI_{jk} est une mesure isocratique qui ne privilégie numériquement aucun des taxats de la matrice de données, l'$IPI(1)_{jk}$ en revanche, en tant que mesure anisocratique, favorise numériquement surtout les taxats diatopiquement moins répandus et défavorise par contre les taxats de grande diffusion spatiale. Cet indice répond à une philosophie très répandue parmi les linguistes selon laquelle les traits linguistiques universellement répandus ont une incidence classificatoire moindre que les traits linguistiques rares. Le poids numérique des taxats de moindre diffusion augmente avec la diminution du facteur de pondération dans la formule de l'$IPI(x)_{jk}$ (cf. à ce sujet Goebl 1987, 70 s.). Les exemples présentés par la suite ont été calculés avec le facteur de pondération maximal, à savoir

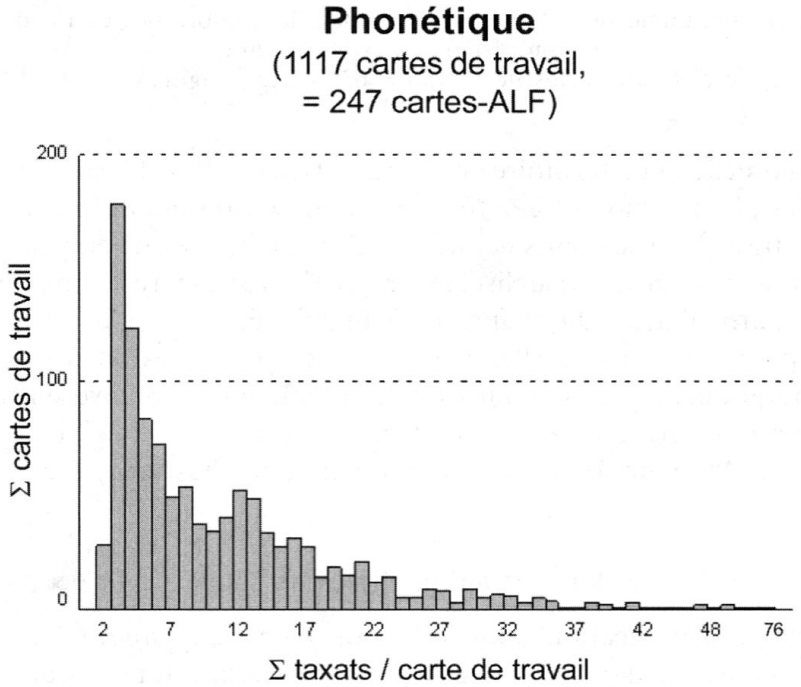


Figure 4: Histogramme relatif à la relation entre le nombre des cartes de travail et leur morcellement taxatoire interne.
Phonétique (1117 cartes de travail = 247 cartes originales de l'ALF).

x = 1. L'indice pondéré $IPI(x)_{jk}$ sert avant tout à l'exploration de paysages dialectaux moins saillants ou d'importance secondaire, et à l'étude d'enchevêtrements géolinguistiques à courte distance.

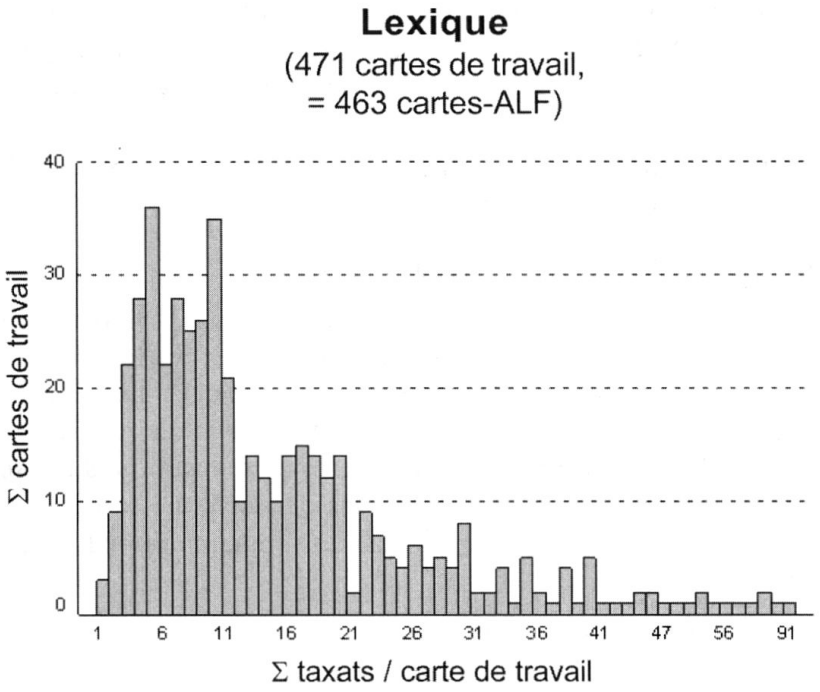


Figure 5: Histogramme relatif à la relation entre le nombre des cartes de travail et leur morcellement taxatoire interne.
Lexique (471 cartes de travail = 463 cartes originales de l'ALF).

4.3. Les exigences statistiques et cartographiques de la mise en carte

Chacune des 24 cartes présentées dans cet article répond à des conventions iconiques précises et bien rodées depuis longtemps. Le but central des cartes est de transposer la variabilité numérique des différentes distributions de similarité ou de distance en une variabilité analogue d'un certain nombre de couleurs ordonnées suivant la logique de l'arc-en-ciel (ou du spectre solaire). Or, la psychologie visuelle nous enseigne que la sensibilité de l'œil humain est le mieux desservie par un jeu différentiel de seulement 6 à 8 paliers chromatiques. La définition des paliers chromatiques utilisés se fait à l'aide de deux algorithmes d'intervallisation (MINMWMAX et MEDMW) élaborés il y a belle lurette: cf. Goebl 1981, 365 s.; 1984 I, 93-97 et 1987, 79 s.). La même remarque vaut pour les histogrammes qui, à l'aide de 2n colonnes verticales, visualisent le rendement numérique de la distribution de fréquence mise en carte

moyennant un algorithme d'intervallisation n-tuple (cf. Goebl 1981, 363 et 1984 I, 97-98).

Le linéament de la courbe gaussienne (ou normale) théorique, calculée à partir de la moyenne arithmétique et de l'écart-type de la distribution de fréquence (i. e. de similarité ou de distance) respective, permet en outre d'en contrôler rapidement certaines propriétés mathématiques.

5. La mensuration de la distance linguistique entre les points contigus de l'ALF (DM interponctuelle): voir la carte 1

5.1. Conventions graphiques et taxométriques

Il y a plus d'un siècle, le germaniste Carl Haag (1898) a démontré que le tracement rationnel d'isoglosses combinées ne pouvait se faire qu'à l'aide de la géométrie de Delaunay-Voronoi. Depuis la mise en place de la cartographie assistée par ordinateur, il est facile d'en réaliser tous les prérequis, tant théoriques que pratiques. La polygonisation du réseau de l'ALF a abouti à la définition de 1792 côtés de polygone le long desquels aura lieu le tracement – d'ailleurs entièrement imaginaire – des différents tronçons d'isoglosse.

Rappelons encore trois choses:

1) que chaque tronçon d'isoglosse est défini en termes de distance (dissemblance, dissimilarité, etc.) entre deux points d'atlas contigus (cf. Goebl 1983, passim);

2) qu'il existe depuis 1953 (Lalanne) le terme heureux d'«interpoint» pour désigner la position intermédiaire, et partant relationnelle, de chaque côté du polygone entre deux points d'atlas contigus;

3) que la contiguïté des points de l'ALF a été définie au cours de la triangulation du réseau de l'ALF, qui constitue la première étape du processus de polygonisation. La triangulation a abouti à la géométrie dite «de Delaunay».

Pour la visualisation des 1792 scores de distance interponctuels nous avons choisi huit couleurs différentes dont la répartition incombe à l'algorithme d'intervallisation MEDMW. Cet algorithme privilégie numériquement les intervalles extrêmes (1-2 et 7-8) et confère ainsi, à la syntaxe iconique du graphique entier, une allure plus accidentée. Rappelons aussi que le profil polygonal de la carte 1 repose sur le jeu interférentiel de 1792 côtés de polygone disposant chacun d'une épaisseur et d'un coloriage particulier qui varient en fonction de huit degrés linéaires et chromatiques préalablement définis.

Les plus grandes distances interponctuelles sont symbolisées par l'apparition concomitante de la plus grande épaisseur linéaire du côté du polygone et du coloriage en bleu foncé (évoquant psychologiquement la «froideur» des relations interindividuelles) alors que le contraire est visualisé par des côtés de polygone minces d'un côté et coloriés en rouge de l'autre. Les intervalles 1-4 se situent, du point de vue numérique (ou statistique), au-dessous de la moyenne arithmétique ($\bar{x}$ = 18,12, voir le seuil supérieur de l'intervalle 4) tout en correspondant iconiquement aux couleurs «chaudes» alors que les intervalles 5-8 occupent l'espace numérique situé au-dessus de la moyenne arithmétique et sont visualisés par les couleurs «froides». L'indice de distance appliqué et mis en carte par la suite est l'«Indice Relatif de Distance» (IRD_{jk}). Il constitue le complément numérique de l'IRI_{jk} selon la formule suivante: $IRI_{jk} + IRD_{jk} = 100$.

5.2. Interprétation de la carte 1

La carte 1 est sillonnée d'un vaste champ de cloisonnements interponctuels très intenses, intercalés entre les domaines d'Oc et d'Oïl et s'étendant aussi en Suisse romande, Lorraine, Wallonie et Picardie. Les scores interponctuels les plus importants se trouvent entre l'italien standard (P. 998) et les points-ALF provençaux voisins, aux environs de Bordeaux autour de l'îlot linguistique de la Petite Gavacherie (P. 635, Andraut) et le long du pourtour du domaine francoprovençal (surtout dans la région du Forez). Les cloisonnements minimaux (symbolisés par les intervalles 1 et 2) par contre, se trouvent au centre du domaine d'Oïl, en Languedoc et en Provence. Une comparaison de la carte 1 avec des synthèses isoglottiques antérieures (voir p. ex. Rosenqvist 1919 et Ettmayer 1924) s'impose. Fort de cette comparaison, le lecteur jugera lui-même des progrès méthodiques et techniques accomplis en quelque soixante-dix ans de recherches et aussi de la dimension de l'erreur commise jadis par Gaston Paris (1888, 435) quand il affirmait: «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 France, qu'aucune limite réelle ne sépare les Français du Nord de ceux du Midi [...].». Il s'agit là d'une prise de position typophobe classique née dans un climat épistémologique peu enclin à la classification rationnelle de données géolinguistiques et exacerbée en plus par certaines exaspérations géopolitiques unitaristes jaillies à la suite de la guerre franco-prussienne de 1870-71.

La syntaxe iconique de la carte 1 nous montre en outre clairement qu'il n'y a aucune délimitation plus ou moins linéaire au sein de la Galloromania et que, de ce fait, la vieille métaphore de la «frontière» linguistique est carrément inutilisable. Ce n'est que la métaphore de la «zone» ou «bande» de transition qui tient debout devant l'évidence des résultats classificatoires.

Soulignons aussi le fait quelque peu paradoxal que d'entiers paysages dialectaux (comme le wallon, le limousin, l'auvergnat ou le francoprovençal) semblent s'estomper complètement dans le chassé-croisé des cloisonnements épais (et bleus). Ceci ne les empêche d'ailleurs pas de réapparaître avec netteté au cours d'autres analyses dialectométriques, telle la mensuration de la similarité et des paramètres caractéristiques ou l'analyse dendrographique. Cette contradiction étrange s'explique par le caractère foncièrement ambivalent du concept de «frontière»: toute frontière marque d'un côté l'arrêt ou le ralentissement de communications interponctuelles, mais aussi – bien que dans une mesure plus ou moins limitée – la persistance de contacts et d'échanges de l'autre. L'effet ralentisseur est saisi par la DM interponctuelle (i. e. relative aux distances ou dissemblances entre deux points contigus d'un réseau interactif) alors que l'effet communicatif constitue l'objet de la DM de similarité. Voir à ce propos aussi les livres stimulants de Nordman 1998 et de Bromberger/Morel 2001.

Mentionnons encore, avant de conclure ce paragraphe, un fait taxométrique tout autre qu'insignifiant: des 205 120 scores de distance répertoriés dans la matrice de distance (selon IRD_{jk}) respective, seulement un pourcentage minime (i. e. 0,87 % = 1792 : 205 120) a été utilisé pour la génération de la carte 1, alors que les visualisations présentées dans les paragraphes 7 et 8 de cet article reposent sur l'effectif total de la même matrice de distance (ou de similarité). Ceci prouve que l'importance taxométrique de la DM interponctuelle est très limitée et qu'il est bon de ne pas en surestimer la valeur exploratoire.

6. La mensuration de la similarité linguistique entre les 641 points-ALF

6.1. Conventions graphiques et taxométriques

Pour des raisons d'espace il est impossible de présenter la numérotation des points d'enquête de l'ALF sur le fond des cartes de similarité. Mais il est facile d'identifier les numéros des différents points ou polygones de l'ALF en juxtaposant les cartes de cet article et une copie de la carte poly-

gonale muette de l'ALF pourvue de la numérotation complète, disponible sur internet: http://ald.sbg.ac.at/dm/germ/Allgemein/ALF_Netz.pdf.

La visualisation des cartes 2-7 et 9-13 repose sur l'algorithme d'intervallisation MINMWMAX pour 6, 8 et 10 paliers chromatiques différents. Les indices de similarité utilisés sont l'IRI_{jk} et l'$IPI(1)_{jk}$ (cartes 4-5); voir aussi – à titre de comparaison – les cartes coloriées dans Goebl 2000.

6.2. Présentation de la carte 2: la francisation de l'espace galloroman

Le point de référence de la carte 2 est le point-ALF factice 999 qui correspond au français standard (défini par la terminologie des titres des cartes-ALF respectives). Le polygone relatif au P. 999 a été inséré à proximité des PP. 226 et 227. Le profil choroplèthe de la carte 1 – qui correspond d'ailleurs de très près à celui établi à partir des autres points-ALF situés en Ile-de-France: voir la figure 1 dans Goebl 2000, 224 – montre la position géolinguistique du français par rapport à la Galloromania entière. L'on discerne bien – en rouge – les zones de la plus forte pénétration du français et aussi quelques provignements tentaculaires vers l'Ouest, le Sud, l'Est et le Nord comme manifestations typiques de l'irradiation spatiale de la langue de la Royauté et de la République françaises. Les plages polygonales en bleu foncé (intervalle 1) constituent en quelque sorte les «antipodes» géotypologiques du français: il s'agit de neuf points-ALF équivalant au catalan du Roussillon, au catalan standard, à l'italien standard et à deux points du provençal maritime. L'ensemble des polygones rouges, orange et jaunes recouvre grosso modo le domaine d'Oïl à l'exception cependant du wallon et de quelques enclaves lorraines. À noter en outre l'étendue et la structuration de la zone verte (intervalle 3) qui recouvre, entre autres, le couloir rhodanien entre Lyon et Marseille, porte d'entrée de nombreuses innovations septentrionales. Signalons aussi la silhouette «crescenti-forme» de l'ensemble des polygones situés entre Bordeaux et Clermont-Ferrand, et aussi la position détachée d'un polygone isolé en jaune (intervalle 4) au sud de Bordeaux (P. 635, Andraut, Petite Gavacherie). De ces deux particularités il sera encore question plus tard (6.5. et 6.6.).

Ajoutons encore que toute carte de similarité peut être interprétée aussi d'un point de vue communicatif et diffusionniste. Dans cette perspective, il s'agit de juger de l'impact communicatif et diffusionniste de la dialecticité du point de référence (définie par l'ensemble des attributs linguistiques respectifs répertoriés dans la matrice de données) sur le reste du réseau de l'ALF (cf. Goebl 1981, 375 et 1984 I, 104-113).

6.3. Présentation des cartes 3 et 4: la position d'un parler wallon au sein du réseau-ALF

Les cartes 3 et 4 servent à la démonstration de deux choses: 1) à la présentation du profil choroplèthe d'un paysage linguistique périphérique à l'aide de l'analyse des similarités, et 2) à la démonstration de l'incidence de la pondération (carte 4). La carte 4 se distingue de la carte 3 par la diminution des plages rouges et orange d'un côté (21 + 7 vs. 12 + 4) et par l'augmentation des plages en bleu foncé et moyen (19 + 166 vs. 167 + 58) de l'autre. C'est ainsi que le profil choroplèthe de la carte 4 devient plus accidenté par rapport à celui de la carte 3. À noter que les plages en couleurs «chaudes» embrassent, sur la carte 3, une bonne partie du francoprovençal, alors que, sur la carte 4, la plupart en est répertoriée en vert et ne dispose donc que de similarités situées au-dessous de la moyenne arithmétique. À relever également la silhouette «crescentiforme» de la frange-nord du domaine d'Oc (en vert sur la carte 3 et en bleu moyen sur la carte 4).

6.4. Présentation de la carte 5: la position d'un parler picard au sein du réseau-ALF

Le contraste entre les cartes 4 et 5, calculées toutes les deux à l'aide d'un indice de similarité pondéré, est relativement petit: la position des plages «chaudes» (intervalles 6-4) d'un côté et des plages «froides» (intervalles 1-3) de l'autre est plus ou moins identique. Les différences entre les deux cartes concernent surtout les intervalles 6 et 5 dont les polygones délimitent le paysage linguistique picard de façon très compacte. Notons encore l'étalement bien ordonné des six paliers chromatiques dans l'espace, et aussi que le type iconique de la carte choroplèthe entière ne changerait guère au cas où l'on déplacerait le point de référence à l'intérieur de la zone «picarde» (= polygones rouges et orange).

6.5. Présentation des cartes 6-8: la position d'un parler francoprovençal au sein du réseau-ALF

Le point de référence des cartes 6 et 7 correspond au P.-ALF 969, L'Étivaz, situé au sud-est de Château d'Oex et à proximité directe de la frontière linguistique entre le francoprovençal de la Suisse romande et des parlers alémaniques de la Suisse allemande. Il s'agit donc d'un parler galloroman en position périphérique. Le «francoprovençal» – c'est-à-dire l'espace disposant du plus haut degré de typicité locale selon les attributs

linguistiques répertoriés dans le vecteur 969 de la matrice de données – est bien marqué par l'ensemble des plages rouges et orange. Le pourtour correspond parfaitement à la définition géolinguistique traditionnelle du francoprovençal (cf. p. ex. Hasselrot 1966, Tuaillon 1972 ou Martin 1990). Soulignons encore que la même zone apparaîtrait aussi à partir d'autres points de référence situés dans la zone rouge et orange.

Remarquons aussi que l'ensemble des polygones rouges, orange et jaunes crée une superficie cohérente qui s'étend de la Normandie et de la Bretagne romane jusqu'en Vallée d'Aoste. La carte 7 en donne une vision encore plus claire: le profil iconique y a été réalisé avec seulement deux paliers chromatiques différents où les plages rouges symbolisent des zones linguistiquement plus ou moins proches de la dialecticité du point de référence alors que les plages bleues en signalent le contraire: à noter les grandes enclaves bleues en Picardie, en Wallonie et en Lorraine!

Or, l'espace constitué par les polygones rouges rappelle de très près l'étendue de l'ancienne «Provincia lugdunensis» (voir la carte 8) dont la latinité – suivant une théorie chère aux francoprovençalistes depuis Bengt Hasselrot (1966) et surtout Pierre Gardette (p. ex. 1967/74) – devrait constituer le noyau historique du francoprovençal actuel. En effet, la ressemblance des espaces rouges sur les cartes 7 et 8 plaide énergiquement en faveur de cette théorie. Rappelons que, du point de vue méthodique, le statut épistémologique de la carte 8 – tirée d'un atlas historique (Motta 1995, 26) – est très haut et rejoint celui d'une carte dialectométrique qui, elle, de par sa nature typologique et partant globale voire généralisante, dépasse de loin celui d'une simple mise en carte d'un trait linguistique isolé. Vu son rang épistémologique majeur, la carte des provinces romaines de la Gaule peut donc être comparée directement avec une carte dialectométrique issue de la synopse quantitative de plus de vingt mille aires linguistiques isolées dont chacune est le reflet d'un microprocessus historique particulier.

En comparant les cartes 7 et 8 il ne faudrait surtout pas s'attendre à une coincidence spatiale parfaite. Évidemment, la vieille mouvance lugdunaise, relayée plus tard par celle de Lutetia/Paris, a été élargie considérablement tant vers le Nord que vers le Sud sans d'ailleurs se confondre avec les latinités belges et germaines d'un côté et leurs pendants aquitains et narbonnais de l'autre. De tout ceci il résulte que les fondements de la fameuse tripartition de la Gaule (Aquitaine et Narbonnaise dans le Sud, Lugdunaise au Centre, Belgique et Germanie dans le Nord) ressortent fort bien de la structuration choroplèthe de la carte 7.

6.6. Présentation des cartes 9-12: la position d'un parler «crescentain» au sein du réseau-ALF

Les cartes 9-12 servent à la discussion dialectométrique de deux phénomènes très intéressants: 1) de la structuration spatiale de plusieurs profils de similarité relatifs à un point-ALF du croissant, et 2) de l'incidence de la catégorie linguistique du corpus utilisé (lexique vs. vocalisme vs. consonantisme) sur les résultats dialectométriques.

La zone du croissant – dont l'historique a été magistralement décrit par Guylaine Brun-Trigaud en 1990 – constitue, un peu à l'instar du franco-provençal, une zone d'intersection complexe entre les latinités d'Oïl d'un côté et les latinités d'Oc de l'autre. La métaphore géographique du «croissant», relative à l'aspect cintré du tracé combiné de plusieurs isoglosses entre les domaines d'Oïl et d'Oc au Nord du Massif Central, a été mise en circulation en 1913 par Jules Ronjat (6-7). Le point-ALF 706 (Merlines, Corrèze) se situe à proximité du secteur central du croissant. La carte 9 en montre le profil de similarité à l'aide de huit (et non plus de six) gradins chromatiques pour mieux en faire ressortir les différents enchevêtrements spatiaux. Le voisinage immédiat du point 706, défini par les plages rouges (intervalle 8), orange (intervalle 7) et ocres (intervalle 6), recouvre grosso modo l'occitan septentrional (limousin et auvergnat), mis à part quelques provignements en ocre projetés vers le couloir rhodanien. L'ensemble des polygones répertoriés dans les intervalles 8-5 (couleurs «chaudes») forme un espace relativement cohérent qui s'inscrit tant dans le Nord que dans le Sud de la Galloromania. Par rapport aux cartes 2-5 (toutes du meilleur «aloi» oïlique) il s'agit en effet d'une conjonction nettement hybride. La même hybridation ressort de la carte 10 qui repose d'ailleurs uniquement sur le corpus lexical (471 cartes de travail). La seule différence consiste dans une légère dérive, sur la carte 10, de la zone «chaude» vers le Nord.

Les cartes 11 et 12 par contre montrent des situations moins ambiguës. La carte 11 qui repose sur des données exclusivement vocaliques (612 cartes de travail), montre un profil foncièrement «sudiste»: les zones «chaudes» y recouvrent la totalité du domaine d'Oc. Ceci signifie que la part vocalique des attributs linguistiques du vecteur matriciel du P.-ALF 706 est d'origine carrément occitane. Tel n'est pas le cas pour la carte 12 où le corpus est constitué de 505 cartes (de travail) uniquement consonantiques. Le profil choroplèthe en est plutôt – mais non pas entièrement – de type oïlique: la zone «chaude» s'étend jusqu'à la Manche, mais évite les Pyrénées sans pour autant se détacher du littoral méditer-

ranéen de la Provence. Il en appert que la part consonantique des attributs linguistiques du P.-ALF 706 est de provenance plutôt oïlique.

La conclusion qui se dégage de ces constellations changeantes est que la nature profonde du croissant consiste en une imbrication complexe de traits linguistiques d'origine occitane et oïlique qui, elle, varie en fonction des différentes catégories grammaticales.

6.7. Présentation de la carte 13: la position de la Petite Gavacherie au sein du réseau-ALF

Il est bien connu que le réseau de l'ALF comprend un point d'enquête «allochtone», à savoir le P. 635, Andraut, qui fait partie de l'îlot linguistique de la Petite Gavacherie (cf. Tourtoulon-Bringuier 1876, 567 s.; Suchier 1888, 598; Ronjat 1930, 24-25). La Petite Gavacherie constitue une colonie linguistique fondée au XVe siècle par des colons d'origine poitevine et saintongeaise. Or, à l'aide d'une carte de similarité finement articulée il est possible d'examiner la question de l'origine (linguistique) des fondateurs de la Petite Gavacherie du point de vue dialectométrique. La carte 13 montre donc que les points-ALF disposant de la plus grande similarité linguistique avec le point de référence 635 se trouvent dans le voisinage immédiat: il s'agit des PP. 632, Alzac ($IRI_{635,632}$ = 80,74) et 630, Saint-Savin ($IRI_{635,630}$ = 78,21), situés tous les deux dans la Gironde. Ceci signifie que la dialecticité du P. 635, au moins telle qu'elle a été enregistrée dans l'ALF, est dans une mouvance beaucoup plus saintongeaise que poitevine.

La carte 13 montre en outre que le «Hinterland» oïlique du point de référence 635, défini par l'intervalle 9 avec 65 points-ALF, est très large et que la dialecticité du P. 635 dispose ainsi d'un ancrage fort solide dans la latinité d'Oïl entière. Ajoutons encore qu'une analyse dialectométrique de la position du P. 635 faite selon les différentes catégories linguistiques, montre que la substance phonétique de la dialecticité d'Andraut continue d'être foncièrement oïlique, mais qu'une certaine occitanisation se manifeste déjà sur le plan du lexique, surtout par l'adoption de lexèmes d'origine gasconne ou gasconnisants.

7. Au-delà de la mensuration de la similarité: la synopse de paramètres caractéristiques des 641 distributions de similarité

Chacune des 641 distributions de similarité stockées dans la matrice de similarité (voir la figure 1) dispose, comme toute distribution de fréquence, de plusieurs paramètres statistiques aptes à la caractériser du

point de vue mathématique: minimum, moyenne arithmétique, médiane, maximum, écart-type, coefficient d'asymétrie de Fisher, etc. Comme tous ces paramètres accusent des variations spatiales bien ordonnées, il est indiqué de les réunir en synopse cartographique et d'en tirer des cartes choroplèthes visualisées en bonne et due forme. En effet, les profils iconiques qui s'y manifestent sont du plus haut intérêt dialectologique et dialectométrique; ils permettent d'approfondir considérablement l'analyse typologique des données de l'ALF. Dans les paragraphes suivants nous présenterons la synopse des écarts-types (carte 14) et celle des coefficients d'asymétrie de Fisher (cartes 15-20).

7.1. Présentation de la carte 14: la synopse des écarts-types (ET)

L'«écart-type» (all. «Standardabweichung», angl. «standard deviation») est un des paramètres classiques de la statistique descriptive. Il sert à la mensuration du degré de dispersion des différents scores d'une distribution de fréquence par la saisie de l'écart entre ces derniers et la moyenne arithmétique respective. En voici la formule en forme simplifiée (cf. Muller 1968, 54; Kreyszig 1968, 40 et Goebl 1984 I, 149):

$$ET_j = \sqrt{\frac{\sum_{i=1}^{n} IRI_{jk}^2 - n \cdot \overline{IRI}_j^2}{n-1}}$$

La signification des symboles est la suivante:

ET_j l'écart-type d'une distribution de similarité relative au point de référence j

n l'ensemble des 640 scores d'une distribution de similarité (IRI_{jk}); ici: n = N – 1, pour N = 641

i l'indicatif (allant de 1 à 640) des n scores d'une distribution de similarité

IRI_{jk} un des 640 scores d'une distribution de similarité relative au point de référence j

j l'indicatif d'un point de référence (allant de 1 à 641)

k l'indicatif d'un point comparé (allant de 1 à 641)

$\overline{IRI}_j$ la moyenne arithmétique d'une distribution de similarité relative au point de référence j

L'utilité dialectométrique de la synopse des écarts-types réside dans sa capacité de détecter, au sein d'un réseau d'atlas donné, les zones de transition entre différents macro-systèmes dialectaux à forte cohésion interne. La carte 14 en fournit une application éloquente et iconiquement très attrayante à la fois. Les domaines d'Oïl et d'Oc sont séparés entre eux par de larges zones de transition, toutes coloriées de différentes teintes de bleu, qui embrassent, grosso modo, la couronne septentrionale de l'occitan avec le francoprovençal, et qui s'étendent de ce fait de l'Atlantique aux Alpes. L'on distingue, sur la carte 14, deux noyaux systémiques d'importance inégale: le noyau central ou oïlique, d'une part, et le noyau secondaire languedocien, de l'autre, à côté duquel l'on devine à grand-peine les restes d'une vieille agglomération provençale engloutie (intervalle 4).

La structuration harmonieuse du profil choroplèthe est d'autant plus étonnante qu'elle est le fruit de la synthèse quantitative (et partant en quelque sorte «automatique») de plus de vingt mille microstructures spatiales (ou aires) d'allure souvent très chaotique! Évidemment, le tracé «crescenti-forme» des plages polygonales en bleu foncé couronnant le domaine d'Oc vers le Nord, pourra contribuer à raviver le débat autour de l'extension et des caractères du «croissant» (cf. Brun-Trigaud 1990, passim).

7.2. Présentation des cartes 15-20: la synopse des coefficients d'asymétrie de Fisher (CAF)

Alors que l'écart-type constitue un indice pour la mesure du degré de dispersion d'une distribution de similarité, le «coefficient d'asymétrie de Fisher» (all. «Schiefe», angl. «skewness») sert à la mensuration d'une des propriétés les plus saillantes de n'importe quelle distribution de fréquence, à savoir de la symétrie. En comparant entre elles les silhouettes des histogrammes des cartes 2, 3, 6 et 9 l'on peut constater que leur symétrie est très variable. Ceci est dû, entre autres, au fait que les 640 scores de similarité sont répartis de façon inégale des deux côtés de la moyenne arithmétique. Or, quand la majorité des scores d'une distribution de similarité se situe au-dessus de la moyenne arithmétique, l'on peut dire, métaphoriquement parlant, que cette distribution de similarité s'insère «bien» dans les données de départ (c'est-à-dire dans les mailles du réseau de l'ALF) et que, partant, ses capacités communicatives sont «bonnes». L'inverse est vrai pour une distribution de similarité dont la plupart des scores se concentrent au-dessous de la moyenne arithmétique. La considération taxométrique du degré de symétrie (ou d'asymétrie) des 641 distributions

de similarités stockées dans notre matrice de similarité, et leur synopse cartographique sont donc du plus haut intérêt dialectologique et aussi dialectométrique.

Un des meilleurs indices pour la saisie de la symétrie d'une distribution de fréquence est le coefficient d'asymétrie de Fisher (CAF): pour la formule cf. Goebl 1981, 394-401; 1982, 44-46 et 1984 I, 150-153. Le CAF prend des valeurs négatives quand la distribution de similarité est asymétrique vers la droite (et accumule de ce fait les scores au-dessus de la moyenne arithmétique), et des valeurs positives quand la distribution de similarité penche vers la gauche (avec un surplus de scores de similarité au-dessous de la moyenne arithmétique). Quand en revanche la distribution de similarité est symétrique, le CAF adopte la valeur 0.

Le CAF est une des pièces fortes de la panoplie méthodique de la dialectométrie. Ceci surtout à cause du fait qu'il revêt une grande importance exploratoire du point de vue diachronique. C'est que le CAF permet de revoir un problème crucial (et partant familier à tous les romanistes) de la linguistique galloromane et de lui conférer une dimension scientifique nouvelle. Il s'agit des modalités de la genèse du faciès géolinguistique de la Galloromania actuelle. À ce propos, l'on sait depuis longtemps que le rôle directeur de Lugdunum/Lyon, vieille capitale politique et culturelle des Gaules, a été éclipsé, dans la seconde moitié du premier millénaire, par celui de Lutetia/Paris et que, de cette façon, l'espace galloroman est devenu le théâtre d'un rayonnement linguistique intense à partir de l'Ile-de-France, accompagné de provignements dans toutes les directions et aussi, par contre-coup, d'actes de persévérance linguistique plus ou moins marqués, d'où, entre autres, la genèse du francoprovençal et du domaine d'Oc avec son articulation intérieure caractéristique.

Ce processus d'expansion, de conservation et de retrait constitue donc, en dernière analyse, un changement linguistique à grande échelle accompagné de multiples interactions et contacts linguistiques à plus ou moins grande distance. Il est bien connu que l'analyse scriptologique des textes oïliques, francoprovençaux et occitans du Moyen Âge permet de saisir beaucoup d'aspects de ce drame bigarré et riche en facettes hétéroclites.

Or, l'analyse dialectométrique du degré de symétrie des distributions de similarité calculables pour un réseau de recherche donné, permet de capter de plus près les effets interactifs à plus ou moins longue distance qui sillonnent – tant dans le présent que, à plus forte raison, dans le passé – le territoire examiné. Il est donc possible de discerner, par la considé-

ration comparative des scores de symétrie, les endroits à forte expansion linguistique de ceux où se manifeste encore une certaine opposition face à l'avancement entropique du «brassage (ou compromis) linguistique» (all. «Sprachausgleich») général.

Regardons maintenant la carte 15! Elle a été dressée à l'aide de huit paliers chromatiques pour mieux en faire ressortir les accidents du profil iconique. Du point de vue linguistique, les couleurs bleues (bleu foncé et bleu moyen) peuvent être associées aux épicentres du compromis linguistique en progression qui – des Iles anglo-normandes jusqu'à la Lorraine, la Franche-Comté, la Bourgogne et le Bourbonnais – affecte des zones où la latinité centrale du domaine d'Oïl se frotte énergiquement aux latinités périphériques tout en les refoulant à l'instar d'une force militaire victorieuse. La même chose se passe aux pourtours du francoprovençal, ou la vieille latinité lugdunaise se heurte au Nord contre la latinité franc-comtoise et celle en provenance du Centre, alors qu'au Sud elle est aux prises avec les latinités narbonnaise et alpine.

Quant aux couloirs jaunes et ocres qui sillonnent le domaine d'Oc, ils représentent également, bien qu'à un degré mineur, des zones d'interaction et d'échange. En ce qui concerne par contre les plages vertes de la carte 16 (surface statistique lissée) ou les polygones verts répertoriés dans la classe 4 sur la carte 15, ils constituent, pour les domaines d'Oïl et francoprovençal, des zones «pacifiées» dans lesquelles le dynamisme du brassage (ou compromis) linguistique à grande échelle a été ralenti après avoir rejoint un certain taux d'enchevêtrements interrégionaux.

Étant donné que les cartes 15 et 16 représentent la synthèse de 20043 aires linguistiques isolées dont chacune dispose d'une géographie et d'une histoire particulière (cf. Christmann 1971), il en résulte, en dernière analyse, un portrait global de la genèse bimillénaire du faciès géolinguistique de la Galloromania.

Comme les programmes VDM et Map Info permettent d'appliquer encore d'autres techniques de visualisation, il est indiqué d'y recourir, surtout pour mieux encore présenter les effets de l'expansion circulaire du compromis linguistique au Nord et de la résistance ponctuelle de quelques grands domaines dialectaux au Sud. Parmi ces techniques figure le calcul de «surfaces statistiques lissées»: voir la carte 16.

Il s'agit d'une carte stéréographique, tridimensionnelle (en «3 D»), issue de calculs de lissage qui présuppose la fiction que la variable étudiée (en l'occurrence les scores-CAF) soit une grandeur à implantation spatiale

continue (à l'instar de la pression atmosphérique, de la pluie, de la température, etc.). Malgré cette entorse faite à la logique discrète du réseau-ALF, il en ressort une excellente vision d'ensemble de la stratification diatopique des scores-CAF au sein de la Galloromania.

7.2.1. Présentation des cartes 17 et 18: l'incidence des catégories linguistiques sur les résultats de l'analyse dialectométrique

L'exemple des cartes 9-12 nous a déjà démontré la manière dont les résultats de l'analyse dialectométrique dépendent de l'appartenance catégorielle (vocalisme, consonantisme, lexique, etc.) du corpus exploré. Les cartes 17 et 18 en offrent un nouvel exemple très éloquent. Dans une perspective cavalière, les profils choroplèthes des deux cartes correspondent de très près à celui de la carte 15 (basée sur le corpus intégral). Le Nord de la Galloromania est caractérisé par un rayonnement expansif circulaire (voir avant tout les polygones en bleu) alors que le Sud montre la persévérance de plusieurs noyaux (ou poches) de résistance (en rouge et orange). Il y a cependant, entre les deux cartes, une différence très importante: c'est la position divergente du francoprovençal. Du côté vocalique (carte 17) le francoprovençal occupe, face au domaine d'Oïl, une position fortement détachée sans se confondre d'ailleurs complètement avec le domaine d'Oc (voir à ce propos la courte barrière verte, formée de cinq polygones enchaînés, au Sud du domaine francoprovençal).

Du côté consonantique en revanche (carte 18), la zone francoprovençale semble avoir participé pleinement au dynamisme expansif d'Oïl. Il s'agit là d'une hybridation linguistique très intéressante qui avait déjà été pressentie vaguement pas G. I. Ascoli en 1874 (74 s.).

Fort de la leçon des cartes 9-12 d'un côté et des cartes 17-18 de l'autre, nous pouvons donc dire que l'hybridation catégorielle est une des particularités saillantes du croissant et de la couronne-nord de l'occitan d'un côté et du francoprovençal de l'autre.

7.2.2. Présentation des cartes 19 et 20: l'incidence de l'ampleur du réseau: la France de l'Ouest face à la France de l'Est

Le programme VDM permet en outre de manipuler avec aisance des corpus partiels comprenant seulement une partie des 641 points-ALF du réseau intégral. De telles manipulations sont indiquées surtout lorsqu'il s'agit d'étudier de plus près certains phénomènes découverts au cours de l'analyse du corpus intégral. Or, les phénomènes d'expansion circulaire

et de résistance ponctuelle dégagés au cours de l'établissement de la carte 15, méritent un examen approfondi. Il s'agit de voir la manière selon laquelle les phénomènes de brassage et de résistance linguistiques se déploient dans la seule moitié occidentale de la France, c'est-à-dire à l'écart du poids des données de la moitié orientale de notre réseau, et vice versa.

Comme le montrent les cartes 19 et 20, l'enjeu de cette expérience en valait vraiment la peine. Sur la carte 19 les zones en bleu foncé (intervalles 1-2) se concentrent surtout dans la région des Marches tout en représentant, une fois de plus, une agglomération oblongue et incurvée, semblable à un croissant. Mais la zone bleue s'étale également le long d'une bande allant de la Vendée jusque dans la Creuse qui, comme il est bien connu, constituait, durant le Haut Moyen Âge, le théâtre d'une forte poussée oïlique vers le Sud. À noter que la latinité limousine (voir les polygones respectifs en rouge) constitue, au sein de la seule moitié occidentale des données de l'ALF, un rempart de résistance comparable au gascon et au catalan du Roussillon.

Sur la carte 20 les zones-charnières du brassage linguistique (intervalles 1-2) délimitent le francoprovençal vers le Nord et aussi vers le Sud tout en sillonnant également la Franche-Comté, la Lorraine et l'espace picardo-wallon. À noter la formation d'un petit noyau dur en pleine Picardie (intervalle 5) qui fait écho à un phénomène déjà vaguement amorcé sur la carte 15 (voir les polygones verts, intervalle 4).

En comparant les cartes 15 (France entière), 19 (France de l'Ouest) et 20 (France de l'Est) et les divergences des profils choroplèthes respectifs qui y figurent, l'on constate que le profil de la moitié orientale de la carte 15 se différencie moins nettement de celui de la carte 20 que ne le fait le profil de la moitié occidentale de la carte 15 de celui de la carte 19. La raison en est que le poids de la variabilité géolinguistique interne de la moitié orientale de la France dépasse celui de la moitié occidentale. De là à dire que les épicentres de l'histoire linguistique de la France se situent plutôt à l'est d'une ligne allant du Havre à Marseille, il y a d'ailleurs loin.

8. La dialectométrie dendrographique: voir les cartes 21-24

Ce type de DM se sert – du point de vue mathématique – de la «classification ascendante hiérarchique» (CAH). On entend par là un certain nombre de processus taxométriques bien rodés depuis plusieurs décennies qui génèrent, par voie statistique, des hiérarchies arborescentes

de classes disjonctives. L'expression iconique (ou cartographique) directe en est l'arbre généalogique. Or, il est bien connu que la métaphore de l'arbre généalogique a été très souvent utilisée, en matière de linguistique, pour la description de relations de filiation et de parenté interdialectales ou interlinguales, et que la tradition en remonte au XIXe siècle (Schleicher 1863). Plus tard, des classifications dendrographiques ont fait leur apparition aussi en matière de lexicostatistique et de glottochronologie: pour plus ample information cf. les ouvrages très stimulants de Hoenigswald/Wiener 1987 et de Dyen/Kruskal/Black 1992 ainsi que nos remarques respectives publiées en 1991, 1992c et 1993c.

Pour une bonne intelligence des cartes 21-24 il faut connaître – au moins dans les grandes lignes – le principe taxométrique des algorithmes utilisés («Complete Linkage» et la méthode de Ward). Ces algorithmes sont décrits dans tous les manuels modernes de taxométrie: pour «Complete Linkage» cf. Sneath/Sokal 1973, 216 s.; Bock 1974, 382 s.; Chandon/Pinson 1981, 106 s.; pour la méthode «de Ward» cf. Bock 1974, 407 s. et Chandon/Pinson 1981, 123-124. Les deux algorithmes régissent un processus de fusions (ou d'agglomérations) qui comprend N – 1 étapes différentes. Il y a donc, pour N = 641, 640 fusions (ou agglomérations) différentes auxquelles correspondent autant de bifurcations (ou embranchements) binaires de la hiérarchie arborescente. Le processus agglomératif commence par considérer les maxima des scores de similarité (ou les minima des scores de distance) emmagasinés dans la matrice de similarité (N^2), et en fusionner – suivant des critères inhérents à l'algorithme respectif – ceux qui se différencient le moins. Après chaque fusion, l'algorithme arborescent génère une nouvelle matrice de similarité ((N – x) (N – x)) avec des dimensions toujours réduites, avant de reprendre son activité agglomératrice. En procédant ainsi, l'algorithme prend en compte les hétérogénéités tant «*intra*-group» (i. e. à *l'intérieur* des différents groupes ou classes d'objets) que «*inter*-group» (i. e. *entre* les différents groupes) qui, toutes les deux, augmentent au fur et à mesure que le processus d'agglomération s'approche de la racine de l'arbre: voir à ce propos, en bas des cartes 21 et 23, la ligne fléchée avec la légende «distance linguistique (hétérogénéité intra-group)».

La taxométrie moderne dispose d'un grand nombre de tels algorithmes agglomératifs. Évidemment, leur valeur classificatoire et leur utilisation heuristique dépendent des présuppositions théoriques et empiriques du classificateur. Des algorithmes «passe-partout» ou «bons à tout faire» n'existent pas. En tant que dialectométricien nous avons dû procéder comme tous les autres taxométriciens empiriques: expérimenter un certain

nombre d'algorithmes dendrographiques dans des conditions taxométriques diversifiées, construire les arbres respectifs et en apprécier l'utilité géolinguistique. Au cours de ces expériences, l'algorithme «Complete Linkage» et celui proposé par Ward se sont avérés très utiles. Les calculs nécessaires ont été exécutés à l'aide du logiciel VDM (développé par Edgar Haimerl) qui permet en outre, rapidement et à peu de frais, la spatialisation de la structure ramifiée des arborescences calculées, et le coloriage consécutif tant des «dendrèmes» de l'arbre que des «chorèmes»(5) de la spatialisation y ayant trait. Nous réitérons ici l'expression de notre profonde gratitude à l'adresse de M. Haimerl.

Les cartes 21 et 23 constituent des représentations simplifiées des arbres originaux qui, eux, disposent évidemment de 641 feuilles terminales. Aux 11 branches classificatoires (ou dendrèmes) de l'arbre de la carte 21 correspondent autant d'agrégations spatiales cohérentes (ou chorèmes) de la carte 22. Sur les cartes 23 et 24 par contre, l'on ne discerne que neuf dendrèmes d'un côté et neuf chorèmes de l'autre. À la différence des cartes 1-7 et 9-20 le coloriage des cartes 21-24 ne reflète que des faits qualitatifs et n'a donc aucune importance quantitative. Ce qui surprend au premier coup d'œil c'est l'étonnante cohésion spatiale des chorèmes des cartes 22 et 24. À cette cohésion territoriale ne font défaut que quelques rejetons détachés des chorèmes 2 et 6 sur la carte 22 et du chorème 2 sur la carte 24.

L'interprétation linguistique des cartes peut se faire de deux manières: 1) du point de vue diachronique, et 2) du point de vue synchronique.

8.1. Interprétation diachronique des cartes 21-24

Il s'agit d'une application de la pensée lexicostatistique qui veut que la richesse phénoménologique observable dans une région donnée, se soit développée, au cours du temps, à partir d'une phase initiale une et non diversifiée. Pour problématique que soit une telle théorie pour la Romania à l'heure de l'implantation de la langue latine, elle connaît de nombreux partisans parmi les linguistes romans. L'interprétation diachronique de nos cartes consistera dans une suite de 640 scénarios de fragmentations successives – d'ordre évidemment théorique – dont nous ne signalons ici que quelques-unes à titre d'exemple. Toutes les fragmentations (ou bifur-

(5) 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 *dendron* «arbre») en 1998 (296).

cations) se feront dans le sens de la flèche «distance diachronique (fragmentation diatopique)» (en haut sur les cartes 21 et 23).

Pour la carte 21 (algorithme *Complete Linkage*):

Scénario 1 (voir la ligne verticale bleue):
branche *A*: dendrème/chorème 1 (= italo-catalan)
branche *B*: dendrèmes/chorèmes 2-11 (= galloroman)

Scénario 2 (voir la ligne verticale verte):
branche *C*: dendrèmes/chorèmes 2-5 (= occitan du Sud)
branche *D*: dendrèmes/chorèmes 6-11 (= domaine d'Oïl avec la couronne-nord de l'occitan et le francoprovençal)

Scénario 3 (voir la ligne verticale rouge):
branche *E*: dendrème/chorème 2 (= provençal alpin oriental)
branche *F*: dendrèmes/chorèmes 3-5 (= occitan du Sud)
branche *G*: dendrème/chorème 6 (= couronne-nord de l'occitan)
branche *H*: dendrème/chorème 7 (= francoprovençal)
branche *I*: dendrèmes/chorèmes 8-11 (= domaine d'Oïl).

Pour la carte 23 (algorithme de *Ward*):

Scénario 1 (voir la ligne verticale bleue):
branche *A*: dendrèmes/chorèmes 1-4 (= domaine d'Oc)
branche *B*: dendrèmes/chorèmes 5-9 (= domaine d'Oïl)

Scénario *2* (voir la ligne verticale verte):
branche *C*: dendrèmes/chorèmes 1-4 (= domaine d'Oc)
branche *D*: dendrème/chorème 5 (= francoprovençal)
branche *E*: dendrèmes/chorèmes 6-9 (= domaine d'Oïl)

Scénario 3 (voir la ligne verticale rouge):
branche *F*: dendrèmes/chorèmes 1-2 (= limousin, auvergnat, provençal)
branche *G*: dendrème/chorème 3 (= languedocien)
branche *H*: dendrème/chorème 4 (= gascon)
branche *I*: dendrème/chorème 5 (= francoprovençal)
branche *J*: dendrèmes/chorèmes 6-7 (= domaine d'Oïl oriental)
branche *K*: dendrèmes/chorèmes 8-9 (= domaine d'Oïl central et occidental).

Dans la lignée de la pensée lexicostatistique, il est d'ailleurs possible d'associer, aux frontières des différents chorèmes des cartes 22 et 24, un âge plus ou moins important. C'est ainsi que, sur la carte 24, la frontière entre les chorèmes 1-4 d'un côté et 5-6 de l'autre est la plus ancienne. La même chose vaut pour l'âge théorique des chorèmes et des espaces dialectaux respectifs. C'est ainsi que, dans la perspective avec laquelle la DM dendrographique (allant de pair, en l'occurrence, avec la lexicostatistique)

regarde les cartes 23 et 24, la genèse du francoprovençal (dendrème/chorème 5, branche *H*) est antérieure à celle du picardo-wallon (dendrème/chorème 6) et aussi à celle du languedocien (dendrème/chorème 3) ou du gascon (dendrème/chorème 4). La raison en est que la bifurcation où pend le francoprovençal, est plus proche de la racine de l'arbre que celle de laquelle dépendent le languedocien et le gascon: nous conseillons vivement aux lecteurs d'analyser à ce propos attentivement la structure ramifiée complexe de la carte 23.

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 de la structure arborescente et aussi l'âge réel des données brutes (= ALF: exploré entre 1897 et 1901!!). Toujours est-il qu'à la richesse statistique souvent déroutante des méthodes dendrographiques modernes correspond une multiplicité iconique non moins troublante des arbres généalogiques qui en dérivent.

À côté des arbres des cartes 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 le lecteur non averti – 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. Interprétation synchronique des cartes 21-24

Le but principal d'une telle interprétation est de saisir les liens de proximité linguistique dans l'espace et de les cartographier par la suite. Prenons, à titre d'exemple, l'affiliation changeante de l'espace francoprovençal et de la couronne-nord de l'occitan (croissant, limousin, auvergnat). Sur les cartes 21 et 22, ces deux espaces (ou macro-chorèmes) s'agrègent – selon la logique de l'algorithme hiérarchique ascendant «Complete Linkage» – aux macro-dendrèmes oïliques (branche *D* sur la carte 21) tout en occupant deux des trois branches supérieures (*G* et *H* sur la carte 21), alors que, sur la carte 23 (établie suivant l'algorithme de Ward), la couronne-nord de l'occitan est incorporée aux autres branches de l'occitan (ramification *F* sur la carte 23) d'un côté, et l'espace francoprovençal (dendrème/chorème 5, ramification *I*) aux branches d'Oïl de l'autre. Mais les deux groupements occupent toujours, dans le cadre de leurs affiliations

dans les arbres respectifs, des positions périphériques tout en gardant d'ailleurs leur remarquable cohérence interne. Ceci n'exclut pas certaines modifications mineures de leur étendue chorématique: c'est ainsi que le chorème francoprovençal (no. 7 sur la carte 22) est légèrement plus grand que celui de la carte 24 (chorème 5) alors que le contraire est vrai pour la couronne-nord de l'occitan: étendue plus réduite sur la carte 22 (chorème 6) par rapport à la carte 24 (chorème 2).

Évidemment, il serait souhaitable de pousser plus avant l'analyse et la spatialisation successive des dendrèmes de nos deux arbres – chose aisément réalisable devant l'écran de l'ordinateur et à l'aide de l'admirable logiciel VDM – mais irréalisable dans le cadre de cet article. Soulignons encore que l'étonnante cohérence spatiale des macro-dendro-/chorèmes des cartes 21-24 vaut également pour des analyses plus fines, c'est-à-dire faites à l'intérieur des macro-chorèmes visualisés sur les cartes 21-24.

9. Épilogue

Nous concluons avec une brève remarque d'ordre purement pratique (9.1.) et une autre plus longue et d'inspiration épistémologique (9.2.)

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

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)[6] 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»[7] – 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.

---

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

(7) Pour l'historique de cette querelle cf. les bilans critiques contemporains d'Adolf Horning (1893) et de Louis Gauchat (1903) ainsi que nos analyses méthodiques rétrospectives de 1986, 1990 (version allemande) et 1995 (version italienne). La paternité de la consigne 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(8) 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(9). En dernière analyse il s'agit là d'un épiphénomène de la fameuse «Querelle des Universaux» du Moyen Âge (cf. Stegmüller 1956/57).

L'exemple des sciences naturelles, méthodiquement plus mûres en la matière que les sciences humaines (y compris aussi les 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 XXe 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 (cf. Rosch/Llloyd 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.

Université de Salzbourg. Hans GOEBL
Slawomir SOBOTA*
Blaustein (Bade-Wurtemberg). Edgar HAIMERL**

(8) 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.

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

* Pour la cartographie assistée par l'ordinateur.

** Pour la programmation du logiciel VDM.

Références bibliographiques

AIS: Jaberg, Karl/Jud, Jakob (eds.) Sprach- und Sachatlas Italiens und der Südschweiz, Zofingen 1928-1940, 8 vol. (réimpression: Nendeln 1971).

ALD-I: Goebl, Hans/Bauer, Roland/Haimerl, Edgar et alii (eds.): Atlant linguistich dl ladin dolomitic y di dialec vejins, 1ª pert. Atlante linguistico del ladino dolomitico e dei dialetti limitrofi, 1ª parte. Sprachatlas des Dolomitenladinischen und angrenzender Dialekte, 1. Teil, Wiesbaden 1998, 7 vol.; accompagné de 3 CD-ROM, Salzbourg 1999-2000.

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

Altmann, Gabriel: Die Entstehung diatopischer Varianten. Ein stochastisches Modell, in: Zeitschrift für Sprachwissenschaft 4 (1985) 139-155.

Altmann, Gabriel/Lehfeldt, Werner: Allgemeine Sprachtypologie. Prinzipien und Meßverfahren, München 1973.

Ascoli, Graziadio Isaia: Schizzi franco-provenzali [1874], in: Archivio glottologico italiano 3 (1878) 61-120.

Ascoli, Graziadio Isaia: P. Meyer e il franco-provenzale, in: Archivio glottologico italiano 2 (1876) 385-395.

Bailey, Kenneth D.: Typologies and Taxonomies. An Introduction to Classification Techniques, Thousand Oaks, Londres, New Delhi 1994.

Berschin, Helmut/Felixberger, Josef/Goebl, Hans: Französische Sprachgeschichte. Lateinische Basis. Interne und externe Geschichte. Sprachliche Gliederung Frankreichs, Munich 1978.

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

Broadfield, A.: The Philosophy of Classification, Londres 1946.

Bromberger, Christian/Morel, Alain (eds.): Limites floues, frontières vives. Des variations culturelles en France et en Europe, Paris 2001.

Brun-Trigaud, Guylaine: Le Croissant: le concept et le mot. Contribution à l'histoire de la dialectologie française au XIXe siècle, Lyon 1990.

Brunet, Roger: La carte. Mode d'emploi, Paris 1987.

Centre National de Recherches de Logique/Société belge de Logique et Philosophie des Sciences (eds.): La classification dans les sciences, Gembloux 1963.

Chandon, Jean-Louis/Pinson, Suzanne: Analyse typologique. Théories et applications, Paris, New York, Barcelone, Milan 1981.

Christmann, Hans Helmut: Lautgesetze und Wortgeschichte. Zu dem Satz «Jedes Wort hat seine eigene Geschichte», in: Stempel, Wolf-Dieter/Coseriu, Eugenio (eds.): Sprache und Geschichte. Festschrift für Harri Meier zum 65. Geburtstag, Munich 1971, 111-124.

Dos Santos, José Rodrigues: La constitution historique des «Cévennes». Réflexions sur le statut des entités culturelles, in: Bromberger/Morel 2001, 183-217.

Dyen, Isidore/Kruskal, Joseph. B./Black, Paul: An Indoeuropean Classification: A Lexicostatistical Experiment, Philadelphia 1991.

Ettmayer, Karl von: Über das Wesen der Dialektbildung, erläutert an den Dialekten Frankreichs, in: Denkschriften [Mémoires] der Akademie der Wissenschaften in Wien, philosophisch-historische Klasse, vol. 66, Vienne 1924, 1-56, 7 planches.

FEW: Wartburg, Walther von: Französisches etymologisches Wörterbuch, Bonn 1922 ss.

Gardette, Pierre: Le francoprovençal. Son histoire, ses origines [1967/74], in: id.: Études de géographie linguistique [Horiot, Brigitte/Simoni, Marie-Rose/Straka, Georges (eds.)], Strasbourg 1983, 569-584.

Gauchat, Louis: Gibt es Mundartgrenzen?, in: Archiv für das Studium der neueren Sprachen und Literaturen 111 (1903) 365-403.

Gilliéron, Jules: Généalogie des mots qui désignent l'abeille d'après l'Atlas linguistique de la France, Paris 1918.

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

Goebl, Hans: Dialektometrie. Prinzipien und Methoden des Einsatzes der Numerischen Taxonomie im Bereich der Dialektgeographie, in: Denkschriften [Mémoires] der Österreichischen Akademie der Wissenschaften, philosophisch-historische Klasse, vol. 157, Vienne 1982, 1-123.

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

Goebl, Hans: Dialektometrische Studien. Anhand italoromanischer, rätoromanischer und galloromanischer Sprachmaterialien aus AIS und ALF, Tübingen 1984, 3 vol.

Goebl, Hans: Typophilie und Typophobie. Zu zwei problembeladenen Argumentationstraditionen innerhalb der Questione ladina, in: Holtus Günter/Ringger, Kurt (eds.): Raetia antiqua et moderna. Wilhelm Theodor Elwert zum 80. Geburtstag, Tübingen 1986, 513-536.

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

Goebl, Hans: «Ma il distintivo necessario del determinato tipo sta appunto nella simultanea presenza o nella particolar combinazione di quei caratteri». Methodische und wissenschaftsgeschichtliche Bemerkungen zum Diskussionskomplex «unità ladina», in: Ladinia 14 (1990) 219-257.

Goebl, Hans: Problèmes et méthodes de la dialectométrie actuelle (avec application à l'AIS), in: Euskaltzaindia / Académie de la langue basque (ed.): Nazioarteko dialektologia biltzarra. Agiriak / Actes du Congrès international de dialectologie (Bilbo/Bilbao 1991), Bilbo/Bilbao 1992a, 429-475.

Goebl, Hans: Una classificazione gerarchica di dati geolinguistici tratti dall'AIS. Saggio di dialettometria dendrografica, in: Rivista italiana di dialettologia 16 (1992b) 67-80.

Goebl, Hans: Dendrogramme im Dienst der Dialektometrie. Zwei hierarchisch-agglomerative Klassifikationen von Daten des Sprachatlasses AIS, in: Klenk, Ursula (ed.): Computatio linguae. Aufsätze zur algorithmischen und quantitativen Analyse der Sprache, Stuttgart 1992b, 54-73.

Goebl, Hans: Probleme und Methoden der Dialektometrie. Geolinguistik in globaler Perspektive, in: Viereck, Wolfgang (ed.): Verhandlungen des Internationalen Dialektologenkongresses Bamberg 1990, Stuttgart 1993a, vol. I, 37-81.

Goebl, Hans: Dialectometry. A Short Overview of the Principles and Practice of Quantitative Classification of Linguistic Atlas Data, in: Köhler, Reinhard/ Rieger, Burghard (eds.): Contributions of Quantitative Linguistics, Dordrecht, Boston, Londres 1993b, 277-315.

Goebl, Hans: Eine hierarchisch-agglomerative Klassifikation von Daten des Sprachatlasses AIS, in: Schmidt-Radefeldt, Jürgen/Harder, Andreas (eds.): Sprachwandel und Sprachgeschichte. Festschrift für Helmut Lüdtke zum 65. Geburtstag, Tübingen 1993c, 63-78.

Goebl, Hans: Che cos'è un geotipo? Il problema dell'unità ladina in chiave ascoliana, in: Banfi, Emanuele/Bonfadini, Giovanni/Cordin, Patrizia/Iliescu, Maria (eds.): Italia settentrionale: crocevia di idiomi romanzi. Atti del Convegno internazionale di studi (Trento 1993), Tübingen 1995, 103-131.

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

Goebl, Hans: La dialectométrisation de l'ALF: présentation des premiers résultats, in: Linguistica 40 (2000) 209-236.

Gordon, A. D.: Classification, Boca Raton/Londres/New York/Washington D. C. 1999.

Haag, Carl: Die Mundarten des oberen Neckar- und Donautales (Schwäbisch-alemannisches Grenzgebiet: Baarmundarten), Reutlingen 1898.

Hasselrot, Bengt: Les limites du francoprovençal et l'aire de *nostron*, in: Revue de linguistique romane 30 (1966) 257-266.

Hoenigswald, Henry M./Wiener, Linda F. (eds.): Biological Metaphor and Cladistic Classification. An Interdisciplinary Perspective, Philadelphia 1987.

Horning, Adolf: Über Dialektgrenzen im Romanischen, in: Zeitschrift für romanische Philologie 17 (1893) 160-187 (réimprimé in: Spitzer, Leo (ed.): Meisterwerke der romanischen Sprachwissenschaft, Munich 1930, vol. II, 264-298).

Kreyszig, Erwin: Statistische Methoden und ihre Anwendungen, Göttingen 1968.

Lalanne, Théodore: Indice de polyonymie, indice de polyphonie, in: Français moderne 21 (1953) 263-274.

Mahmood, Cynthia K./Armstrong, Sharon L.: Do Ethnic Groups exist? A Cognitive Perspective on the Concepts of Cultures, in: Ethnology 31 (1992) 1-14.

Martin, Jean-Baptiste: Francoprovençal, in: Holtus, Günter/Metzeltin, Michael/ Schmitt, Christian (eds.): Lexikon der romanistischen Linguistik, Tübingen, vol. V/1 (1990) 671-685.

Meyer, Paul: compte rendu de: Ascoli 1874, in: Romania 4 (1875) 293-296.

Moriconi, Anna: La classificazione nelle scienze sociali, Formello 2000.

Motta, Giuseppe (ed.): Atlante storico, Novara 1995.

Muller, Charles: Initiation à la statistique linguistique, Paris 1968.

Needham, Rodney: Polythetic Classifications: Convergence and Consequences, in: Man 10 (1975) 349-369.

Nordman, Daniel: Frontières de France. De l'espace au territoire. XVIe-XIXe siècle, Paris 1998.

Okabe, Atsuyuki/Boots, Barry/Sugihara, Kokichi: Spatial Tesselations. Concepts and Applications of Voronoi Diagrams, Chichester/New York/Brisbane/Toronto/ Singapour 1992.

Paris, Gaston: compte rendu de: Adam, Lucien: Les patois lorrains, Paris 1881, in: Romania 10 (1881) 601-609.

Paris Gaston: Les parlers de France [1888], in: id.: Mélanges linguistiques, Paris 1909, 432-448.

Parrochia, Daniel (ed.): Mathématiques & Existence. Ordres. Fragments. Empiétements, Seyssel 1991.

Ronjat, Jules: Essai de syntaxe des parlers provençaux modernes, Mâcon 1913.

Ronjat, Jules: Grammaire Istorique des Parlers Provençaux modernes, Montpellier 1930-1932, 2 vol. (réimpression: Genève/Marseille 1980).

Rosch, Eleanor/Lloyd, Barbara B. (eds.): Cognition and Categorization, Hillsdale N. J. 1978.

Rosenqvist, Arvid: Limites administratives et division dialectale de la France, in: Neuphilologische Mitteilungen 20 (1919) 87-119.

Sankoff, David: Mathematical Developments in Lexicostatistic Theory, in: Current Trends in Linguistics 11 (1973) 93-113.

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

Sériot, Patrick: Structure et totalité. Les origines intellectuelles du structuralisme en Europe centrale et orientale, Paris 1999.

Sneath, Peter A./Sokal, Robert R.: Numerical Taxonomy. The Principles and Practice of Numerical Classification, San Francisco 1973.

Stegmüller, Wolfgang: Das Universalienproblem einst und jetzt, in Archiv für Philosophie 6 (1956) 192-225 et 7 (1957) 45-81 (réimpression: Darmstadt 1965).

Suchier, Hermann: Die französische und provenzalische Sprache und ihre Mundarten, in: Gröber Gustav (ed.): Grundriß der romanischen Philologie, Strasbourg 1888, vol. I, 561-668.

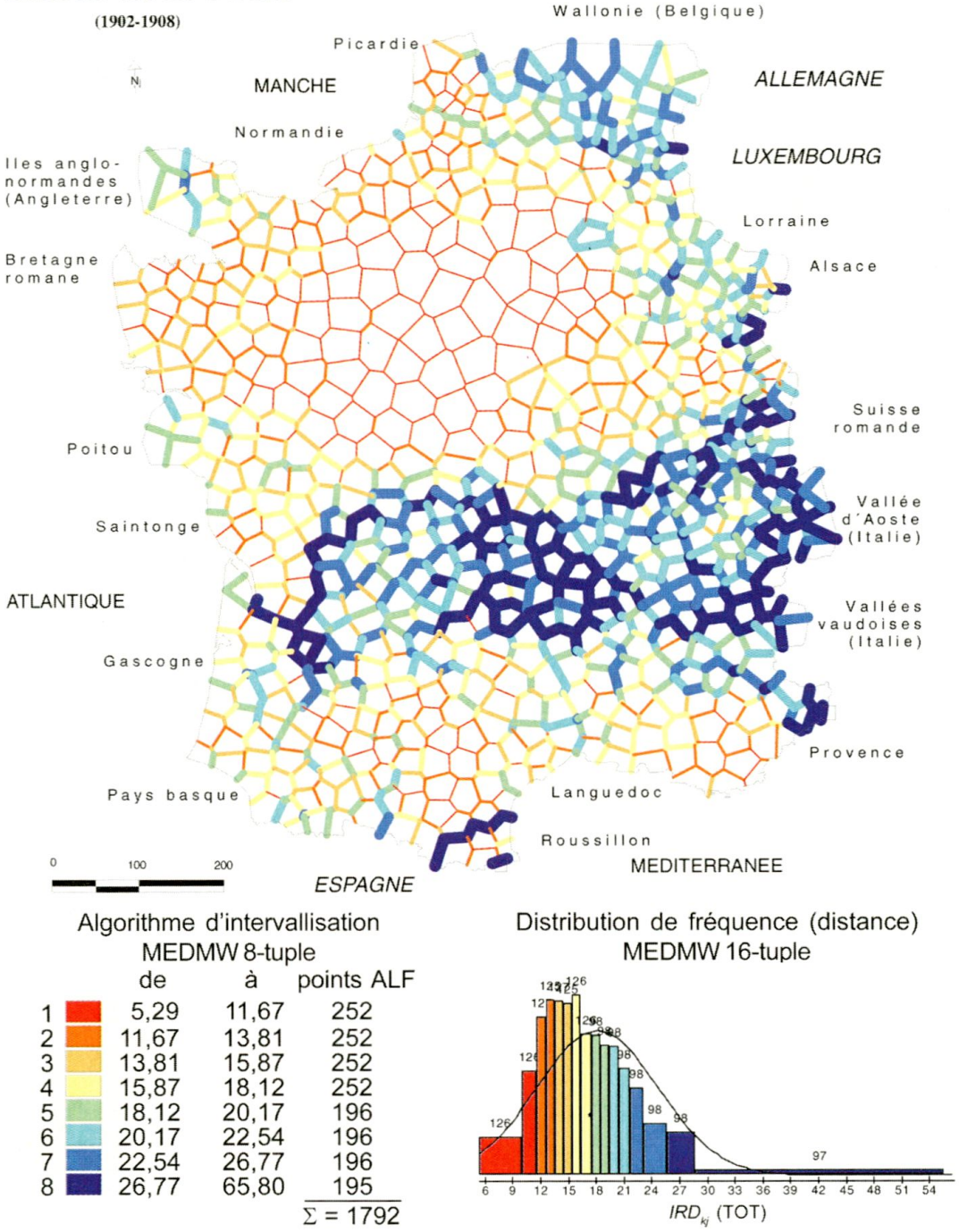
Tourtoulon, Charles de/Bringuier, Octavien: Rapport sur la limite géographique de la langue d'oc et de la langue d'oil [sic], in Archives des Missions scientifiques et littéraires III/3 (1876) 544-605, 1 carte.

Tuaillon, Gaston: Le francoprovençal: progrès d'une définition, in: Travaux de linguistique et de littérature 10/1 (1972) 293-339 (réimpression: Saint-Nicolas, Vallée d'Aoste 1983, 1994).

ALF

Série A: cartes 1-1421

(1902-1908)

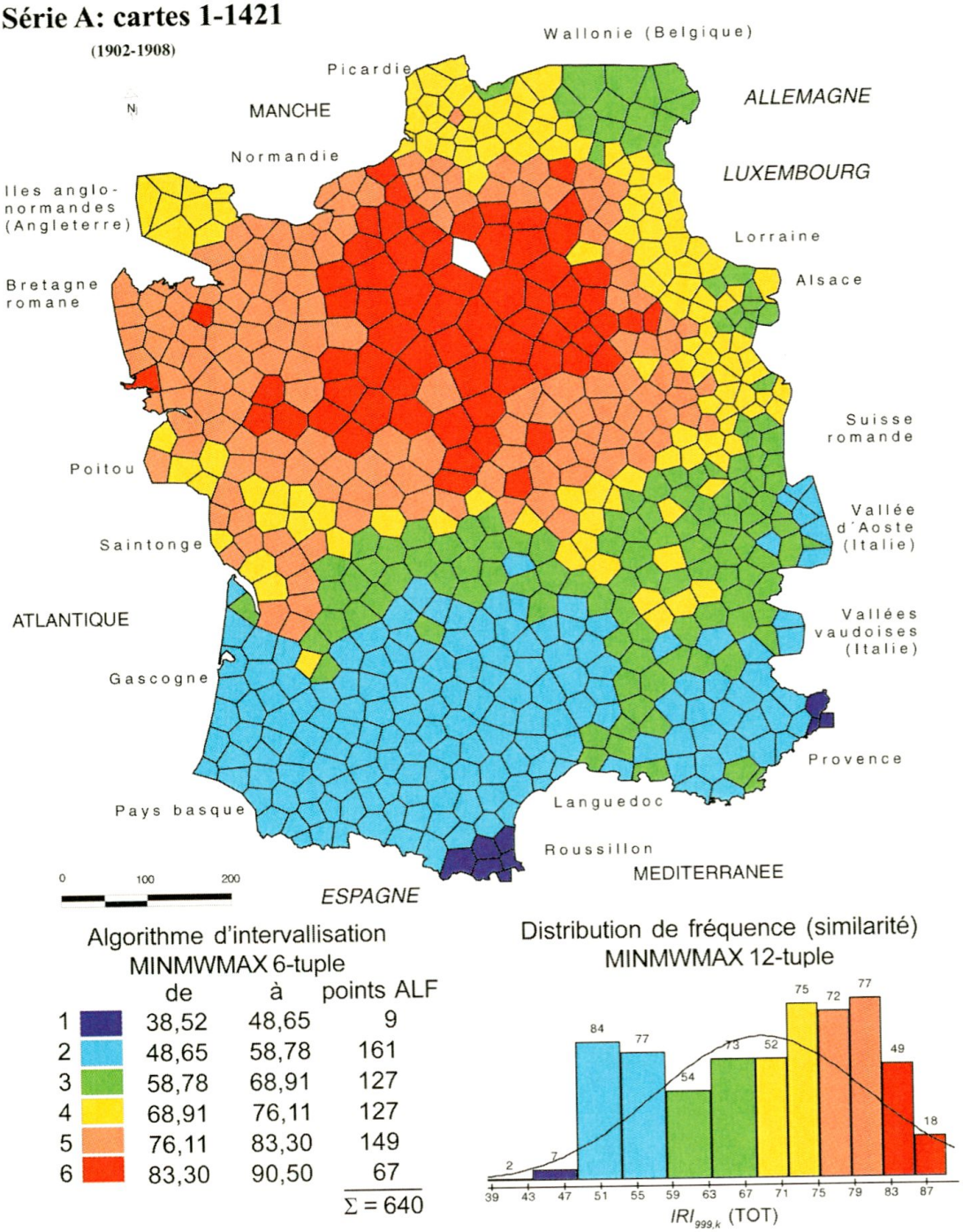


Carte 1: Carte à interpoints en fonction discriminatoire (isoglosses quantitatives).
Corpus: 1687 cartes de travail.
Nombre des paliers d'épaisseur variable des segments de polygone: 8.
Algorithme d'intervallisation: MEDMW. Voir aussi le chapitre 5.2.

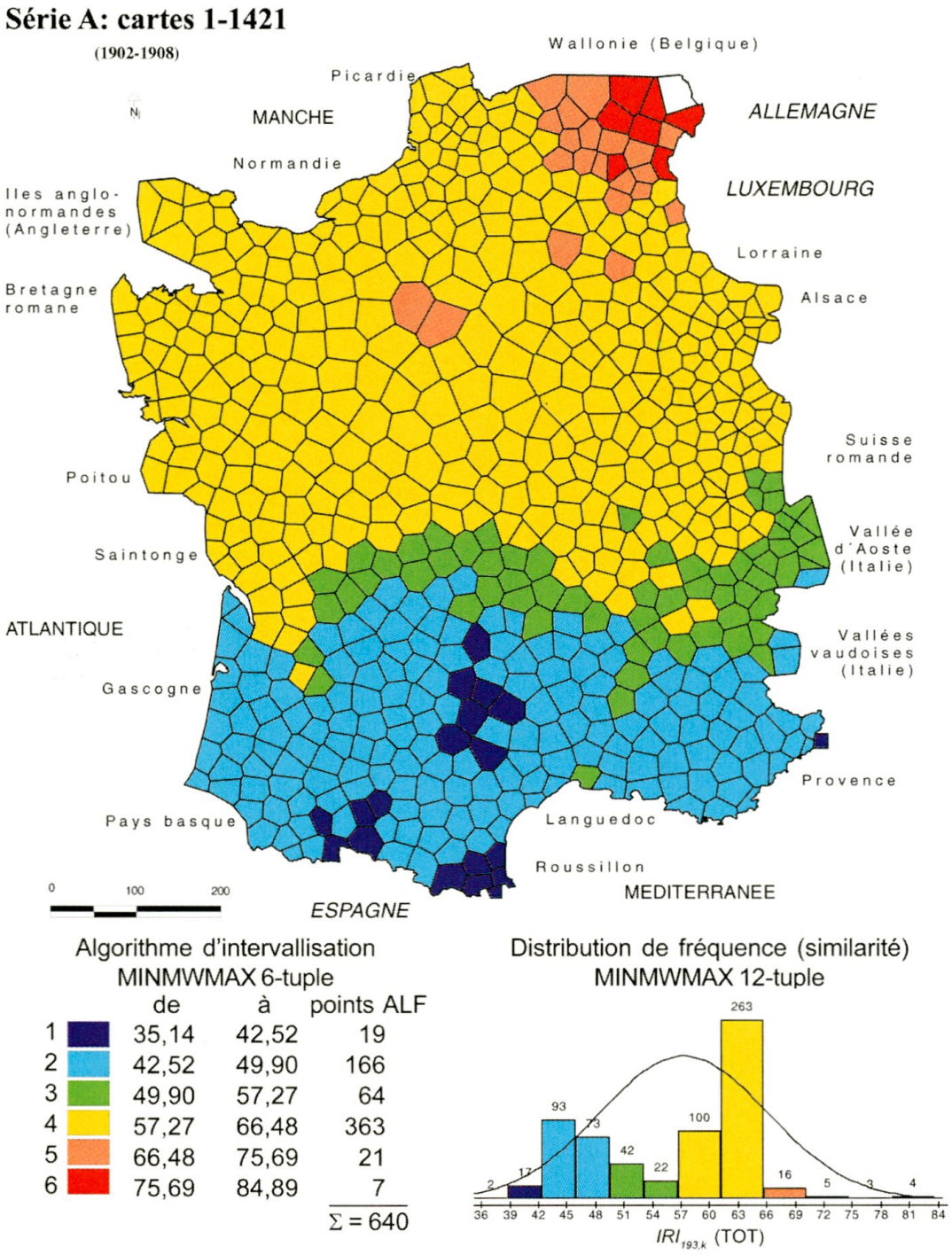
ALF

Série A: cartes 1-1421

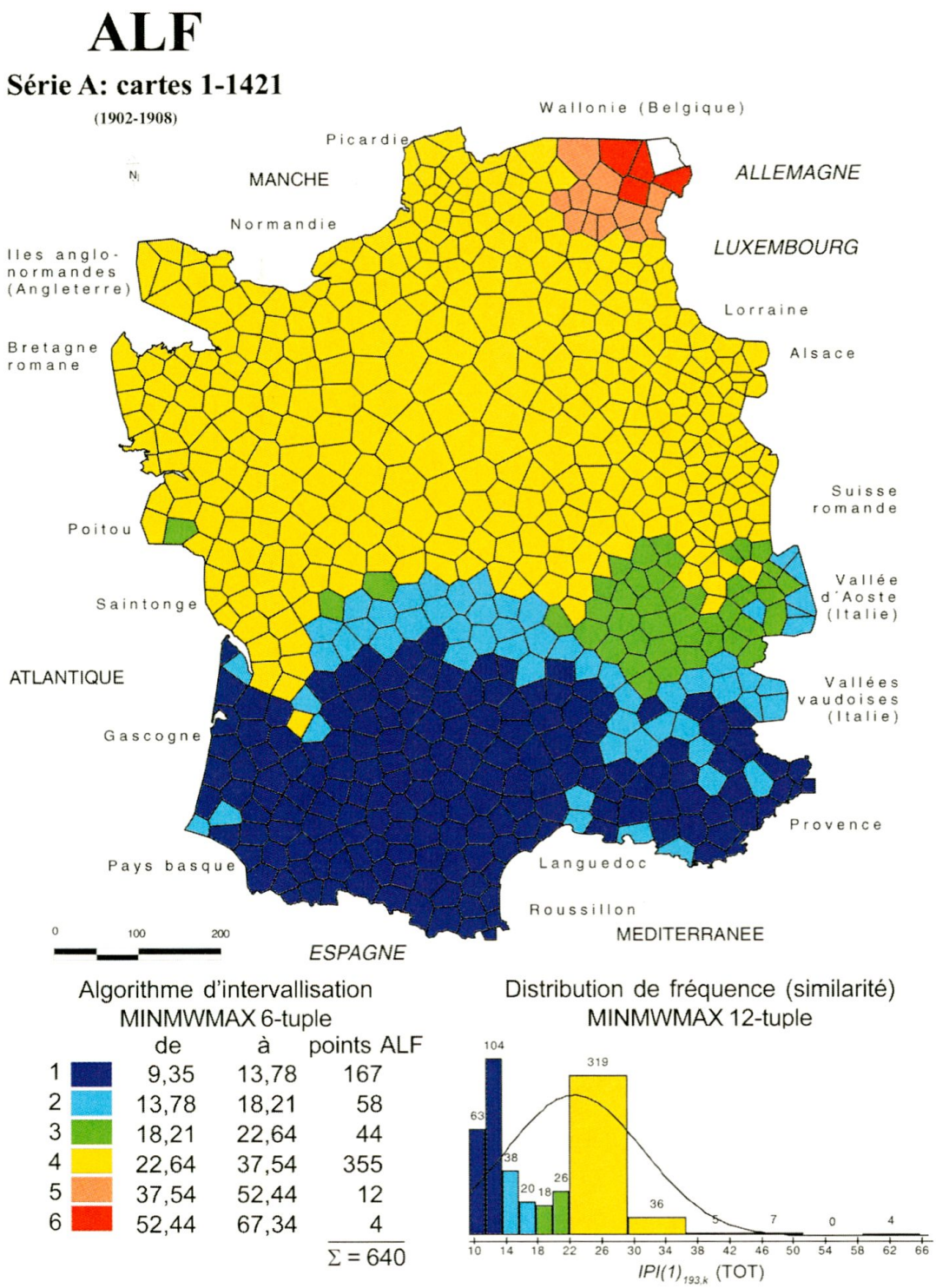
(1902-1908)



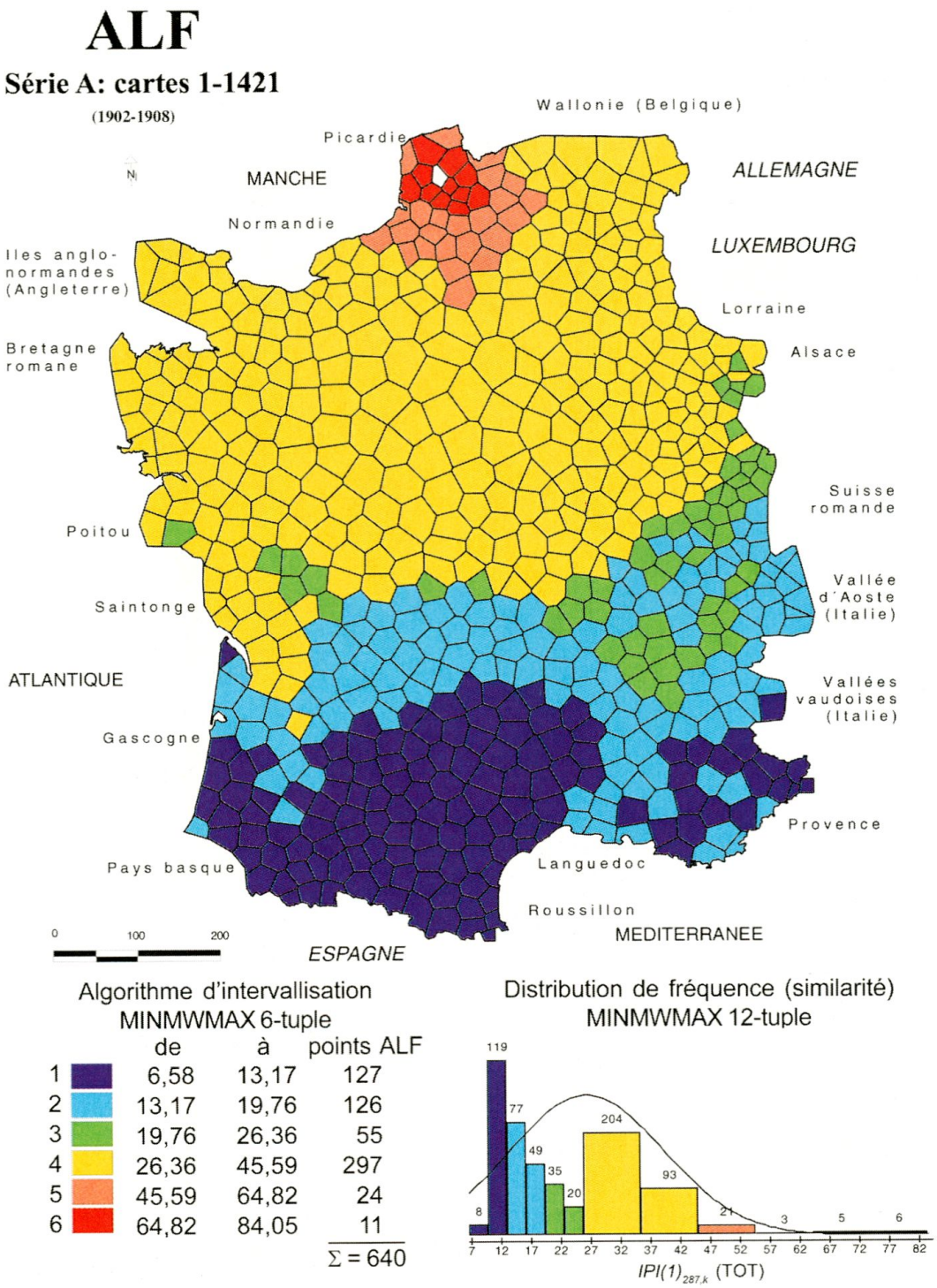
Carte 2: Carte choroplèthe de la distribution de similarité ($IRI_{999,k}$) relative au point de référence factice 999 (français standard). Corpus: 1687 cartes de travail. Voir aussi le chapitre 6.2.



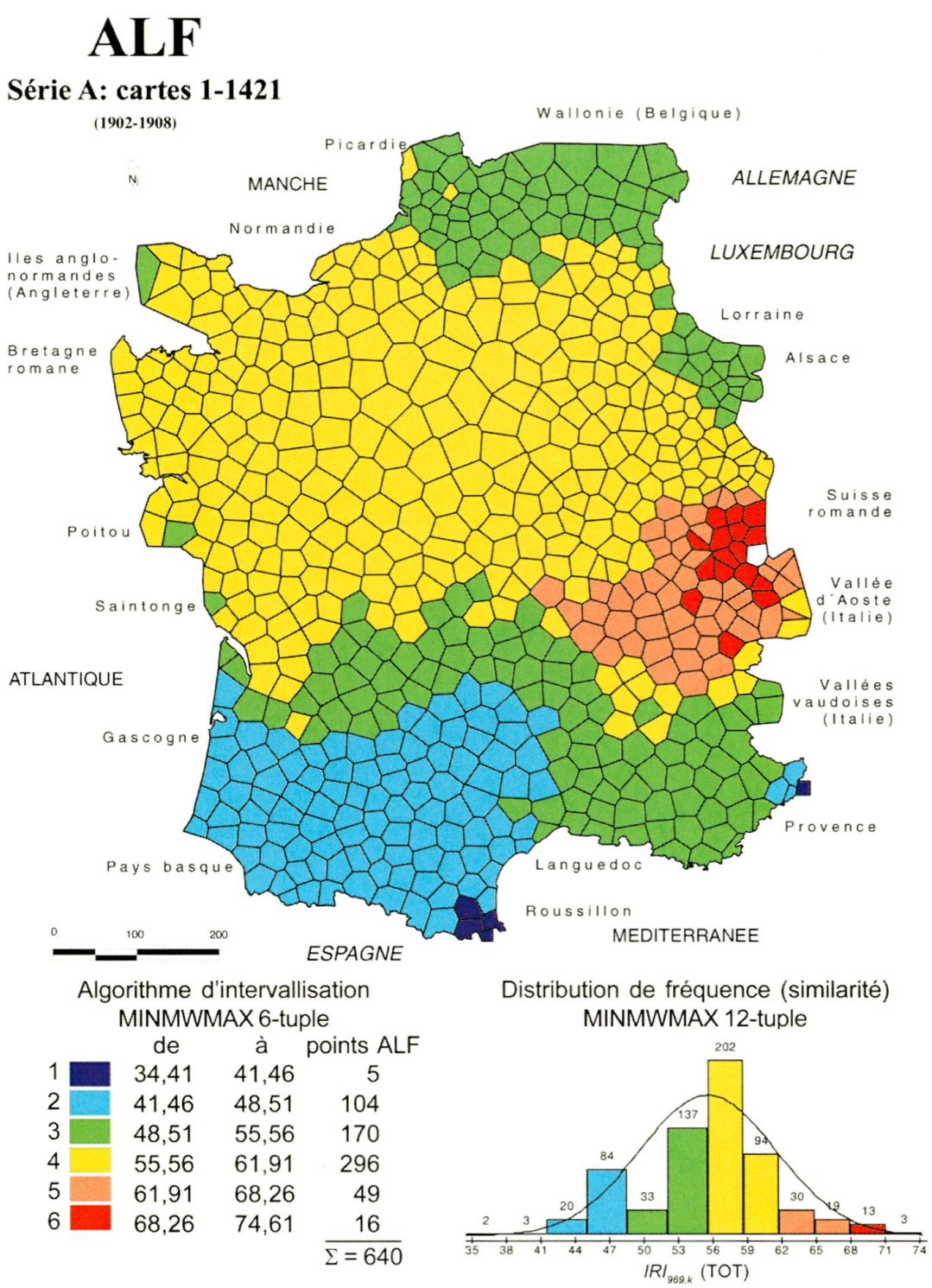
Carte 3: Carte choroplèthe de la distribution de similarité ($IRI_{193,k}$) relative au point-ALF 193 (Dolhain, Wallonie).
Corpus: 1687 cartes de travail. Voir aussi le chapitre 6.3.



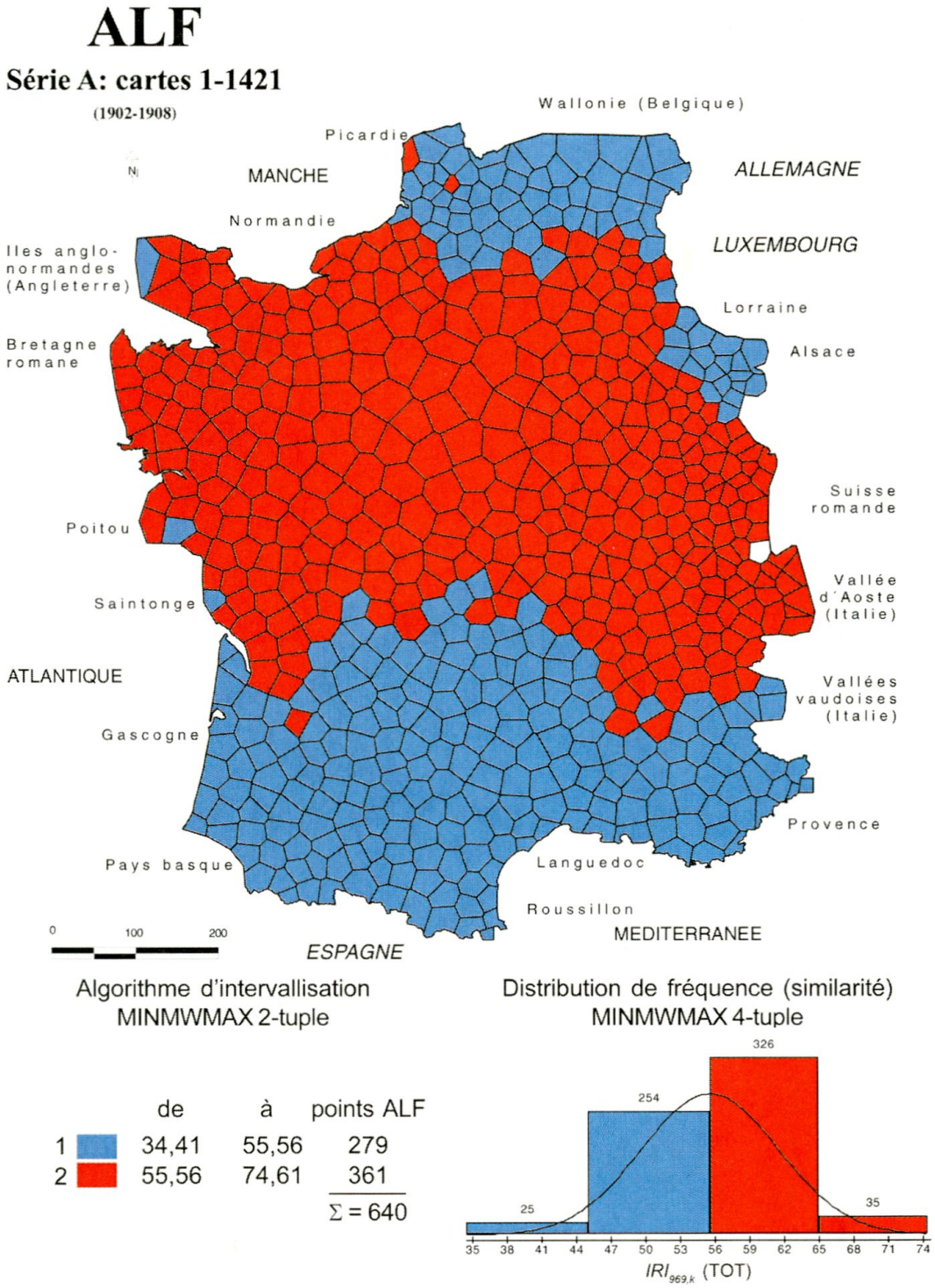
Carte 4: Carte choroplèthe de la distribution de similarité ($IPI(1)_{193,k}$) relative au point-ALF 193 (Dolhain, Wallonie).
Corpus: 1687 cartes de travail. Voir aussi le chapitre 6.3.



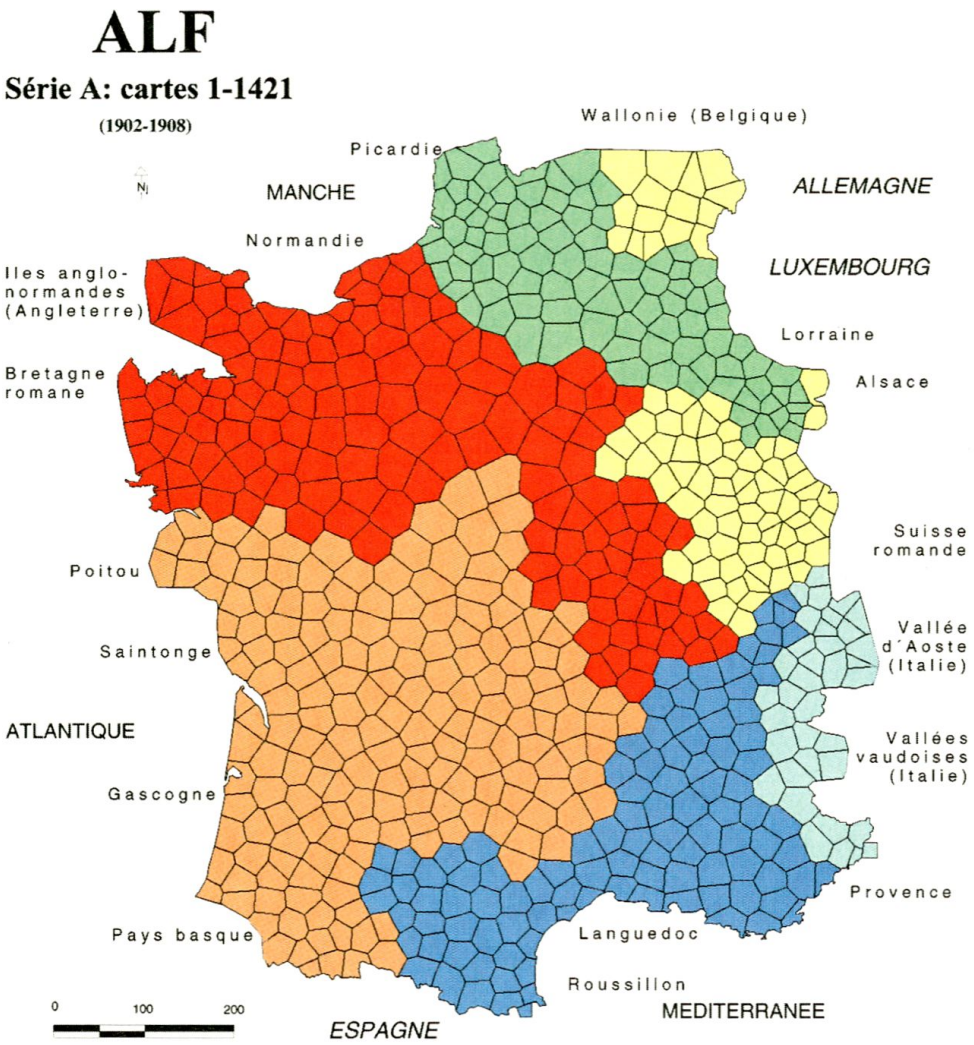
Carte 5: Carte choroplèthe de la distribution de similarité ($IPI(1)_{287,k}$) relative au point-ALF 287 (Teneur, Pas-de-Calais).
Corpus: 1687 cartes de travail. Voir aussi le chapitre 6.4.



Carte 6: Carte choroplèthe de la distribution de similarité ($IRI_{969,k}$) relative au point-ALF 969 (L'Étivaz, canton de Vaud).
Corpus: 1687 cartes de travail. Algorithme d'intervallisation: MINMWMAX 6-tuple.
Voir aussi la carte 7 et le chapitre 6.5.



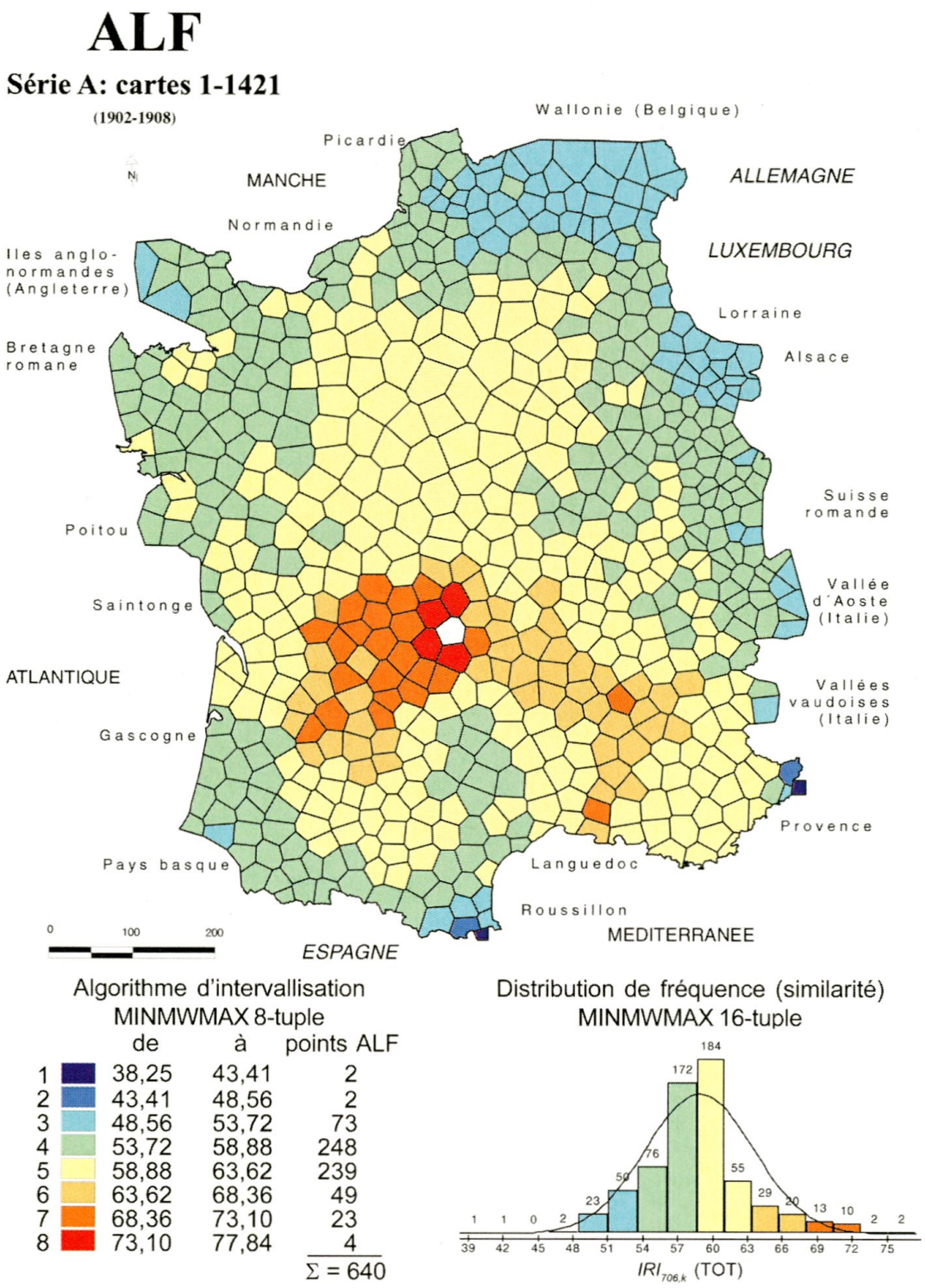
Carte 7: Carte choroplèthe de la distribution de similarité ($IRI_{969,k}$) relative au point-ALF 969 (L'Étivaz, canton de Vaud). Corpus: 1687 cartes de travail. Algorithme d'intervallisation: MINMWMAX 2-tuple. Voir aussi la carte 6 et le chapitre 6.5.



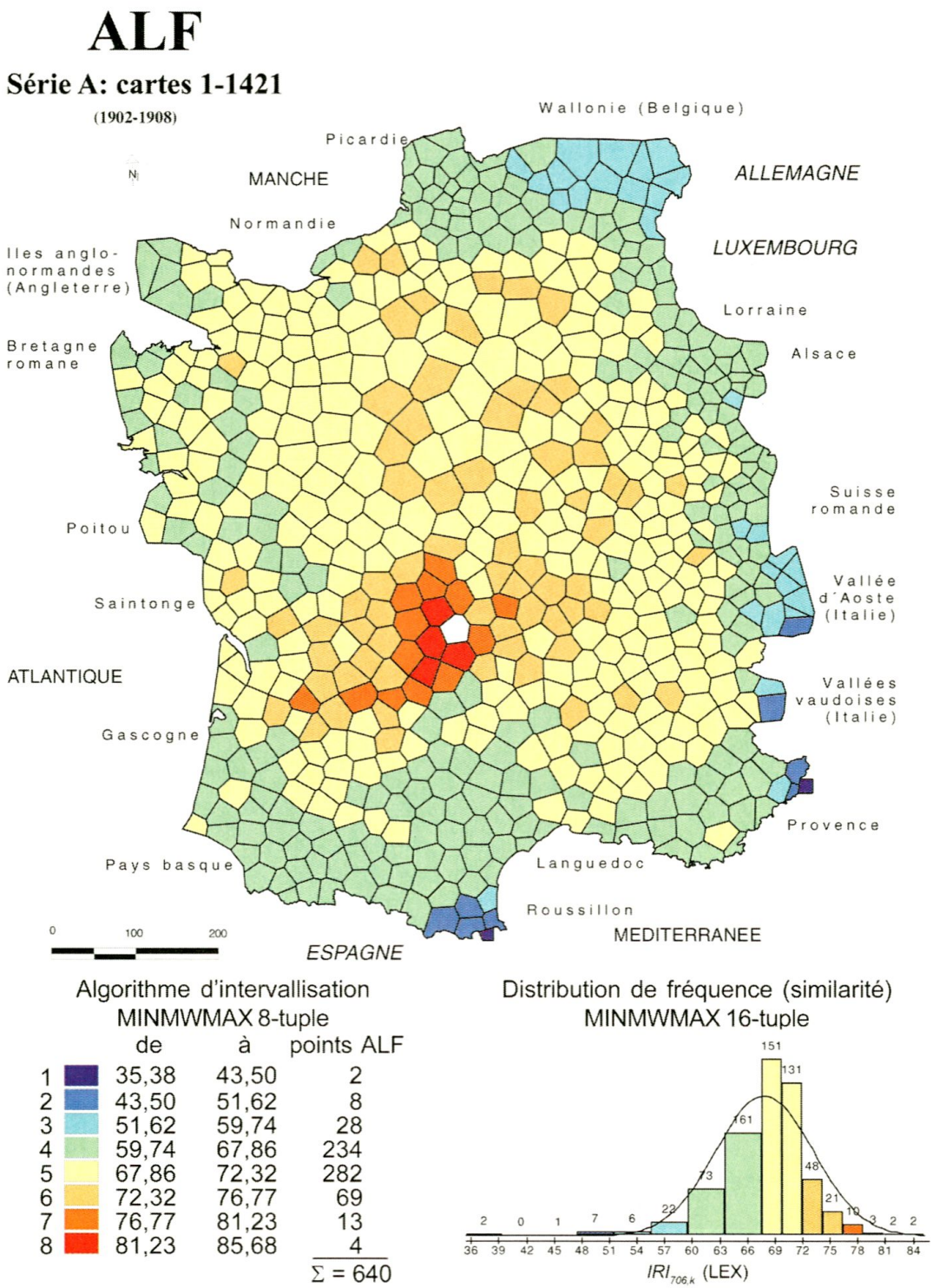
Les provinces romaines de la Gaule au début de 2ème siècle après Jésus-Christ

1	Lugdunensis	144
2	Aquitania	181
3	Germania (inferior et superior)	81
4	Belgica	94
5	Narbonensis	105
6	Gallia cisalpina	36
		Σ = 641 points ALF

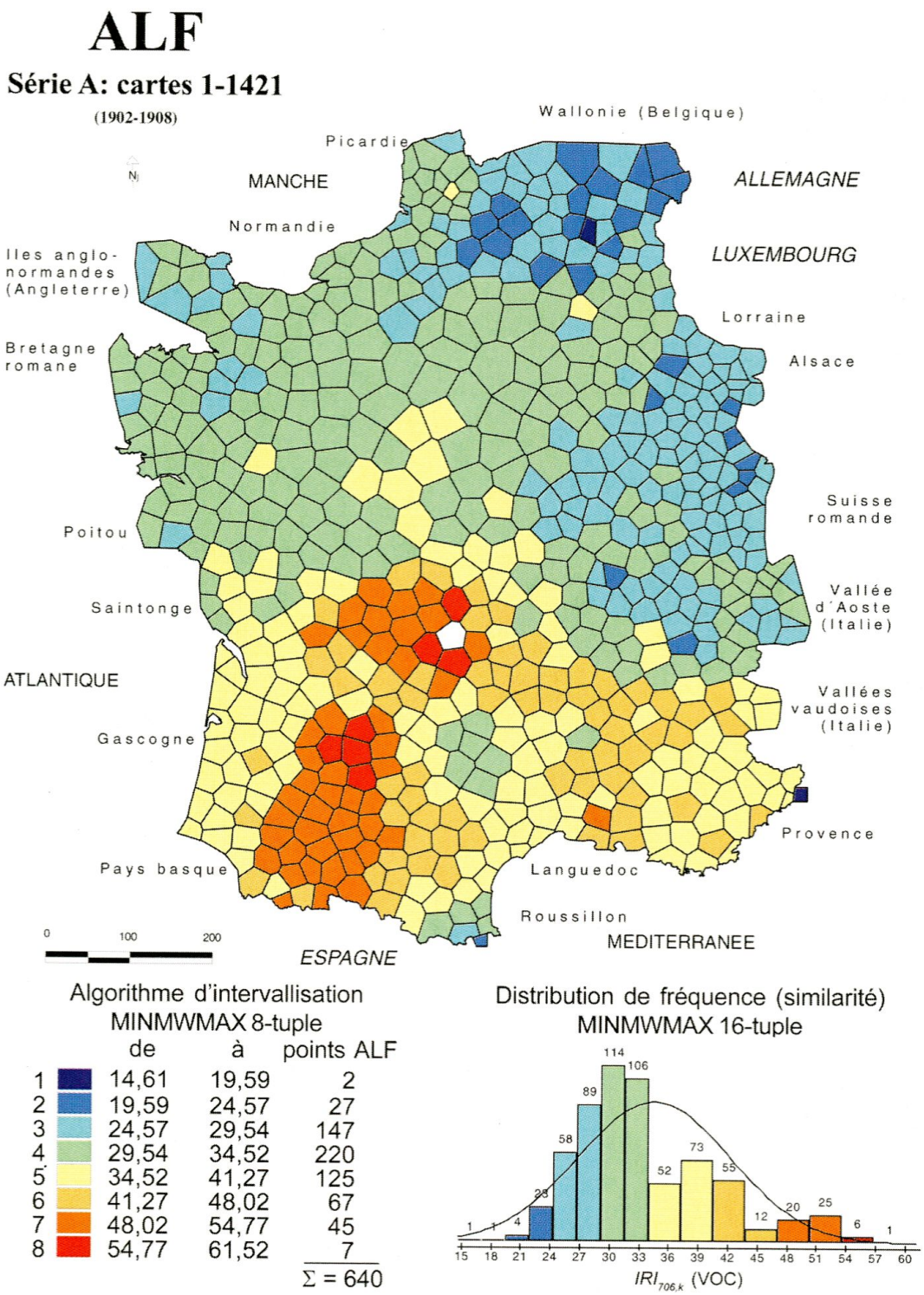
Carte 8: Répartition des provinces romaines de la Gaule au début du 2e siècle après Jésus-Christ au sein du réseau de l'ALF (d'après Motta 1995, 26).



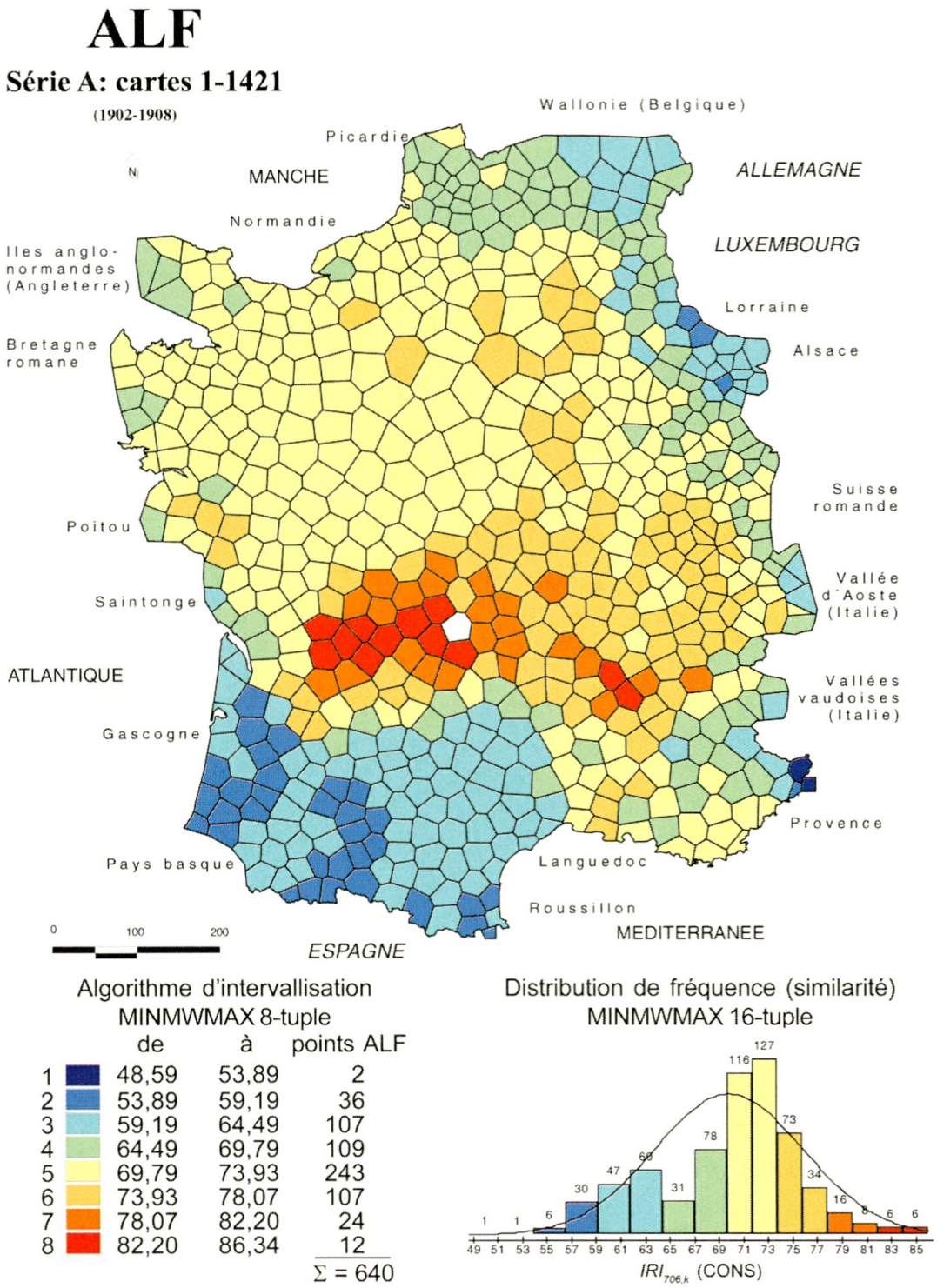
Carte 9: Carte choroplèthe de la distribution de similarité ($IRI_{706,k}$) relative au point-ALF 706 (Merlines, Corrèze). Corpus: 1687 cartes de travail. Voir aussi les cartes 10-12 et le chapitre 6.6.



Carte 10: Carte choroplèthe de la distribution de similarité ($IRI_{706,k}$) relative au point-ALF 706 (Merlines, Corrèze). Corpus: 471 cartes de travail (lexique). Voir aussi les cartes 9, 10-12 et le chapitre 6.6.



Carte 11: Carte choroplèthe de la distribution de similarité ($IRI_{706,k}$) relative au point-ALF 706 (Merlines, Corrèze).
Corpus: 612 cartes de travail (vocalisme).
Voir aussi les cartes 9-10, 12 et le chapitre 6.6.

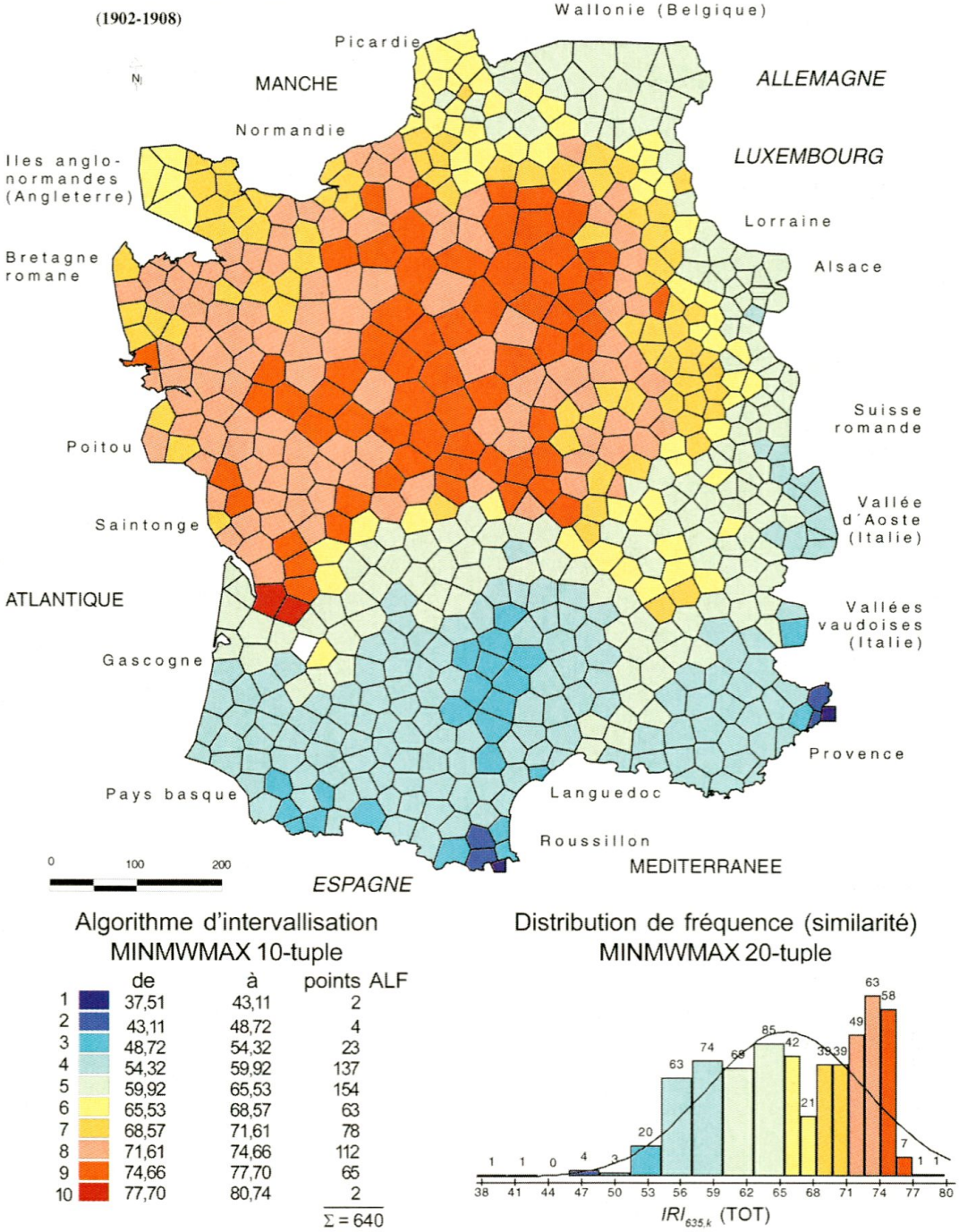


Carte 12: Carte choroplèthe de la distribution de similarité ($IRI_{706,k}$) relative au point-ALF 706 (Merlines, Corrèze). Corpus: 505 cartes de travail (consonantisme). Voir aussi les cartes 9-11 et le chapitre 6.6.

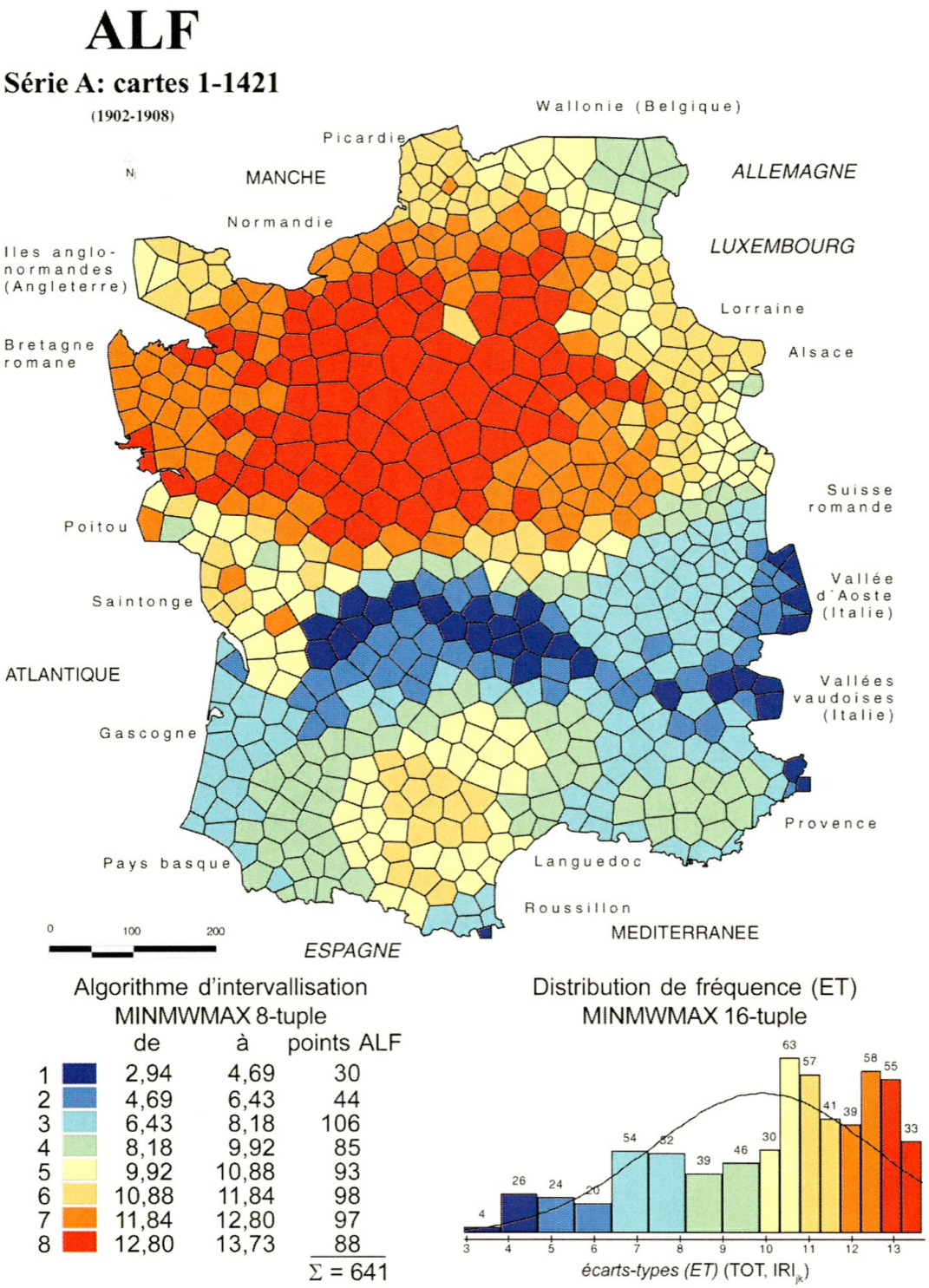
ALF

Série A: cartes 1-1421

(1902-1908)



Carte 13: Carte choroplèthe de la distribution de similarité ($IRI_{635,k}$) relative au point-ALF 635 (Andraut, Gironde). Corpus: 1687 cartes de travail. Voir aussi le chapitre 6.7.

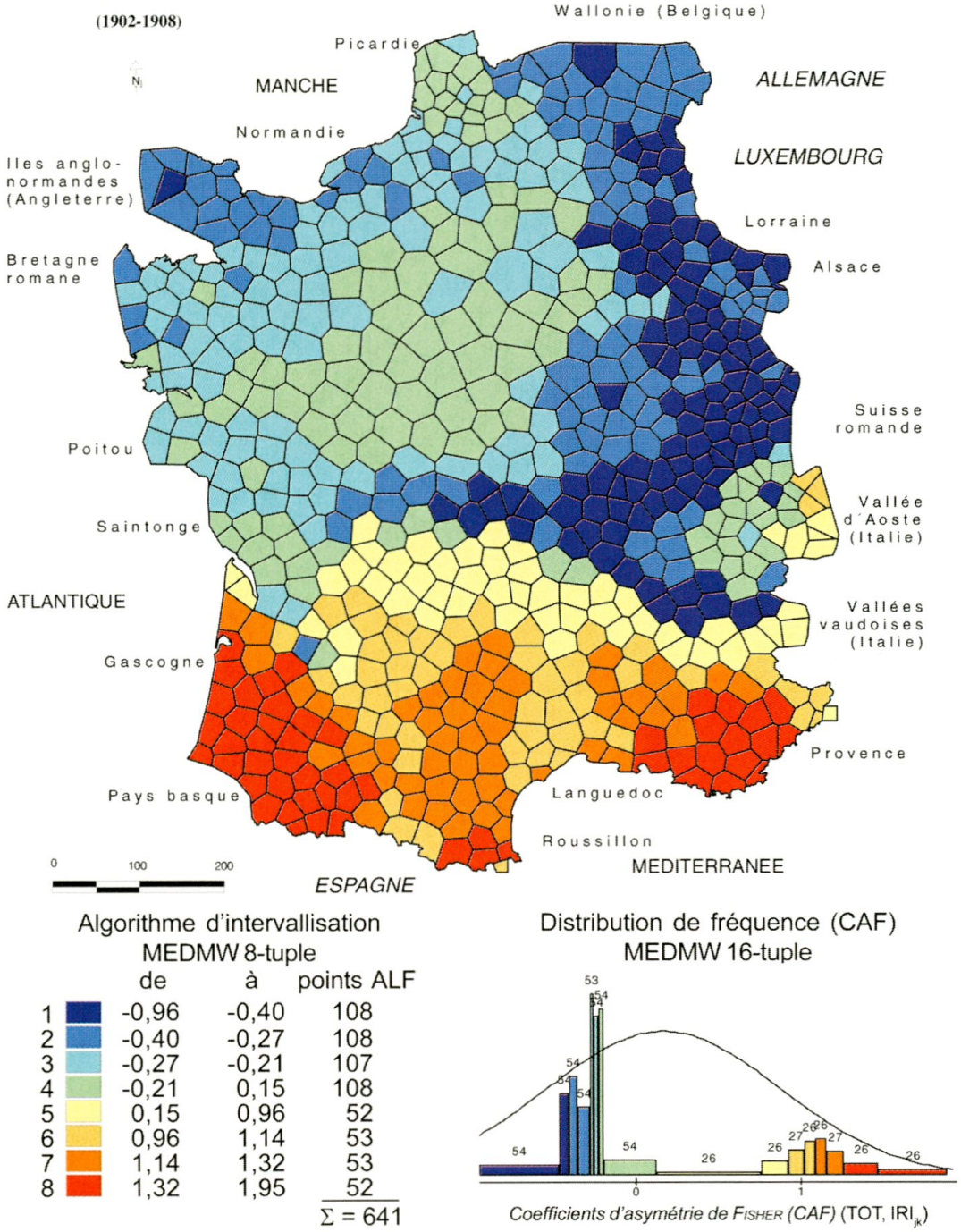


Carte 14: Carte choroplèthe de la synopse de 641 écarts-types (ET).
Corpus: 1687 cartes de travail. Indice de similarité: IRI_{jk}.
Voir aussi le chapitre 7.1.

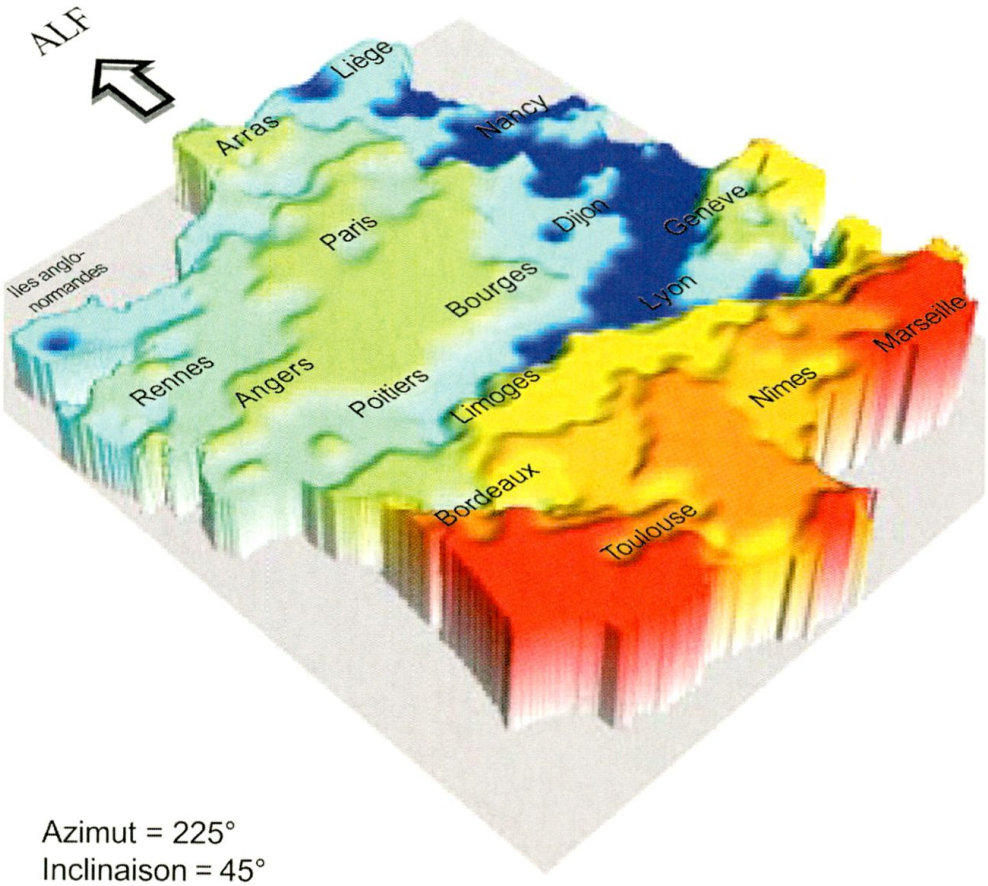
ALF

Série A: cartes 1-1421

(1902-1908)



Carte 15: Carte choroplèthe de la synopse de 641 coefficients d'asymétrie de Fisher (CAF). Corpus: 1687 cartes de travail. Indice de similarité: IRI_{jk}. Voir aussi le chapitre 7.2.



Azimut = 225°
Inclinaison = 45°
MEDMW 8-tuple
Altitude = 1,5

Surface statistique lissée

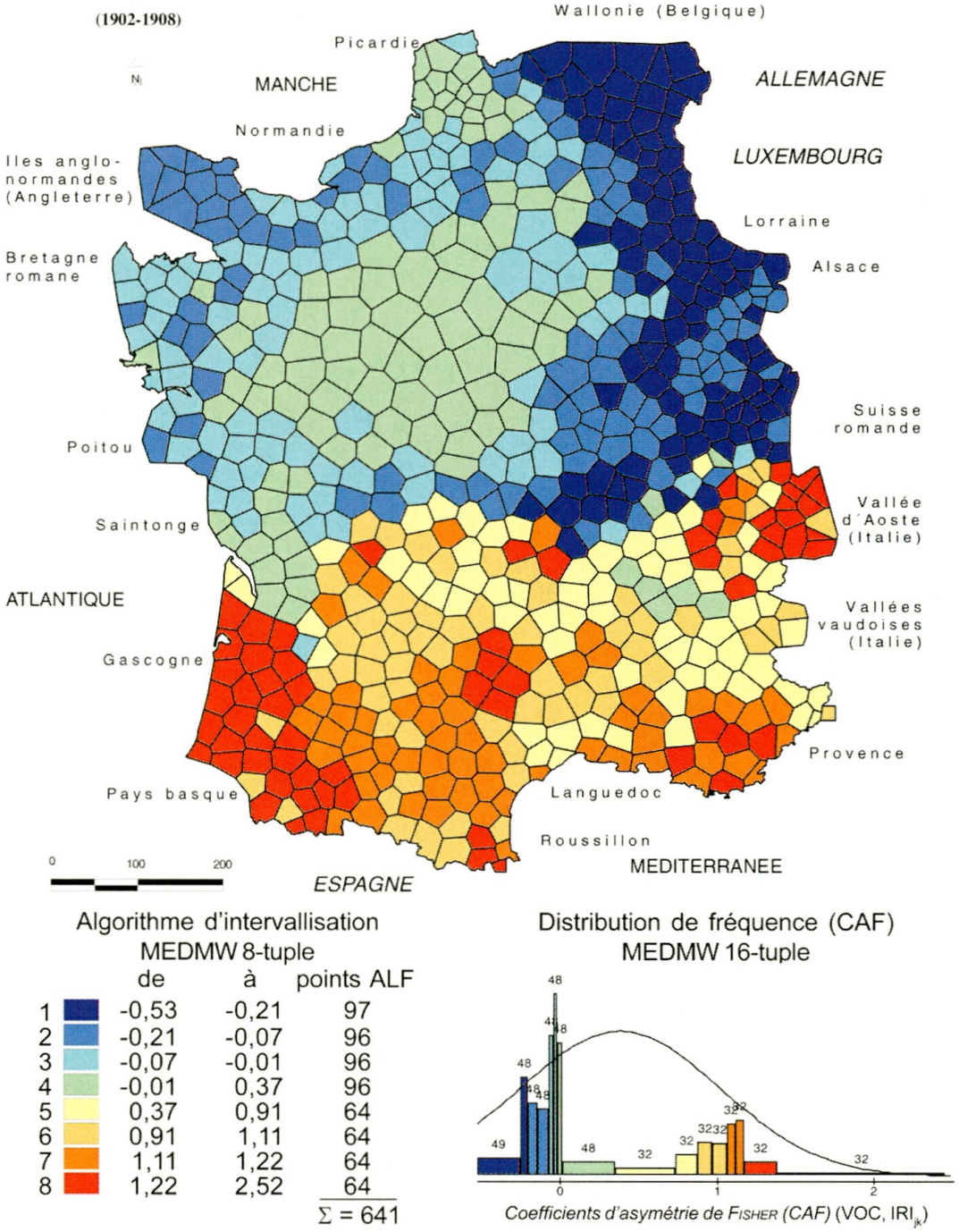
Synopse de 641 *Coefficients d'asymétrie de* FISHER *(CAF)*
(TOT, IRI_{jk})
Corpus: 1687 cartes de travail

Carte 16: Surface statistique lissée de la synopse
de 641 coefficients d'asymétrie de Fisher (CAF).
Mêmes données que dans la carte 15.
Point de vue de l'observateur: à partir du Nord-Ouest (225°).
Angle d'élévation de l'observateur: 45°. Corpus: 1687 cartes de travail.
Voir aussi les cartes 15, 17-18 et le chapitre 7.2.

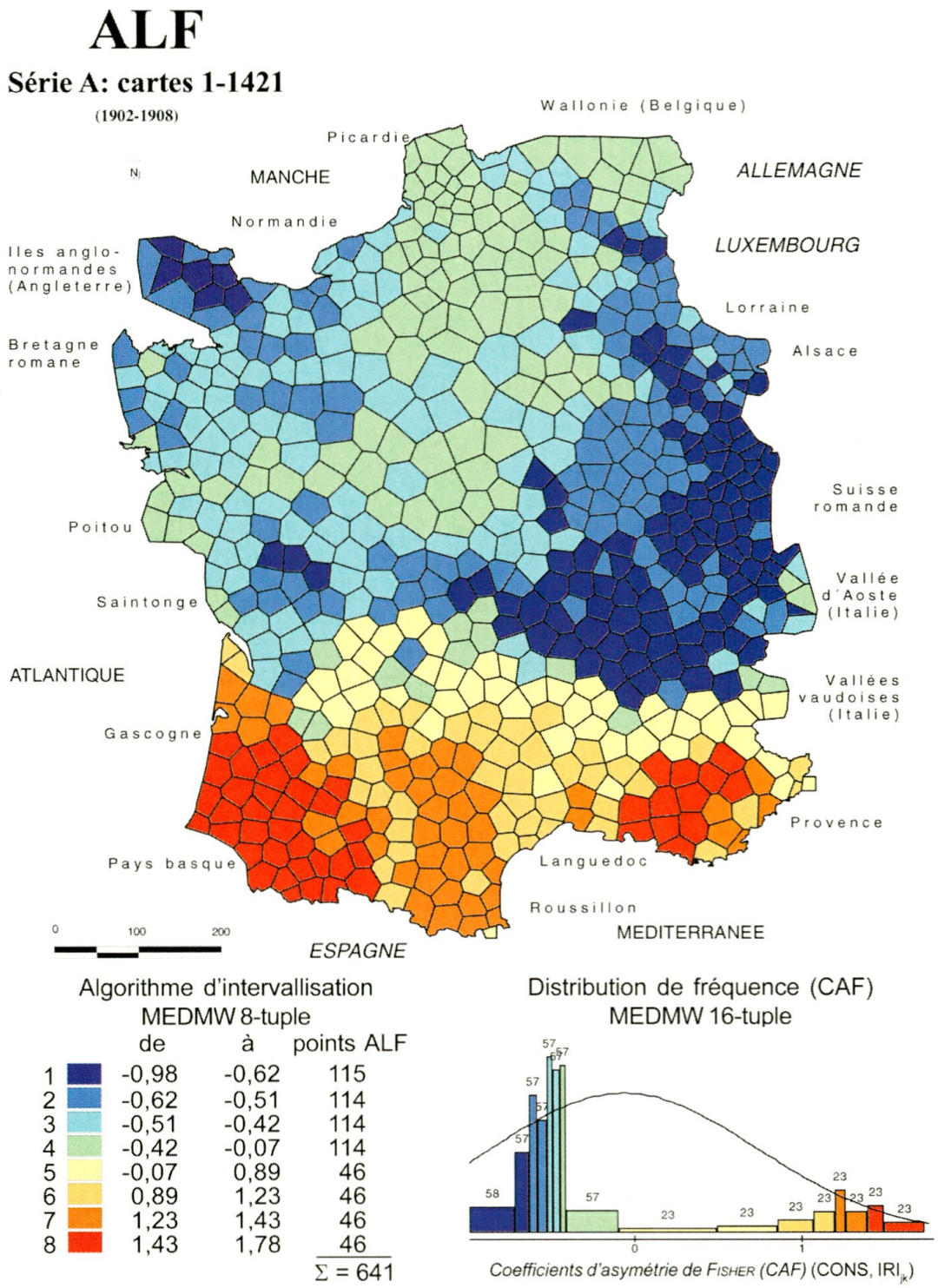
ALF

Série A: cartes 1-1421

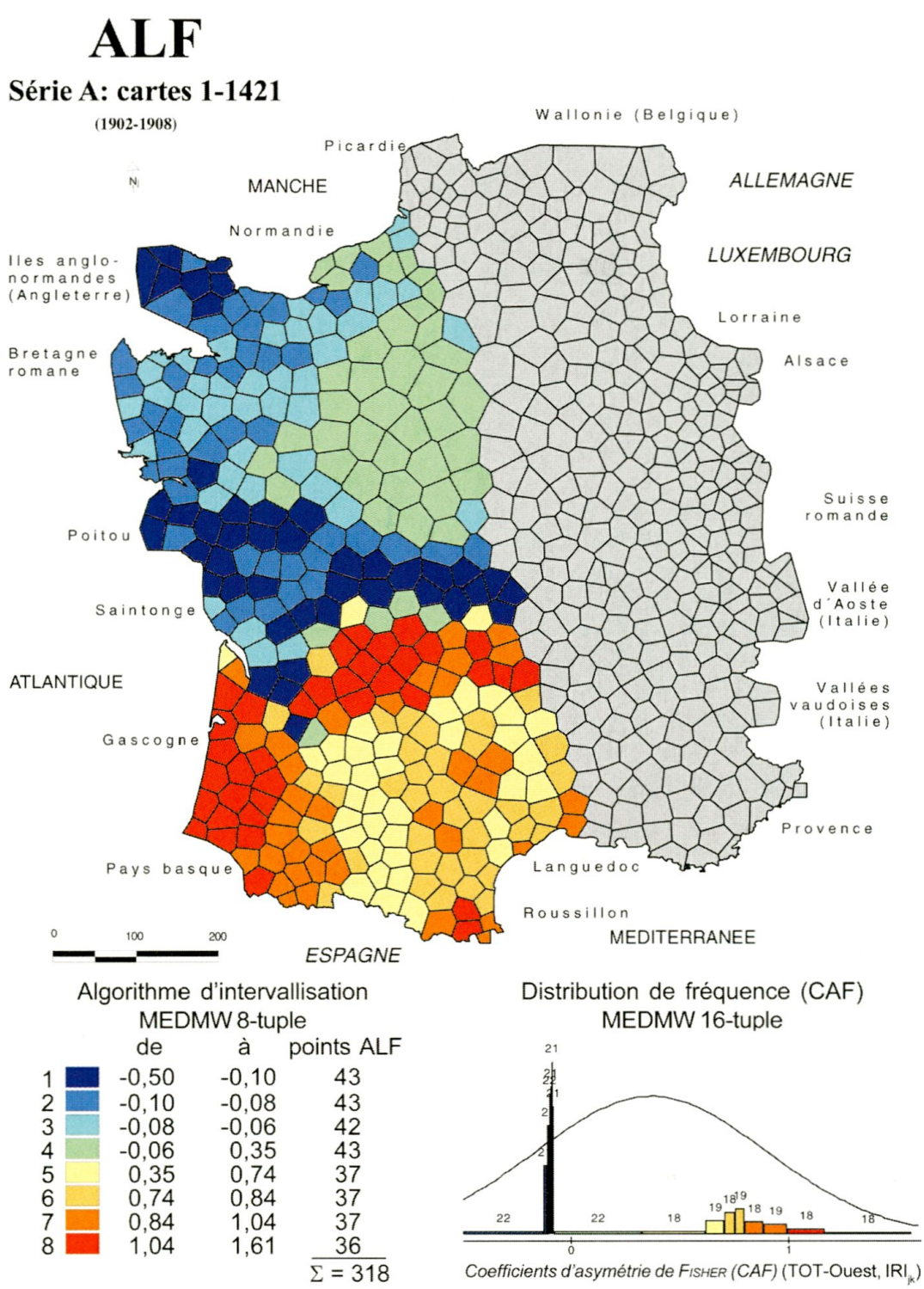
(1902-1908)



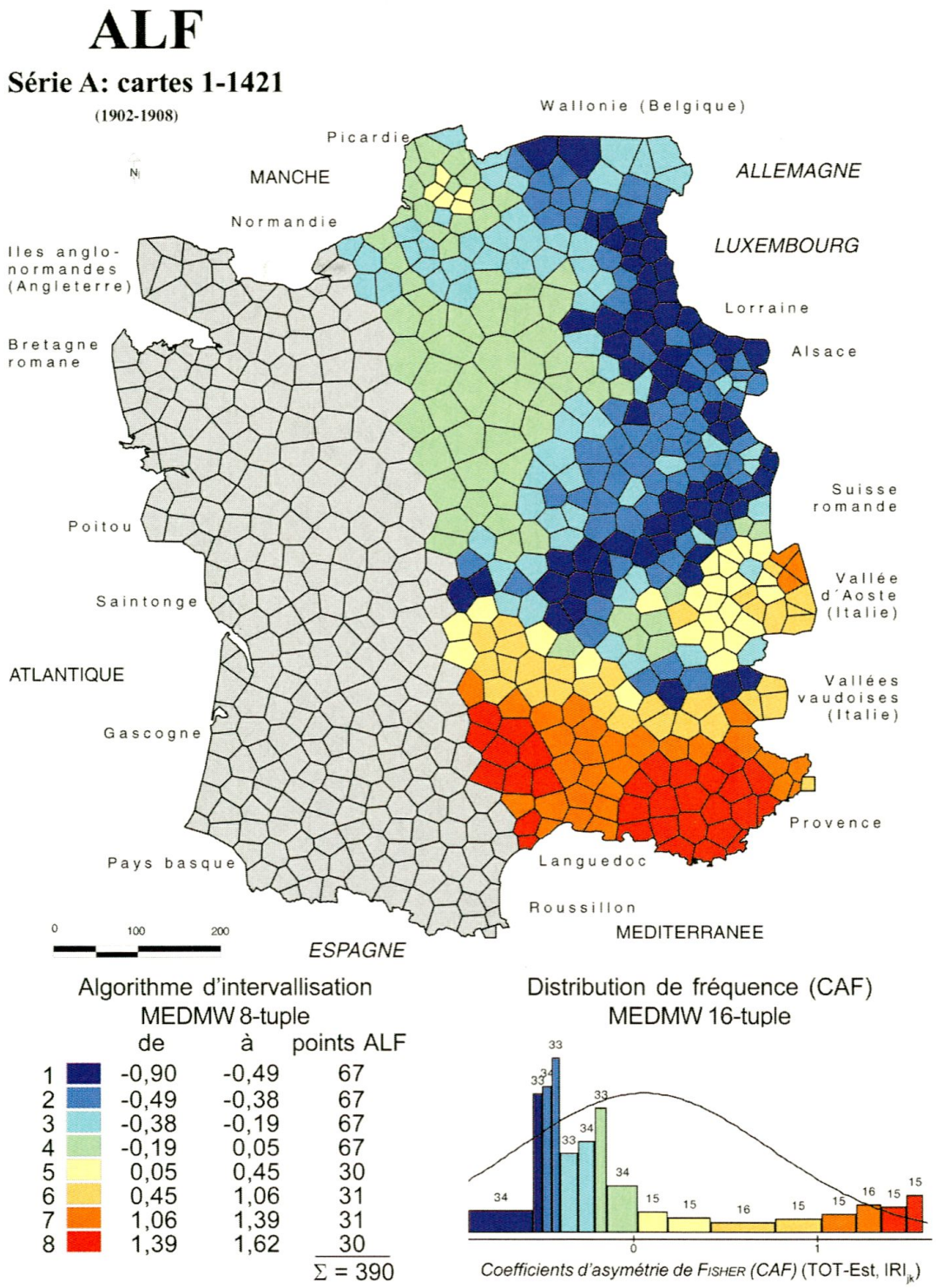
Carte 17: Carte choroplèthe de la synopse de 641 coefficients d'asymétrie de Fisher (CAF). Corpus: 612 cartes de travail (vocalisme). Indice de similarité: IRI_{jk}. Voir aussi les cartes 15-16, 18 et le chapitre 7.2.1.



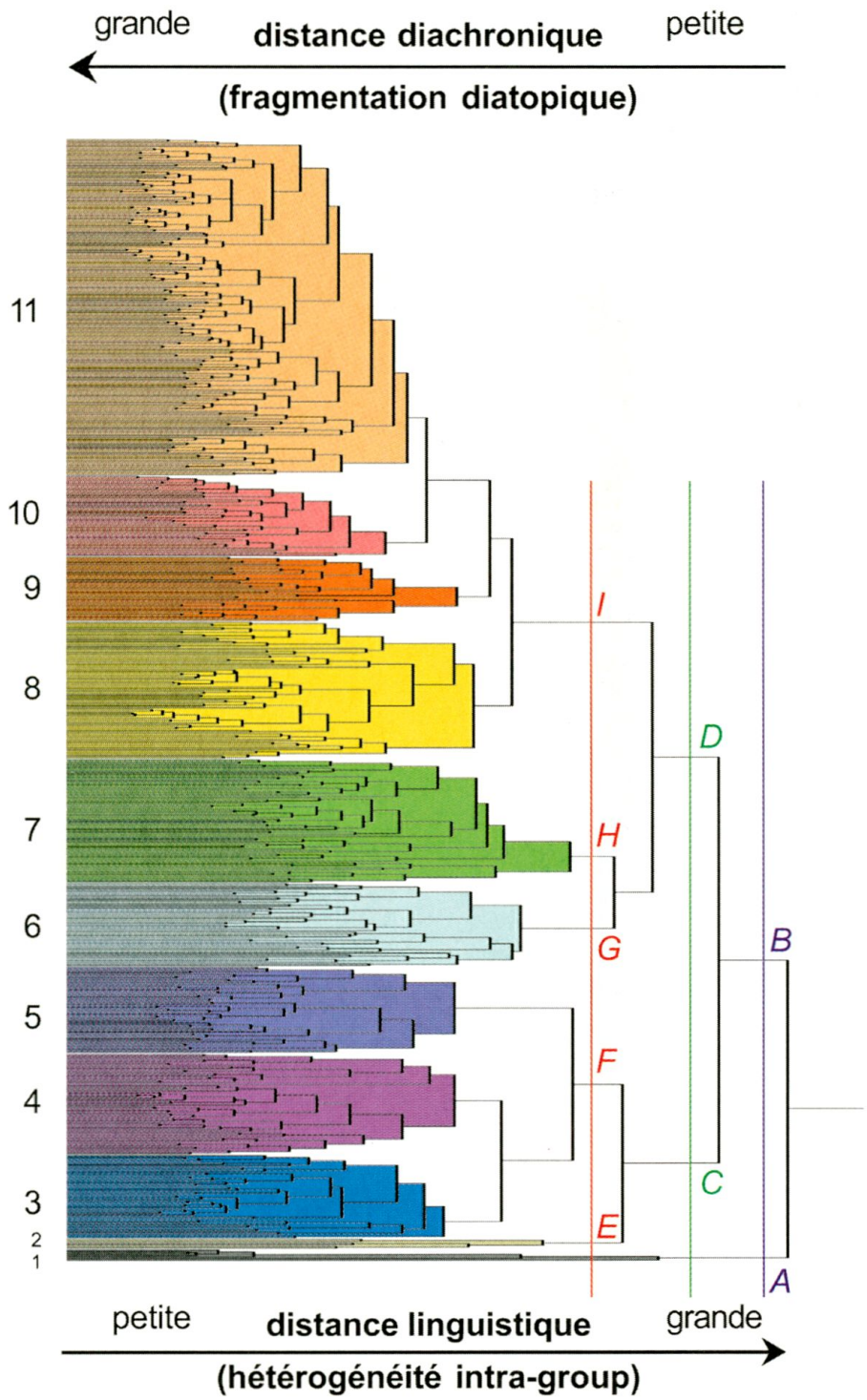
Carte 18: Carte choroplèthe de la synopse de 641 coefficients d'asymétrie de Fisher (CAF). Corpus: 505 cartes de travail (consonantisme). Indice de similarité: IRI_{jk}. Voir aussi les cartes 15-17 et le chapitre 7.2.1.



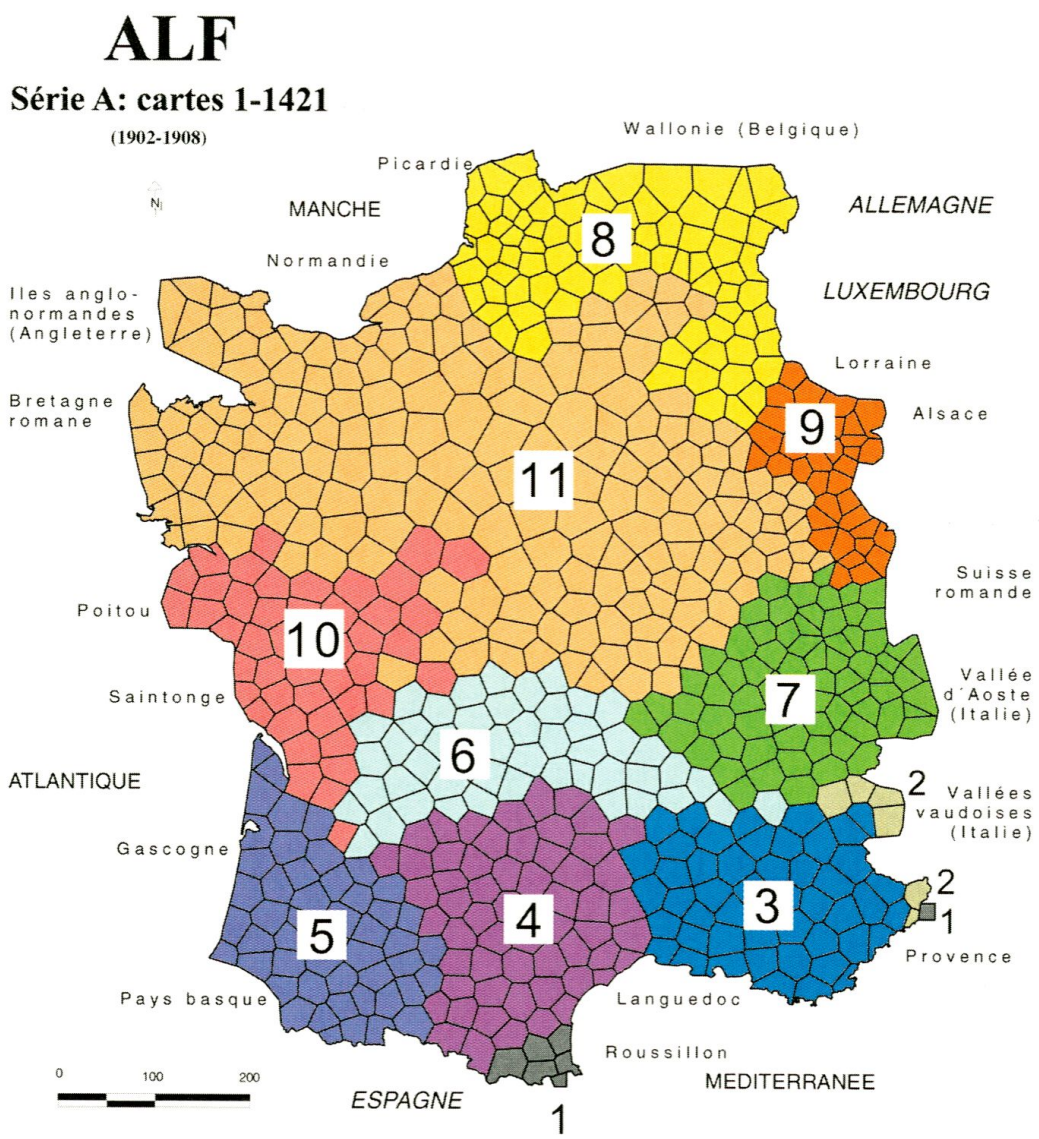
Carte 19: Carte choroplèthe de la synopse de 318 coefficients d'asymétrie de Fisher (CAF). Corpus: 1687 cartes de travail. Indice de similarité: IRI_{jk}. Voir aussi les cartes 15, 20 et le chapitre 7.2.2.



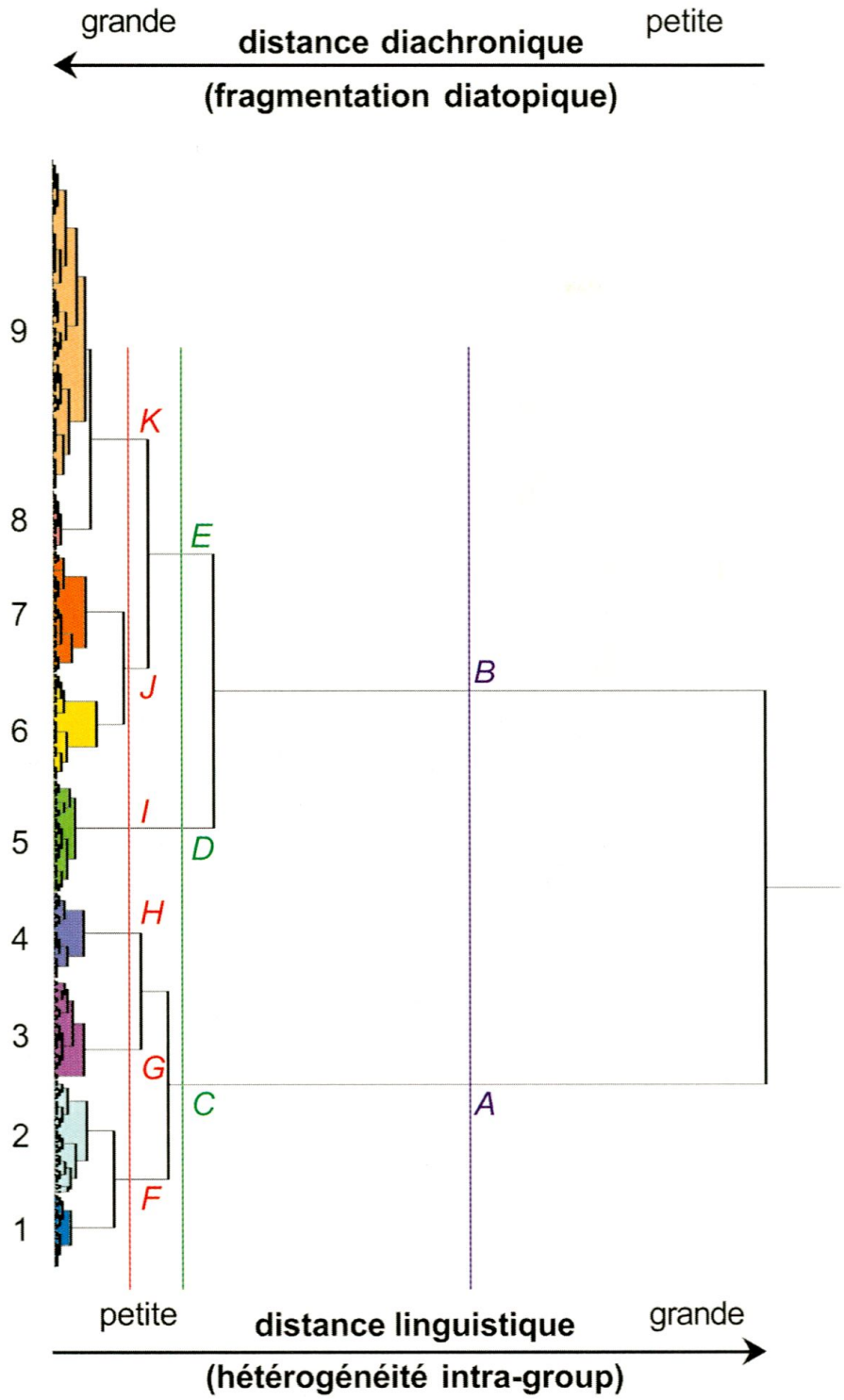
Carte 20: Carte choroplèthe de la synopse de 390 coefficients d'asymétrie de Fisher (CAF). Corpus: 1687 cartes de travail. Indice de similarité: IRI_{jk}. Voir aussi les cartes 15, 19 et le chapitre 7.2.2.



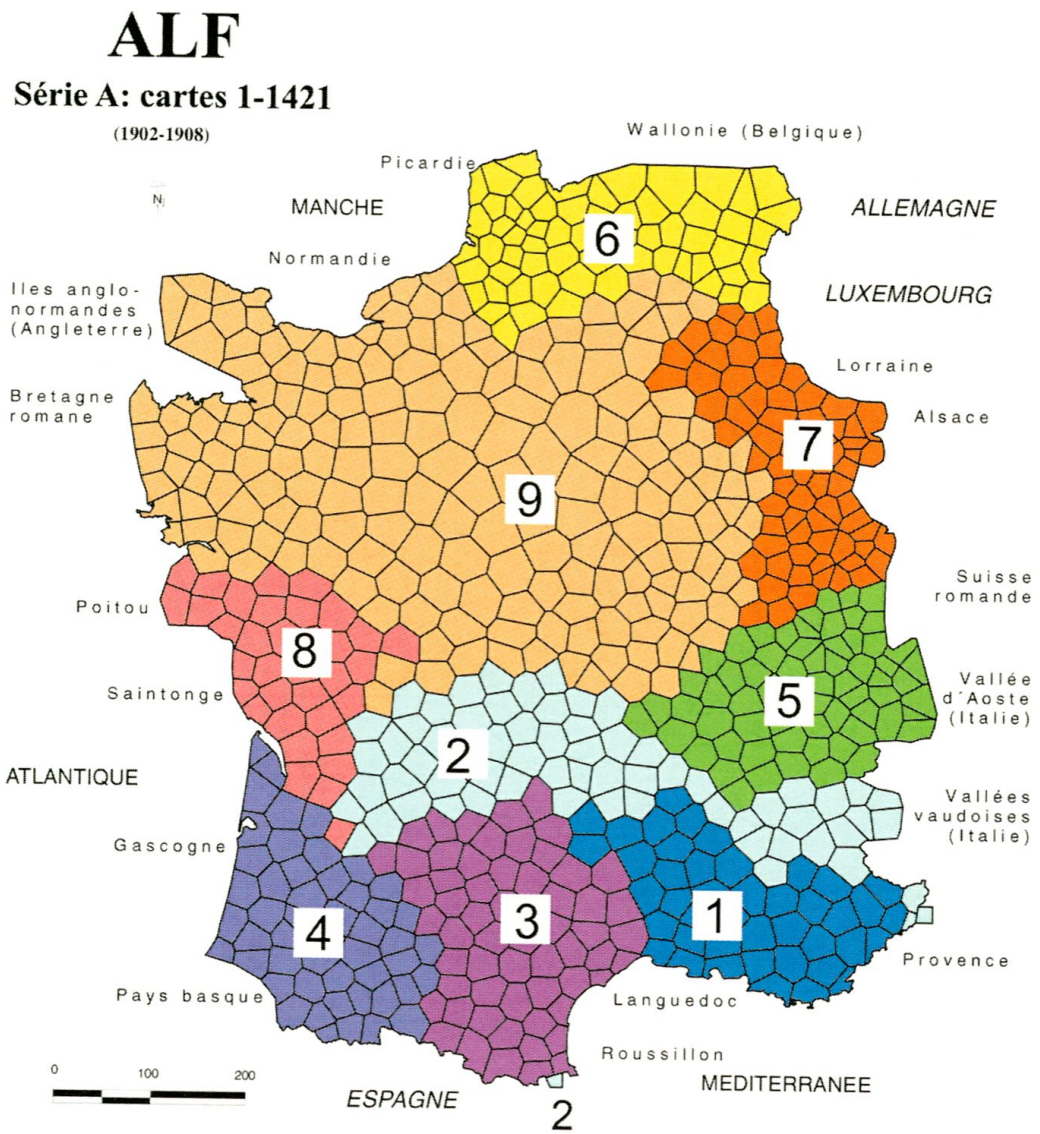
Carte 21: Classification dendrographique des vecteurs matriciels de 641 points-ALF. Algorithme agglomératif: Complete Linkage. Indice de similarité: IRI_{jk}. Corpus: 1687 cartes de travail. Nombre des dendrèmes coloriés: 11. Voir aussi la spatialisation de la carte 22 et le chapitre 8.



Carte 22: Spatialisation des données de la carte 21.
Voir aussi l'arbre de la carte 21 et le chapitre 8.



Carte 23: Classification dendrographique des vecteurs matriciels de 641 points-ALF. Algorithme agglomératif: méthode de Ward. Indice de similarité: IRI_{jk}. Corpus: 1687 cartes de travail. Nombre des dendrèmes coloriés: 9. Voir aussi la spatialisation de la carte 24 et le chapitre 8.



Carte 24: Spatialisation des données de la carte 23.
Voir aussi l'arbre de la carte 23 et le chapitre 8.