Zeitschrift: Revue de linguistique romane
Herausgeber: Société de Linguistique Romane
Band: 78 (2014)
Heft: 311-312

Artikel: L'analyse dialectométrique des cartes de la série B de l'ALF
Autor: Goebl, Hans / Smeka, Pavel
DOI: https://doi.org/10.5169/seals-842258

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: 07.07.2025

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

L'analyse dialectométrique des cartes de la série B de l'ALF

1. Remarques préliminaires

Le but de cet article est double : d'un côté, il s'agit de présenter les résultats de l'analyse dialectométrique (« dialectométrisation ») de 294 des 326 cartes originales de la série B de l'« Atlas linguistique de la France » (ALF) – limitée, comme cela est bien connu, à la moitié sud du réseau de l'ALF[1] – et, de l'autre, de les comparer, à l'aide de cinq méthodes dialectométriques (DM) différentes, avec les résultats de la dialectométrisation de 626 des 1 421 cartes originales de la série A de l'ALF[2], qui, elle, a déjà été faite à la fin du siècle dernier. Évidemment, cette comparaison présuppose l'utilisation parallèle du réseau de la série B qui comprend 326 des 638 points d'enquête originaux de l'ALF[3].

L'initiative d'entreprendre la dialectométrisation des cartes de la série B de l'ALF est entièrement due à Pavel Smečka qui s'est attelé, en 2010 et de son propre gré, à la tâche ardue – inévitable dans ce genre de recherches – de la préparation matérielle des cartes originales de la série B, de leur taxation consécutive, de la saisie informatique des données en découlant, et, finalement, de l'exécution des calculs-DM nécessaires[4].

Comme les *calculs* en question de même que la *visualisation* de leurs résultats ont été réalisés à l'aide du logiciel VDM[5], à savoir le programme-DM standard de la dialectométrie salzbourgeoise (DM-S), et que ce dernier est d'un maniement facile tout en disposant d'une grande efficacité informatique, les étapes du *calcul* et de la *visualisation* étaient les moins onéreuses que Pavel Smečka a dû parcourir.

---

1 Voici la numérotation des trois séries de cartes de l'ALF : série A, 1-1 421 ; série B, 1 422-1 747 (avec 326 cartes) ; série C, 1 748-1 920 (avec 173 cartes).

2 Le nombre de cartes originales de l'ALF utilisées pour la première dialectométrisation du réseau entier de l'ALF est de 626. Il est vrai cependant qu'en établissant le corpus « A-to-B », deux de ces 626 cartes-ALF n'ont plus pu être utilisées parce que leur partie méridionale n'offrait aucune variation géolinguistique. C'est pourquoi les chiffres fournis à ce sujet dans les tableaux 1 et 2 divergent légèrement.

3 Le réseau de la série A de l'ALF (appelée par Jules Gilliéron aussi 'atlas général') comprend 638 points ou localités (où ont été réalisées 639 enquêtes) ; celui de la série B contient 326 points, alors que le réseau de la série C, situé au sud-est de la France, dispose de 204 localités.

4 Voir à ce sujet les étapes 1-5, visualisées sur la figure 3.

5 VDM : Visual DialectoMetry. Ce programme a été créé, en 1999-2000, par notre ami commun Edgar Haimerl (résidant, actuellement, à Seattle, USA).

2. Genèse et contenu des cartes de la série B de l'ALF

Vu l'esprit « géométrique » avec lequel Jules Gilliéron a planifié l'ALF[6], il peut sembler étrange, même après plus de cent ans, qu'il n'ait pas utilisé, d'entrée de jeu, un seul type de questionnaire. Les informations techniques qu'il nous a laissées, dans la fameuse « Notice servant à l'intelligence des cartes », nous disent (Notice, 56) qu'à partir de la 80e enquête (= P.-155, Belval, Dommartin-sur Yèvre, Marne), il a commencé à élargir continuellement le nombre des questions de son questionnaire initial. Le moment de cette décision se situe au mois de janvier 1898, juste au milieu de la première des huit missions de recherche effectuées par Edmond Edmont, qui, elle, a eu lieu entre le 1er août 1897 et le 18 juin 1898.

Précisons que les 639 enquêtes de l'ALF se sont déroulées entre le 1er août 1897 et le 10 août 1901 et que, à en juger par les informations recueillies par la regrettée Marie-Rose Simoni-Aurembou à la Bibliothèque Nationale de Paris, E. Edmont n'a employé, pour l'accomplissement de ces 639 enquêtes au cours des huit missions, qu'un total de 1 351 jours[7]. Ceci donne une moyenne de 2,11 jours par enquête dont, évidemment, l'ampleur empirique pouvait varier entre 1 421 (série A) *au début* et 1 920 (séries A, B et C) questions *à la fin* de cette grande geste empirique.

Gilliéron était bien conscient du fait que l'élargissement *ex post* du questionnaire de base cadrait mal avec l'intention originale de son entreprise où l'intercomparabilité parfaite de tous les matériaux recueillis était un dogme de base. À la page 6 de la « Notice », il s'exprime comme suit :

> « Au cours de l'enquête, nous avons ajouté au questionnaire bon nombre de mots nouveaux, – quelquefois pour combler des lacunes apparues, souvent aussi par la nécessité de mettre dans un rapport plus équitable, et plus conforme à notre projet primitif, d'une part le temps pris par l'interrogation (temps ne dépassant pas nos prévisions), et d'autre part le temps consacré à la recherche d'un lieu et d'un sujet favorables (les excédant, celui-ci, de près de moitié). À la vérité nous cédions encore à une troisième considération, peut-être plus impérieuse que les deux premières : loin de s'émousser à la tâche déjà considérable que nous nous étions imposée, *notre appé-*

---

6 Voir à ce sujet nos contributions historiques et interprétatives de 2006c et 2013.

7 Voir la carte des huit missions de recherche de l'ALF publiée dans Brun-Trigaud / Le Berre / Le Dû 2005, 21 (document 6). Les informations présentées dans cette carte reposent sur la lecture détaillée du dossier Gilliéron, déposé à la BN de Paris sous la cote *NAF 11791-12030*, faite par les soins de Marie-Rose Simoni-Aurembou. Au cours de cette prise de connaissance attentive, Mme Simoni-Aurembou a dûment annoté les pages 25-28 de la « Notice » (« Concordance des numéros d'ordre chronologique ») tout en leur conférant des datations précises. Les feuilles respectives ont été remises ensuite par M.-R. Simoni-Aurembou à G. Brun-Trigaud (Nice) qui, elle, nous en a fourni une copie. Nous l'en remercions chaleureusement.

tit grandissait au fur et à mesure qu'avançait le travail [c'est nous qui soulignons]. Nous devons nous excuser de cette dérogation *au principe que nous avions adopté de ne soumettre au lecteur que des cartes complètes* [c'est nous qui soulignons], car on éprouvera peut-être un peu du dépit que nous éprouvons nous-mêmes à voir dans certaines de nos cartes de grands espaces vides de formes, et à ne trouver, vers la fin de l'ouvrage[8], que des cartes du midi de la France seulement au lieu de cartes complètes du territoire exploré. Nous donnons plus loin un tableau de ces mots nouveau-venus, avec la date de leur apparition dans le questionnaire, en tant du moins qu'ils intéressent l'*atlas général*, c'est-à-dire l'atlas à cartes de la France entière.

Il nous paraît superflu pour le moment – et il nous serait bien difficile, vu l'état dans lequel se trouvent actuellement nos matériaux – de signaler tous ceux qui ne rempliront que des cartes de la moitié sud du territoire et qui feront suite à l'atlas général. Le nombre de ces cartes est de 400 au moins (occupant 200 feuilles du format adopté) qui constitueront 4 fascicules au moins[9]».

Ce n'est donc que la *soif empirique* qui a poussé Gilliéron à augmenter sans cesse l'ampleur du questionnaire. On ne peut plus déterminer si cette augmentation s'est opérée continuellement ou par à-coups. Toujours est-il que quelques-unes des questions répertoriées à la fin de la «Notice» (56) ne figurent sur aucune des cartes des trois séries de l'ALF: ceci vaut, p. ex., pour les items *étable*, *gratte-cul*, *if*, *merisier* et *toit à porcs*[10]. Les raisons pour lesquelles Gilliéron n'en a pas publié les réponses resteront, très vraisemblablement, dans l'ombre à tout jamais.

L'écrasante majorité des titres des 326 cartes de la série B se réfèrent à des substantifs (= 218 unités) alors que 83 de ces cartes sont dédiées à la morphologie verbale, 3 à des adjectifs et 8 s'occupent des pronoms, adverbes et prépositions. Ce qui frappe, c'est le grand nombre de questions relatives à des plantes très particulières telles que la *bryone* (1479), la *cardère* (1487), le *cornouiller sanguin* (1517), la *digitale* (1536), le *gratteron* (1584), l'*hièble* (1595) ou la *pariétaire* (1655). Au vu des vastes espaces vides qui apparaissent sur les cartes correspondantes nous pensons que, très souvent, les sujets interrogés par Edmont ou bien ne comprenaient pas le sens de la question ou bien ne connaissaient pas la plante concernée.

---

8 L'on trouve ces indications à la page 56 de la «Notice».

9 Il est bien évident qu'en rédigeant ces lignes, Gilliéron ne pensait qu'aux besoins et dimensions des cartes de la série B tout en laissant complètement de côté les quelque 170 cartes de la série C.

10 Ce n'est que pour les stimuli *étable*, *if* et *merisier* qu'il y a des réponses dans la partie alphabétique des «Suppléments» de l'ALF. Dans les matériaux publiés de l'ALF, il n'y a aucune trace des réponses aux deux autres questions. Ajoutons que l'écrasante majorité des items répertoriés dans les «Suppléments» n'a aucune contre-partie dans les items des séries A, B et C de l'ALF. Il doit donc s'agir de matériaux 'épars' qui remontent à des élargissements 'de fortune' du questionnaire initial.

En regardant la carte des huit missions d'exploration d'Edmont (publiée chez Brun-Trigaud / Le Berre / Le Dû 2005, 21) et la liste diachronique de ces missions dressée par M.-R. Simoni-Aurembou, l'on constate qu'une augmentation considérable du questionnaire initial de 1 421 à 1 747 items a dû s'opérer à la fin de la 4e mission qui s'est conclue à la mi-mai 1899[11].

En effet, des 326 enquêtes du réseau de la série B, 320 ont été effectuées durant les missions 5-8. Seulement six de ces enquêtes ont été accomplies déjà pendant la 4e mission: il s'agit des explorations faites aux points-ALF 418 [296], 419 [297], 479 [314], 459 [315], 429 [317] et 540 [318][12]. Selon la liste de Simoni-Aurembou, le P.-ALF 418 a fait l'objet d'une enquête au mois de mars 1899 et le P.-ALF 419 au mois d'avril, alors que les quatre points restants ont été visités par Edmont en mai 1899. De ces indications, il ressort fort bien que seulement un petit nombre des enquêtes faites entre mars et mai 1899 a été inséré par Gilliéron dans le réseau définitif de la série B.

Les raisons de ce curieux triage ne sont pas claires. Peut-être Gilliéron voulait-il donner, par cette prise en compte sélective des matériaux empiriques à sa disposition, un aspect plus rectiligne au tracé de la frontière *septentrionale* du réseau de la série B[13].

Sur la carte géographique, tous les points mentionnés ci-dessus se trouvent dans le Poitou occidental (dans les Départements de la Vendée et des Deux-Sèvres) et représentent, de ce fait, la périphérie nord-occidentale du réseau de la série B.

Toujours est-il que la totalité des missions 5 à 8, exécutées entre le 1er juin 1899 et le 10 août 1901, s'est déroulée sous le signe d'un questionnaire comprenant d'abord au moins 1 747 items (= séries A et B) et, dès le début de la

---

11 Curieusement, le problème de l'existence parallèle de trois séries de cartes dans l'ALF n'a reçu, dans les ouvrages de Sever Pop, de loin le plus éminent biographe de Jules Gilliéron, qu'un traitement très sommaire: cf. Pop 1950 I, 134*sq.* et Pop / Pop 1959, 86*sq.*

12 Les numéros entre crochets renvoient à la place que ces enquêtes occupent dans le tableau chronologique de toutes les enquêtes réalisées pour l'ALF (cf. Notice, 25-28). Le relevé effectué au P.-ALF 540 [318] est la dernière des 68 explorations faites au cours de la 4e mission.

13 Il semble bien que la genèse de la frontière *occidentale* du réseau de la série C s'explique par une raison 'géométrique' analogue: bien que la plupart des localités-ALF qui la constituent sont issues de la 6e mission, deux points situés en plein milieu de cette frontière (PP.-ALF 713 et 715) n'ont été explorés qu'au cours de la 8e (et dernière) mission. Toujours est-il que la première enquête-ALF admise dans le réseau C détient, sur la liste de Simoni-Aurembou, la 400e place (= P.-ALF 704); elle a été réalisée au mois de décembre 1899. C'est la première enquête dont on peut dire avec certitude qu'elle portait, au moins, sur 1 920 questions.

6e mission (décembre 1899), avec un questionnaire pourvu d'au moins 1 920 questions (= séries A, B et C).

Comme l'ensemble des enquêtes faites par Edmont au cours de la 6e mission ne se reflète pas complètement dans le réseau de la série C, il est évident qu'une partie des enquêtes faites avec le maxi-questionnaire de 1 920 items n'a pas été publiée par Gilliéron. Il est cependant fort probable que les carnets alors remplis se trouvent encore dans le dossier Gilliéron déposé à la Bibliothèque Nationale.

Il est par contre hors de doute que les cartes de la série B ont plus de lacunes que la majorité des cartes de la série A. Il est non moins évident que l'ensemble des questions des séries B et C constitue des apports supplémentaires qui ont été ajoutés après coup, individuellement et selon le cas, au tronc commun du questionnaire de base.

L'intérêt du dialectométricien – qui s'occupe des *structures de profondeur* qui sous-tendent des stocks de données dialectales – est de savoir si, et surtout dans quelle mesure, les structures de profondeur des cartes de la série A et de la série B se ressemblent. Or, les résultats dont il va être question plus tard montrent qu'elles se ressemblent beaucoup. Et ceci malgré l'ampleur divergente des deux corpora dialectométrisés[14], et aussi malgré l'application de deux «philosophies taxatoires» différentes[15].

3. Description et rendement des deux campagnes de dialectométrisation de données de l'ALF

3.1. La campagne réalisée entre 1997 et 2000, relative à 626 cartes originales de la série A de l'ALF

La première analyse-DM de données originales tirées de l'ALF que nous ayons faite concernait uniquement le réseau de la Normandie, élargi légèrement

---

[14] Effectif total des CT retranchées de notre corpus-DM tiré de la série A de l'ALF: 1 650; effectif total des CT extraites de la série B de l'ALF: 675. Rappelons qu'au-dessus d'un seuil de quelque 200 à 250 CT, aléatoirement tirées d'un corpus-DM plus grand, les structures-DM globales ainsi calculées restent très stables. Ceci signifie que la comparaison des structures globales tirées, d'un côté, de 1 650 CT et, de l'autre, de 675 CT est non seulement légitime, mais aussi absolument fiable du point de vue taxométrique: cf. à ce sujet Goebl 1984 I, 206*sqq*.

[15] L'équipe de taxation qui, de 1997 à 2000, s'est attelée à la tâche de la taxation de 626 cartes originales de l'ALF, obéissait à des consignes taxatoires plutôt globales et enveloppantes («lumping»), alors que la taxation opérée par Pavel Smečka suivait des principes plus différenciateurs («splitting»). Pour ces deux manières d'analyser ou de capter la réalité empirique, cf. Sneath / Sokal 1973, 7 et 431.

vers l'ouest, le sud et l'est[16]. Les résultats-DM correspondants ont été utilisés dans notre thèse d'habilitation (Goebl 1984) et un certain nombre de publications postérieures. Ce n'est qu'en 1997, sur la base d'expériences faites au préalable avec des données d'atlas de nature *phonétique*[17], que nous avons osé nous attaquer à une analyse-DM plus universelle de l'ALF, qui devait embrasser, d'un côté, la totalité de son réseau et, de l'autre, un nombre aussi grand que possible de ses cartes avec inclusion de toutes les catégories linguistiques.

Il fallait alors définir les normes de taxation sous la forme d'un véritable manuel de taxation, en communiquer les principes à une équipe de jeunes collaborateurs, et surtout mettre à leur disposition les « moyens matériels » pour l'exécution ordonnée de leur travail. Par « moyens matériels », nous entendons un exemplaire de la réimpression de l'ALF, parue en 1968 chez l'éditeur Forni à Bologne (Italie), lequel, une fois acheté, avait été décomposé en feuilles détachées. Ceci signifiait que chacune des 1 421 cartes de la série A de l'ALF était à la disposition individuelle de nos collaborateurs.

Évidemment, il fallait encore préparer convenablement le travail d'extraction des données dialectales des cartes de l'ALF. Pour ce faire, nous avons fait imprimer chez un imprimeur salzbourgeois, sur chacune des 1 421 cartes détachées de la série A, un réseau multicolore d'une douzaine de « parcours d'épreuve » lesquels, en zigzagant à travers le réseau de l'ALF, reliaient entre eux, à l'instar d'un chapelet à égrener, l'ensemble des 638 points à dépouiller.

Un jeu de modules appropriés, arrangés selon la logique spatiale des parcours d'épreuve mentionnés ci-dessus, permettait de fixer et de saisir à la main les réflexions et les résultats des analyses taxatoires. Ce n'est qu'après la saisie manuelle des taxats qu'a eu lieu l'intégration du contenu des modules dans l'ordinateur. Ce travail avait déjà été appuyé par le logiciel salzbourgeois VDM. Inutile de dire que les données, une fois informatisées, ont encore subi une nouvelle vérification avec, au besoin, les corrections nécessaires.

Les cartes détachées de l'ALF ont été distribuées à cinq collaboratrices[18] dans l'ordre de leur arrangement alphabétique original. Vu les contraintes de

---

16 Le réseau utilisé alors comprenait en tout seulement 71 points-ALF (dont un point artificiel, relatif au français standard).

17 Il s'agit de la dialectométrisation de la moitié des quelque 400 colonnes des « Tableaux phonétiques des patois de la Suisse romande » (TPPSR) : cf. Goebl 1985, *passim*.

18 Il s'agissait des romanistes salzbourgeoises suivantes : Barbara Aigner, Irmgard Dautermann, Hildegund Eger, Susanne Oleinek et Annette Schatzmann. Ma profonde gratitude pour leur travail précis et tenace leur est acquise pour toujours.

la disponibilité personnelle de nos collaboratrices et aussi l'ampleur limitée des subsides financiers alors à notre disposition, il n'était pas possible de traiter la totalité des planches de la série A de l'ALF, si bien que nous ne pouvions analyser, en fin de compte, que 626 des 1 421 cartes disponibles.

L'unité de base, issue d'une analyse taxatoire des données d'un atlas linguistique et faite selon les modalités en usage à Salzbourg, s'appelle «carte de travail» (CT): voir les cartes 1-4 de l'annexe et aussi les positions 1 et 2 de la figure 3. Le travail analytique qui mène à l'établissement d'une CT correspond parfaitement à la vieille tradition des romanistes de «typiser», à l'aide de cartes muettes et d'une cartographie souvent très rudimentaire, le contenu d'une planche originale d'atlas. En d'autres mots: la fameuse carte à hachures – ajoutée par J. Gilliéron à la fin de son livre bien connu sur l'abeille de 1918 – a été dressée, par son auteur, exactement dans le même «esprit» que les cartes 3 et 4 de cet article[19]. Dans tous les cas, il s'agit d'une typisation *lexicale* (ou *onomasiologique*). La DM salzbourgeoise repose donc, de par ses données de départ, sur des prérequis méthodiques rodés et notoires en matière de romanistique.

Pour décrire convenablement une CT, il faut en retenir l'*appartenance catégorielle*, le *nombre des taxats* ainsi que la *taille* et la *structuration* de ses aires. Comme l'analyse-DM vise l'étude de la variation dialectale et que cette dernière présuppose, sur une CT, la présence d'au moins deux taxats se partageant le réseau, l'éventail de la variation taxatoire (*poly*-nymie) des CT s'inscrit toujours entre 2 (CT *bi*-nyme) et *n* (CT *n*-nyme)[20]. La polynymie et l'appartenance catégorielle des cartes 1-4 se présente donc comme suit:

	Catégorie linguistique	polynymie
Carte 1	phonétique	11-nyme
Carte 2	phonétique	9-nyme
Carte 3	lexique	8-nyme
Carte 4	lexique	9-nyme

[19] La même remarque vaut, en guise d'exemple, pour toutes les cartes contenues dans le livre de K. Jaberg sur l'ALF de 1908.

[20] Pour la désignation du degré changeant de la polynymie des CT, nous utilisons le triplet terminologique grec suivant: CT avec *très peu* de taxats, *oligo*-nyme; CT avec un nombre *moyen* de taxats, *méso*-nyme; CT avec *beaucoup* de taxats, *poikilo*-nyme (< grec *poikilós* 'bigarré').

Un autre problème digne d'être mentionné est le nombre des lacunes sur les planches originales d'atlas. Comme la prise en compte dans la saisie taxatoire des données originales d'un atlas linguistique d'un nombre trop grand de lacunes cause inévitablement, au cours des calculs-DM, des distorsions numériques – souvent très fâcheuses – des résultats-DM, il est indiqué d'en contrôler rigoureusement la quantité. Dans les analyses taxatoires faites entre 1997 et 2000, nous n'avions admis, pour les lacunes, que quelque 10% des 638 points du réseau de base de l'ALF. Cela signifiait que toutes les cartes-ALF disposant de plus de 63 «cases vides» avaient été écartées de l'analyse-DM, indépendamment de leur intérêt ou de leur valeur linguistiques.

Pour le bilan global des analyses-DM effectuées entre 1997 et 2000, voir le tableau suivant[21]:

	∑ cartes-ALF originales dépouillées	∑ cartes de travail (CT)	∑ taxats (aires taxatoires)	Variation polynymique (taxats / carte de travail)
Phonétique	247	1 117	10 642	de 2 à 72
Vocalisme[22]	242	477	7 533	de 2 à 72
Consonantisme	227	479	2 746	de 2 à 26
Lexique	457	465	6 514	de 2 à 90
Morphosyntaxe	84	99	891	de 2 à 55
Total	626	1 681	18 047	de 2 à 90

Tableau 1: Synopse quantitative du corpus dialectométrique tiré de la série A de l'ALF entre 1997 et 2000 (comprenant le réseau intégral de l'ALF)

[21] Les valeurs du tableau 1 divergent légèrement de celles que l'on trouve dans nos publications antérieures de 2000, 2002 et 2003. Ceci s'explique par le fait que nous avons rectifié, depuis 2009, les décomptes originaux en éliminant, de nos calculs, toutes les «données manquantes» (= lacunes) ainsi que les CT mononymes, dont certaines avaient encore été incluses dans le décompte initial du total des CT prises en considération. Précisons, en outre, que les calculs-DM à proprement parler n'ont jamais souffert de ces incongruités et que, de ce fait, les résultats-DM présentés dans nos travaux de 2000, 2002 et 2003 (et après), restent entièrement valables.

[22] Comme les exemples iconiques de cet article se réfèrent uniquement aux corpora *total*, *phonétique* et *lexical*, les indications relatives aux sous-catégories linguistiques du *vocalisme*, du *consonantisme* et de la *morphosyntaxe* ont été mises en italique et réalisées, en outre, en caractères plus petits.

En règle générale, une planche-ALF *lexicalement* mononyme – dont les réponses reposent toutes, par conséquent, sur le même étymon – a pu être analysée plusieurs fois, de sorte qu'il était possible de dériver, de 247 cartes-ALF de ce genre, un total de 1 117 CT de type *phonétique*. Par contre, la même diversification était impossible pour les planches *lexicales* de l'ALF où, le plus souvent, nous ne pouvions déduire qu'une seule CT d'une carte originale.

Notons, en outre, que c'est toujours le lexique qui fournit les éventails polynymiques les plus larges, alors que, par rapport à la richesse taxatoire du lexique, la phonétique est toujours plus ou moins en retrait. Ceci s'explique, avant tout, par le fait bien connu que l'inventaire des unités de base de la phonétique est fermé, alors que ce n'est pas le cas pour le lexique.

Les scores présentés dans le tableau 1 peuvent être mis en relation avec la position 2 de la figure 3, relative à la « matrice de données ». Le point d'arrivée de l'analyse taxatoire est toujours, selon les consignes de la DM-S, l'établissement d'une *matrice de données* qui, d'un côté, constitue le dernier maillon *qualitatif* de notre chaîne méthodique et, de l'autre, servira comme point de départ pour toutes les analyses *quantitatives* de nature dialectométrique à proprement parler.

3.2. La réduction des données-ALF taxées en 1997-2000 aux dimensions de la série B de l'ALF (corpus « A-to-B »)

Voir le tableau 2 et les figures 1 et 2.

Évidemment, une comparaison directe des structures de profondeur des cartes de la série B de l'ALF avec celles émanant de la série A présuppose l'utilisation du même réseau. Comme celui de la série B comprend 326 points-ALF, nous avons dû réduire les dimensions de la matrice originale en en enlevant quelque 300 vecteurs de points d'enquête. Dans ce processus, nous avons néanmoins conservé nos trois points artificiels « traditionnels », relatifs aux langues standard du *français* (P. 999), de l'*italien* (P. 998) et du *catalan* (P. 997). De cette mesure, il résulte un total de 329 points-ALF.

Le retranchement de la moitié nord du réseau de l'ALF a également été pratiqué dans la structure polygonale de la carte de base pour conférer plus de clarté à la comparaison visuelle des deux niveaux d'analyse.

	Corpus-ALF réduit du sud (série A [641 P.-ALF] → série B [329 P.-ALF])				Corpus-ALF de la série B [329 P.-ALF]			
	∑ cartes-ALF originales dépouillées	∑ cartes de travail (CT)	∑ taxats (aires taxatoires)	Variation polynymique (taxats / carte de travail)	∑ cartes-ALF originales dépouillées	∑ cartes de travail (CT)	∑ taxats (aires taxatoires)	Variation polynymique(taxats / carte de travail)
Phonétique	247	1 096	8 812	de 2 à 58	62	313	2 588	de 2 à 47
Vocalisme	242	477	6 043	de 2 à 58	62	142	1 781	de 3 à 47
Consonantisme	222	466	2 431	de 2 à 20	54	121	701	de 2 à 18
Lexique	449	457	4 624	de 2 à 67	274	300	4 616	de 2 à 87
Morphosyntaxe	84	97	725	de 2 à 47	36	62	330	de 2 à 56
Total	624	1 650	14 161	de 2 à 67	294	675	7 534	de 2 à 87

Tableau 2: Synopse quantitative des corpora dialectométriques tirés des séries A et B de l'ALF (comprenant 329 points méridionaux de l'ALF)

La moitié de gauche du tableau 2 montre les nouvelles relations quantitatives qui se sont instaurées à l'intérieur de la matrice de données «amputée», alors que les figures 1 et 2 en visualisent certains aspects quantitatifs très intéressants. Il s'agit, en l'occurrence, de la démonstration des rapports quantitatifs qui existent entre la fragmentation taxatoire des cartes de travail (CT) – voir l'abscisse (à l'horizontale) – et le nombre absolu de ces dernières – voir l'ordonnée (à la verticale): plus cette fragmentation est grande, plus le nombre des CT respectives diminue.

La remarquable régularité de la déclivité continue des deux courbes de type exponentiel renvoie à l'existence d'un ordre quantitatif sous-jacent commun. Soit dit entre parenthèses, de telles régularités ont pu être décelées dans toutes nos études-DM[23]. Elles constituent la contre-partie *géolinguistique* des lois quantitatives qui régissent la production *grammaticale* de l'activité langagière de l'homme[24]. Dans notre cas, cette dernière se déploie non pas dans les

[23] Cette régularité a le rang d'une véritable loi mathématique. Ceci a été prouvé en 1985 par le linguiste-mathématicien allemand Gabriel Altmann. Il a d'ailleurs baptisé cette loi (all. *Gesetz*) du nom de «Goebl-Gesetz».

[24] Des lois statistiques similaires régissent également le rendement lexical de l'activité langagière de l'*homo loquens*. Il s'agit là de la «loi Zipf-Mandelbrot». Pour une vue

dimensions de la *grammaire*, mais dans celles de l'*espace*. Depuis longtemps[25], nous qualifions cette activité humaine de « gestion basilectale de l'espace par l'*homo loquens* ».

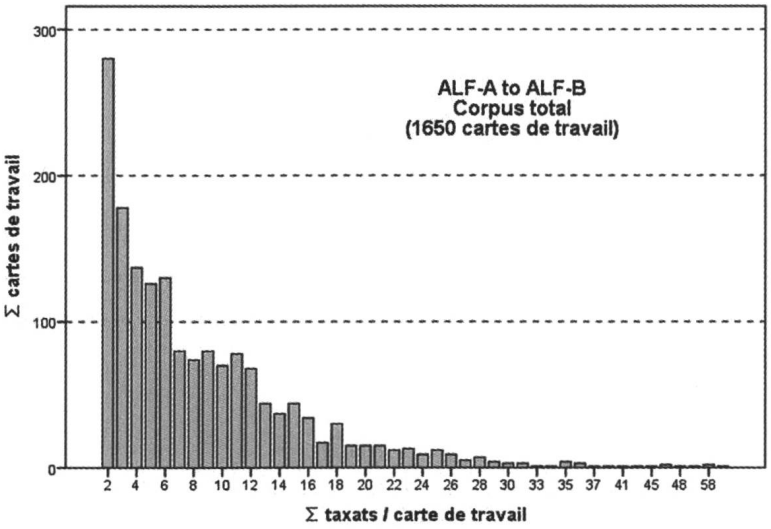


Figure 1 : Histogramme – avec 44 barres verticales de hauteurs inégales – montrant, pour 1 650 cartes de travail et 14 161 taxats tirés de la moitié sud de la série A des cartes de l'ALF, la relation entre le nombre de taxats par carte de travail (en abscisse) et le nombre absolu des cartes de travail (en ordonnée)

Le dernier tronçon de barre (à droite) se réfère à la valeur 67 (= taxats par carte de travail) et à une seule carte de travail.

3.3. La campagne réalisée entre 2010 et 2012 par Pavel Smečka, relative à 294 des 326 cartes originales de la série B de l'ALF

Les conditions extérieures de ce travail ressemblaient beaucoup à celles de la taxation de la série A réalisée entre 1997 et 2000. D'entrée de jeu, le 9e volume de la réimpression de l'ALF a été décomposé en cartes détachées si bien qu'il était facile de scanner d'abord les 326 cartes originales de la série B et de leur superposer ensuite – par voie informatique (et non plus par voie « analogique ») – un réseau polychrome avec plusieurs « parcours d'épreuve ».

d'ensemble des lois statistiques relatives à la linguistique, cf. Best 2003 (*passim*) et le manuel de linguistique quantitative (HSK 27) publié en 2005 par R. Köhler, G. Altmann et P. Piotrowski.

25 Cf. Goebl 1993, 277.

Par la suite, Pavel Smečka a mis au point un manuel de taxation et un jeu de modules appropriés. Étant donné la nature souvent très lacunaire des cartes de la série B, il a fallu en écarter une bonne trentaine d'entrée de jeu. C'est ainsi que seulement 294 des 326 planches originales de la série B ont été soumises à l'analyse taxatoire typique de la DM-S. Pour les résultats, voir la moitié droite du tableau 2 et la figure 2 (histogramme).

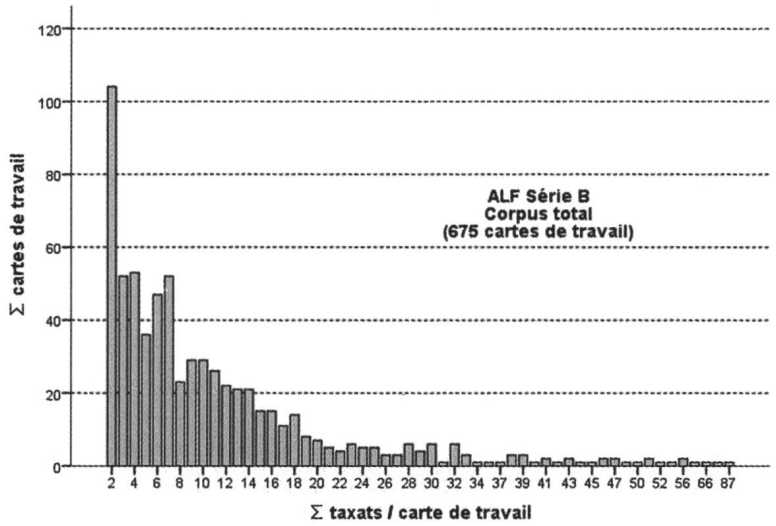


Figure 2: Histogramme – avec 55 barres verticales de hauteurs inégales – montrant, pour 675 cartes de travail et 7 534 taxats tirés de la série B des cartes de l'ALF, la relation entre le nombre de taxats par carte de travail (en abscisse) et le nombre absolu des cartes de travail (en ordonnée)

La déclivité de la courbe exponentielle de la figure 2 ressemble de très près à celle de la courbe de la figure 1. Cela prouve que les structures internes des deux corpora sont les reflets de deux réalités empiriques très similaires, sinon identiques.

La lecture et l'interprétation des chiffres du tableau 2 permettent cependant de faire ressortir une certaine divergence qui existe, entre nos deux corpora, dans leur degré de polynymie. Par la multiplication du nombre des *cartes de travail* avec *le nombre des points d'atlas* (= 329), l'on obtient la grandeur de la matrice de données respective. Exemple pour le corpus-ALF « A-to-B »: 1 650 × 329 = 542 850. Moyennant la division de cet effectif par le *nombre de taxats* (ou *aires taxatoires*) respectif, l'on obtient la grandeur (ou: l'étendue) d'une aire taxatoire moyenne: 542 850 ÷ 14 161 = 38,33 points d'atlas.

En jetant un coup d'œil sur les cartes 1 et 3 où figurent des aires taxatoires de taille très inégale, il est possible de mieux évaluer cet état des choses.

Le même calcul, fait avec les effectifs de la moitié droite du tableau 2, relative à la taxation de la seule série B de l'ALF et effectuée sous la responsabilité de P. Smečka, aboutit au résultat suivant: (675 × 329) ÷ 7 534 = 29,47. Cela signifie que la matrice de données de ce corpus comprend, en moyenne, des aires taxatoires sensiblement plus petites.

La raison de cette divergence tient à deux facteurs: à la structuration linguistique interne des matériaux géolinguistiques de base, c'est-à-dire des cartes originales de la série B de l'ALF, et à la précision, voire la finesse, de l'analyse taxatoire. Au vu des chiffres calculés, il semble bien que le travail taxatoire de P. Smečka se soit déroulé sous le signe du «splitting», alors que le même travail effectué par l'équipe salzbourgeoise active entre 1997 et 2000, ait obéi plutôt au principe du «lumping[26]».

On comprend mieux la fonction et les conséquences du «lumping» et du «splitting» en regardant les cartes 3 et 4. La légende de la carte 3 nous dit que les types linguistiques *baquer* et *vacher*, dérivés pourtant tous les deux d'un étymon latin commun (< *VACCÁRIU), correspondent à deux taxats différents. Ceci est dû au fait que le taxat *baquer* est porteur de deux évolutions phonétiques très «marquées» (1. bétacisme du V-initial, 2. conservation de la vélaire latine C devant A). Une acception taxatoire rigoureusement «étymologisante» aurait, par contre, rangé ces deux types dans la même classe tout en les définissant comme taxat individuel et isolé.

Une observation similaire vaut pour les types *rito* et *tiro* de la carte 4 dont le premier est issu d'une simple inversion de la charpente consonantique (*t-r*) du second type. Comme l'origine étymologique des deux taxats est de nature onomatopéique et que la position géographique de leurs aires respectives suggère la scission postérieure d'une aire à l'origine cohérente, P. Smečka a préféré en tirer deux taxats au lieu d'un seul.

Les conséquences de cette divergence sur les résultats-DM finals sont d'ailleurs minimes: voir à ce sujet notre contribution de 2014 où nous avons pu démontrer que l'utilisation comparée de corpora-DM *oligo-*, *méso-* et *poikilo-*nymes conduit à des résultats sensiblement égaux. Il n'y a donc lieu de s'inquiéter: la fiabilité statistique de la comparaison-DM de nos deux corpora est pleinement assurée.

---

26 Les désignations de «splitting» et «lumping» pour une mensuration plus ou moins fine des attributs d'entités empiriques à classifier ont été empruntées à la classification numérique (et à la biologie): cf. Sneath / Sokal 1973, 7 et 431.

4. Notices servant à la bonne intelligence des cartes dialectométriques de cet article

Dans ce paragraphe, nous tâcherons de fournir les outils méthodiques et conceptuels de base pour une bonne intelligence des 32 cartes-DM de l'annexe.

4.1. Observations méthodiques générales

Le but central de la DM-S est de déceler les structures de *profondeur* qui, invisibles d'entrée de jeu, existent dans (ou: sous) la masse des données d'un atlas linguistique. La DM-S admet en outre que les cartes d'atlas à proprement parler ainsi que leurs structures géolinguistiques directement accessibles à l'œil de l'observateur en constituent les structures de *superficie*. Comme la démarche exploratoire de la DM-S est inductive, elle se dirige du niveau des faits *particuliers* vers celui des faits *généraux*. Dans cette perspective, les différentes cartes d'un atlas linguistique ainsi que les cartes de travail (CT) (voir les cartes 1-4) qui en sont dérivées se situent au niveau du *particulier*, alors que les visualisations globales de type *choroplèthe* et *isarithmique* (voir les cartes 5-32) occupent le niveau épistémologique du *général*.

Il va de soi que la DM-S postule au préalable l'existence réelle de régularités finement articulées à l'intérieur des données d'un atlas linguistique, tout en repoussant l'idée de vouloir les inventer après coup. Sa démarche vise donc à la détection de ces structures préexistantes: elle est, de ce fait, *exploratoire*.

Nous supposons, en outre, que les régularités spatiales en question sont l'émanation directe d'une activité spéciale de l'homme, à savoir de la «gestion basilectale de l'espace par l'*homo loquens*». Ce concept, que nous utilisons depuis 1993[27], confère non seulement une place centrale à l'activité langagière des dialectophones, mais ouvre aussi un grand nombre de perspectives interdisciplinaires. Ceci surtout à cause du fait que le même acteur, à savoir l'homme (ou: *homo faber*), est appelé à gérer l'espace qu'il habite non seulement de manière *linguistique* mais aussi de beaucoup d'autres manières, qu'elles soient de nature *matrimoniale*, *génétique*, *agricole*, *économique*, *écologique*, *politique*, *ecclésiastique* ou autre.

---

[27] Voir nos contributions de 1993, 277; 2002, 5; 2003, 60 et 2005a, où ce concept figure même dans le titre de l'article en question.

4.2. Des données d'atlas aux cartes dialectométriques en passant par la taxation, la mensuration de similarités et de distances et les défis d'une visualisation appropriée

Voir la figure 3.

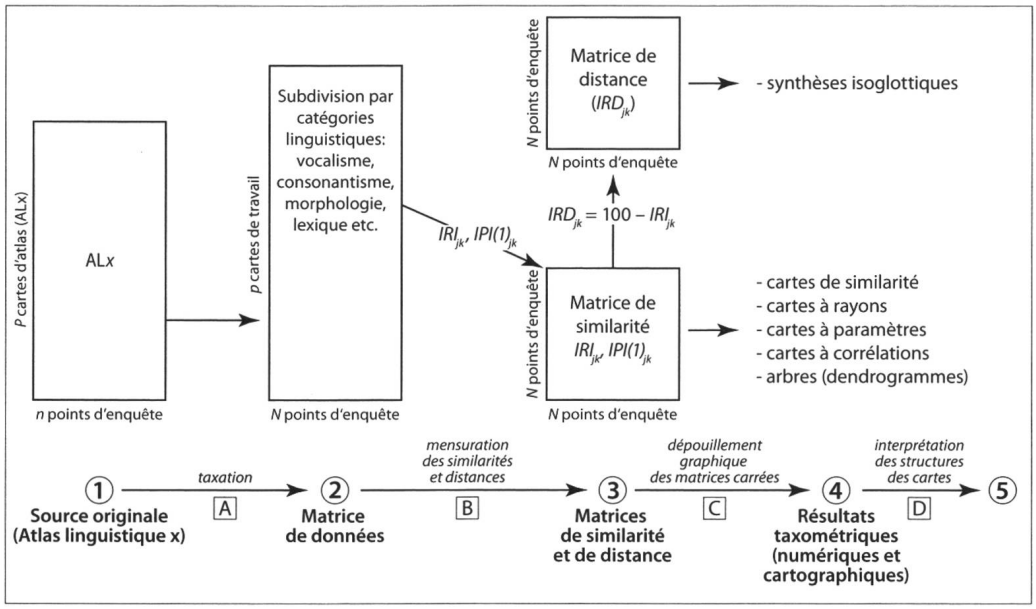


Figure 3: Diagramme de la chaîne méthodique utilisée par l'« Ecole Dialectométrique de Salzbourg »

Les méthodes de la DM-S constituent les maillons d'une chaîne qui s'étend entre les données brutes (figure 3: position 1) et les différentes visualisations finales (figure 3: position 5).

L'étape A, qui ne comprend que le traitement typologisant (ou: taxation) des données d'atlas, correspond largement à la vieille technique de la typisation des données originales d'un atlas linguistique examiné, appliquée couramment au sein de la romanistique depuis la période de Jules Gilliéron: voir à titre d'exemple la carte typisée (en noir et blanc) de l'abeille (ALF 1) dans Gilliéron 1918, ou les cartes (en couleurs) présentées en appendice dans Jaberg 1908. Précisons que ce genre de travail présupposait – et continue de le faire – le recours à des *cartes muettes* du réseau d'atlas en question. Le produit final de l'étape A est la matrice de données (N × p) dont le contenu se situe, du point de vue métrologique, au niveau de l'échelle *nominale* (ou: *cardinale*).

Le volet 'innovateur' de la DM-S débute avec l'étape B où il s'agit de mesurer, à l'aide de différents indices taxométriques, ou bien les *similarités* ou bien les *distances* qui existent entre les N vecteurs d'attributs de la matrice de données. Pour des raisons d'économie, nous n'utilisons, dans cet article, que l'« Indice Relatif d'Identité » (IRI_{jk}) pour la mensuration des *similarités*, et l'« Indice Relatif de Distance » (IRD_{jk}) pour celle des *distances* entre deux points d'enquête (mieux: vecteurs d'attributs) – *j* et *k* – de la matrice de données. Dans les deux cas, il s'agit d'indices très 'robustes' qui ont prouvé leur utilité à d'innombrables reprises. Tous les deux sont du type 'adansonien': ceci signifie qu'ils traitent tous les taxats de la matrice de données sur un pied d'égalité (numérique) et ne pratiquent, de ce fait, aucune pondération quantitative. Les scores calculés à l'aide de ces deux indices sont complémentaires, selon la formule suivante: $IRI_{jk} + IRD_{jk} = 100$.

Informations supplémentaires: les présentations les plus détaillées se trouvent dans Goebl 1984 I, 74 (en allemand); pour des explications plus fouillées en français voir Goebl 1981 (*passim*), 2002 et 2003; en anglais: 1993 et 2006 (tous les deux: *passim*).

Le but de l'étape C consiste dans le dépouillement *graphique* du contenu *numérique* des matrices de similarité et de distance, selon les besoins de la géolinguistique. Il s'agit donc d'extraire, des deux matrices carrées calculées, certains secteurs censés être *utiles* ou *féconds* pour les propos de la géolinguistique, et de les convertir en images ou graphiques appropriés. Ce faisant, la DM-S s'efforce depuis toujours d'opérer un *choix judicieux* et de ne pas déverser, dans un réflexe irréfléchi d'innovation technologique, toute la panoplie de la visualistique moderne sur ses lecteurs.

À cet égard, le programme VDM est très souple parce qu'il permet – à travers la possibilité de modifier à peu de frais la *finesse*, les *teintes* et les *modalités* de l'étalement spatial du symbolisme chromatique utilisé – d'adapter les graphiques-DM générés aux besoins *cognitifs* et aux capacités *physiologiques* de l'œil de l'observateur.

L'étape finale D est la plus exigeante et, en même temps, la plus prometteuse. Elle sert à l'interprétation linguistique des cartes établies, c'est-à-dire au mariage du rendement heuristique de l'analyse-DM à proprement parler avec les prérequis théoriques et pratiques de la recherche dialectologique en cours.

Informations supplémentaires: cf. Goebl 1984 I, 86*sqq.*; pour des visualisations réalisées en noir et blanc: cf. Goebl 1981 et 2006 (*passim*); et en couleurs: 1993, 2002, 2003 (et après). Pour le fonctionnement et l'application d'algorithmes de visualisation: cf. 1984 I, 93*sqq.* ainsi que 1981 et 1983 (*passim*).

4.3. La carte de similarité (CS)

Voir 5.2. et les cartes 9-16.

La CS est le pilier heuristique central de la DM-S. C'est un dessin relationnel – pivotant autour d'un point de repère à définir au préalable – dont l'utilité est double :

- montrer la position géotypologique de la dialecticité d'un point d'enquête au sein du réseau examiné, et
- visualiser l'étonnante régularité avec laquelle les scores de similarité diminuent dans l'espace.

C'est par le biais des CS qu'a été découverte la relation intime (et toujours très ordonnée) qui existe entre la distance kilométrique et la similarité linguistique par rapport à un point de repère donné : c'est que la similarité linguistique décroît avec l'augmentation de la distance kilométrique. Pour la dimension de l'*espace*, cette relation a la force d'une loi *géo-linguistique*, comparable aux *lois phonétiques* (« Lautgesetze »), découvertes par les néo-grammairiens de Leipzig à la fin du XIXe siècle, lesquelles régissent le déploiement des faits linguistiques dans la dimension du *temps*.

Du point de vue cartographique, la CS appartient à la classe des cartes *choroplèthes*. Un tel type de carte se sert d'un fond de carte polygonisé et d'un certain nombre de couleurs graduées (toutes extraites du spectre solaire) pour symboliser, à travers la variation de leur étalement sur la carte, la variation respective des données numériques à représenter.

La logique d'une CS veut qu'elle visualise le contenu numérique d'un des N vecteurs de la matrice de similarité. Il existe donc, pour une seule matrice de similarité (des dimensions N × N), le nombre N de CS. La même logique veut que le polygone du point de repère[28] reste toujours en blanc et que la visualisation à proprement parler concerne donc seulement N – 1 polygones. Pour la répartition algorithmique des couleurs (ou teintes) – dont le nombre peut varier, dans VDM, entre 2 et 20 – VDM offre trois solutions différentes, marquées par les sigles MINMWMAX, MEDMW et MED. Ces algorithmes permettent de conférer, au même potentiel numérique, trois sortes de visualisations qui se distinguent par le caractère plus ou moins accidenté de leurs profils choroplèthes (ou isarithmiques).

Les cartes de cet article n'utilisent que les algorithmes MINMWMAX et MEDMW. Tous les deux répartissent les N – 1 scores à visualiser des deux

[28] Le score-IRI attribué au point de référence (j), est toujours réflexif (IRI_{jj}) et, de ce fait, s'élève toujours à 100(%).

côtés de la moyenne arithmétique (MA) issue de ces valeurs. Les scores situés au-*dessus* de la MA seront toujours visualisés par des couleurs «chaudes» (jaune, orange, rouge), alors que les scores situés au-*dessous* de la MA recevront un coloriage «froid» (vert, bleu moyen, bleu foncé).

Pour l'algorithme MINMWMAX, les seuils qui délimitent les intervalles des deux côtés de la MA se calculent à partir de la différence qui existe, d'un côté, entre la MA et le score *minimal*, et, de l'autre, entre le score *maximal* et la MA. Au cas où *n* intervalles (ou couleurs) sont prévus pour la visualisation, chacune de ces deux différences est divisée par *n/2*.

L'algorithme MEDMW par contre, crée, des deux côtés de la MA, des intervalles dont l'ampleur – définie par le nombre de polygones marqués par la même couleur – est sensiblement égale. Cet algorithme génère des profils plus accidentés.

Toute CS comprend trois parties: le *cartogramme* à proprement parler, la *légende numérique* (en bas, à gauche) et l'*histogramme* (en bas, à droite). Dans la légende, l'on trouve les indications relatives aux seuils numériques des *n* intervalles et au nombre des points d'enquête répertoriés dans chaque intervalle.

Précisons, en outre, que toutes les cartes quantitatives de cet article disposent de (seulement) six paliers chromatiques. En général, cet effectif est suffisamment grand pour permettre une bonne détection des structures globales de la carte choroplèthe (ou isarithmique) en question.

Les contours des aires marquées par une des six couleurs mentionnées ne doivent d'ailleurs pas être confondus avec des «limites (ou: frontières) linguistiques». Ils correspondent plutôt aux concepts géographiques des *isohypses*, *isobares* ou *isohyètes*[29] qui représentent trois variantes du concept général de l'*isoligne*.

L'histogramme, constitué toujours de *2n* barres verticales à hauteurs inégales, sert à visualiser le caractère statistique de la distribution de fréquence en question. Il montre surtout si cette dernière est symétrique ou non, et aussi, si elle dispose d'un ou de plusieurs sommets, ou même des endroits de discontinuité.

Le recours au message de l'histogramme est indispensable pour la définition et aussi pour la bonne intelligence des «cartes à paramètres».

Informations supplémentaires: Goebl 1984 I, 90*sqq.*; 1981 et 1983 (*passim*).

---

[29] Il s'agit respectivement, dans les trois cas, de lignes qui, sur un graphe ou une carte météorologique, relient les points où l'altitude (*-hypse*), la pression atmosphérique (*-bare*) et la quantité de précipitations (*-hyète*) sont égales.

4.4. L'analyse des distances interponctuelles (cartes isoglottiques)

Voir 5.3. et les cartes 17-20.

L'élément iconique central de ce type de cartes n'est plus la *superficie* d'un polygone, mais l'ensemble des *côtés* qui l'entourent. Au cours de la visualisation, chacun des côtés de polygone deviendra porteur de deux symbolismes graphiques variables: de l'*épaisseur* et de la *couleur*. Le message global des cartes 17-20 dépend donc de l'interaction optique de 890 côtés de polygone, équivalant à autant de scores-IRD interponctuels, qui auparavant avaient été extraits de la matrice de distance.

La relation cartographique entre l'épaisseur et la couleur est la suivante: plus les scores-IRD sont grands, plus les côtés respectifs seront épais et tireront vers le bleu. L'épaisseur des côtés de polygone rappelle donc l'épaisseur des faisceaux d'isoglosses de la géographie linguistique classique.

En matière de cartographie, ce genre de cartes, basées sur des éléments linéaires, est appelé (carte) *isarithmique*. La perception visuelle du profil réticulaire de cartes isarithmiques pose plus de problèmes à l'œil que celle de cartes choroplèthes, qui ont l'avantage de lui présenter une surface (pseudo-)continue. Néanmoins, la compréhension intuitive de synthèses isoglottiques est facile. Ceci vaut également pour leur comparaison visuelle directe avec des superpositions isoglottiques de type traditionnel: voir à ce sujet les cartes respectives publiées par le romaniste suédois Arvid Rosenqvist en 1919 (réseau entier de l'ALF) ou par le romaniste autrichien Karl von Ettmayer en 1924 (moitiés nord et sud de l'ALF).

Il ne faut cependant pas surestimer le rang taxométrique de la synthèse isoglottique. Les cartes 17-20 reposent sur 890 scores de l'IRD. Cet effectif ne correspond qu'à 1,64% des scores de distance déposés dans la matrice de distance à disposition[30]. La portée classificatoire de cette analyse est donc relativement limitée. Précisons que cette limitation est due, avant tout, au principe de l'interponctualité des mensurations qui n'admet la considération de différences linguistiques qu'à très courte distance, tout en écartant toutes les autres possibilités de comparaison.

Informations supplémentaires: Goebl 1984 I, 183*sqq.*; 1983 (*passim*).

---

[30] En voici les calculs respectifs: dimensions de la matrice de distance (N × N): 329 × 329. Nombre des scores de distance stockés dans cette matrice et prêts à être utilisés (suivant la formule N·(N – 1) : 2): 53 956. Ensuite: 890 (valeurs interponctuelles): 53 956 = 1,64%.

4.5. Une carte à paramètres: l'analyse des «coefficients de Fisher»

Voir 5.4. et les cartes 21-24.

Dans le paragraphe 4.3., nous avons déjà évoqué le fait que les distributions de similarité, sur lesquelles reposent les cartes de similarité, servent non seulement à révéler l'étonnante régularité de l'étalement de similarités linguistiques dans l'espace, mais rendent aussi d'autres services taxométriques. C'est que la considération synoptique de leurs «paramètres caractéristiques» (comme, p. ex., le *minimum* et le *maximum*, la *moyenne arithmétique* (MA), l'*écart-type*, différents *coefficients d'asymétrie* etc.) permet la saisie ponctuelle de certains aspects quantitatifs du fonctionnement des dialectes dans l'espace.

Un des paramètres les plus intéressants du point de vue de la géolinguistique est le «coefficient d'asymétrie de Fisher» (CAF). Il permet de mesurer la symétrie (ou asymétrie) d'une distribution de fréquence qui peut revêtir aussi une certaine importance *linguistique*.

Prenons l'exemple des cartes 7 (CS depuis la Gascogne: P. 684) et 13 (CS depuis le Languedoc: P. 753). À côté du fait que l'histogramme de la carte 13 offre une silhouette plus symétrique que celui de la carte 7, l'on peut également constater – en contrôlant la position des scores de l'IRI des deux côtés de la moyenne arithmétique respective (MA) – que, sur la CS gasconne (carte 7), seulement 147 (= 44,8%) des 328 scores en question la dépassent, alors que sur la CS de type languedocien (carte 13) ce sont 170 (= 51,8%) scores-IRI sur 328 qui signalent la présence d'une *bonne*, voire *excellente* proximité linguistique avec le reste du réseau. En adoptant une perspective *communicative*, l'on est donc en droit de dire que l'insertion communicative du vecteur d'attributs du P.-ALF 753 dans le réseau-ALF de la série B est meilleure que celle du P.-ALF 684.

Remarquons, en outre, que chacun des 329 vecteurs d'attributs de nos deux réseaux est caractérisé par la présence d'un certain nombre d'aires taxatoires dont la taille est très variable. Parmi ces aires, il y en a dont la grandeur est *petite* (= *micro*-chorique), *moyenne* (= *méso*-chorique) ou *grande* (= *méga*-chorique). La spécificité communicative d'un vecteur d'attributs dépend donc de la *combinaison particulière* de ces trois types d'aires.

Plus un tel vecteur comprendra des aires méso- et mégachoriques, plus l'impact communicatif en sera grand. Or, la présence d'aires plutôt grandes est le résultat de certaines évolutions historiques, marquées par une suite d'activités diffusionnistes ou expansives intenses. La mensuration globale de cette *combinaison particulière* constitue donc un moyen très efficace pour saisir le caractère «dynamique» d'un locolecte donné. Or, nous savons depuis longtemps que cette mensuration est assurée de façon idéale par le «coefficient

d'asymétrie de Fisher» (CAF)[31] qui compte parmi les meilleurs indices pour la saisie numérique de l'asymétrie d'une distribution de fréquence.

La propriété *linguistique* captée et mesurée par le CAF a été appelée par nous d'abord «Sprachausgleich» en allemand, et baptisée par la suite «compromis (ou brassage) linguistique» en français.

Dans le cas de la symétrie parfaite d'une distribution de fréquence (ou de similarité), le score du CAF est de 0, alors qu'une asymétrie vers la *gauche* (= cumul des scores au-dessous de la MA) entraîne des scores *positifs* du CAF, et une asymétrie vers la *droite* (= cumul des scores au-dessus de la MA) des scores *négatifs*.

Sur les cartes 21-24, les scores minimaux (et négatifs) sont marqués en bleu (foncé et moyen), alors que les scores maximaux (ou positifs) sont visualisés en rouge. La signification *linguistique* de ces couleurs est la suivante:

- *Bleu*: zone caractérisée par un grand dynamisme linguistique et beaucoup de contacts linguistiques (passifs et actifs) à plus ou moins grande distance. Présence d'un grand nombre d'aires taxatoires *méso-* et *méga*choriques.
- *Rouge*: zone caractérisée par un comportement linguistique isolationniste, récessif ou défensif et par un nombre réduit de contacts linguistiques à grande distance. Présence de beaucoup d'aires taxatoires *micro*choriques.

La carte choroplèthe du CAF résulte de la mise en carte synoptique de N scores-CAF, émanant de N distributions de similarité qui, elles, sont stockées dans les N vecteurs d'une matrice de similarité carrée[32].

En règle générale, les profils choroplèthes des cartes-CAF offrent des structurations très claires dont l'importance est, avant tout, d'ordre *diachronique*. À côté de la carte à corrélations, c'est un des points forts de la DM-S.

Informations supplémentaires: la valeur et l'importance du CAF ressortent le mieux de nos études relatives au réseau complet de l'ALF: cf., avant tout, nos contributions de 2000, 2002 et 2003. Voir aussi nos travaux de 2004a, 2005b, 2006a, 2007a, 2007b et 2008a.

4.6. L'analyse dendrographique

Il s'agit là de l'utilisation du schéma ramifié de l'arbre pour saisir les relations (généalogiques) de parenté dont on pense qu'elles existent entre les dialectes. Or, il est bien connu que le recours à la métaphore de l'arbre a une

---

31 Cf. Goebl 1981, 394-401 et 1984 I, 150-153 (présentation de la formule du CAF et explication de ses fonctions). Ronald A. Fisher (1890-1962) était un biologiste et statisticien anglais.

32 Voir la figure 3.

longue (et heureuse) tradition en matière de linguistique. Ceci vaut également pour la géolinguistique, à condition toutefois que l'outil heuristique de l'*arbre* soit utilement combiné avec celui de la *carte*.

La construction algorithmique de schémas dendrographiques constitue, au sein de la classification numérique, une vieille pratique, tant sur le plan purement statistique que sur celui de l'application à des fins classificatoires. Il en résulte un nombre relativement grand d'algorithmes dendrographiques souvent très différents, dont l'utilité classificatoire est également très variable. Le dialectométricien-classificateur se trouve donc dans la nécessité d'effectuer, devant cette richesse méthodique, un choix judicieux, tout en considérant toujours les prérequis et finalités de la géolinguistique.

Parmi les algorithmes dendrographiques à disposition, ce sont ceux qui opèrent suivant les principes de la «classification ascendante hiérarchique» (CAH) qui sont les plus prometteurs pour les propos de la géographie linguistique. Ils procèdent, en se basant sur les données d'une matrice de similarité carrée (N × N), par une suite d'agglomérations binaires d'éléments aussi similaires que possible, et ceci en commençant au niveau des «feuilles» de l'arbre pour finir par en rejoindre la racine (ou: le tronc). Il s'agit donc, métaphoriquement parlant, d'une «ascension», à l'intérieur d'un arbre retourné, du *bas* (à partir de N «feuilles») vers le *haut* (vers une seule «racine»). Les agglomérations se traduisent graphiquement par la création de *ramifications* (ou: *bifurcations*) toujours *binaires* qui s'organisent sous la forme de hiérarchies disjonctives.

Un des algorithmes les plus utiles en matière de DM a été proposé, en 1963, par Joe Ward, Jr.[33] L'avantage en réside dans sa capacité de créer, à l'intérieur des ramifications («branchages» ou «ramages») de l'arbre, des groupements à la fois très homogènes vers l'intérieur («intra-group similarity») et très hétérogènes vers l'extérieur, c'est-à-dire face à des groupements environnants («inter-group distance»).

Dans la pratique concrète du travail-DM, il s'agit d'abord d'identifier, à l'intérieur de l'arbre calculé, certains branchages («dendrèmes») linguistiquement «significatifs» et, par la suite, d'en projeter leurs éléments dans l'espace. C'est de là que naissent, sur la carte polygonisée du réseau en question, des aires en général très cohérentes («chorèmes») dont le pavage offre un grand intérêt linguistique.

Évidemment, le logiciel VDM permet non seulement d'utiliser un grand nombre d'algorithmes dendrographiques différents, mais aussi de colorier à

[33] Joe H. Ward, Jr. (1926-2011) était un statisticien américain.

discrétion les arborescences ainsi générées, et de transmettre, par la suite, les dendrèmes ainsi définis sur la carte. Nous appelons ce processus « spatialisation » (du contenu dendrématique de l'arbre).

L'interprétation linguistique des arbres-DM est double : elle peut être faite dans la direction de la racine vers les feuilles (interprétation *dia*chronique) et dans le sens inverse (interprétation *syn*chronique).

La perspective *diachronique* part de l'idée, chère à la lexicostatistique et à la glottochronologie, que tout domaine linguistique avait été, d'entrée de jeu, exempt de différences (spatiales) et partant homogène, et que les structurations observables de nos jours sont le résultat d'une suite de partages et scissions successifs.

La perspective *synchronique* vise à l'intelligence des enchevêtrements et hiérarchies géographiques qui existent entre les différentes parties (branchages) de l'arbre et leurs pendants chorématiques.

Informations supplémentaires : cf. Goebl 1984 I, 172*sqq.* et 2003, 84-89. Pour une description détaillée du fonctionnement des algorithmes dendrographiques, cf. nos contributions de 1991 et 1992 (en allemand). Pour l'algorithme de Ward, nous renvoyons aux explications qu'en donnent trois grands manuels de classification numérique : Sneath / Sokal 1973, 204*sqq.* (en anglais) ; Bock 1974, 407*sqq.* (en allemand) et Chandon / Pinson 1981 (en français), 123*sq.*

4.7. Les cartes à corrélations

Voir 5.6. et les cartes 29-32.

Le but central de la DM corrélative est d'étudier le degré de convergence qui existe entre deux façons différentes de gérer le même secteur de l'espace, et d'en tirer des leçons d'ordre, avant tout, diachronique. Évidemment, cette finalité présuppose l'existence de « forces ordonnatrices » dans l'espace, émanant de l'action d'agents *anthropiques* ou *non*, et qui sont censées gérer l'espace de façon *non chaotique*. Dans cette perspective, les dyades en question peuvent comprendre deux catégories linguistiques différentes (p. ex. la *phonétique* et le *lexique*), deux mensurations-DM différentes (p. ex. la mensuration-*IRI* et la mensuration-*IPI*[34]) ou le contraste qui s'interpose entre la gestion basilectale de l'espace par l'*homme* et par des relations *non-anthropiques* (p. ex. par la géométrie euclidienne).

Du point de vue statistique, la DM corrélative demande la mise en place, pour chacune des deux forces à comparer par voie de corrélation, d'une

[34] Pour la formule de l'IPI (« Indice Pondéré d'Identité ») et son utilité dialectométrique, cf. Goebl 1984 I, 83-86 et 1987 (*passim*).

matrice (de *similarité*, etc.) carrée pourvue des mêmes dimensions extérieures (N × N). À l'aide de ces deux matrices carrées, il est désormais possible de comparer quantitativement – moyennant l'application d'indices de corrélation appropriés – N paires de vecteurs (disposant chacun de N scores de *similarité* etc.), d'en tirer N scores de corrélation et de les visualiser, en fin de compte, dans la forme habituelle d'une synopse choroplèthe.

L'indice de corrélation le plus approprié à cet effet est l'indice proposé successivement par Auguste Bravais (1811-1863, physicien et minéralogue français) et Karl Pearson (1857-1936, polymathe d'origine anglaise) : r(BP).

Le processus de corrélation, pour la réalisation duquel le logiciel VDM dispose d'un module adéquat à partir de 2004, mène à la détection de structures spatiales étonnamment bien ordonnées qui, en plus, peuvent être aisément interprétées du point de vue linguistique.

À l'instar de la synopse choroplèthe des CAF, l'intérêt linguistique des cartes à corrélations réside, avant tout, dans la diachronie. En effet, les structurations spatiales des cartes à corrélation constituent, en dernière analyse, un miroir finement articulé de dynamismes qui se sont déroulés, avec ou sans la responsabilité directe des humains, dans l'espace dans un passé plutôt lointain.

Informations supplémentaires: la contribution fondatrice pour la DM corrélative est de 2005a. À part cela, l'on trouve des applications et des discussions de la DM corrélative dans presque toutes nos publications postérieures à 2005; voir, en outre, nos travaux, centrés exclusivement sur la DM corrélative, de 2007c et 2008b (en allemand).

5. Présentation et interprétation comparatives des résultats-DM

L'agencement des cartes 1-32 est tel, à l'exception de la carte 15 (relative à la série B de l'ALF), que toutes les cartes de *gauche* se réfèrent à la dialectométrisation de l'ALF entreprise entre 1997 et 2000 et reposent donc sur les CT issues de la série A de l'ALF. Par contre, les cartes situées à *droite* se réfèrent à la dialectométrisation de la série B de l'ALF, faite par Pavel Smečka entre 2010 et 2012.

5.1. La taxation: définition et établissement des cartes de travail (CT)

Voir 4.2. et les cartes 1-4 de l'annexe.

Les cartes 1-4 sont des spécimens d'autant de cartes de travail d'ordre *phonétique* (cartes 1-2) et *lexical* (cartes 3-4). Pour des raisons d'économie cartographique, nous avons choisi, parmi les 1 650 CT du corpus « A-to-B » et les

675 CT du corpus de la série B, des échantillons dont la structuration interne est relativement simple : il s'agit donc de CT *méso*-nymes. Comme il importe, avant tout, de rappeler aux lecteurs la configuration et la fonction des CT comme éléments de base de la DM-S, la présentation de ces quatre CT sert à des fins uniquement illustratives.

La carte 1, qui est une CT *phonétique* 11-nyme, dispose, dans le corpus « A-to-B », de 52 pendants alors que la carte 2, également d'ordre *phonétique*, en a, dans le corpus de la série B, seulement treize. La carte 3, 8-nyme et d'ordre *lexical*, appartient à un groupe de 23 CT de la même structure polynymique et linguistique. De la carte 4, qui est 9-nyme tout en se référant également au *lexique*, il existe, au sein du corpus issu de la série B, 12 spécimens en tout qui partagent avec elle le même degré de polynymie et la même appartenance linguistique.

Les cartes 1 et 2 se réfèrent au même micro-contexte phonétique. Elles signalent la répartition spatiale des différents succédanés du nexus latin C devant A, *accentué* (carte 1 : MERCÁTU) et *prétonique* (carte 2 : CARRICÁTU). Bien que l'inventaire phonétique des deux séries de succédanés soit très similaire (voir les taxats 1-7), les deux profils choroplèthes ne se ressemblent que dans les grandes lignes.

La configuration de l'aire taxatoire rouge, relative au taxat 1 (C+A > š), offre une curiosité qui a déjà fait couler beaucoup d'encre. Elle consiste dans un provignement linéaire qui s'insinue, dans la direction du nord-ouest au sud-est, entre les domaines d'Oc et francoprovençal comme une sorte de « cordon sanitaire ». Il s'agit là du réflexe spatial d'un phénomène d'insécurité linguistique de la part de locuteurs trilingues (francoprovençal, occitan, français) qui habitent dans la zone de contact entre la langue d'Oc et le francoprovençal. Tiraillés entre deux registres chancelants et menacés (= occitan et francoprovençal), ils ont fini par adopter, dans certains contextes phonétiques, la solution proposée par le registre considéré comme étant le plus solide (= français).

Cet effet n'est pas passé inaperçu de la part des dialectologues du XXe siècle, parmi lesquels on comptait aussi Pierre Gardette (1906-1973), grand spécialiste du francoprovençal. Au sujet de cette curieuse intrusion d'éléments francisants le long d'un bande continue allant de Thiers (= P.-ALF 806, Puy-de-Dôme) à Sassenage (= P.-ALF 806, Isère, situé à la périphérie occidentale de Grenoble), il s'exprimait de la façon suivante : « On s'étonne parfois du chemin que certaines aires de francisation dessinent sur la carte de notre pays, spécialement dans le domaine francoprovençal, qu'il s'agisse du lexique, de la morphologie ou de la phonétique. C'est ainsi qu'Albert Dauzat [1877-1955 ;

HG / PS] a été intrigué par le long et large chemin qu'a pris le *ch* français de Roanne à Grenoble. Je voudrais montrer que ce chemin n'est pas seulement celui du *ch* français, mais qu'il est aussi celui d'autres traits de francisation; qu'il est moins un chemin qu'une zone francisante, allongée le long de la frontière qui sépare, à l'ouest et au sud-ouest, le francoprovençal du provençal; qu'il est le résultat d'une sorte d'effondrement de la région francoprovençale entière, effondrement produit par le climat d'incertitude régnant au sujet des mots et des formes francoprovençales et occitanes.» (Gardette, 1970, 291).

Des points de vue étymologique et onomasiologique, les cartes 3 et 4 n'offrent que peu de particularités. C'est pourquoi nous réservons nos remarques étymologiques aux formes (taxats) pour lesquelles il n'y a pas d'équivalents directs dans les langues standards.

Carte 3:

- taxat 4 *aoulier*: cf. FEW 7, 446*sq.* (appartenant à la famille de OVÍCULA «schäfchen»)
- taxat 7 *bouerin*: cf. FEW 1, 445-447 (appartenant à la famille de BOS «ochse»)

Carte 4:

- taxat 1 *cane*: d'origine onomatopéique: cf. REW 4671a (*kan*, Schallwort, Schnattern der Ente)
- taxat 3: *caná*: même origine que pour le taxat 1, avec rotation de l'accent sur la finale
- taxat 4: *guite*: d'origine onomatopéique: cf. FEW 4, 138 (*git*, schallwort) et REW 3768 (**git*, Schallwort)
- taxat 4 *canard*: d'origine onomatopéique: cf. FEW 2/1, 164-167 (*kann*) et REW 4671a (*kan*, Schallwort, Schnattern der Ente)
- taxat 6 *rito*: d'origine onomatopéique: cf. FEW 10, 420*sq.* (*rit*, schallwort) et REW 7337 (*rit*, Schallwort, Lockruf für die Ente)
- taxats 7 et 8: même origine que pour le taxat 6, avec métathèse de *r-t* à *t-r*
- taxat 7 *tirouno*: avec suffixe augmentatif: cf. FEW 10, 421a

5.2. L'analyse des cartes de similarité (CS)[35]

Voir 4.3. et les cartes 5-16 de l'annexe.

Comme les six corpora pris en compte (deux fois: corpus *total*, *phonétique* et *lexical*) disposent, chacun, de 329 points-ALF et que ceux-ci n'ont été soumis qu'à une seule application de l'indice de similarité IRI_{jk}, il existe 987

[35] Au prix d'un effort oculaire particulier, il est possible, de repérer, sur certains polygones des cartes 5-24 et 29-32, la localisation des scores *minimaux* et *maximaux*: voir, à cet effet, les hachures blanches qui recouvrent le coloriage de base (bleu foncé et rouge) des polygones respectifs. En règle générale, le maximum se trouve à proximité immédiate du point de repère alors que le minimum se trouve aux 'antipodes' de la CS en question.

paires de CS[36] dignes d'être présentées. Il était donc nécessaire d'un tirer judicieusement un échantillon aussi illustratif que possible. C'est ainsi que nous avons décidé de choisir, pour chacun des grands domaines dialectaux du sud de la France, une paire de CS basée sur le corpus *total*, et de présenter en plus, pour le domaine du languedocien (= P.-ALF 753: cartes 14-16) et le corpus de la série B, le contraste qui existe entre des CS relatives au corpora *total*, *phonétique* et *lexical*.

Il va de soi que, pour les cartes 5-16, les modalités visualisatrices sont toujours les mêmes.

Le point de repère des cartes 5 et 6 est le français standard, si bien que les deux cartes constituent, en dernière analyse, un bilan cartographique de la francisation du sud de la Galloromania. Bien que le profil de gauche, relatif au corpus de la série B, ressemble de très près à celui de droite, le profil choroplèthe de ce dernier est plus accidenté. Tel va être le cas aussi pour les autres CS de ce paragraphe. Remarquons en outre que, dans les légendes numériques des cartes de droite, apparaissent très souvent des scores-IRI_{jk} plus petits pour les *minima*, les *MA* et les *maxima*. Du point de vue statistique, ceci est dû, d'un côté, à la complexité onomasiologique des cartes de la série B et, de l'autre, à la philosophie taxatoire «splitting » de P. Smečka. Il est d'ailleurs difficile d'en préciser la part exacte.

Les cartes 7 et 8 sont des CS typiquement 'gasconnes'. À remarquer, sur la carte 8, la plus grande netteté avec laquelle le francoprovençal s'écarte du reste du réseau.

Les cartes 9 et 10 visualisent la position relationnelle du *francoprovençal* par rapport au reste du réseau du sud.

Les cartes 11 et 12 montrent la position du *provençal maritime* alors que les cartes 13 et 14 en font autant pour le *languedocien*.

Jusqu'ici, tous les profils choroplèthes s'appuyaient sur les effectifs *totaux* des corpora de base. En revanche, la carte 15, basée sur 313 CT *phonétiques* du corpus de la série B, et la carte 16, issue de 300 CT *lexicaux* du même corpus, offrent des perspectives relationnelles plus particulières[37]. En comparant les cartes 14 et 15 d'un côté, et la carte 16 de l'autre, l'on constate que, sur cette dernière, le décalage des similarités *lexicales* dans l'espace s'opère d'une

---

36 En voici le calcul: 2 corpora-ALF («A-to-B», série B) × 329 points d'enquête × 3 (sous-)corpora = 1 974 CS ÷ 2 = 987 paires de CS.

37 Dans notre contribution de 2003, nous avions déjà présenté des analyses-DM relatives uniquement au sud du réseau de l'ALF et à différentes sous-catégories linguistiques: voir les cartes 9, 10, 17, 23 et 24 de cet article (de 2003).

manière plus circulaire que dans les deux autres cas: voir, sur la carte 16, la silhouette plutôt circulaire de la zone jaune.

5.3. L'analyse des distances interponctuelles (cartes isoglottiques)

Voir 4.4. et les cartes 17-20 de l'annexe.

Quant à l'analyse-DM des *distances* interponctuelles (c'est-à-dire en fonction *discriminatoire*[38]), nous savons depuis longtemps que les résultats basés sur un corpus *phonétique* s'écartent légèrement de ceux qui s'appuient sur un corpus *lexical*. C'est la raison pour laquelle, dans ce paragraphe, les deux paires de cartes se réfèrent, d'un côté, à la *phonétique* (cartes 17 et 18) et au *lexique* de l'autre (cartes 19 et 20).

Les cloisonnements interponctuels des cartes 17 et 18 (relatives à la phonétique) se ressemblent beaucoup. Dans une certaine mesure, ceci vaut également pour les cartes 19 et 20 (lexique). Mais quand on compare les cartes 17 et 19 (ou 18 et 20) entre elles, l'on constate certaines divergences dans la trame du cloisonnement interponctuel, surtout en ce qui concerne la netteté avec laquelle se dessine le pourtour septentrional du domaine d'Oc (au nord du Limousin et de l'Auvergne). Sur les deux cartes lexicales (19 et 20), c'est surtout l'Auvergne qui – de même que la frange orientale du francoprovençal – devient le théâtre de cloisonnements interponctuels très intenses (symbolisés par des côtés de polygone épais coloriés en bleu foncé). Du point de vue de la géographie physique, cette zone correspond au cœur du Massif central qui semble avoir préservé l'ancienne bigarrure du lexique mieux que des zones géographiquement plus basses et moins escarpées.

5.4. L'analyse des « coefficients d'asymétrie de Fisher » (CAF)

Voir 4.5. et les cartes 21-24 de l'annexe.

Rappelons que le CAF mesure le degré du *brassage linguistique* qui existe à l'intérieur d'un réseau examiné, et aussi que la couleur *bleue* signale un brassage linguistique très intense alors que la couleur *rouge* renvoie à un grand isolement linguistique de la zone en question. Rappelons aussi que l'action de *brasser* s'applique au mélange des aires taxatoires dont il y a, dans le vecteur d'attributs de chaque point du réseau-ALF, des spécimens de *petite*, *moyenne* et *grande* taille. Ce qui compte, *dia-* et *syn*chroniquement parlant, c'est la

---

38 L'analyse interponctuelle peut être utilisée aussi en fonction *communicative*, c'est-à-dire pour le calcul (et la visualisation successive) de *similarités* interponctuelles: il en résulte les «cartes à rayons». Pour une présentation détaillée de ce genre de carte-DM, cf. Goebl 1983 (*passim*).

combinaison particulière de ces trois composantes. Un locolecte de l'ALF sera d'autant mieux connecté avec le reste du réseau qu'il aura, parmi les aires taxatoires de son vecteur d'attributs, plus de spécimens très grands[39].

Une fois de plus, les deux paires de cartes se réfèrent d'abord à la *phonétique* (cartes 21 et 22) et ensuite au *lexique* (cartes 23 et 24). La carte 21 fournit la structuration la plus claire: les polygones en bleu foncé forment un système continu de dépressions qui s'intercalent entre les grands domaines dialectaux du domaine d'Oc: entre le gascon et le languedocien, entre celui-ci, le provençal et l'auvergnat, et aussi entre le provençal et le francoprovençal. Rappelons qu'il s'agit là de zones d'échanges et d'interactions de vieille date.

Signalons aussi que, parmi les grands domaines dialectaux d'Oc, seul le languedocien occupe, de par le coloriage en bleu, une position privilégiée d'*interaction* et d'*échange* (et non pas de *retrait* et d'*isolation*).

En revanche, les zones marqués, sur la carte 21, en rouge et orange se distinguent par leur position «isolationniste» par rapport à l'ensemble du réseau: ce sont la Gascogne atlantique, le bloc pictavo-saintongeais (augmenté du Limousin), l'Auvergne, le Roussillon, la Provence et la partie orientale (romande) du domaine francoprovençal.

La carte 22, basée sur 313 CT *phonétiques* du corpus de la série B reflète, en principe, la même structure que la carte 21 bien qu'avec un certain nombre de distorsions et de déplacements. Par rapport à la carte 21, le champ du languedocien a été rétréci, alors que le champ du provençal s'est élargi vers l'ouest. Du reste, les délimitations du Limousin et du francoprovençal ressortent mieux qu'auparavant.

Les cartes 23 et 24 se réfèrent au *lexique*. Comme celui-ci se compose, dans le sud de la Galloromania, d'éléments spatialement souvent bien circonscrits, la synthèse des CAF reflète cet état des choses. C'est ainsi que la carte 23 montre une bande transversale d'échange qui s'étend de la Gironde aux Alpes, et qui semble séparer, au sein de notre réseau, la mouvance *oïlique* de sa contre-partie *occitane*. Les domaines du gascon, du languedocien, du provençal, du francoprovençal et du saintongeais-poitevin se démarquent, par contre, comme autant de foyers d'individualisme linguistique (mieux: lexical).

La structure du profil choroplèthe de la carte 24 obéit aux mêmes principes que celle de la carte 23, tout en exhibant une allure moins claire.

---

[39] Rappelons, pour saisir cet état des choses, le triplet terminologique *micro-*, *méso-* et *macro*-chorique que nous utilisons couramment pour qualifier l'étendue spatiale des différents taxats et de leurs aires.

5.5. L'analyse dendrographique

Voir 4.6. et les cartes 25-28 de l'annexe.

Il s'agit de quatre analyses arborescentes-CAH selon l'algorithme proposé par Joe Ward, Jr. dont le rendement dendrographique n'a été subdivisé – à titre d'exemple – qu'en six *dendrèmes* qui, eux, ont été projetés («spatialisés») ensuite sur le fond de carte (polygonal) de notre réseau, sous la forme d'autant de *chorèmes*. Pour faciliter la comparaison visuelle des structures des quatre pavages chorématiques, nous avons toujours attribué les mêmes couleurs aux différents secteurs de notre réseau. En outre, cette mesure permet de mieux saisir la position hiérarchique des six dendrèmes à l'intérieur des arbres respectifs.

Quant à l'architecture binaire de l'arbre, il est recommandé de toujours contrôler le rendement classificatoire de la première ramification située près de la racine de l'arbre. En l'occurrence, il s'agit de savoir si le dendrème teinté en bleu moyen (correspondant au limousin et à l'auvergnat, avec, le cas échéant, des provignements se projetant jusque dans les Alpes) a été agglomméré à la branche du nord (avec le poitevin-saintongeais et le francoprovençal) ou à celle du sud (avec le gascon, le languedocien et le provençal).

En regardant la carte 25, l'on constate que ce dendrème a été agrégé à la branche du sud, alors que, sur les cartes 26, 27 et 28, le même dendrème a été affilié à la branche du nord. La structure quantitative interne du dendrème-chorème en question est donc hybride; son affiliation s'avère donc comme *variable* ou carrément *précaire*. La même remarque vaut également pour le français qui a été agrégé trois fois (voir les cartes 25-27) au dendrème pictavo-saintongeais (en rouge) et seulement une fois (voir la carte 28) au dendrème «précaire» en bleu moyen.

La dépendance directe de deux dendrèmes du même nœud de l'arbre signale leur grande similarité réciproque. À remarquer la grande cohésion spatiale des chorèmes qui tous disposent d'aires parfaitement continues. Quant aux pourtours des frontières des chorèmes, il est bon de les comparer aussi avec le message géo-classificatoire d'autres cartes-DM, quitte d'ailleurs à ne pas les confondre avec le concept traditionnel des «limites (ou: frontières) linguistiques». Parmi les analyses-DM présentées dans cet article, ce n'est que l'analyse interponctuelle qui se rapproche de l'idée traditionnelle de la «limite ou frontière linguistique». Voir à ce sujet notre contribution théorique et méthodologique de 2011.

Évidemment, il ne faut pas oublier que les différents dendrèmes se distinguent par la taille de leur variabilité quantitative interne («intra-group-

variability»). Plus un dendrème se trouve proche de la racine, plus son hétérogénéité interne est grande. Cette hétérogénéité décroît au fur et à mesure que le point de rattachement du dendrème en question s'approche des feuilles de l'arbre. De ces automatismes, d'ailleurs inévitables en la matière, il résulte également un certain rapport entre la «fiabilité classificatoire» d'un dendrème et son hétérogénéité interne. Plus cette dernière est grande, plus la première est petite. Ceci, hélas, est une tare qui planera toujours au-dessus de la CAH.

Ceci n'empêche que l'analyse dendrographique s'avère très utile et efficace en matière de géolinguistique.

5.6. L'analyse corrélative

Voir 4.7. et les cartes 29-32 de l'annexe.

Rappelons que la DM corrélative sert, avant tout, à déterminer le degré de *con-* et *di*vergence de deux phénomènes qui se développent dans le même espace. Ces phénomènes peuvent être d'ordre *linguistique*, *taxométrique* ou de nature carrément *extralinguistique* comme les distances (ou proximités) kilométriques.

Comme les CS ont démontré clairement que la similarité linguistique diminue dans l'espace avec l'augmentation de la distance kilométrique par rapport à un point de référence préalablement établi, une des applications les plus intéressantes de la DM corrélative est l'analyse de la corrélation qui existe entre l'univers *anthropique* des *similarités* linguistiques et le monde *géométrique* des distances, mieux: *proximités* kilométriques. L'indice de corrélation le plus approprié à ces besoins est le coefficient de corrélation de Bravais et Pearson [r(BP)][40] qui, tout en oscillant entre les valeurs -1 et +1, saisit la corrélation linéaire – positive ou négative – entre deux variables quantitatives.

Depuis 2004, le logiciel VDM dispose d'un module de corrélation qui permet l'application rapide du r(BP) sur deux matrices carrées de la même structure extérieure (N × N) et la visualisation non moins rapide des résultats qui en découlent.

Après leur mise en carte, les scores-r(BP) doivent être interprétés comme suit:

- polygones coloriés en *bleu*:
  Les scores-r(BP) sont négatifs ou oscillent autour de zéro. Ceci signifie que les deux dimensions «en lice» ne vont pas dans la même direction et que, partant, il y a (mieux: avait) de fortes dissonances entre elles, dues à des conflits historiques,

[40] Pour la formule du r(BP) et tous les détails techniques ayant trait à la DM corrélative, voir Goebl 2005a (*passim*).

politiques, sociaux ou autres. En termes de linguistique, il y a lieu d'admettre, pour les régions coloriées en bleu, l'existence, dans le passé, de beaucoup de frictions et tensions entre ces deux dimensions.

- polygones coloriés en *rouge*:
  Les scores-r(BP) sont positifs et se situent aux environs de +1. Ceci signifie qu'entre les deux dimensions corrélées, il existe (mieux: existait) une sorte d'harmonie ou de «pas cadencé». L'étalement des deux dimensions dans l'espace obéit donc aux mêmes principes ou lois.

Les cartes 29 et 30 se réfèrent à la corrélation qui existe entre la *similarité* linguistique (selon l'IRI_{jk}) et la *proximité* kilométrique, calculée – à partir des coordonnées *x* et *y* des points d'enquête disponibles dans VDM – selon le théorème bien connu de Pythagore ($a^2 + b^2 = c^2$).

Les zones rouges suggèrent une évolution linguistique «pacifique», en accord avec les prémisses géométriques de l'espace, alors que les zones bleues laissent deviner la présence, dans un passé plus ou moins lointain, de contrastes et conflits linguistiques. C'est dans cette perspective que la Gascogne centrale et méridionale et l'ensemble du Languedoc et aussi de l'Auvergne semblent avoir mené une vie linguistique «naturelle», alors que l'espace pictavo-saintongeais (avec le Médoc) et surtout la zone de passage entre le provençal et le francoprovençal étaient le théâtre de dissonances entre le développement naturel des langues locales et les conditions euclidiennes de l'espace. Ces dissonances pouvaient se traduire, évidemment, en contacts et conflits linguistiques de toute sorte.

Les profils choroplèthes des cartes 29 et 30 se ressemblent beaucoup. Comme les cartes à corrélations visent, parmi les structures de profondeur dévoilées par la DM-S, celles avec la plus grande profondeur diachronique, la grande similarité entre les profils choroplèthes des cartes 29 et 30 n'a pas de quoi surprendre. C'est qu'à la grande *profondeur* diachronique de ces structures répond leur grande *robustesse* cartographique.

Les cartes 31 et 32 se réfèrent à la corrélation qui existe entre les dimensions linguistiques de la *phonétique* et du *lexique*. Il est, en effet, légitime de se demander s'il y a, entre les gestions *phonétique* et *lexicale* de l'espace par l'*homo loquens*, des différences spatialement bien ordonnées. Évidemment, cette question peut être posée aussi pour d'autres catégories linguistiques telles que le *vocalisme*, le *consonantisme*, la *morphologie* ou la *syntaxe*, toujours dans la mesure où ces catégories sont suffisamment bien documentées dans les atlas linguistiques. De telles questions touchent des chapitres de la grammaire variationnelle, restés jusqu'alors inexplorés ou carrément inconnus.

Le message iconique[41] des cartes 31 et 32 est très clair. À l'intérieur du réseau-ALF du sud, il existe, quant à la di- et convergence entre la *phonétique* et le *lexique*, une vaste zone de dissonance qui s'étend, à l'instar d'une bande ondulée, de la Gironde aux Alpes piémontaises. Au nord et au sud de cette bande de dissonance, s'ouvrent, par contre, des zones de convergence où, en termes de géolinguistique, les deux dimensions linguistiques vont de pair.

Il ne faut toutefois pas confondre la bande bleue ondulée avec le 'croissant' de mémoire ronjatienne. Celui-ci se situe plus au nord et est beaucoup plus étroit: voir à ce sujet notre étude historique de 2004c et le livre bien connu de G. Brun-Trigaud de 1990.

Signalons, en outre, que la configuration de base des cartes 31 et 32 apparaît aussi à l'issue d'une corrélation-DM entre nos deux corpora (« A-to-B », série B) *intégraux*. Elle semble donc avoir un caractère plus universel[42].

Une fois de plus, l'ensemble des CT tirées de la série B de l'ALF emboîte le pas de celles extraites de la série A. La série B semble donc contenir des matériaux dont l'orientation géolinguistique de base est très proche de celle de la série A.

6. Épilogue

Arrivés au terme de cette étude méthodiquement complexe, il nous semble indiqué d'en souligner encore une fois les grandes lignes argumentatives, et ceci à partir de trois points de vue, qui se réfèrent (1) à la dialectométrie à proprement parler, (2) à l'historique de l'ALF et (3) à la géolinguistique et dialectologie romane en général.

Commençons par une brève reconsidération du rôle voire du statut de la DM dans cet article.

La DM, bien loin d'être uniquement un 'amas de formules', présuppose l'existence d'une certaine conception (ou attitude) théorique de la part du chercheur par rapport aux données géolinguistiques qu'il est en train d'examiner. Cette conception embrasse, d'un côté, l'hypothèse de l'existence de structures spatiales bien ordonnées dans les profondeurs des données géolinguistiques en question, et, de l'autre, l'idée que ces structures sont, en der-

---

41 Les profils choroplèthes des cartes 29-32 ont été réalisés à l'aide de l'algorithme d'intervallisation MEDMW pour leur conférer une apparence plus claire et plus accidentée.

42 Voir aussi les cartes 23 et 24 dans Goebl 2005a (367) qui montrent également, bien que sur la base du réseau intégral de l'ALF, les effets de la corrélation entre deux dimensions *linguistiques* différentes.

nière analyse, l'émanation directe de l'activité communicative de l'homme dans l'espace. Depuis un certain temps, nous appelons cette activité 'gestion basilectale de l'espace par l'*homo loquens*'.

Précisons en outre que ces deux idées ne sont nullement 'révolutionnaires'. Tout au contraire : il est possible d'en examiner la lente éclosion entre le XVIIe et le XXe siècle, à commencer par les activités des géodètes et géographes royaux, déployées sous les règnes de Louis XIV (1654-1715) et de Louis XV (1715-1774). À la lumière de ces antécédents historiques, nous sommes en droit de dire que la genèse de la DM est intimement liée à l'action d'un substrat intellectuel et culturel *typiquement français* dont il n'y avait, dans le passé, aucune contre-partie dans d'autres régions de l'Europe (cf. Goebl 2006c).

De nos jours, la DM, telle qu'elle est pratiquée à Salzbourg, dispose des caractéristiques suivantes :

(1) C'est une discpline 'inductive' : ceci signifie qu'elle part d'une multitude de 'petits faits' empiriques épars, situés tous sur le plan du *particulier*, les métabolise ensuite du *qualitatif* au *quantitatif* par le biais de méthodes statistiques, et finit par émettre des résultats numériques (et aussi visuels) synthétiques qui se trouvent au niveau du *général*.

Cette procédure généralisante, bien connue et discutée déjà dans l'Antiquité et durant le Moyen Âge, fournit des résultats dont la pertinence se situe en dehors et à côté de celles des données de départ. En outre, ces résultats synthétiques permettent une meilleure maîtrise intellectuelle de la bigarrure souvent très déroutante des données de départ.

(2) Du point de vue *méthodique*, la DM est une discipline au moins bicéphale. C'est qu'elle dispose de deux volets méthodiques qui concernent, d'un côté, un certain nombre de procédures *mathématiques*, judicieusement extraites, pour les propos de la DM, de l'immense panoplie de la statistique contemporaine, et, de l'autre, un jeu bien circonscrit de méthodes *cartographiques* – toutes assistées par ordinateur – qui également ont été choisies avec beaucoup de circonspection pour les besoins de la géolinguistique.

Ces méthodes cartographiques visent, d'une part, la préparation visuelle adéquate du fond de la carte du domaine linguistique examiné avec les moyens de la géométrie de Voronoi (ici : de la France du sud dans l'optique de l'ALF) et, de l'autre, la visualisation efficace des résultats numériques calculés en recourant à la vertu cognitive (et classificatrice) des couleurs du spectre solaire.

Précisons que tous les requis de ces deux étapes (*numérique* et *visuelle*) peuvent être pris en compte et convenablement résolus à l'aide de notre logiciel DM-ique VDM (« Visual DialectoMetry »), en usage à Salzbourg depuis 1999/2000.

Quand on dit que la DM sert avant tout à la détection de structures géolinguistiques cachées dans la profondeur des données d'un atlas linguistique, cette constation devrait être complétée par une autre disant que la connaissance de telles structures de profondeur nous fournit de très précieuses informations sur le fonctionnement interne de réseaux géolinguistiques en général, et aussi sur le comportement sémiotique (souvent de nature sub-consciente) de l'*homo loquens* dans l'espace. C'est donc une recherche qui se situe dans la mouvance des 'universaux linguistiques' tout court.

Ceci signifie que n'importe quel atlas linguistique – qu'il soit de type roman, germanique, anglais, slave ou autre – peut devenir le point de départ de réflexions et recherches analogues.

Quant à l'historique des séries A et B de l'ALF, il est indéniable que nous sommes encore loin d'en connaître tous les secrets. Ceci n'empêche qu'à la distance de plus de cent ans de la publication de l'ALF, il est légitime de se demander si cette ignorance n'est pas due plutôt à l'effet de l'oubli dont on sait qu'il entache inexorablement l'évolution de toutes les sciences.

À la lumière de ce qui a été trouvé par Marie-Rose Simoni-Aurembou dans le dossier Gilliéron déposé à la BN de Paris, il semble évident que J. Gilliéron a été pris par une soif empirique toujours plus grande à mesure que les pérégrinations d'E. Edmont se développaient dans l'espace. L'ajout de plus de 300 questions aux 1400 items du questionnaire de base, semble s'être opéré au milieu de 1899, donc deux ans après le démarrage des enquêtes dialectales de l'ALF sur le terrain.

Malgré le caractère souvent très spécifique de ces 300 questions 'hors série', leur analyse DM-ique révèle, avec une netteté très claire, que les structures de profondeur de ce corpus supplémentaire s'alignent harmonieusement sur celles de notre corpus-ALF de base.

Soulignons en outre que dans n'importe quelle analyse DM-ique faite selon les principes de Salzbourg, le pivot central est la *matrice de données*: voir la figure 3, position 2. Elle constitue une sorte de *reproduction* (ou *image*) des données de l'atlas examiné, lesquelles, en dernière analyse, représentent également la reproduction d'une autre réalité, extrêment mouvante et bigarrée, à savoir des manifestations phoniques des sujets interrogés.

La responsabilité reproductrice incombe, pour le passage de l'oral au questionnaire, à l'explorateur, alors que, pour le transfert des données répertoriées sur les planches de l'atlas respectif dans les lignes et colonnes d'une matrice de données, elle reste du côté du dialectométricien à l'œuvre (ici: P. Smečka).

Chacun de ces deux métabolismes se fait sous les auspices d'une théorie de recherche bien déterminée. Soulignons, à ce propos, que l'écart diachronique qui se creuse entre l'effort du couple Gilliéron-Edmont et celui de P. Smečka, est relativement grand: il s'agit de plus d'un siècle!

N'oublions pas non plus qu'en y regardant de près, il manque encore un *maillon* très important dans cette *chaîne* tendue entre 1902 (*sqq.*) et 2014: c'est le *comportement métalinguistique* des sujets interrogés par E. Edmont. Car ces individus, tous plus au moins bilingues, ont dû commettre, à d'innombrables reprises, un remarquable exploit intellectuel, à savoir opérer un tri judicieux, dans leurs propres compétences linguistiques multiples, pour satisfaire à la curiosité linguistique d'E. Edmont désireux de ne collecter que des formes prétendument *basilectales*. Cette curiosité devait se manifester, le plus souvent, dans une question du type suivant: «Comment appelle-t-on ici tel objet?» N'oublions pas non plus que ce travail exploratoire a dû être fait sous l'impératif, cher à J. Gilliéron, de l'*intercomparabilité* aussi grande que possible des réponses recueillies.

En glissant en arrière le long de l'axe du temps depuis 1900, l'on peut se demander à partir de quelle date la collecte de données basilectales, pour un atlas linguistique du type-ALF, n'aurait plus été possible. Déjà en 1800, à cause de la diffusion encore trop faible du français (entendu comme langue générale d'une telle entreprise)? Ou seulement en 1700 (et avant) quand la connaissance du français et, partant, la présence d'un bilinguisme conscient (*dialecte local – français*) étaient encore circonscrites aux grandes villes et à quelques milieux privilégiés qui, en plus, habitaient souvent loin des provinces 'profondes' qui, dans ce contexte, nous intéressent tout particulièrement.

Pour stimulante qu'elle soit, cette question factice doit rester, par manque de données, sans réponse.

En ce qui concerne la 'plus-value' de cette étude pour les propos de la dialectologie et géolinguistique romanes, nous nous limiterons à quelques remarques sommaires. A ce propos, il semble intéressant d'évoquer trois aspects majeurs qui valent pour toutes les analyses DM-iques.

(i) Le premier concerne l'importance diachronique de telles études, étant donné le fait qu'elles suggèrent nécessairement de nombreuses questions sur l'évolution *diachronique* des structures synthétiques de profondeur découvertes. N'oublions pas

que les méthodes utilisées dans cet article ont déjà été appliquées plus d'une fois à des données scripturaires médiévales et qu'il a pu être démontré, de cette manière, qu'entre les structures de profondeur du domaine d'oïl de 1300 et celles de l'ALF (1902*sqq.*), il existe d'étonnantes similarités (cf. Goebl / Schiltz 2001 et Goebl 2008c). Il serait donc intéressant de réitérer cette expérience pour le domaine d'oc, quitte à en recueillir (et dépouiller), d'entrée de jeu, la totalité des textes vulgaires médiévaux disponibles.

(ii) Un autre aspect majeur qui est digne d'intérêt concerne la fonction 'signalétique' des apports de la DM pour d'autres secteurs de la linguistique romane. Certaines de nos mensurations-DM ont démontré l'existence de zones *hybrides* (ou de *passage*), où deux géotypes linguistiques se jouxtent ou se fondent d'une façon plus ou moins douce: voir surtout les cartes 21-24 de cet article. Il serait donc intéressant (et très prometteur à la fois) d'examiner de plus près les dialectes parlés dans ces parages de *transition*.
La même remarque vaut pour les cartes 31-32 qui montrent très clairement que les gestions *phonétique* et *lexicale* de la Galloromania du sud divergent considérablement à l'intérieur d'une bande relativement large qui s'étend entre l'Atlantique et les Alpes piémontaises (voir les polygones en bleu foncé sur les cartes 31 et 32). Ceci signifie que, dans cette zone, les *grammaires* des dialectes locaux sont *mixtes*, et, partant, *hybrides*. Ce fait pourrait constituer un défi intéressant pour les spécialistes de la *grammaire variationnelle*, d'autant plus que la même dissonance a pu être observée aussi entre toutes les autres sous-catégories grammaticales prises par nous en considération, comme, p. ex., le *vocalisme* vs. le *consonantisme*, les réalisations phonétiques *avant*, *sous* et *après* l'accent tonique, etc.

(iii) Le dernier des trois aspects majeurs évoqués ci-dessus, concerne d'éventuelles comparaisons (ou coopérations) interdisciplinaires qui pourraient être organisées, sous le signe de la DM et de la synthèse quantitative des données, entre la géolinguistique romane d'un côté et, de l'autre, les secteurs *géo-relationnels* de beaucoup de disciplines *humaines* (et au delà) qui, en dehors de la linguistique, s'occupent d'autres manières anthropiques de gérer l'espace. Parmi ces disciplines, je ne cite que la *géographie humaine*, l'*anthropologie*, l'*ethnographie*, la *génétique des populations*, la *démographie*, l'*économie*, etc.

Qu'on ne dise donc pas que la DM est une discipline auto-suffisante ou isolationniste: c'est qu'il reste assez de pain sur sa planche...

7. Remerciements

- dialectométrisation des cartes de la série B de l'ALF, établissement du corpus réduit «A-to-B» de l'ALF et confection des figures 1 et 2: Pavel Smečka, Salzbourg
- confection de la figure 3 et des 32 cartes de cet article: Werner Goebl, Vienne
- mise à disposition d'un extrait fait par Marie-Rose Simoni-Aurembou du dossier Gilliéron à la BN de Paris: Guylaine Brun-Trigaud, Nice

– toilette stylistique de notre texte français: Stéphanie Lescure, Marbourg sur Lahn (Allemagne)

Que Mesdames Brun-Trigaud et Lescure et Messieurs Smečka et Werner Goebl reçoivent ici l'expression de ma parfaite reconnaissance pour leur précieuse collaboration.

8. Abréviations techniques souvent utilisées

CAF	coefficient d'asymétrie de Fisher: cf. 4.5.
CAH	classification ascendante hiérarchique: cf. 4.6.
CS	carte de similarité: cf. 4.3.
CT	carte de travail: cf. 4.1.
DM	dialectométrie, dialectométrique, relatif à la dialectométrie
DM-S	Dialectométrie / École dialectométrique de Salzbourg
$IPI(x)_{jk}$	Indice Pondéré d'Identité (avec le poids x) (entre les vecteurs d'attributs *j* et *k*): cf. 4.7.
IRD_{jk}	Indice Relatif de Distance (entre les vecteurs d'attributs *j* et *k*): cf. 4.2.
IRI_{jk}	Indice Relatif d'Identité (entre les vecteurs d'attributs *j* et *k*): cf. 4.2.
MA	moyenne arithmétique: cf. 4.3.
MED	algorithme d'intervallisation: cf. 4.3.
MEDMW	algorithme d'intervallisation: cf. 4.3.
MINMWMAX	algorithme d'intervallisation: cf. 4.3.
r(BP)	coefficient de corrélation de Bravais et Pearson: cf. 4.5.
VDM	Visual DialectoMetry (logiciel dialectométrique créé par Edgar Haimerl): cf. 3.1.

Université de Salzbourg

Hans GOEBL
Pavel SMEČKA

9. Références bibliographiques

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

ALF = Gilliéron, Jules / Edmont, Edmond, 1902-1010. *Atlas linguistique de la France*, 10 vols., Paris, Champion (réimpression: Bologne, Forni, 1968).

Altmann, Gabriel, 1985. «Die Entstehung diatopischer Varianten», *Zeitschrift für Sprachwissenschaft* 4, 139-155.

Berschin, Helmut / Felixberger, Josef / Goebl, Hans, 2008[2]. *Französische Sprachgeschichte. Lateinische Basis, interne und externe Geschichte, sprachliche Gliederung Frankreichs. Mit einer Einführung in die historische Sprachwissenschaft*, Hildesheim, Olms.

Best, Karl-Heinz, 2003. *Quantitative Linguistik. Eine Annäherung*, Göttingen, Peust & Gutschmidt Verlag (2e édition).

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

Brun-Trigaud, Guylaine (1990). *Le Croissant: le concept et le mot. Contributions à l'histoire de la dialectologie française au XIXe siècle,* Lyon, Centre d'Études linguistiques Jacques Goudet.

Brun-Trigaud, Guylaine / Le Berre, Yves / Le Dû, Jean, 2005. *Lectures de l'Atlas linguistique de la France de Gilliéron et Edmont. Du temps dans l'espace. Essai d'interprétation des cartes de l'*Atlas linguistique de la France *de Jules Gilliéron et Edmond Edmont augmenté de quelques cartes de l'*Atlas linguistique de la Basse-Bretagne *de Pierre Le Roux*, Paris, CTHS.

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

Ettmayer, Karl von, 1924. «Über das Wesen der Dialektbildung, erläutert an den Dialekten Frankreichs», *Denkschriften der Akademie der Wissenschaften in Wien, philosophisch-historische Klasse* 66, 1-56 (avec 7 tableaux).

FEW = Wartburg, Walther von, *Französisches etymologisches Wörterbuch. Eine darstellung des galloromanischen sprachschatzes,* Leipzig/Bonn/Bâle, Schroeder/Klopp/Teubner/Helbing & Lichtenhahn/Zbinden, 25 vol., 1922-2002.

Gardette, Pierre, 1970. «Rencontre de synonymes et pénétration du français dans les aires marginales», *RLiR* 34, 280-305 (réimpression: Gardette, Pierre, 1983. *Études de géographie linguistique*, Strasbourg, Société de linguistique romane, 521-546).

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

Goebl, Hans, 1981. «Eléments de l'analyse dialectométrique (avec application à l'AIS)», *RLiR* 45, 349-420.

Goebl, Hans, 1983. «Parquet polygonal et treillis triangulaire: les deux versants de la dialectométrie interponctuelle», *RLiR* 47, 353-412.

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

Goebl, Hans, 1985. «Coup d'œil dialectométrique sur les *Tableaux phonétiques des patois suisses romands*», *Vox Romanica* 44, 189-233.

Goebl, Hans, 1987. «Points chauds de l'analyse dialectométrique: pondération et visualisation», *RLiR* 51, 63-118.

Goebl, Hans, 1991. «Una classificazione gerarchica di dati geolinguistici tratti dall'AIS. Saggio di dialettometria dendrografica», *Linguistica* 31 [Paulo Tekavčić sexagenario in honorem oblata], 341-352.

Goebl, Hans, 1992. «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 von Sprache*, Stuttgart, Steiner, 54-73 (*Zeitschrift für Dialektologie und Linguistik*, Beiheft 73).

Goebl, Hans, 1993. «Dialectometry. A Short Overview of the Principles and Practice of Quantitative Classification of Linguistic Atlas Data», in: Köhler, Reinhard / Rieger, Burghard B. (ed.), *Contributions to Quantitative Linguistics*, Dordrecht, Kluwer, 277-315.

Goebl, Hans, 2000. «La dialectométrisation de l'ALF: présentation des premiers résultats», *Linguistica* 40, 209-236 (avec 12 cartes en couleurs).

Goebl, Hans, 2002. «Analyse dialectométrique des structures de profondeur de l'ALF», *RLiR* 66, 5-63 (avec 24 cartes en couleurs).

Goebl, Hans, 2003. «Regards dialectométriques sur les données de l'Atlas linguistique de la France (ALF): Relations quantitatives et structures de profondeur», *Estudis Romànics* 25, 59-120 (avec 24 cartes en couleurs).

Goebl, Hans, 2004a. «Sprache, Sprecher und Raum: Eine kurze Darstellung der Dialektometrie. Das Fallbeispiel Frankreich», *Mitteilungen der Österreichischen Geographischen Gesellschaft* 146, 247-286 (avec 24 cartes en couleurs).

Goebl, Hans, 2004b. «Bref aperçu sur les problèmes et méthodes de la dialectométrie (avec application à l'ALF)», in: Éloy, Jean-Michel (ed.), *Des langues collatérales. Problèmes linguistiques, sociolinguistiques et glottopolitiques de la proximité linguistique.* Actes du Colloque international réuni à Amiens, du 21 au 24 novembre 2001, Paris/Budapest/Turin, L'Harmattan/Université de Picardie Jules Verne, vol. I, 39-60 (avec 8 cartes en couleurs).

Goebl, Hans, 2004c. «Das "Croissant" – eine Nachschau im Abstand von 90 Jahren (mit zwei dialektometrisch erstellten Farbkarten)», in: Noll, Volker / Thiele, Sylvia (ed.), *Sprachkontakte in der Romania. Zum 75. Geburtstag von Gustav Ineichen*, Tübingen, Niemeyer, 159-172 (avec 2 cartes en couleurs).

Goebl, Hans, 2005a. «La dialectométrie corrélative. Un nouvel outil pour l'étude de l'aménagement dialectal de l'espace par l'homme», *RLiR* 69, 321-367 (avec 24 cartes en couleurs).

Goebl, Hans, 2005b. «Dialektometrie (Art. 37)», in: Köhler, Reinhard / Altmann, Gabriel / Piotrowski, Rajmund G. (ed.), *Quantitative Linguistik / Quantitative Lin-*

guistics. Ein internationales Handbuch / An International Handbook (*Handbücher zur Sprach- und Kommunikationswissenschaft* [HSK], vol. 27), Berlin/New York, de Gruyter, 498-531 (avec 8 cartes en couleurs relatives à l'ALF).

Goebl, Hans, 2006a. «Recent Advances in Salzburg Dialectometry», *Literary and Linguistic Computing* 21, 411-435.

Goebl, Hans, 2006b. «DM und VDM: kurzgefasste Vorstellung der Dialektometrie (DM) und des Programmpakets Visual DialectoMetry (VDM) mit Beispielen aus dem "Atlas linguistique de la France" (ALF)», in: Timuška, A. (ed.), *Proceedings of the 4th International Congress of Dialectologists and Geolinguists* [Riga 2003], Riga, Latvian Language Institute, 153-190 (avec 16 cartes en couleurs).

Goebl, Hans, 2006c. «Warum die Dialektometrie nur in einem roman(ist)ischen Forschungskontext entstehen konnte», in: Dahmen, Wolfgang / Holtus, Günter / Kramer, Johannes / Metzeltin, Michael / Schweickard, Wolfgang / Winkelmann, Otto (ed.), *Was kann eine vergleichende romanische Sprachwissenschaft heute (noch) leisten?* [Romanistisches Kolloquium XX], Tübingen, Narr, 291-317.

Goebl, Hans, 2007a. «Dialectometry: Theoretical Prerequisites, Practical Problems, and Concrete Applications (mainly with Examples drawn from the "Atlas linguistique de la France", 1902-1010)», in: *Geolinguistics around the World. Proceedings of the 14th NIJLA International Symposium* (Tokyo, August 22-23, 2007), Tokyo, National Institute of Japanese Language [NIJLA], 65-74 (aussi in: *Dialectologia* [Barcelone], special issue I, 2010) (avec 6 cartes en couleurs).

Goebl, Hans, 2007b. «A Bunch of Dialectometric Flowers: a brief Introduction to Dialectometry», in: Smit, Ute / Dollinger, Stefan / Hüttner, Julia / Kaltenböck, Gunther / Lutzky, Ursula (ed.), *Tracing English through Time. Explorations in Language Variation. In Honour of Herbert Schendl on the Occasion of his 65th Birthday*, Wien, Braumüller, 133-171 (avec 12 cartes en couleurs relatives à l'ALF).

Goebl, Hans, 2007c. «Kurzvorstellung der Korrelativen Dialektometrie», in: Grzybek, Peter / Köhler, Reinhard (ed.), *Exact Methods in the Study of Language and Text. Dedicated to Gabriel Altmann on the Occasion of his 75th Birthday*, Berlin, Mouton de Gruyter, 165-180.

Goebl, Hans, 2008a. «Le laboratoire de dialectométrie de l'Université de Salzbourg», *ZfSL* 118, 35-55 (avec 16 cartes en couleurs).

Goebl, Hans, 2008b. «Die korrelative Dialektometrie. Eine Kurzvorstellung anhand von Beispielen aus AIS und ALF», in: Bernhard, Gerald / Siller-Runggaldier, Heidi (ed.). *Sprache im Raum - Raum in der Sprache* [Spazi comunicativi / Kommunikative Räume, vol. 4], Frankfurt/Main etc., Peter Lang, 67-90 (avec 6 cartes en couleurs relatives à l'ALF).

Goebl, Hans, 2008c. «Sur le changement macrolinguistique survenu entre 1300 et 1900 dans le domaine d'Oïl. Une étude diachronique d'inspiration dialectométrique», *Dialectologia* (Revista electrònica, Barcelona) 1, 3-43.

Goebl, Hans, 2009. «Quelques coups d'œil dialectométriques sur l'Atlas linguistique de la France: structures de surface et structures de profondeur», in: Dalbera-Stefanaggi, Marie-José / Simoni-Aurembou, Marie-Rose (ed.), *Images de la langue: représentations spatiales, sémantiques et graphiques*, Paris, Editions du CTHS, 39-60 (avec 12 cartes en couleurs).

Goebl, Hans, 2010. «Dialectometry and quantitative mapping», in: Lameli, Alfred / Kehrein, Roland / Rabanus, Stefan (ed.), *Language and Space. An International Handbook of Linguistic Variation*, vol. 2: Language Mapping (*Handbücher der Sprach- und Kommunikationswissenschaft* [HSK] 30.2.), Berlin, de Gruyter, 1re partie (texte) 433-457, 2e partie (cartes), 2201-2212 (avec 12 cartes en couleurs).

Goebl, Hans, 2011. «Áreas, fronteras, similitudes y distancias: lección breve de geolingüística cuantitativa», in: de Andrés Díaz, Ramón (coord.), *Lengua, ciencia y fronteras,* Uviéu / Oviedo, Ediciones Trabe/Universidá d'Uviéu [Anexos de *Revista de Filoloxía Asturiana*, II], 11-33 (avec 31 cartes en couleur).

Goebl, Hans, 2012. «Introduction aux problèmes et méthodes de l'"École dialectométrique de Salzbourg" (avec des exemples gallo-, italo- et ibéroromans)», in: Álvarez Pérez, Afonso / Carrilho, Ernestina / Magro, Catarina, *Proceedings of the International Symposium on Limits and Areas in Dialectology (LimiAr)*, Lisbonne 2011 (‹http://limiar.clul.ul.pt›), Lisbonne, Centro de Linguística da Universidade de Lisboa, 117-166 (avec 28 cartes en couleurs relatives à l'ALF).

Goebl, Hans, 2013. «Le Baiser de la Belle au bois dormant ou: des péripéties encourues par la géographie linguistique depuis Jules Gilliéron», in: Caprini, Rita (ed.), *Dialectologie. Corpus, atlas, analyses*, Alessandria, Edizioni dell'Orso [= Corpus 12, 2013], 61-84.

Goebl, Hans, 2014. «L'impact de la polynymie des cartes d'atlas sur le résultat de calculs dialectométriques», in: *Linguistique romane et linguistique indo-européenne. Mélanges offerts à Witold Mańczak à l'occasion de son 90e anniversaire*, Cracovie, PAU-UJ, 243-260 (avec 10 cartes en couleur).

Goebl, Hans / Schiltz, Guillaume, 2001. «Der "Atlas des formes et des constructions des chartes françaises du 13e siècle" von Anthonij Dees (1980) – dialektometrisch betrachtet», in: Gärtner, Kurt / Holtus, Günter / Rapp, Andrea / Völker, Harald (ed.): *Skripta, Schreiblandschaften und Standardisierungstendenzen. Urkundensprachen im Grenzbereich von Germania und Romania im 13. und 14. Jahrhundert.* Beiträge zum Kolloquium vom 16. bis 19. September 1998 in Trier, Trier, Kliomedia, 169-221.

Jaberg, Karl, 1908. *Sprachgeographie. Beitrag zum Verständnis des Atlas linguistique de la France*, Aarau, Sauerländer (traduction espagnole: *Geografía lingüística. Ensayo de interpretación del «Atlas lingüístico de Francia».* Traducción de A. Llorente y M. Alvar, Granada, Universidad de Granada. Secretariado de Publicaciones, 1959).

Notice = Gilliéron, Jules, 1902. *Atlas linguistique de la France. Notice servant à l'intelligence des cartes*, Paris, Champion (réimpression: Bologne, Forni, 1968).

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

Pop, Sever, 1950. *La dialectologie. Aperçu historique et méthodes d'enquêtes linguistiques*, Louvain, chez l'auteur/Gembloux, Duculot, 2 vol.

Pop, Sever / Pop, Rodica Doina, 1959. *Jules Gilliéron. Vie, enseignement, élèves, œuvres, souvenirs*, Louvain, Centre international de dialectologie générale.

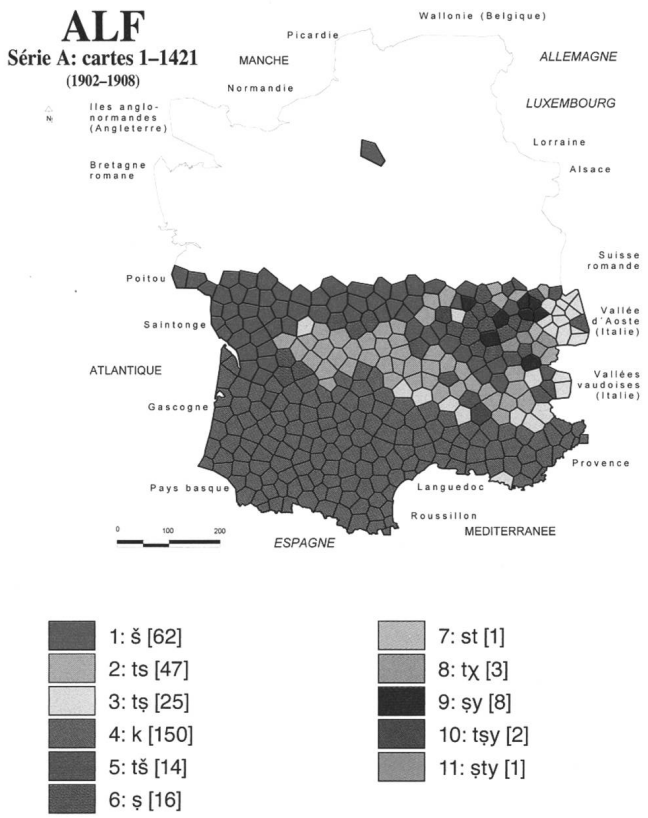
REW = Meyer-Lübke, Wilhelm, 1935³. *Romanisches etymologisches Wörterbuch*, Heidelberg, Winter.

Rosenqvist, Arvid, 1919. «Limites administratives et division dialectale de la France», *Neuphilologische Mitteilungen* 20, 87-119 (carte publiée à nouveau in: Berschin / Felixberger / Goebl 2008², 261).

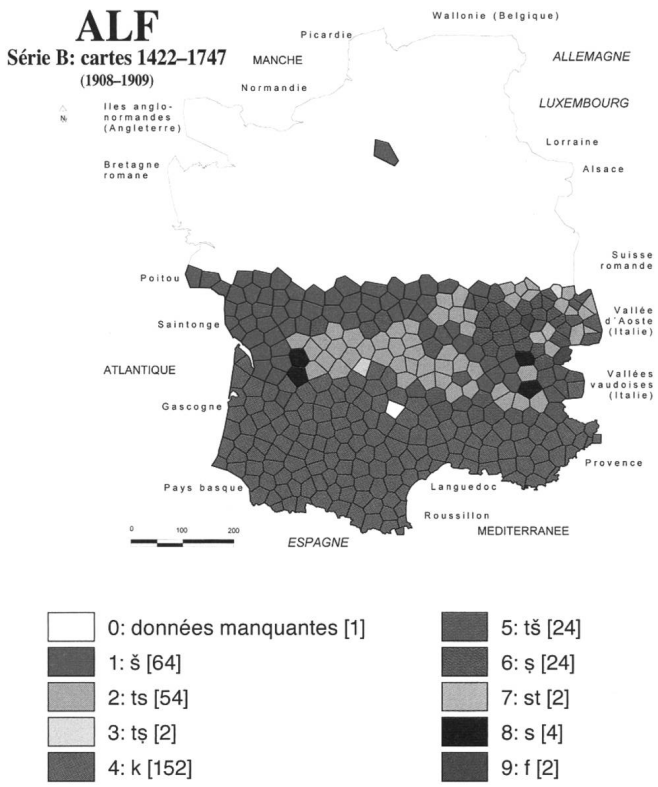
Sneath, Peter A. / Sokal, Robert R., 1973. *Numerical Taxonomy. The Principles and Practice of Numerical Classification*, San Francisco, Freeman.

Suppléments = Gilliéron, Jules / Edmont, Edmond, 1920. *Atlas linguistique de la France. Suppléments. Tome premier*, Paris, Champion (réimpression: Bologne, Forni, 1969).

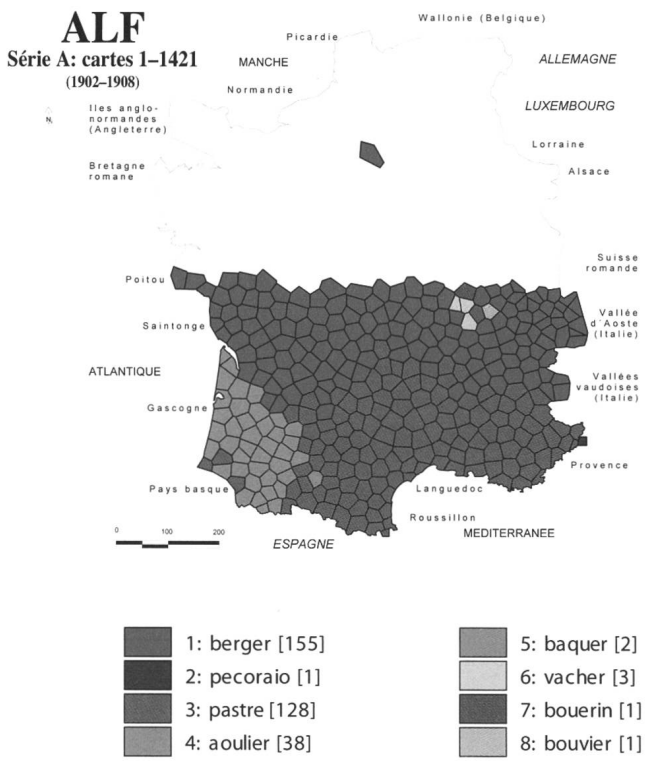
Ward, Joe, Jr., 1963. «Hierarchical Grouping to optimize an Objective Function», *Journal of the American Statistical Association* 58, 236-244.



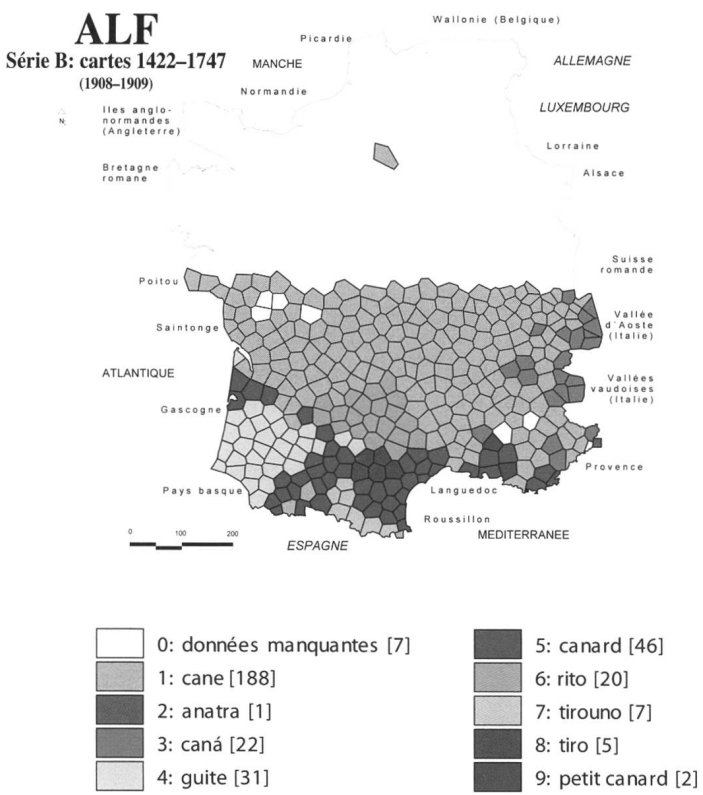
Carte 1: Exemple d'une carte de travail *11*-nyme de nature *phonétique*: les succédanés de C latin devant A latin accentué dans MERCÁTU (selon ALF 812 *marché*)
Réseau-ALF réduit: 329 PP. (selon la série B de l'ALF).
Le corpus-ALF „A-*to*-B" dispose de 53 CT de la même polynymie et catégorie linguistique.



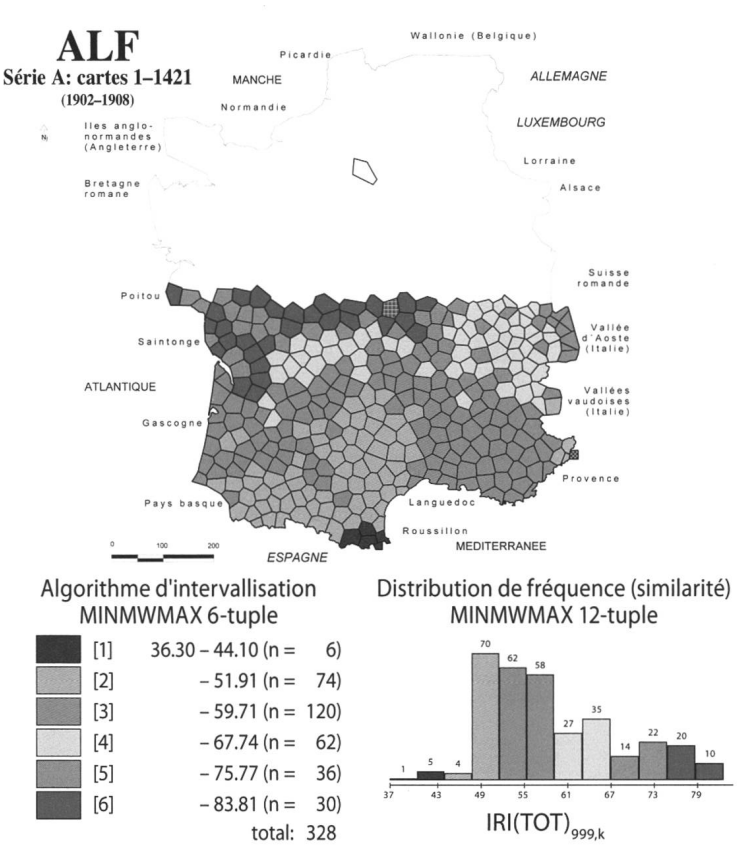
Carte 2: Exemple d'une carte de travail *9*-nyme de nature *phonétique*: les succédanés de C latin devant A latin prétonique dans CARRICÁTU (selon ALF 1494 *chargé* [*chargée*])
Réseau-ALF réduit: 329 PP. (selon la série B de l'ALF).
Le corpus-ALF „série B" dispose de 14 CT de la même polynymie et catégorie linguistique.



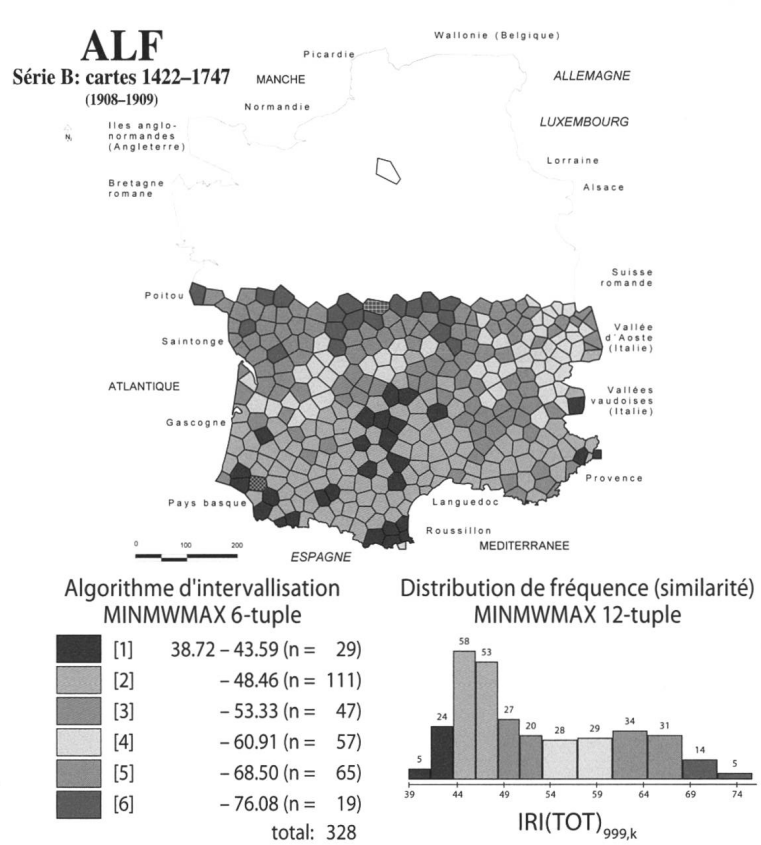
Carte 3: Exemple d'une carte de travail *8*-nyme de nature *lexicale*: les dénominations du berger dans la moitié sud de la Galloromania (selon ALF 128 *berger* [*bergère*])
Réseau-ALF réduit: 329 PP. (selon la série B de l'ALF).
Le corpus-ALF «A-*to*-B» dispose de 23 CT de la même polynymie et catégorie linguistique.



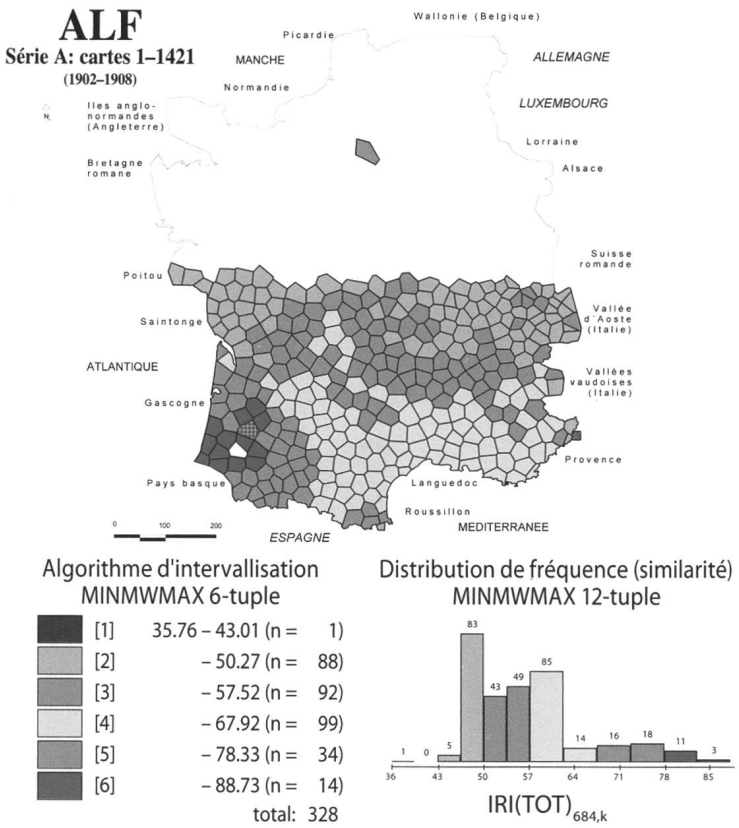
Carte 4: Exemple d'une carte de travail *9*-nyme de nature *lexicale*: les dénominations de la cane dans la moitié sud de la Galloromania (selon ALF 1486 *cane* [*caneton*]).
Réseau-ALF réduit: 329 PP. (selon la série B de l'ALF).
Le corpus-ALF „série B" dispose de 12 CT de la même polynymie et catégorie linguistique.



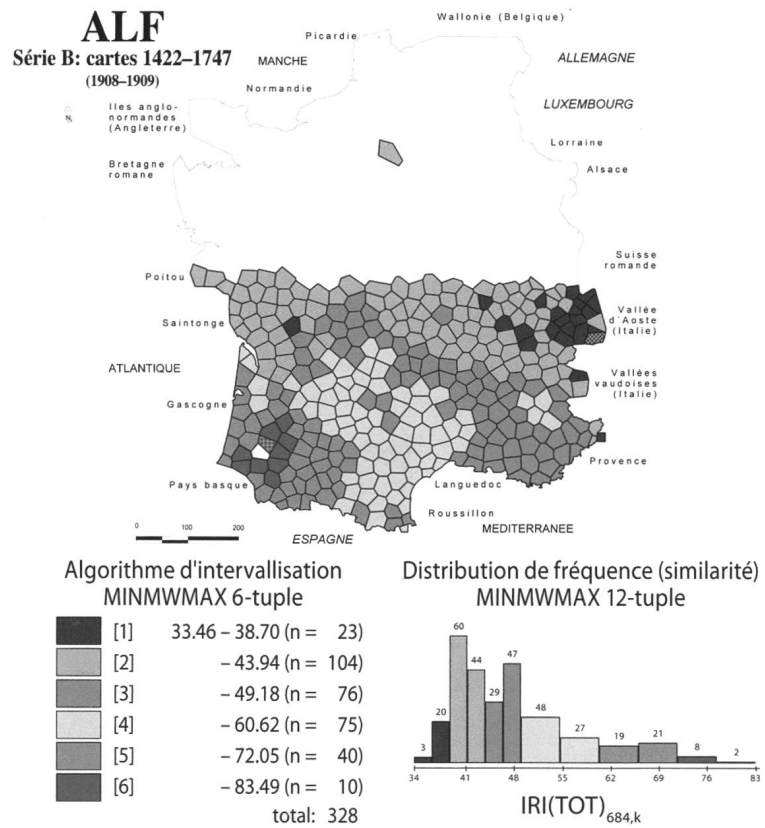
Carte 5: Profil choroplèthe d'une carte de similarité (selon IRI_{jk}) relative au point-ALF factice 999 (français standard)
Corpus: 1650 cartes de travail (série A: toutes les catégories linguistiques)
Algorithme d'intervallisation: MINMWMAX 6-tuple



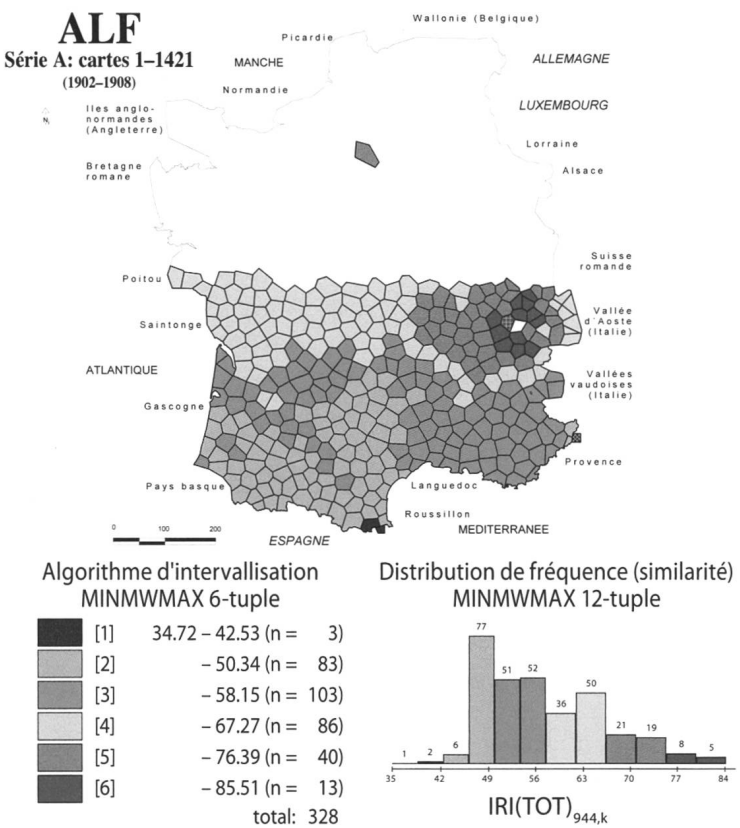
Carte 6: Profil choroplèthe d'une carte de similarité (selon IRI_{jk}) relative au point-ALF factice 999 (français standard)
Corpus: 675 cartes de travail (série B: toutes les catégories linguistiques)
Algorithme d'intervallisation: MINMWMAX 6-tuple



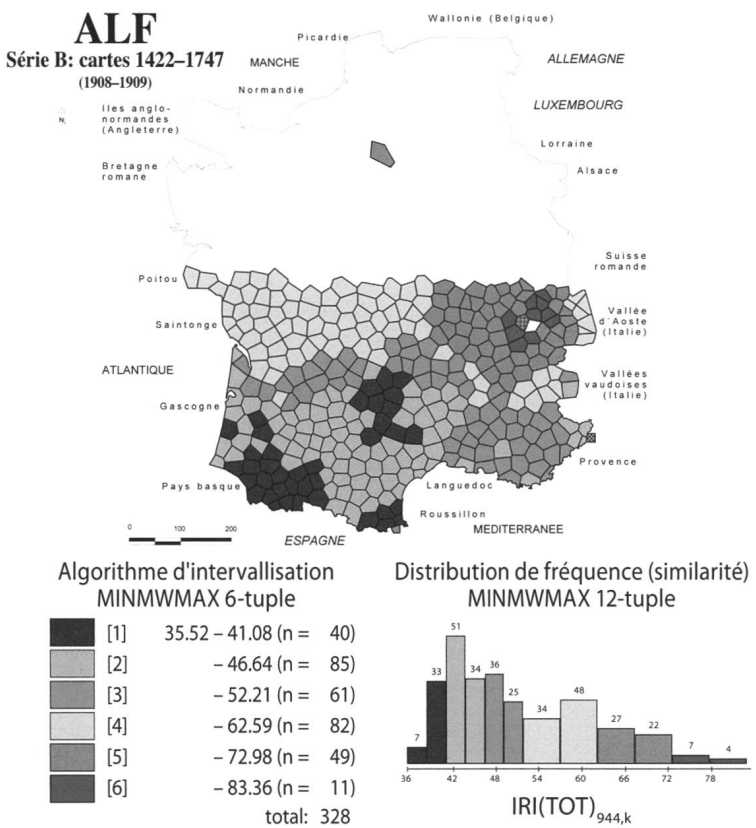
Carte 7: Profil choroplèthe d'une carte de similarité (selon IRI_{jk}) relative au point-ALF 684 (Hagetmau, Landes)
Corpus: 1650 cartes de travail (série A: toutes les catégories linguistiques)
Algorithme d'intervallisation: MINMWMAX 6-tuple



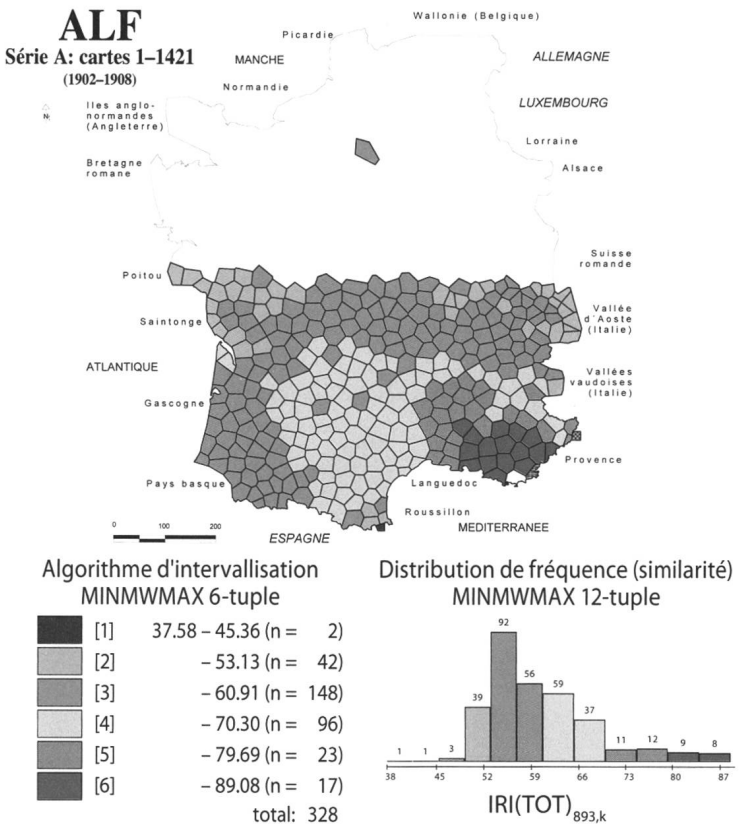
Carte 8: Profil choroplèthe d'une carte de similarité (selon IRI_{jk}) relative au point-ALF 684 (Hagetmau, Landes)
Corpus: 675 cartes de travail (série B: toutes les catégories linguistiques)
Algorithme d'intervallisation: MINMWMAX 6-tuple



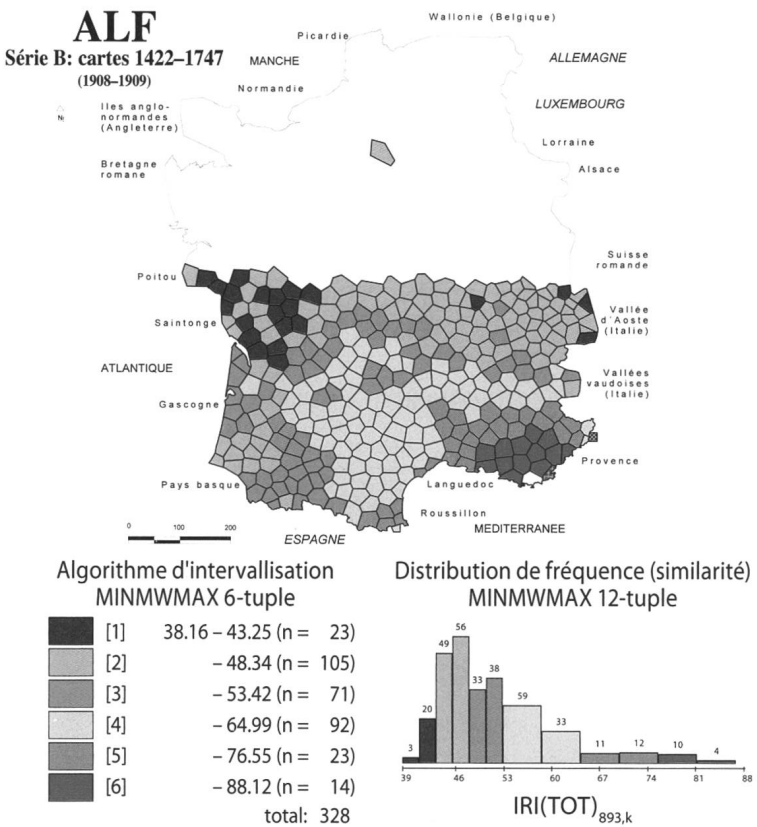
Carte 9: Profil choroplèthe d'une carte de similarité (selon IRI_{jk}) relative au point-ALF 944 (Thônes, Haute-Savoie)
Corpus: 1650 cartes de travail (série A: toutes les categories linguistiques)
Algorithme d'intervallisation: MINMWMAX 6-tuple



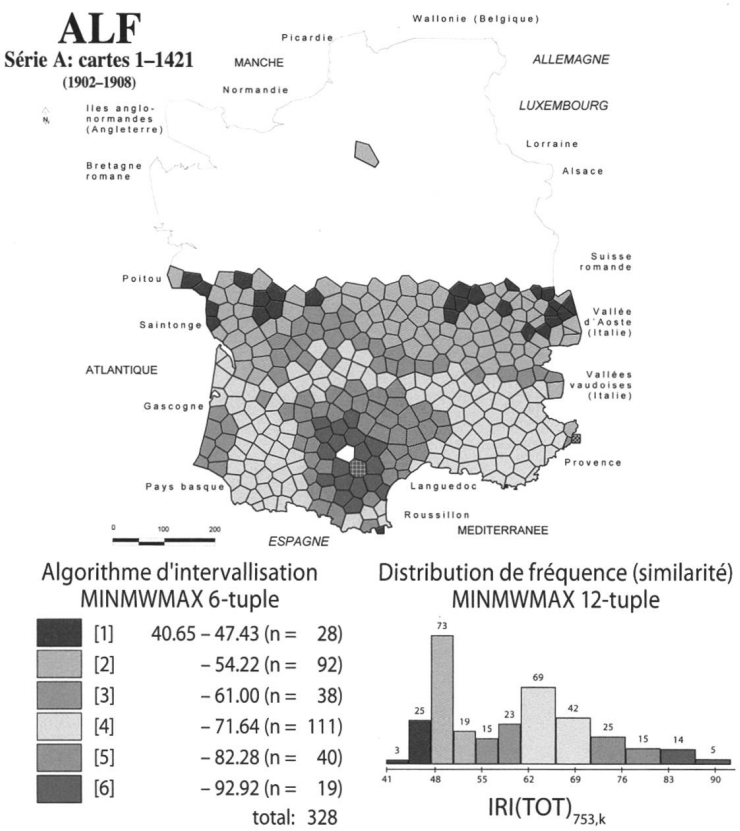
Carte 10: Profil choroplèthe d'une carte de similarité (selon IRI_{jk}) relative au point-ALF 944 (Thônes, Haute-Savoie)
Corpus: 675 cartes de travail (série B: toutes les catégories linguistiques)
Algorithme d'intervallisation: MINMWMAX 6-tuple



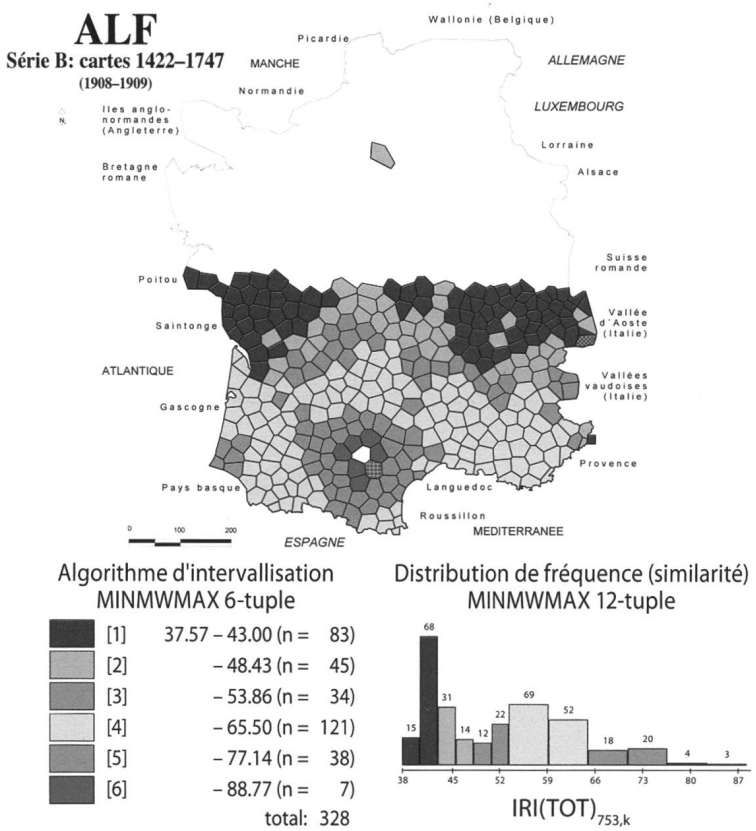
Carte 11: Profil choroplèthe d'une carte de similarité (selon IRI_{jk}) relative au point-ALF 893 (Hyères, Var)
Corpus: 1650 cartes de travail (série A: toutes les catégories linguistiques)
Algorithme d'intervallisation: MINMWMAX 6-tuple



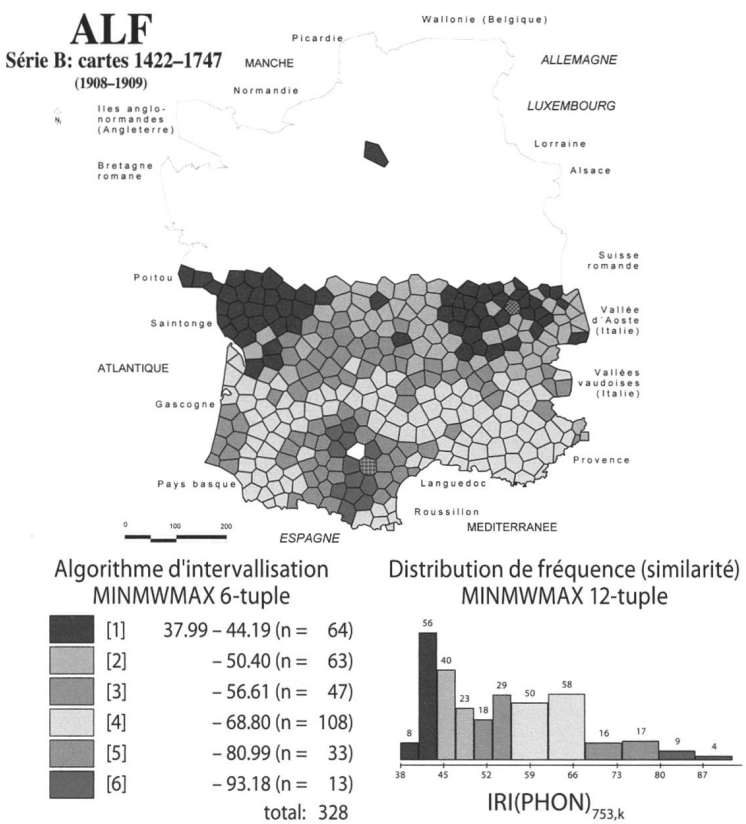
Carte 12: Profil choroplèthe d'une carte de similarité (selon IRI_{jk}) relative au point-ALF 893 (Hyères, Var)
Corpus: 675 cartes de travail (série B: toutes les catégories linguistiques)
Algorithme d'intervallisation: MINMWMAX 6-tuple



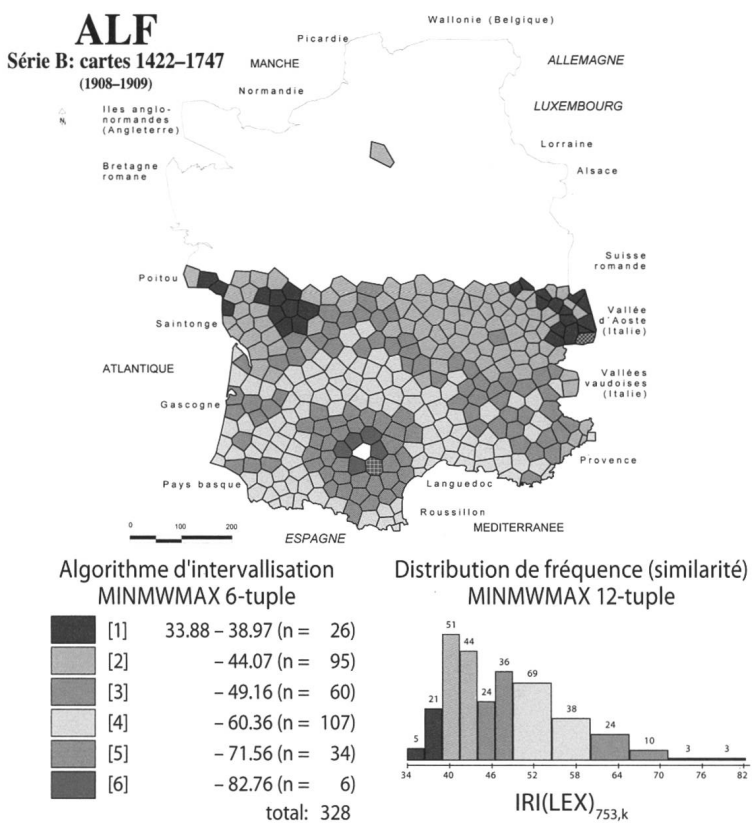
Carte 13: Profil choroplèthe d'une carte de similarité (selon IRI_{jk}) relative au point-ALF 753 (Brousse [Lautrec], Tarn)
Corpus: 1650 cartes de travail (série A: toutes les catégories linguistiques)
Algorithme d'intervallisation: MINMWMAX 6-tuple



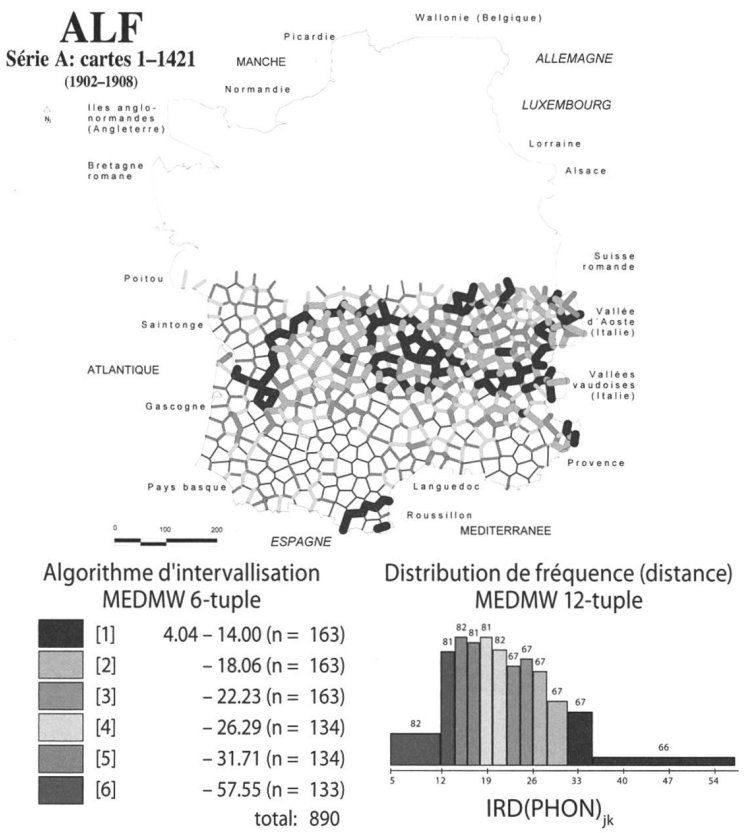
Carte 14: Profil choroplèthe d'une carte de similarité (selon IRI_{jk}) relative au point-ALF 753 (Brousse [Lautrec], Tarn)
Corpus: 675 cartes de travail (série B: toutes les catégories linguistiques)
Algorithme d'intervallisation: MINMWMAX 6-tuple



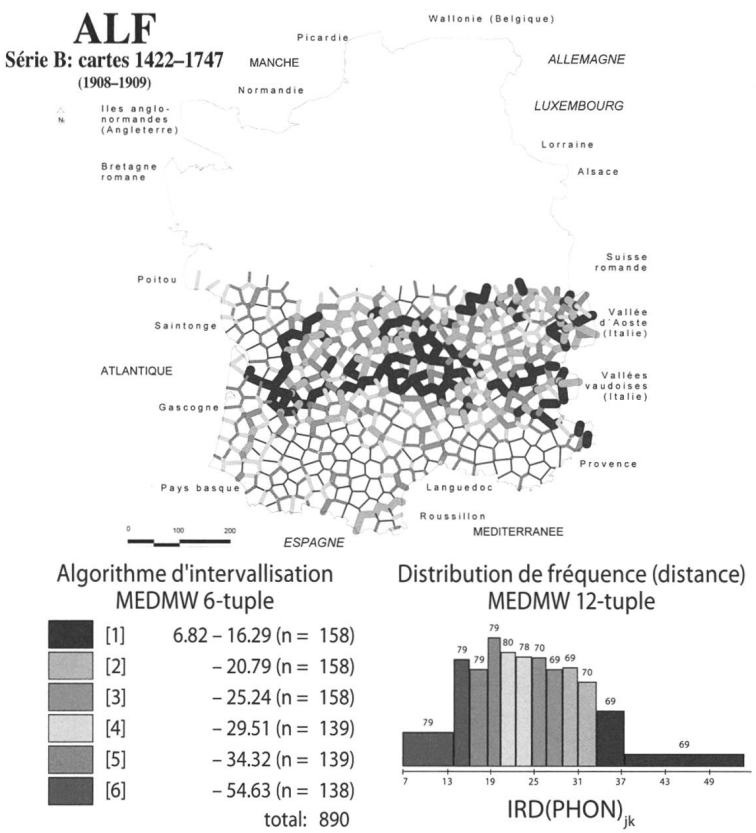
Carte 15: Profil choroplèthe d'une carte de similarité (selon IRI_{jk}) relative au point-ALF 753 (Brousse [Lautrec], Tarn)
Corpus: 313 cartes de travail (série B: phonétique)
Algorithme d'intervallisation: MINMWMAX 6-tuple



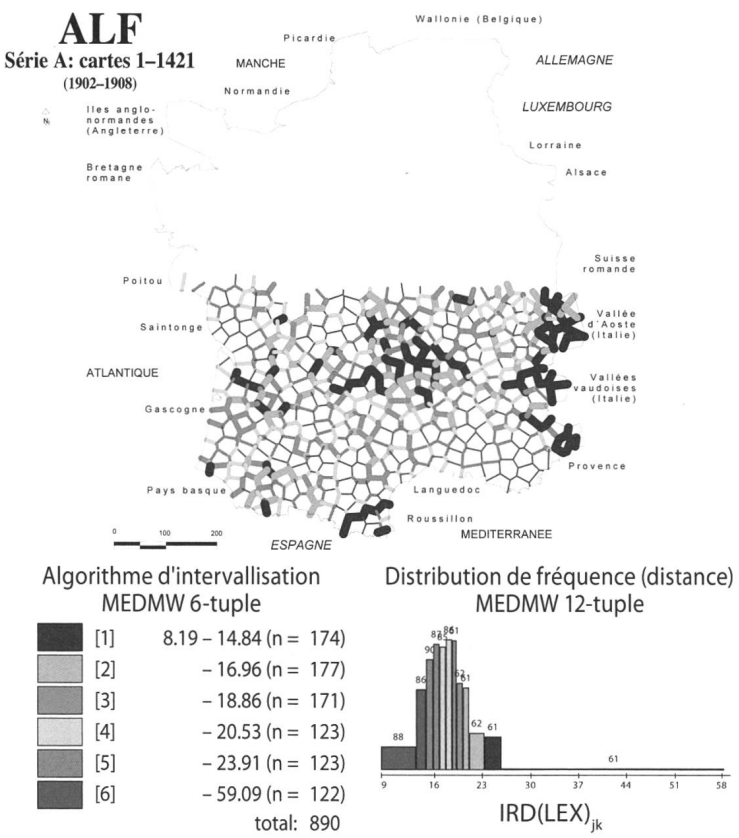
Carte 16: Profil choroplèthe d'une carte de similarité (selon IRI_{jk}) relative au point-ALF 753 (Brousse [Lautrec], Tarn)
Corpus: 300 cartes de travail (série B: lexique)
Algorithme d'intervallisation: MINMWMAX 6-tuple



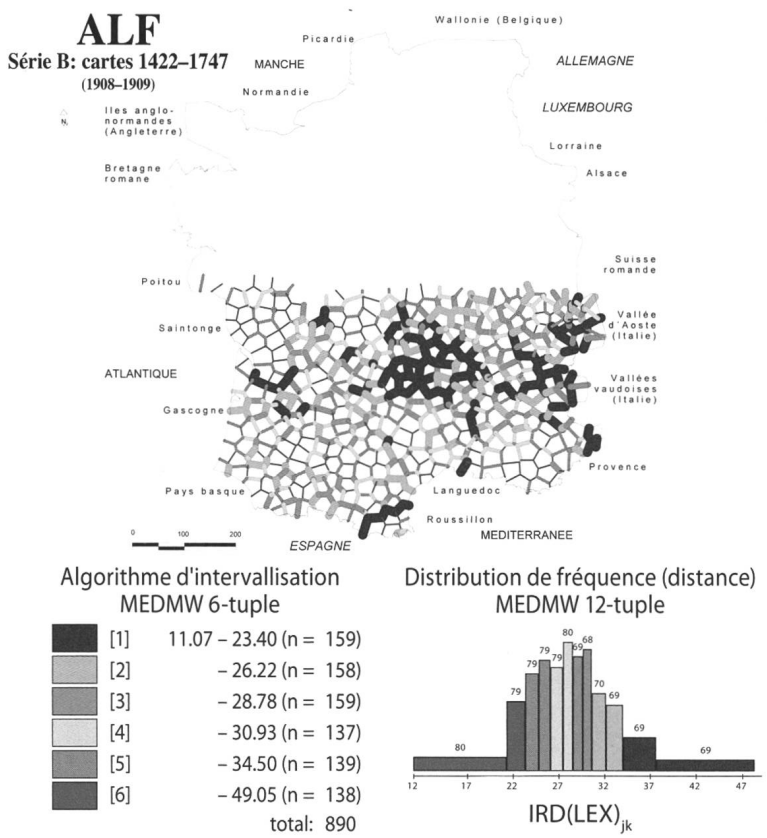
Carte 17: Synthèse isoglottique („carte à cloisons“)
Corpus: 1096 cartes de travail (série A: phonétique)
Indice de distance: IRD_{jk}
Algorithme d'intervallisation: MEDMW 6-tuple



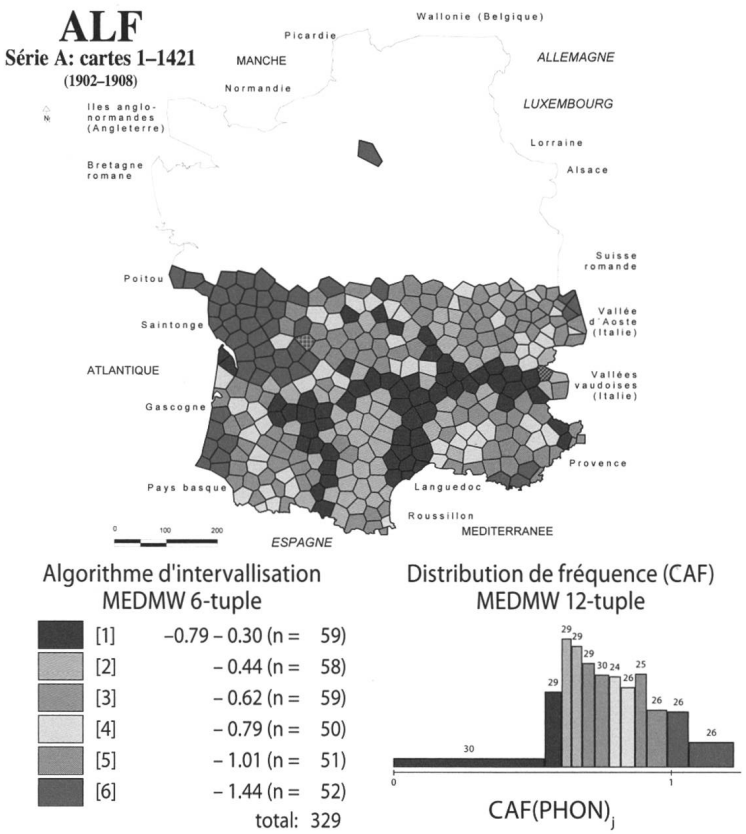
Carte 18: Synthèse isoglottique („carte à cloisons“)
Corpus: 313 cartes de travail (série B: phonétique)
Indice de distance: IRD_{jk}
Algorithme d'intervallisation: MEDMW 6-tuple



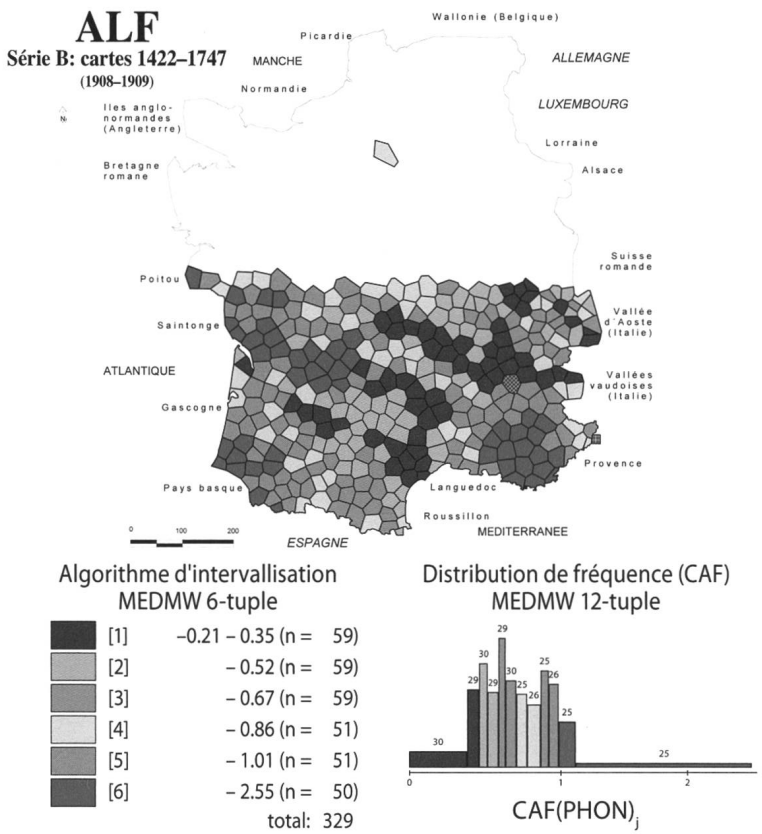
Carte 19: Synthèse isoglottique („carte à cloisons")
Corpus: 465 cartes de travail (série A: lexique)
Indice de distance: IRD_{jk}
Algorithme d'intervallisation: MEDMW 6-tuple



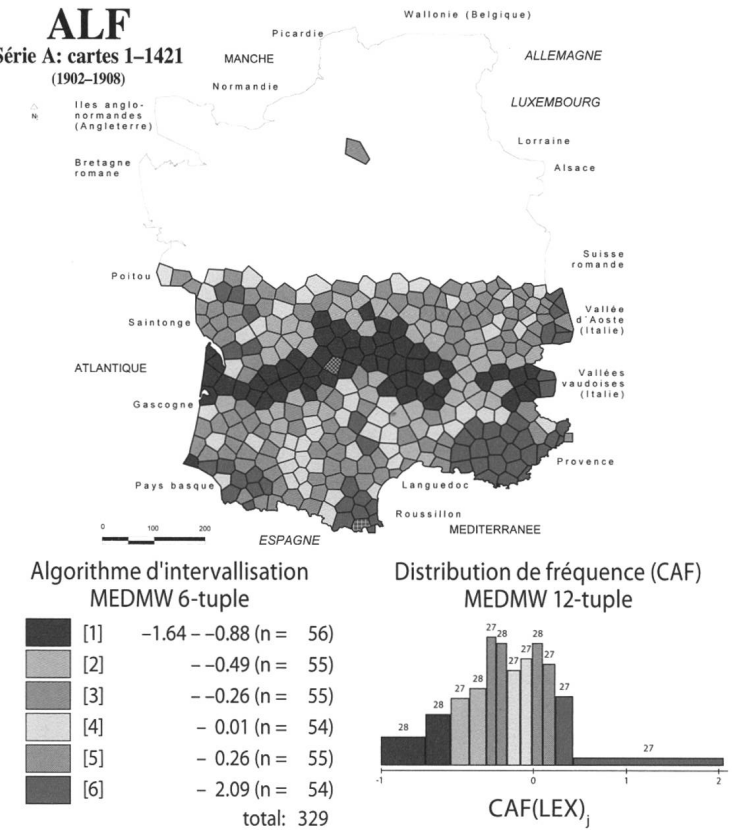
Carte 20: Synthèse isoglottique („carte à cloisons")
Corpus: 300 cartes de travail (série B: lexique)
Indice de distance: IRD_{jk}
Algorithme d'intervallisation: MEDMW 6-tuple



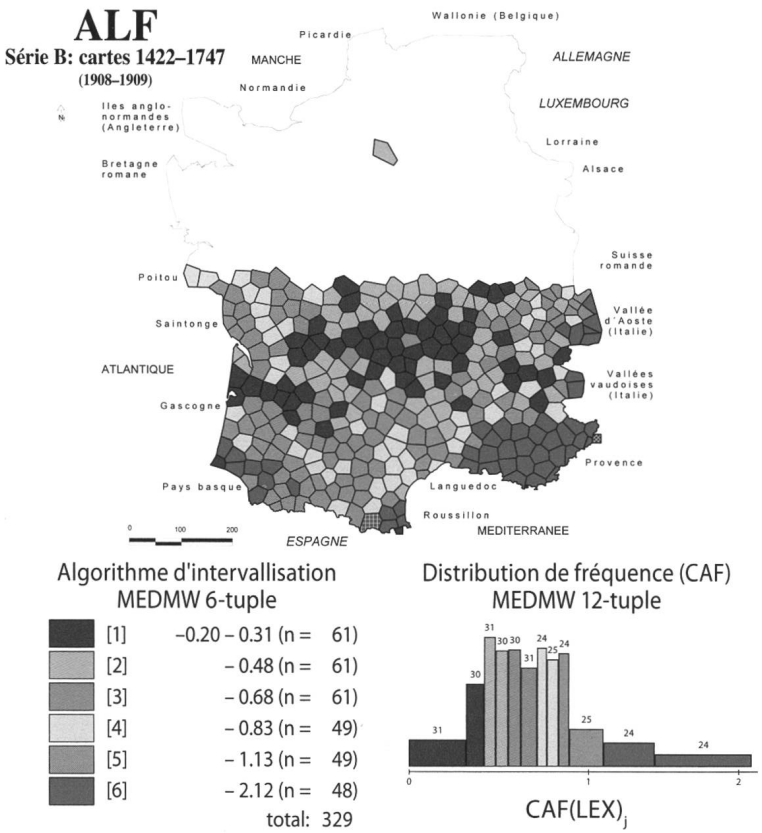
Carte 21: Profil choroplèthe de la synopse de 329 scores-CAF („coefficient d'asymétrie de Fisher")
Corpus: 1096 cartes de travail (série A: phonétique)
Indice de similarité: IRI_{jk}
Algorithme d'intervallisation: MEDMW 6-tuple



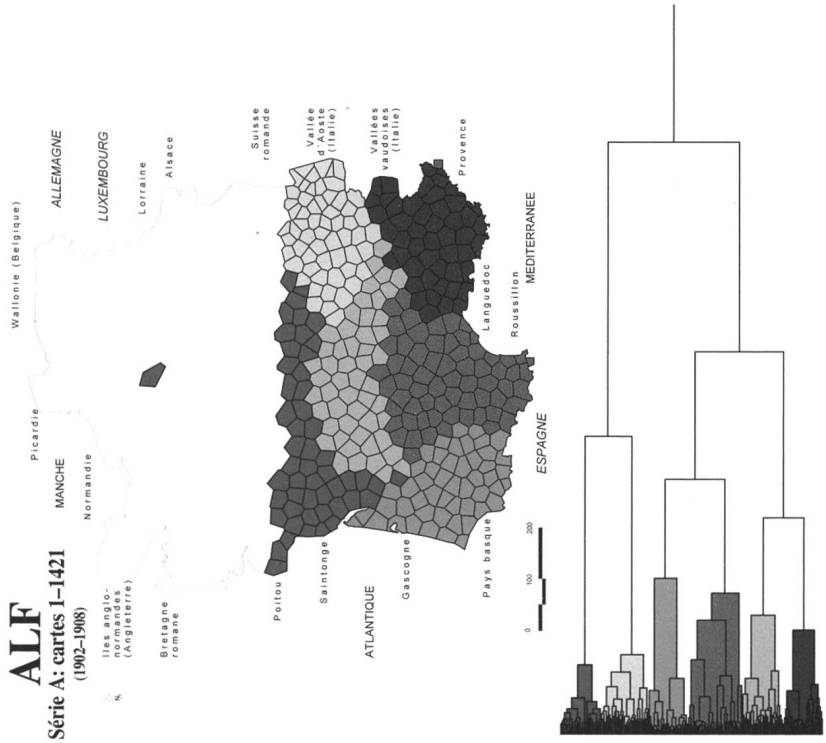
Carte 22: Profil choroplèthe de la synopse de 329 scores-CAF („coefficient d'asymétrie de Fisher")
Corpus: 313 cartes de travail (série B: phonétique)
Indice de similarité: IRI_{jk}
Algorithme d'intervallisation: MEDMW 6-tuple



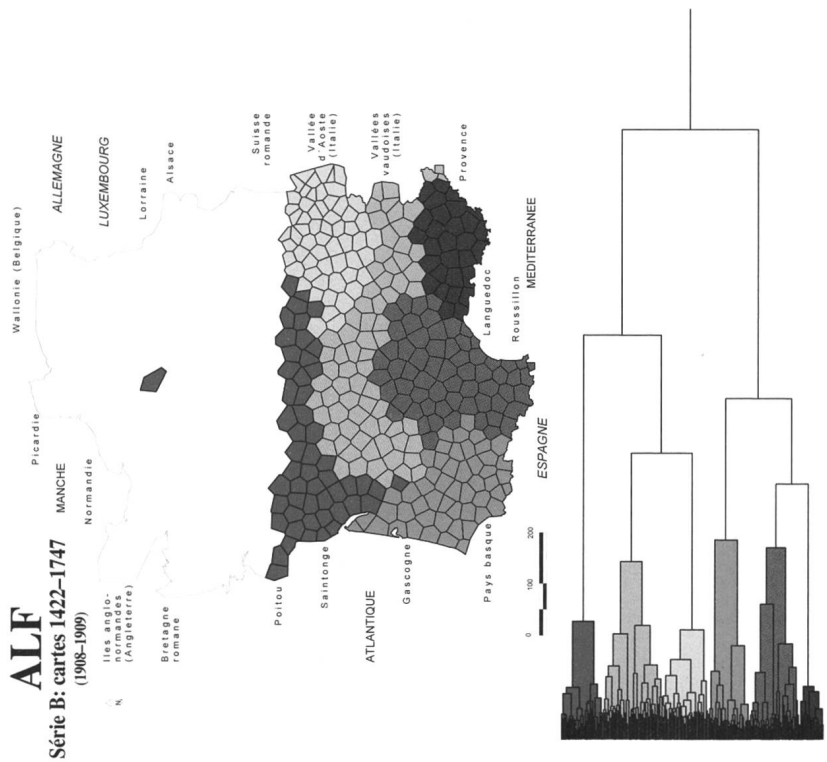
Carte 23: Profil choroplèthe de la synopse de 329 scores-CAF („coefficient d'asymétrie de Fisher")
Corpus: 465 cartes de travail (série A: lexique)
Indice de similarité: IRI_{jk}
Algorithme d'intervallisation: MEDMW 6-tuple



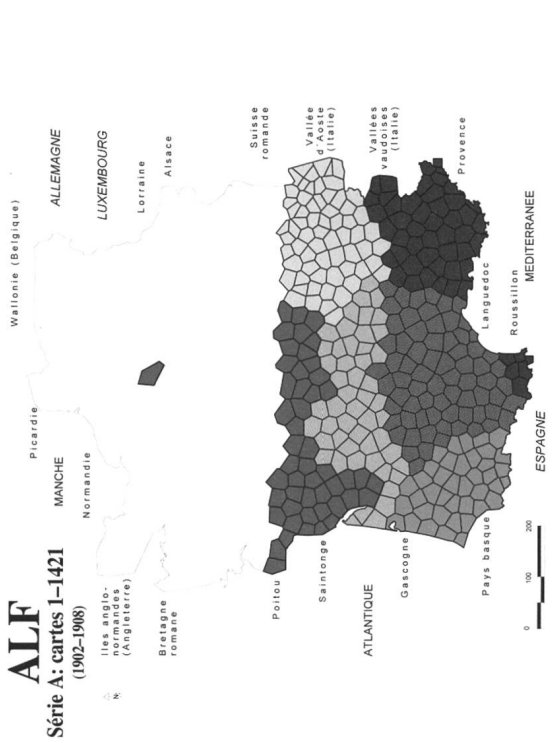
Carte 24: Profil choroplèthe de la synopse de 329 scores-CAF („coefficient d'asymétrie de Fisher")
Corpus: 300 cartes de travail (série B: lexique)
Indice de similarité: IRI_{jk}
Algorithme d'intervallisation: MEDMW 6-tuple



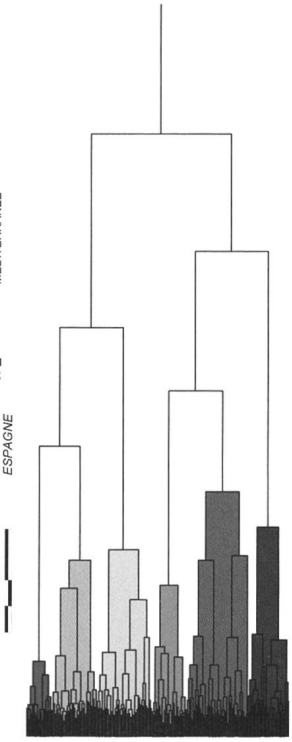
Carte 25: Classification ascendante hiérarchique (méthode de Joe Ward, Jr.)
Corpus: 1096 cartes de travail (série A: phonétique)
Indice de similarité: IRI_{jk}
Nombre des chorèmes (en haut) et dendrèmes (en bas) coloriés: 6



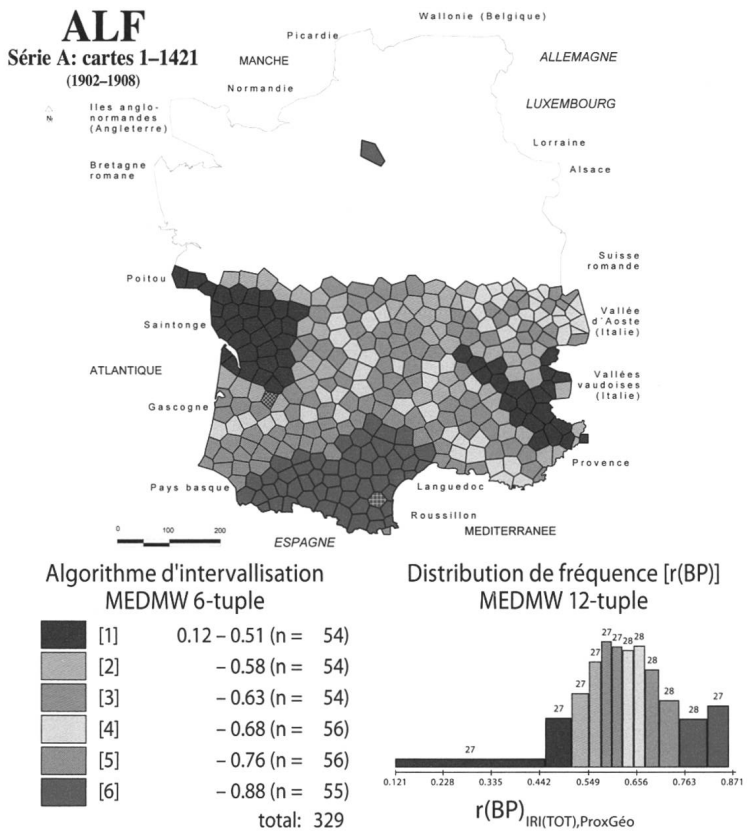
Carte 26: Classification ascendante hiérarchique (méthode de Joe Ward, Jr.)
Corpus: 313 cartes de travail (série B: phonétique)
Indice de similarité: IRI_{jk}
Nombre des chorèmes (en haut) et dendrèmes (en bas) coloriés: 6



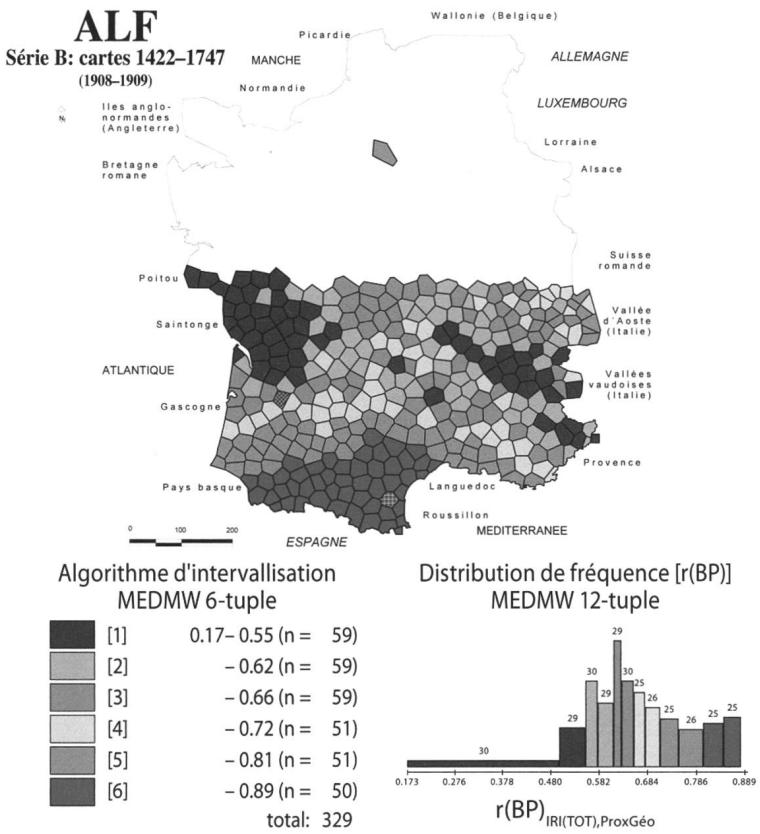
Carte 27: Classification ascendante hiérarchique (méthode de Joe Ward, Jr.)
Corpus: 465 cartes de travail (série A: lexique)
Indice de similarité: IRI_{jk}
Nombre des chorèmes (en haut) et dendrèmes (en bas) coloriés: 6



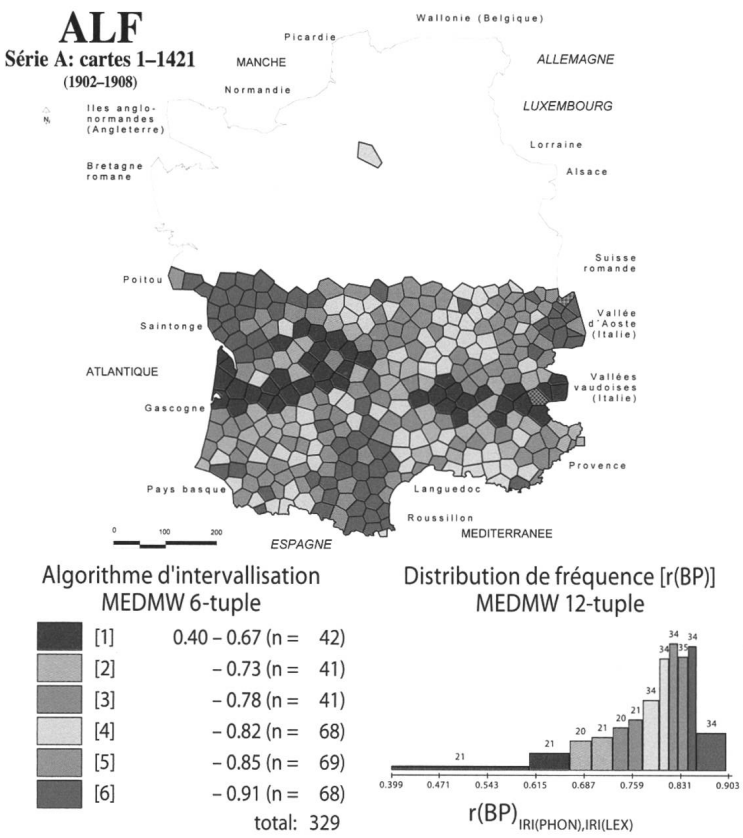
Carte 28: Classification ascendante hiérarchique (méthode de Joe Ward, Jr.)
Corpus: 300 cartes de travail (série B: lexique)
Indice de similarité: IRI_{jk}
Nombre des chorèmes (en haut) et dendrèmes (en bas) coloriés: 6



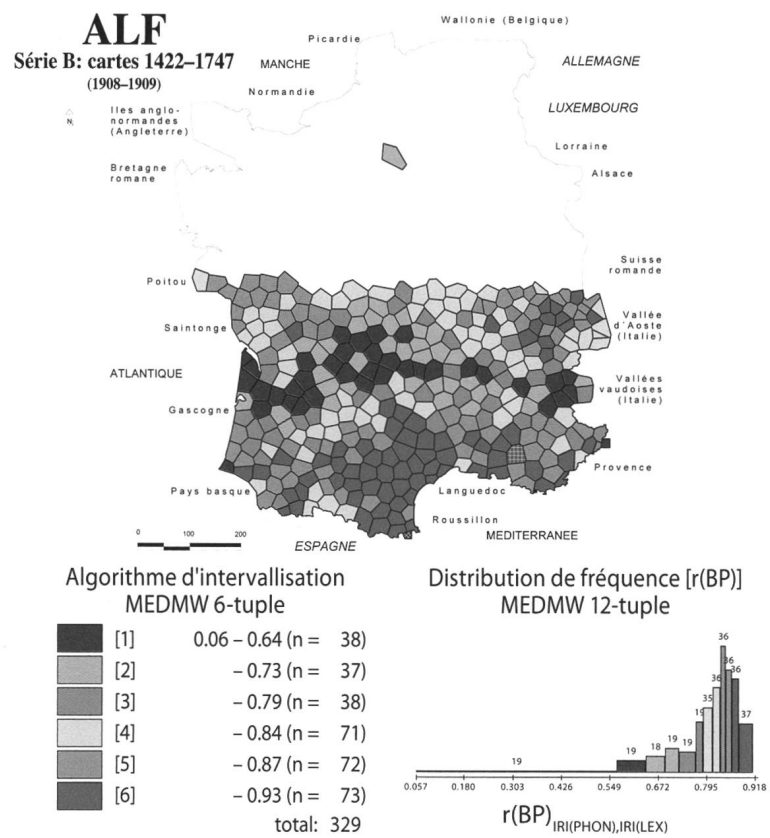
Carte 29: Profil choroplèthe d'une carte à corrélations
Matrices carrées corrélées:
matrice de similarité (selon IRI_{jk})
matrice de proximité géographique (selon le théorème de Pythagore)
Corpus linguistique: 1650 cartes de travail (série A: toutes les catégories linguistiques)
Algorithme d'intervallisation: MEDMW 6-tuple



Carte 30: Profil choroplèthe d'une carte à corrélations
Matrices carrées corrélées:
matrice de similarité (selon IRI_{jk})
matrice de proximité géographique (selon le théorème de Pythagore)
Corpus linguistique: 675 cartes de travail (série B: toutes les catégories linguistiques)
Algorithme d'intervallisation: MEDMW 6-tuple



Carte 31 : Profil choroplèthe d'une carte à corrélations
Matrices carrées de similarité corrélées (selon IRI_{jk}) :
matrice A : 1069 cartes de travail (série A : phonétique)
matrice B : 465 cartes de travail (série A : lexique)
Algorithme d'intervallisation : MEDMW 6-tuple



Carte 32 : Profil choroplèthe d'une carte à corrélations
Matrices carrées de similarité corrélées (selon IRI_{jk}) :
matrice A : 313 cartes de travail (série B : phonétique)
matrice B : 300 cartes de travail (série B : lexique)
Algorithme d'intervallisation : MEDMW 6-tuple