In Buchi, Éva/Chauveau, Jean-Paul/Greub, Yan/Pierrel, Jean-Marie (éd.), 2015 : *Actes du XXVIIe Congrès international de linguistique et de philologie romanes (Nancy, 15-20 juillet 2013). Allocutions d'ouverture, conférences plénières, tables rondes, conférences grand public.* Nancy : ATILF : http://www.atilf.fr/cilpr2013/actes/section-0.html

Du chemin parcouru entre Coquebert de Montbret père et fils et la plus récente dialectométrie : une reconsidération critique

1. De Coquebert de Montbret père et fils à Jules Gilliéron : deux points de repère de la genèse de l'analyse empirique de l'espace de la France

Depuis un certain temps je défends l'idée que l'ALF et l'œuvre entière de Jules Gilliéron (1854-1926) dépendent d'un substrat intellectuel « typiquement français » et qu'ils doivent donc être considérés dans une perspective historique particulière. La même remarque vaut pour la genèse de la dialectométrie qui, de par son souci enveloppant de généralisation spatiale, a navigué dans le même sillage[1].

Rappelons brièvement les jalons les plus importants de ces implications. L'ALF est caractérisé, entre autres, par trois propriétés qui ont grandement favorisé sa réussite fulgurante :

- un maillage d'enquête « économique » et « rationnel », basé sur le réseau départemental de la France et partageant avec lui, de ce fait, tous les avantages y ayant trait, comme l'équidistance entre les mailles de la grille et le nombre « opérationnel » (= limité) des points d'enquête[2].
- un questionnaire d'accessibilité « pan-galloromane », c'est-à-dire contenant des items universellement compréhensibles et partant – métrologiquement parlant – pourvus d'un haut degré de reliabilité et de validité.
- les consignes très strictes données à l'enquêteur Edmond Edmont qui consistaient avant tout à n'enregistrer que la première réponse des sujets interviewés et d'éviter toute sorte de mensuration.

Avec la curiosité de l'historien, l'on peut se demander deux choses :

Tout ceci a-t-il été inventé de toutes pièces par Gilliéron lui-même,

ou l'a-t-il conçu sous le coup d'une tradition spirituelle ou intellectuelle antérieure dans laquelle il avait baigné dès 1876, l'année de son arrivée en France?

---

1 Voir aussi nos contributions de 2002(b) et 2006.

2 Rappelons que la limitation programmatique des travaux de sondage à quelque 600 points d'enquête par rapport au total des 36 000 communes de la France équivaut, en dernière analyse, au tirage d'un échantillon aléatoire et représentatif.

Une autre certitude est son indépendance intellectuelle complète de l'entreprise géolinguistique allemande « Deutscher Sprachatlas » de Georg Wenker (1852-1911), pourtant antérieure à la sienne de quelque 20 ans[3].

Mais d'où avait-il pris l'idée heureuse de construire le réseau d'enquête de l'ALF suivant le compartimentage départemental de la France? Toujours est-il que cette décision s'est révélée, après coup, comme étant vraiment *heureuse*, vu son excellente maniabilité, qui se détache considérablement de la contrepartie allemande lourde – et non pas riche, en fin de compte – de plus 50 000 points d'enquête.

Or, je pense que les idées linguistiques de Gilliéron se nourrissaient – au cours de la période de gestation de l'ALF – d'un substrat intellectuel préexistant dont on parle peu en matière de linguistique ou de géolinguistique.

L'esprit de la départementalisation de la France – qui, elle, était une des premières mesures prises par la Révolution française, devenue par la suite très efficace – a favorisé, entre autres, une nouvelle considération de l'espace de la Nation, tout en la définissant non pas comme un continuum amorphe, mais bien plutôt comme la somme d'un nombre relativement petit de vecteurs, entre eux égaux et équidistants[4]. En bref, c'était un acte de mensuration visant à une discrétisation rationnelle de l'espace de la France, lourd de conséquences. Précisons que les préambules intellectuels de cette réorganisation de l'espace de la France se situaient déjà aux XVIIe et XVIIIe siècles.

Mais, ce qui compte le plus en l'affaire, c'est la dissémination à tous azimuts de l'usage de la nouvelle grille départementale de la France pour toutes sortes de relevés et d'études empiriques qui ont été réalisés au cours du XIXe siècle. Il en est né ce qu'on appelle la « Statistique Départementale ».

Dans ce contexte, marqué par une curiosité empirique toujours portant sur l'ensemble de la France, brillent – en ce qui concerne nos propos de linguistes – les figures et l'œuvre de Coquebert de Montbret père[5] et fils[6], devenus les auteurs, entre autres, de la première enquête linguistique du territoire entier de la France (et aussi bien au-delà), réalisée entre 1806 et 1812[7].

Une autre des émanations les plus connues de cette nouvelle veine d'investigation est la « Carte figurative de l'instruction populaire de la France », dressée en 1827 par le mathématicien Charles Dupin (1784-1873) : voir la figure 1[8].

---

3 Cf. Herrgen (2001).

4 Cf. Ozouf-Marignier (1989).

5 Il s'agit de Charles-Etienne Coquebert de Montbret (1755-1831).

6 Il est question d'Eugène-Barthélémy Coquebert de Montbret (1785-1847), un des cinq enfants de Charles-Etienne.

7 Au sujet de Coquebert de Montbret père et fils voir surtout la somme magistrale d'Isabelle Laboulais-Lesage de 1999, la thèse récente de Sven Ködel soutenue en 2013 aux Universités de Bamberg et Paris 7 (Diderot) et les indications fournies par S. Pop (1950, 19-26).

8 Pour la personne et l'œuvre de Ch. Dupin, cf. Palsky (1996, 62-67).

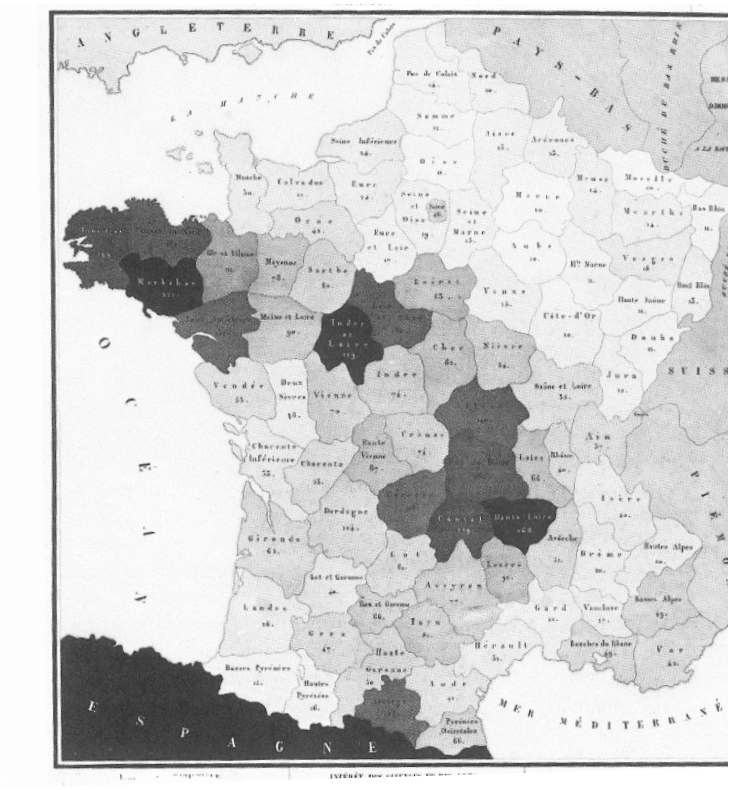


Figure 1 : Charles Dupin (1784-1873) : Carte figurative de l'instruction populaire de la France en 1827.

Cette carte présente déjà quelques particularités empiriques et techniques qui caractérisent également la production cartographique de la dialectométrie. Elle repose sur des données judicieusement collectées, visualise un sujet thématique bien défini (= le degré d'alphabétisation) et a été dressée moyennant la technique visuelle de la carte en plages (= carte choroplèthe). Son message cartographique repose sur une dizaine de teintes grises, réparties entre le noir complet et le blanc, dont le principe sémiotique est simple : plus c'est blanc, meilleure est l'alphabétisation des habitants des différents départements de la France.

Une considération sommaire du profil choroplèthe de la carte fait ressortir une ligne séparatrice située approximativement entre St-Malo et Genève. En 1827 et après, le message de cette carte a provoqué d'énormes conflits, voire des scandales, sur la scène intellectuelle, culturelle et politique de la France, et ceci à cause de l'inégalité évidente entre le sud et le nord de la Nation[9].

[9] Cf. à ce sujet la contribution de Roger Chartier (1992).

La « Statistique Départementale » reposait sur des piliers méthodiques très solides : il y avait déjà une théorie de départ – une sorte d'idée d'une « Belle au bois dormant », dont il fallait trouver la cachette ou les traces –, des techniques bien rodées pour la collecte des données et aussi une maîtrise cartographique toujours en essor pour mettre au point la meilleure des visualisations des recherches entreprises[10].

Évidemment, la galerie de telles cartes pourrait être facilement élargie. La production française respective en est très riche, pour ne pas dire exubérante[11].

À partir de 1860, il y avait même une cartographie statistique institutionnalisée, et pratiquée sous les auspices du Ministère des Travaux Publics, dont l'émanation la plus connue était l' « Album des statistiques graphiques[12] ». C'était une collection de cartes thématiques à couleurs, toutes en format in folio, qui avait le privilège d'être omniprésente dans tous les organismes et bureaux de l'État. En bref, la vie intellectuelle et politique de la France était imprégnée – au moment de la gestation de l'ALF – de cartes, d'idées cartographiques et de tous les prérequis méthodiques, théoriques et pratiques y ayant trait.

Et Jules Gilliéron : combien de tout ceci connaissait-il? Lui, qui pourtant était le fils d'un géologue? Evidemment, il est possible qu'il n'en sût rien, mais est-ce réellement vraisemblable?

Ma conviction est que l'entreprise de l'ALF représente, en dernière analyse, une des dernières émanations de la « Statistique départementale » telle qu'elle a été pratiquée tout au long du XIXe siècle – depuis la Révolution française et le Premier Empire – pour beaucoup d'autres variables, qu'elles soient d'origine humaine ou non.

Évidemment, depuis la période de l'ALF les méthodes de la géographie linguistique ont changé. Le grand dilemme devant lequel se trouvaient les auteurs d'atlas linguistiques du XXe siècle était la nature foncièrement bipolaire du travail d'atlas en tant que tel : ou bien collecter les données dialectales à l'aune de *l'espace* tout en mettant l'accent – dans le binôme *géographie linguistique* – sur le mot *géographie*, ou bien collecter les données à l'aune de la *faculté langagière* des locuteurs (dont on connaît le caractère hétéroclite) tout en mettant l'accent sur l'épithète *linguistique*.

Pour Gilliéron, la collecte de données exclusivement *basilectales* – issues d'un acte conscient de traduction de la part de témoins bilingues – était un prérequis de fer. Cette procédure lui semblait être la garante de la bonne *intercomparabilité* des données collectées. Cette idée – absolument normale au sein de toutes les sciences empiriques telles que la sociologie ou la psychologie – fut encore partagée et défendue par Albert Dauzat (1877-1955) au moment de la planification du « Nouvel Atlas linguistique de la France » (NALF)[13]. Or, nous savons bien que les premières réalisations

---

10 Voir à ce sujet l'excellente analyse de Marie-Noëlle Bourguet de 1988.

11 Pour une description très détaillée de ce contexte, je renvoie expressément à l'excellent livre du géographe français Gilles Palsky de 1996.

12 Cf. Palsky (1996, 141 sq.).

13 Pour le programme du NALF cf. Tuaillon (1983).

concrètes du NALF – toutes conçues et réalisées par d'éminents dialectologues de renom tels que Pierre Gardette (1906-1973), Jean Séguy (1914-1973) ou Pierre Nauton (1917-1970) – n'ont pas été faites sous le signe de l'intercomparabilité des données recueillies : ni entre elles, ni avec celles de l'ALF ou d'un atlas voisin, mais bien plutôt sous le signe de *l'authenticité*, tant des enquêtes – faites souvent en *patois* et partant sans le biais de la traduction bilingue consciente – que des réponses, recueillies non pas en forme standardisée, mais au cours d'une *conversation dirigée*. D'où l'inévitable augmentation, chaleureusement saluée par les dialectologues, de réponses doubles, triples et même au-delà, toutes hautement idiosyncrasiques, mais dépourvues de toute intercomparabilité.

Au début, la géographie linguistique romane n'a nullement été entravée par cette bipolarité. A preuve de cela, je cite la double dédicace, formulée par Walther von Wartburg (1888-1971) au début du premier volume du FEW, à l'intention de Wilhelm Meyer-Lübke (1861-1936), le traditionnaliste néogrammairien, et à Jules Gilliéron (1854-1926), l'innovateur empirique et théorique.

À la différence de Pierre Gardette, Jean Séguy, en tant qu'auteur de l' « Atlas linguistique de la Gascogne » (ALG), voulut longtemps respecter les principes prônés par Albert Dauzat et produire, de ce fait, « de l'intercomparable ». Il a fini par succomber d'abord à la tentation de *l'authenticisme* empirique, pour se convertir ensuite – surtout après l'expérience faite avec le quatrième volume de l'ALG – en *dialectométricien*.

Par le biais de l'analyse *quantitative* des données *qualitatives* débordantes contenues dans son ALG, il voulait en extraire une sorte de sève vitale cachée, comparable à la « Belle au bois dormant », dont on devine l'existence, mais ignore le domicile.

Aux dires de Xavier Ravier, son fidèle compagnon de travail, Jean Séguy voulait faire tenir (ou réduire) le gascon – entendu comme une masse de « 47 millions de chiures de mouche » – « dans une formule ou un schéma[14] ».

Ce qu'il est devenu de la *dialectométrie* depuis 1973, année du lancement du mot et aussi de la chose, j'en parlerai tout de suite. D'entrée de jeu, j'aimerais bien souligner encore une chose. Les études de dialectométrie, dont il existe de nos jours beaucoup de variantes, ont démontré que sous les enchevêtrements souvent inextricables ou carrément chaotiques d'un nombre inouï d'aires linguistiques appartenant à toutes les catégories linguistiques imaginables, il se cache d'étonnantes régularités spatiales qui peuvent être considérées comme autant d'émanations de lois géolinguistiques très claires et simples qui, elles, constituent la contrepartie des « Lautgesetze » (lois linguistiques) d'ordre diachronique découvertes par les néogrammairiens de Leipzig dans le dernier quart du XIXe siècle.

[14] Cf. Ravier (1976, 390).

2. La dialectométrie : une méthode inductive et exploratoire

Au début des années 70 du XXe siècle, la dialectométrie (DM) est née sous l'impulsion de la curiosité de quelques géolinguistes[15] qui voulaient savoir si, derrière (ou : sous) l'apparent désordre des structures spatiales des cartes de n'importe quel atlas linguistique (AL), il existe un ordre enveloppant et, partant, ou bien « majeur » ou bien « caché ». Ce but ne pouvait être rejoint que par la synthèse, faite à la main[16] ou à l'aide de l'informatique alors en état de gestation[17], de beaucoup de cartes d'atlas. Par une vieille tradition de la philosophie occidentale, une telle procédure – qui part du plan du « particulier » pour arriver à celui du « général » – est appelée « induction ». Comme elle se lance aussi à la recherche de choses jusqu'alors inconnues, elle a par définition aussi une vocation « exploratoire ».

Après plus de quatre décennies d'intenses recherches-DM appliquées à des atlas linguistiques romans[18], allemands[19] et anglais[20], il est clair que l'ordre sous-jacent (c'est-à-dire « caché » ou « de profondeur ») assumé d'entrée de jeu, existe réellement et qu'il est l'émanation (ou la création) d'une activité particulière des humains. Au sein de l' « École dialectométrique de Salzbourg » (EDMS), il est question, dans ce contexte, depuis belle lurette de la « gestion basilectale de l'espace par l'homo loquens », considérée comme une des multiples facultés innées des humains et à laquelle est attribué le rang d'un « universal » du langage *à plein titre*.

C'est ainsi qu'en face d'un atlas linguistique (AL) à dialectométriser, nous nous sommes toujours posé les questions suivantes :

- Quelle est la variation géolinguistique *globale* répertoriée sur toutes les planches de l'AL en question, en l'occurrence de l'ALF ?
- Comment se reflète, sur toutes les planches de l'ALF, la *gestion basilectale de la Galloromania* de la part des Français, des Suisses romands et des Wallons? Pour préciser : nos recherches-DM relatives à l'ALF englobaient, certes, le réseau entier de l'ALF, mais ont dû se limiter, pour des raisons d'ordre économique, à l'analyse de quelque 40% des 1421 planches originales de la série A de l'ALF[21]. Ce pourcentage est pourtant suffisamment grand pour pouvoir en tirer des conclusions représentatives et, de ce fait, fiables.

---

15 Je renvoie, à ce sujet, à deux articles de Jean Séguy (1971 et 1973 : avec la mention du néologisme *dialectométrie*) ainsi qu'à mes propres contributions de 1971 (avec N. Winterleitner) et 1976.

16 C'était le cas pour Jean Séguy, décédé déjà en 1973.

17 J'avais le privilège de pouvoir me servir, dès le début, de l'informatique.

18 L'essentiel de nos recherches-DM romanes porte sur les grands atlas nationaux ALF, AIS (Italoromania), ALPI (Ibéroromania) et ALDC (domaine catalan).

19 Cf Goebl/Scherrer/Smečka (2013).

20 Cf. Goebl/Schiltz (1997).

21 Pour plus de détails sur la dialectométrisation des données de l'ALF, cf. Goebl (2002a et 2003).

3. Préparation cartographique du réseau d'enquête de l'ALF

Rappelons d'abord trois caractéristiques élémentaires de l'ALF :

- Le réseau de l'ALF comprend 638 localités[22] où Edmond Edmont[23] a mené 639 enquêtes.
- Jules Gilliéron a inauguré, par les modalités typographiques de la publication des données de l'ALF, la tradition – devenue plus tard caractéristique de la géographie linguistique *romane* en général – de présenter, sur chaque planche d'atlas, une synopse de toutes les transcriptions recueillies sur le terrain, relatives à un item particulier du questionnaire utilisé[24]. C'est le principe du « full text map » qui s'oppose à celui du « symbol map » utilisé par les *germanistes* et les *anglicistes*.
- Dès le début, la présentation des données transcrites originales sur la superficie d'une planche d'atlas a rendu nécessaire – pour les besoins de leur dépouillement rationnel – le recours à un instrument heuristique spécial, à savoir aux « cartes muettes ».

Ces faits expliquent pourquoi la « géographie linguistique » créée par Jules Gilliéron ne pouvait se faire sans un recours continu et massif à la pensée et aux méthodes d'une discipline située en dehors de la linguistique, à savoir de la *cartographie*. Le maniement et le remplissage des *cartes muettes* demandaient en outre, de la part des adeptes de la nouvelle méthode, le déploiement de *réflexions géo-classificatoires* visant à subdiviser les données originales d'une planche d'atlas en groupement spatiaux utiles suivant la logique d'un principe classificateur préalablement choisi. Les annales des études romanes fournissent une quantité énorme d'applications très réussies de cette démarche géo-classificatoire[25].

De même qu'aux temps de la publication de l'ALF quand l'imprimeur Protat Frères (de Mâcon) a mis en vente des cartes muettes du réseau de l'ALF pour permettre aux lecteurs de ce dernier d'y reporter le rendement de leurs réflexions géo-classificatoires, il est également de nos jours indispensable de créer, pour le réseau d'un AL en train d'être dialectométrisé, un support cartographique capable de rece-

---

22 Le nombre réel des *points d'enquête* de l'ALF est 638 et non pas 639 comme il est indiqué très souvent par erreur. L'ALF comporte cependant, sur chacune des 1421 planches de la série A, les résultats transcrits de 639 *enquêtes* faites par Edmond Edmont. La solution de cette « énigme » est qu'Edmont avait réalisé deux enquêtes au point-ALF 284, Saint-Pol-sur-Ternoise, sa ville natale.

23 Rappelons qu'Edmond Edmont (1849-1926) était l'enquêteur de l'ALF.

24 Du temps de la publication des données de l'ALF, le principe combien heureux du « full text map » était bien loin d'être hors conteste. Un des adversaires les plus farouches de ce principe était Paul Meyer qui s'employait énergiquement pour une publication des données de l'ALF sous la forme de tableaux : cf. Storost (1994, 77). Quelle catastrophe pour la géographie linguistique si P. Meyer avait réussi à imposer ses visées saugrenues!!

25 Ceci concerne surtout les apports de l'onomasiologie romane qui, elle, se basait avant tout sur les atlas linguistiques.

voir les informations spatiales tant *qualitatives* que *quantitatives* utilisées au cours de l'analyse-DM. Un tel support présuppose l'établissement d'un compartimentage polygonal du réseau en question, dit aussi « tessélation de Voronoï ».

La polygonisation de Voronoï se fait à travers les étapes suivantes[26] :

- triangulation du réseau des points d'enquête
- construction successive des médiatrices sur les côtés de triangle du réseau
- De la fusion de toutes les médiatrices ainsi construites naît le réseau polygonal dont nous avons besoin.

4. La taxation des données d'un atlas linguistique

Voir les cartes 1 et 2.

Au sein de l'EDMS, l'on entend par la « taxation » (des planches originales d'un AL) l'analyse d'un secteur déterminé des informations originales contenues dans une carte originale de l'AL examiné, et l'établissement successif d'une « image » des données originales respectives sous la forme d'une « carte de travail » (CT). Précisons que ce genre de travail classificatoire s'aligne parfaitement sur une tradition déjà ancienne en matière de linguistique romane qui avait été inaugurée par Jules Gilliéron lui-même[27].

Les CT issues d'une campagne de taxation disposent de structurations internes très variables qui, elles, peuvent osciller entre 2 et *n* « aires taxatoires ». L'élément linguistique de base qui régit chaque aire taxatoire, est appelée « taxat ». Au cours de la dialectométrisation de plus de 600 des 1421 planches originales de l'ALF entreprise peu avant le tournant du millénaire à Salzbourg[28], nous en avons extrait 1681 CT appartenant aux catégories de la *phonétique*, du *lexique* et de la *morpho-syntaxe*, dont la figure 2 montre la distribution quantitative.

---

26 Pour plus amples détails cf. Goebl (1981, 363 sq.) et Goebl (1984 I, 90 sq.).

27 Cf. Gilliéron (1918) : à la fin de cet ouvrage se trouve une typisation lexicale relative à la carte 1 (abeille) de l'ALF. Voir également Jaberg (1908), où l'on trouve toute une gamme de typisations coloriées de différentes planches de l'ALF.

28 Voir nos contributions de 2000, 2002a et 2003 qui en présentent les principes et quelques résultats.

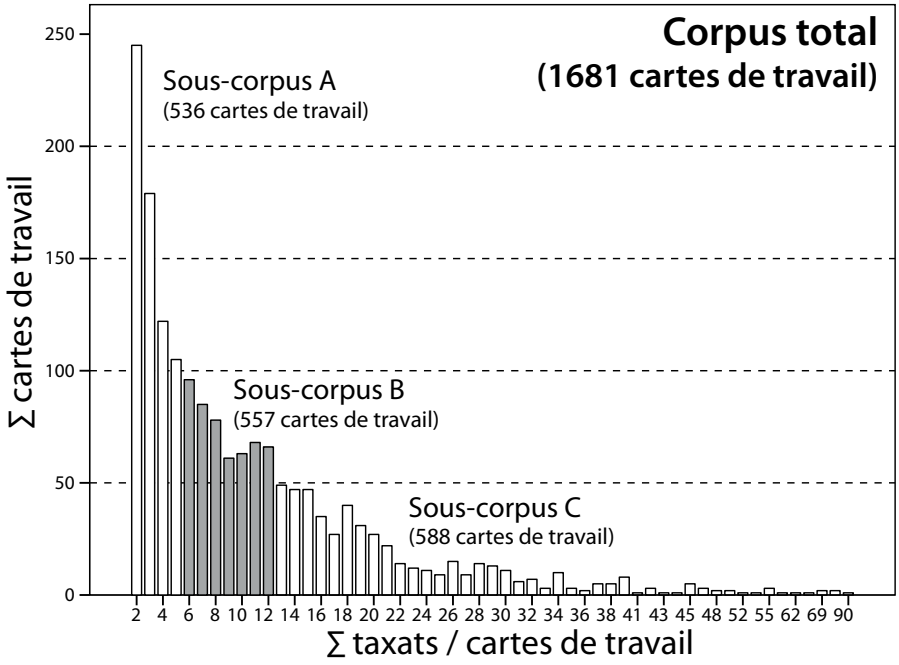


Figure 2 : Histogramme – pourvu de 55 colonne

s – montrant, pour notre corpus-DM tiré 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)

Les colonnes de l'histogramme en question représentent une courbe décroissante de gauche à droite[29]. Il s'agit de 55 colonnes de hauteurs inégales qui couvrent un espace « polynymique » allant de 2 taxats/CT jusqu'à 90 taxats/CT : voir l'abscisse sur la figure 2.

Les hauteurs de 55 colonnes signalent les nombres respectifs des CT analysées. Il y a donc 245 CT « *bi*-nymes » (= avec *deux* types linguistiques ou taxats), 179 CT « *tri*-nymes » (= avec *trois* taxats), 122 CT « *tétra*-nymes » (= avec quatre taxats) et seulement une CT de nature *90*-nyme. Elle se réfère d'ailleurs à la carte 545 de l'ALF, relative aux dénominations de la *fauvette*.

[29] Il s'agit d'une courbe exponentielle. Nous avons d'ailleurs découvert les mêmes régularités dans toutes nos analyses dialectométriques qui portent sur tous les grands atlas romans ainsi que sur ceux des études anglaises. Le caractère exponentiel de ces courbes a été utilisé, en 1985, par le linguiste quantitatif allemand Gabriel Altmann pour la définition de ce qu'il a appelé alors la « loi de Goebl ».

Le bilan taxatoire de notre analyse-DM de l'ALF se présente donc comme suit :

corpus	∑ cartes de travail (CT)	∑ aires taxatoires / taxats
total	1.681	18.047
phonétique	1.117	10.642
lexique	465	6.514
morpho-syntaxe	99	891

Dans l'ensemble, la richesse structurale des quelque 18 mille aires taxatoires analysées est très grande. N'oublions pas que la structure spatiale de chaque carte de travail (CT) peut varier dans quatre directions :

1. du nombre des taxats
2. de la taille des aires taxatoires
3. de la configuration des aires taxatoires et
4. de la localisation des aires taxatoires à l'intérieur du réseau d'atlas en question.

Il en résulte une variabilité vraiment déroutante. Une des conséquences de cette grande variabilité est le fait que les différentes aires taxatoires ne se recouvrent presque jamais. Il en résulte un effet d'*enchevêtrement* très complexe entre les aires taxatoires qui semble être typique de beaucoup de phénomènes spatialement répartis.

La carte 1 montre une des 1117 CT *phonétiques* analysées. Avec ses 17 taxats *phonétiques* – qui symbolisent autant de succédanés du nexus vocalique latin -ātu (dans merc-ātu « marché ») – elle représente une CT déjà relativement complexe.

La carte 2, en revanche, appartient à notre corpus *lexical*, riche de seulement 465 CT : elle est franchement moins complexe[30].

Une des premières expériences que nous ayons faites au cours de nos travaux-DM concernait l'impact du nombre absolu des CT synthétisées par voie-DM. Nous avons constaté alors que les structures-DM de profondeur obtenues à l'aide de la synthèse de 1681 CT (= corpus total) ressemblent de très près à celles obtenues avec 50%, 25% et 15% de cet effectif, à condition que l'extraction de l'urne pleine en ait été faite par voie aléatoire[31]. La même expérience a pu être réitérée avec d'autres atlas linguistiques romans et non romans.

De cette réalité, étonnante au premier coup d'œil, nous avons conclu que les structures-DM de profondeur mises au jour, sont imprégnées d'une grande *redondance* interne qui, elle, serait comparable à celle dont disposent tous les systèmes ou signaux communicatifs, qu'ils soient de nature linguistique ou non.

---

30 La distribution spatiale et l'origine des noms de jours de la semaine de la Galloromania ont été analysées à plusieurs reprises : cf. Gilliéron/Roques (1912, 85-108), Rohlfs (1949), Wartburg (1949) et Baehr (1952).

31 Cf. Goebl (1984 I, 206-219).

Une autre réflexion concernait les conditions de la genèse de cette redondance dont l'origine était indubitablement humaine. Depuis un certain temps, nous sommes convaincus que toutes ces régularités quantitatives sont l'émanation directe d'un comportement particulier des dialectophones par rapport aux défis communicatifs de l'espace qu'ils habitent. Nous lui avons conféré l'appellation de « gestion (ou : aménagement) basilectale de l'espace par l'homo loquens »[32].

5. Brève présentation de la « chaîne dialectométrique » de l'EDMS

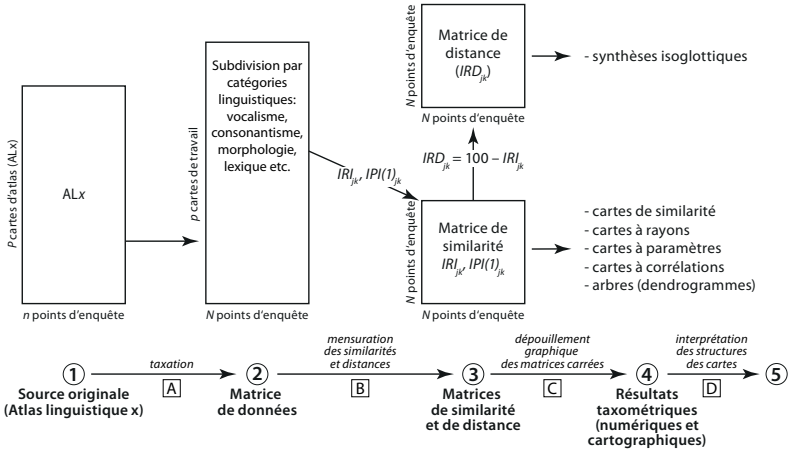


Figure 3 : Diagramme de la chaîne méthodique utilisée par l' « Ecole Dialectométrique de Salzbourg »

Les méthodes de l'EDMS sont alignées en forme de « chaîne » : voir la figure 3. Le tableau suivant renseigne sur les fonctions des différents « chaînons » ou étapes de cet alignement[33].

---

32 Comparer à cela le titre de notre contribution de 2005 et la définition de ce concept dans Goebl (2002a, 5).

33 Pour une introduction détaillée aux buts et méthodes de la DM de Salzbourg nous renvoyons à Goebl (1984), vol. I-III (en allemand) et à nos contributions de 2000, 2002a, 2003 et 2012 (toutes en français). Voir en outre les adresses-web suivantes :

- bibliographie complète de nos travaux-DM :https ://www.sbg.ac.at/rom/people/prof/goebl/dm_publi.htm
- trois sites-DM de travail et de présentation :

http ://www.dialectometry.com/
http ://www.dialectometry.com/dmdocs/index.html
http ://ald.sbg.ac.at/dm/

34 VDM : „Visual DialectoMetry", logiciel dialectométrique créé par Edgar Haimerl (Seattle, EUA) entre 1997 et 1999, mis à jour continuellement depuis cette date.

35 L'indice de similarité standard de la DM de Salzbourg est l' « Indice Relatif d'Identité » (IRI_{jk}) auquel fait écho l' « Indice Relatif de Distance » (IRD_{jk}) selon la formule IRIjk + IRD_{jk} = 100.

Étapes méthodiques	États des données en voie de traitement-DM	Commentaires
A	1-2	Choix d'un atlas linguistique [1] pour la dialectométrisation ; définition et application des principes taxatoires respectifs ; taxation [A] d'un nombre aussi grand que possible de cartes originales de l'atlas linguistique retenu ; saisie électronique successive des cartes de travail (CT) analysées et établissement de la matrice de données [2] (avec N points d'enquête et p cartes de travail) à l'aide du logiciel VDM[34]. Les données taxées et la matrice de données se situent au niveau de l'échelle métrologique nominale.
B	2-3	Choix et application d'une mesure *quantitative* pour déterminer les similarités (quantitatives) entre les N points d'enquête : ici : choix de l'IRI_{jk} (« Indice Relatif d'Identité »)[35] ; établissement consécutif de la matrice de similarité (carrée : N × N) [3, en bas]. Celle-ci contient la variation *qualitative* de la matrice de données après sa conversion dans une forme *quantitative*.
C	3-4	Transformation (partielle ou intégrale) du contenu *quantitatif* de la matrice de similarité en représentations et schémas visuels. Le choix des outils de visualisation dépend uniquement des besoins heuristiques du géolinguiste. En utilisant le logiciel VDM : choix de l'algorithme de visualisation approprié et du nombre des paliers chromatiques (intervalles).
D	4-5	Surtout par voie optique : interprétation et évaluation théoriques et pratiques [D] des structures spatiales des mises en cartes effectuées. Genèse de nouvelles représentations conceptuelles et théoriques [5] des données de départ [1].

Dans la panoplie salzbourgeoise des indices de similarité, il y a, entre autres, aussi l'indice IPIjk (« Indice Pondéré d'Identité ») qui dispose de certains avantages taxométriques et visualisateurs auxquels le dialectométricien peut avoir recours le cas échéant.

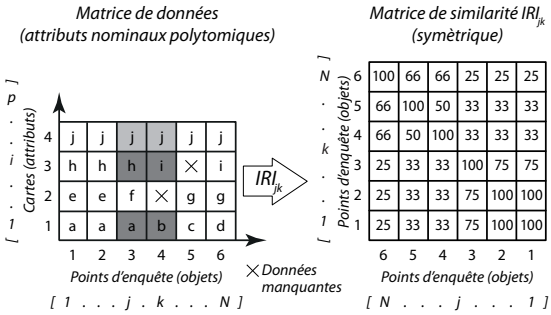
Précisons encore que dans cette démarche pluridisciplinaire, le savoir nécessaire pour l'établissement de la matrice de données et pour la définition de l'indice de similarité est à rechercher du côté de la « classification numérique » (dite aussi « automatique »)[36] alors que les méthodes pour la confection et la visualisation appropriée des cartes en couleurs sont fournies par la cartographie numérisée[37], appelée de nos jours aussi « visualistique » (en anglais : « visualistics »). Ces facilités n'existaient d'ailleurs pas encore en tant que telles au commencement de nos activités-DM.

Une des particularités de la procédure pratique de l'EDMS est de toujours privilégier la dimension heuristique traditionnelle de la géographie linguistique : à savoir le plan des cartes. Tous les résultats quantitatifs des calculs-DM effectués sont directement convertis en schémas visuels appropriés et offrant un accès aussi suggestif (et facile !!) que possible aux géolinguistes. Soulignons en outre que l'EDMS pratique également une sélection judicieuse des procédures statistiques utilisées tout en évitant l'application de celles dont la *complexité mathématique* n'est pas à la hauteur de leur *utilité géolinguistique*.

À partir de 1997, tous les calculs et toutes les visualisations-DM s'effectuent à l'aide d'un logiciel–DM appelé « Visual DialectoMetry » (VDM), qui avait été créé par notre ami Edgar Haimerl et qui, depuis cette date, a connu plusieurs remaniements et élargissements[38].

6. La mesure de similarité entre deux vecteurs de la matrice des données

Voir la figure 4.



$$\mathrm{IRI}_{3,4} = 100 * \frac{\Sigma(\mathrm{COI})}{\Sigma(\mathrm{COI} + \mathrm{COD})} = \frac{1}{1+2} = 33\%$$

Figure 4 : Schéma de la mensuration de la similarité linguistique entre les différents points d'enquête d'un atlas linguistique

36 Cf., entre autres, les grands manuels taxométriques de Hans Hermann Bock (1974) ainsi que de Peter H. A. Sneath et Robert R. Sokal (1973).

37 Voir les manuels cartographiques de Roger Brunet (1987) et de Jean-Paul Bord (2012).

38 Le logiciel VDM est à la libre disposition de qui en fait la demande.

$IRI_{3,4}$: Indice Relatif d'Identité entre les points d'enquête 3 et 4
COI : co-identité (entre les paires de taxat j et j)

COD : co-différences (entre les paires de taxat a et b ainsi que h et i)

L' « Indice Relatif d'Identité » (IRI_{jk}) mesure la proportion de taxats identiques entre deux vecteurs-points d'enquête (j et k) de la matrice de données. Théoriquement, les scores de l'IRI_{jk} peuvent varier entre 1 et 100 (%). Suivant la formule combinatoire N/2 (N - 1) et pour l'effectif N = 641, le bilan total des calculs à effectuer s'élève à 205'120 scores qui, par la suite, seront stockés dans la matrice de similarité carrée (aux dimensions N × N).

Signalons encore que nous avons ajouté, aux 638 points d'enquête *originaux* de l'ALF, trois points *artificiels* relatifs aux langues standard du *français*, de l'*italien* et du *catalan*[39] qui permettent d'étudier l'impact de ces trois registres linguistiques sur le tissu des dialectes de l'ALF. La taille de N s'élève donc à 641 unités.

La matrice de similarité, outre d'être carrée, est aussi parfaitement symétrique. La valeur-IRI de 33% calculée pour les localités factices 3 et 4 se retrouve donc des deux côtés de la diagonale de la matrice de similarité : voir à nouveau la figure 4.

Une fois la matrice de similarité établie, il faut que le dialectométricien-classificateur se mette à la dépouiller judicieusement : voir à ce propos le moment C sur la figure 3. En plus, c'est le premier des nombreux pas *exploratoires* qui mènent des structures « de surface » aux structures de « profondeur ».

7. La mise en carte (visualisation) de scores quantitatifs : à l'exemple de plusieurs cartes de similarité

Voir les cartes 3-8.

Le défi heuristique à relever en la matière, est double[40] :

- du point de vue *statistique* :
  Il s'agit d'extraire un des N vecteurs de la matrice de similarité, de le dépouiller de la valeur relative à la diagonale qui est toujours égale de 1 (ou 100), et de visualiser les N-1 valeurs (ou scores) restantes.
- du point de vue *cartographique* :
  Il s'agit de *convertir* la *variabilité numérique* des N-1 scores de similarité extraits en une *variabilité visuelle analogue* pour fournir au géolinguiste-dialectométricien des informations aptes à être rapidement saisies par voie mentale et interprétées après du point de vue linguistique.

La « machine visualisatrice » employée à cet effet opère avec un certain nombre d'algorithmes de visualisation dont le plus utilisé à Salzbourg est l'algorithme MINMWMAX. Il répartit les N-1 valeurs à visualiser équitablement des deux côtés de la

[39] Cf. à ce propos Goebl 2000, 212.

[40] Voir à ce propos Goebl 1981 et 1984 (vol. I, passim).

moyenne arithmétique (MA) de la distribution de similarité respective, tout en définissant des intervalles numériquement toujours égaux des deux côtés de la MA[41].

Il en résulte des profils iconiques remarquablement stables et bien ordonnés auxquels le logiciel VDM permet de conférer des silhouettes plus ou moins fines, allant de 2 à 18 paliers chromatiques. Les couleurs que nous utilisons obéissent à l'agencement chromatique du spectre solaire qui est d'ailleurs celui de l'arc-en-ciel.

En général, le bleu est associé aux valeurs numériques petites alors que le rouge renvoie aux scores maximaux. Chaque carte de similarité – comme d'ailleurs toutes nos visualisations – dispose, en outre, d'une légende numérique et d'un histogramme qui fournit des informations qui intéressent plutôt le *statisticien* que le *linguiste*.

8. Structure de profondeur A : l'importance et le rôle heuristiques des cartes de similarité

Voir les cartes 3-8.

Les cartes 3-8 constituent six spécimens de cartes de similarité relatives à différents sous-corpus (équivalents à autant de matrices de données) tirés de notre matrice de données-ALF intégrale.

Les cartes 3 et 4 montrent deux des 641 cartes de similarité calculées à partir de notre corpus intégral (contenant 1681 CT). Elles disposent, comme chaque carte de similarité, d'un point de référence dont le polygone est toujours en blanc. Par la structuration – toujours très régulière – de leur profil choroplèthe, elles indiquent le décalage spatial des similarités linguistiques des N-1 points d'atlas restants par rapport aux points de référence préétablis : 394 (Auderville, Manche) et 778 (Agde, Hérault).

Une des grandes découvertes de la DM de Salzbourg est d'ailleurs l'étonnante régularité avec laquelle ce décalage s'inscrit dans l'espace. Et ceci malgré l'indéniable « chaotisme » qui se dégage de la considération particularisante de quelque 18 000 aires taxatoires réparties sur les 1681 CT extraites de l'ALF[42].

En ligne générale et dans les conditions de visualisation déjà mentionnées, l'entourage immédiat du point de référence prend les couleurs *rouge* (= intervalle 6) et *orange* (= intervalle 5), qui constituent toujours une aire (quantitative) cohérente. Par contre, les couleurs *bleu foncé* (= intervalle 1) et *bleu clair* (= intervalle 2) caractérisent les « antipodes » de la carte de similarité. Si la visualisation a été effectuée à l'aide de six paliers chromatiques, les polygones en *jaune* (= intervalle 4) et *vert* (= intervalle 3) représentent des zones-*tampon* (ou de *transition*) entre – métaphoriquement parlant – les « meilleurs amis », d'un côté, et, de l'autre, les « pires ennemis » du basilecte du point de référence.

[41] Voir à ce sujet surtout Goebl (1981, 1983 et 1984).

[42] Cf. Goebl (2003, 65).

Du point de vue linguistique, toute carte de similarité donne une réponse quantitative et finement nuancée à la question de savoir quelle est la position de tel dialecte à l'intérieur d'un espace dialectal examiné. J'ajoute que de telles questions ont été posées très souvent (et depuis longtemps) au sein de la romanistique et d'autres philologies modernes.

Les cartes 3, 4 et 5 montrent des profils géotypologiques pour les domaines du normand, du languedocien et du berrichon. Les profils choroplèthes respectifs sont très différents, surtout en ce qui concerne les cartes 3 (de type d'oïl) et 4 (de type d'oc).

Évidemment, les trois profils de similarité que nous venons de voir ont démontré que le décalage de la similarité linguistique dans l'espace peut être très variable. Toujours est-il qu'il s'agit là d'un phénomène *humain* ou, pour être plus précis, d'une émanation directe de la « gestion basilectale de l'espace par l'homo loquens », évoquée au début de cet article.

Or, il serait intéressant de comparer le décalage quantitatif produit dans l'espace par une activité d'origine *humaine* avec celui issu d'un modèle *géométrique* relatif à la même superficie. Un tel modèle consiste à définir les proximités aériennes réciproques entre les points de l'ALF à l'aide de la formule bien connue de Pythagore ($a^2 + b^2 = c^2$). La visualisation des 641 vecteurs de la matrice des proximités euclidiennes crée autant de profils choroplèthes absolument réguliers qui, eux, peuvent servir d'étalon pour une comparaison avec les calculs de similarité.

La carte 6 en constitue un spécimen relatif au point de référence 103 (Bué, Cher). En la comparant avec la carte 5, l'on comprend aisément que la gestion *basilectale* (= humaine) du réseau de l'ALF s'écarte considérablement de la gestion *géométrique* (= *euclidienne*) du même espace.

Or, la comparaison *visuelle* des deux profils choroplèthes multicolores peut être complétée par une évaluation *quantitative* des deux séries de 641 scores, relatives, d'un côté, aux *valeurs de similarité* (selon l'IRI) et, de l'autre, aux *scores de proximités* (selon la formule de Pythagore). Un tel calcul de corrélation peut être fait moyennant le coefficient de corrélation de [Auguste] Bravais et [Karl] Pearson r(BP). Pour les 2 × 641 scores qui sont à la base des cartes 5 et 6, il en résulte la valeur +0,697. Prise comme telle, cette valeur ne nous dit absolument rien. Ce n'est que dans le cadre de la DM corrélative que de telles valeurs peuvent revêtir une signification *linguistique* à proprement parler : voir le paragraphe 11.

Les cartes 7 et 8 reposent sur deux sous-corpus linguistiquement différents : alors que la carte 7 se réfère à notre sous-corpus *phonétique* (avec 1 117 CT), la carte 8 a été construite à l'aide de notre sous-corpus *lexical* (avec 465 CT).

L'allure de leurs profils choroplèthes est – malgré certaines similarités de base – quelque peu différente. On a l'impression que la stabilité interne du profil choroplèthe de la carte 7 (phonétique) dépasse celle du profil choroplèthe de la carte 8 (lexique).

9. Structure de profondeur B : l'analyse « interponctuelle » et la carte à cloisons

Voir les cartes 9 et 10.

L'analyse dite « interponctuelle » connaît deux versants : celui de la visualisation de *distances* et celui de la visualisation de *similarités* interponctuelles, c'est-à-dire relevées entre deux points d'enquête contigus. La première des deux visualisations, appelée « carte à cloisons », correspond exactement à la synthèse traditionnelle d'isoglosses destinée à dégager des bourrelets de frontière plus ou moins épais[43]. Il s'agit donc de la modélisation, faite avec les moyens de la DM, d'une technique déjà ancienne et, par conséquent, bien connue des géolinguistes.

La seconde visualisation, dite « carte à rayons », est, en revanche, de date récente. Du point de vie taxométrique, elle constitue le complément de la carte à cloisons, puisqu'elle visualise des *similarités* interponctuelles (s) dont le rapport avec les *distances* interponctuelles (d) est régi par la formule suivante : s + d = 100.

Le message linguistique des deux types de visualisation est également complémentaire. Alors que la carte à *cloisons* visualise, à travers l'épaisseur et le coloriage froid des côtés de *polygone*, le degré d' « aversion » qui existe entre deux locolectes contigus, la carte à *rayons* représente, par l'épaisseur et le coloriage chaud des côtés de *triangle*, des relations d' « amitié » à courte distance[44].

La structure iconique des cartes 9 et 10 ressort de l'interaction visuelle de 1791 éléments graphiques (i. e. de côtés de polygone) et fournit une vision très nuancée des enchevêtrements interponctuels qui existent entre le Nord et le Sud de la France. Il n'offre cependant aucune trace d'une confrontation *linéiforme* – c'est-à-dire revêtant les apparences d'une « véritable frontière » – entre les domaines d'oïl et d'oc. Sur les deux cartes, l'œil perçoit, par contre, une vaste ceinture de passage, sillonnée d'effets de cloisonnements finement gradués.

Une fois de plus, la traversée du Nord au Sud semble être spatialement plus « ordonnée » sur la carte 9 (basée sur le sous-corpus *phonétique*) que sur la carte 10 (basée sur le sous-corpus *lexical*).

Encore une petite remarque d'ordre quantitatif : l'effectif total des scores de distance (selon l'IRD_{jk}) pris en compte sur les cartes à cloisons 9 et 10, comprend seulement 0,873% du total des scores disponibles dans les matrices de distance respectives[45]. La clarté très suggestive de ces deux cartes repose donc sur des effectifs très réduits de la matrice de distance sous-jacente. En termes de *taxométrie*, la *valeur classificatoire* des deux cartes est donc relativement *faible*.

---

43 Voir les tableaux en appendice du travail séminal de Karl von Ettmayer (1924) où se trouve un grand nombre de schémas isoglottiques exécutés à la main et contenant des bourrelets de frontière de toute sorte. Voir aussi les cartes isoglottiques reproduites dans Jochnowitz 1973.

44 Pour des spécimens de la carte à rayons, cf. Goebl (1983 et 2012, 137-138).

45 Cette relation se calcule comme suit : 1791 : 205 120 × 100.

Heureusement, il y a des méthodes taxométriques beaucoup plus « enveloppantes » et, de ce fait, plus sûres : voir les paragraphes suivants.

10. Structure de profondeur C : au-delà des cartes de similarité : les cartes à paramètre

Voir les cartes 11-14.

Par « paramètre » l'on entend, en matière de *statistique descriptive*, une des nombreuses valeurs « caractéristiques » qui dénotent une quelconque distribution de fréquence, tel que le *minimum*, le *maximum*, la *MA*, l'*écart-type*, la *symétrie*, etc. En considérant les histogrammes au bas des cartes 3 et 4, l'on constate aisément que la symétrie des distributions de similarité peut être très variable.

À y regarder de près, cette propriété *statistique* revêt aussi une grande importance *linguistique*. Sur la carte 4, l'histogramme de droite, relatif à un parler du languedocien maritime, est fortement biaisé vers la gauche. Ceci signifie que des 640 scores de similarité 432 (= 53 + 245 + 134) se trouvent au-dessous de la MA[46] et seulement 208 (= 100 + 76 + 32) au-dessus du même seuil[47].

Le bilan analogue pour l'histogramme de la carte 3 se présente comme suit :

- valeurs situées au-dessous de la MA : 272 (= 9 + 168 + 95)
- valeurs situées au-dessus de la MA : 368 (= 13 + 198 + 157)

Quand on admet qu'une carte de similarité reflète – au moins en théorie – la position communicative du dialecte d'un point d'atlas, les valeurs mentionnées ci-dessus suggèrent que la position communicative *générale*[48] du dialecte du point-ALF 778 (situé en Languedoc méridional) est moins bonne que celle du point-ALF 394 (situé dans le nord de la Manche) et que, par conséquent, la plus ou moins grande symétrie des histogrammes respectifs renvoie, en dernière analyse, à cette potentialité interactive.

Or, il s'est avéré que le meilleur indice pour saisir la *symétrie* d'une distribution de similarité et, avec elle, sa vertu *communicative*, est le « coefficient d'asymétrie de [Ronald A.] Fisher » (CAF). En calculant le CAF de chacune des N distributions de similarité et en visualisant par la suite les N scores ainsi dégagés, l'on obtient une carte choroplèthe finement articulée (voir les cartes 11-14) dont la valeur linguistique est considérable. En établissant la synthèse des CAF nous poussons – métaphoriquement parlant – encore plus avant notre « plongée exploratoire dans les profondeurs » des données de l'ALF.

---

46 Il s'agit des valeurs appartenant aux intervalles 1-3 : sur la carte 4, elles sont toutes marquées par des couleurs « froides » (bleu foncé, bleu clair et vert).

47 Il s'agit des valeurs appartenant aux intervalles 4-6 : sur la carte 4, elles sont toutes marquées par des couleurs « chaudes » (jaune, orange et rouge).

48 Il s'agit de toujours mettre en rapport les différents locolectes de l'ALF avec l'ensemble de son réseau.

Du point de vue statistique, les scores du CAF oscillent des deux côtés de la valeur 0 : ils prennent des valeurs *négatives* quand les scores de similarité (selon IRI_{jk}) s'accumulent entre le maximum et la MA, et des valeurs *positives* quand le cumul intervient entre le minimum et la MA.

Sur les cartes 11-14, les polygones en *bleu foncé* – équivalant à des scores négatifs du CAF – renvoient à des dialecticités *bien connectées* avec le reste du réseau, alors que les polygones *rouges* – correspondant toujours à des scores *positifs* du CAF – se réfèrent à des variétés *détachées*, *conservatrices* et *peu enclines* à s'ouvrir à des interactions et échanges linguistiques.

Depuis un certain temps, j'utilise le concept de *compromis linguistique* (dérivé du terme allemand *Sprachausgleich*) pour désigner le degré d'interaction qu'un locolecte entretient avec le reste de son réseau. Au vu des cartes 9 et 10, l'on peut donc dire que les locolectes marqués en bleu (foncé et clair) sont très interactifs, alors que le contraire est vrai pour les locolectes marqués en rouge.

L'agencement spatial des deux tonalités est d'ailleurs très éloquent. Alors que les polygones en bleu foncé forment, d'un côté, un grand cercle fermé (autour du domaine d'oïl) et, de l'autre, un petit cercle à demi ouvert (relatif au francoprovençal) et évoquent par là l'idée d'une *irradiation spatiale* de grande envergure, les polygones rouges constituent des agglomérations *ponctuelles* suggérant l'idée d'immobilisme et de défense « tous azimuts ».

En effet, nous sommes en présence d'une analyse-DM très efficace dont l'utilité est avant tout d'ordre *diachronique*. C'est ainsi que le grand agencement circulaire bleu dans le nord évoque l'action irradiatrice continue des dialectes du type francilien, alors que la pince entrouverte bleue autour du domaine francoprovençal renvoie aux contacts et conflits linguistiques que ce géotype entretenait tant vers le nord que vers le sud.

Que le coloriage en bleu foncé – équivalent à des scores négatifs du CAF – soit en relation directe avec l'interaction linguistique tout court, est corroboré par le comportement du seul îlot linguistique compris dans le réseau de l'ALF, à savoir du P. 635 (Andraut, Gironde). L'on y trouve un dialecte oïlique de type saintongeais qui avait été transféré en territoire gascon au cours du XVe siècle. Comme il avait été exposé à toute sorte d'interactions au cours des derniers six siècles, il va de soi qu'il figure en bleu foncé.

Cette analyse-DM permet donc de repérer, au sein d'un réseau donné, des zones linguistiquement « mouvementées » et des régions relativement calmes ou même « inertes ».

Ajoutons encore que les quatre cartes en question reposent sur autant de sous-corpus différenciés du point de vue *quantitatif* et *qualitatif*.

- Pour les cartes 11 et 12 :

Les deux sous-corpus utilisés se distinguent par leur degré de *polynymie*. Alors que la carte 11 repose sur 536 CT dont les structures internes sont relativement simples, la carte 12 est basée sur 588 CT dont les structures sont très complexes. Pour mieux comprendre cet effet, il suffit de jeter un regard sur la figure 2. L'on y perçoit l'emplacement de trois sous-corpus (A, B et C) dont la structuration interne varie entre 2-5 taxats par CT (= sous-corpus *oligo*-nyme[49] A) et 13-90 taxats/CT (= sous-corpus *poikilo*-nyme[50] C).

Entre parenthèses : la carte 1, pourvue de 17 taxats et, partant, d'une structure plutôt complexe, appartient au sous-corpus C, alors que la carte 2, disposant de seulement trois taxats et, de ce fait, d'une structuration très simple, fait partie du sous-corpus A.

Or, il s'est avéré, au cours de nos analyses-DM, que de la synthèse de beaucoup de CT *oligo*-nymes (= simples ou « claires ») ressortent pratiquement les mêmes structures de profondeur que l'on obtient moyennant la synthèse de beaucoup de CT *poikilo*-nymes (= complexes ou « chaotiques »). Cet effet quelque peu surprenant montre que les structures-DM de profondeur disposent d'une grande redondance interne, comparables en cela au comportement d'autres signaux anthropiques.

Il n'est donc pas étonnant que les cartes 11 et 12 montrent pratiquement la même structuration choroplèthe caractérisée par une circularité « oïlique » et une pince entrouverte « francoprovençale » en bleu foncé dans le nord, et par trois « balouards de résistance » en rouge situés dans le sud.

Pour les cartes 13 et 14 :

Alors que le profil choroplèthe de la carte 13 – dressé à l'aide de 1 117 CT *phonétiques* – s'aligne grosso modo sur celui des cartes 11 et 12, le profil choroplèthe de la carte 14, reposant sur 465 CT *lexicales*, s'en écarte un peu. L'effet oïlique de circularité y est moins saillant : ce qui ressort avec beaucoup de clarté, c'est, d'un côté, une large bande d'*interaction* et de *compromis linguistique* qui se creuse, d'une façon très compacte, au beau milieu de la Galloromania, et, de l'autre, la réduction de la zone coloriée en rouge, orange et jaune, causée par le détachement du Limousin et de l'Auvergne du bloc occitano-catalan.

11. Structure de profondeur D : la dialectométrie corrélative

Voir les cartes 15-18.

La DM corrélative, qui est une des pièces fortes de l'arsenal méthodique de l'EDMS, existe depuis 2005[51] et a fourni depuis beaucoup de preuves de sa grande utilité heuristique. Le module de VDM relatif à la DM corrélative permet de calculer la corrélation entre deux matrices de similarité (etc.) à dimensions extérieures égales

[49] Grec *olígos* „petit, peu".

[50] Grec *poikílos* „bigarré, polychrome".

[51] Cf. Goebl (2005, passim).

(= deux fois N × N), et de visualiser rapidement le rendement de ces calculs sous la forme d'une carte choroplèthe à couleurs avec N points (ou polygones).

La *nature phénoménologique* des deux variables à corréler peut varier à loisir : l'on peut corréler, entre elles, différentes *catégories linguistiques*, les mettre en relation, l'une après l'autre, avec l'*espace euclidien*, ou comparer, par voie de corrélation, les résultats numériques de différentes *mensurations de similarité*. Ce qui importe au cours de ces opérations statistico-visualisatrices, c'est de ne jamais perdre de vue leur finalité et signification *linguistiques* à proprement parler. Ajoutons encore que la profondeur des structures synthétiques rejointes par la DM corrélative, est très grande.

Nous commençons la description de l'efficacité de la DM corrélative par la présentation d'une corrélation cartographique entre les *similarités linguistiques* d'un côté et les *proximités euclidiennes* de l'autre : voir la carte 15. Le but en est d'étudier de plus près les multiples facettes du décalage spatial que subissent les scores de similarité linguistique des différentes cartes de similarité.

En reprenant la leçon donnée par les cartes 5 (= de *similarité*) et 6 (= de *proximité*) il est possible de calculer 641 scores de corrélation – toujours selon le coefficient de corrélation r(BP) – et de les visualiser par la suite suivant les normes habituelles de l'EDMS. Il en résulte la carte 15 qui, d'un côté, repose sur le corpus-ALF intégral (= 1681 CT) et, de l'autre, sur les proximités euclidiennes calculées selon le théorème de Pythagore.

Le profil choroplèthe de la carte 15 est d'une clarté et régularité exemplaires. Pour bien pouvoir le comprendre en tant que *linguiste*, il faut avant tout connaître la signification *linguistique* de la variabilité *numérique* calculée, et du symbolisme chromatique y ayant trait.

Or, les teintes *chaudes* renvoient à une *harmonie intrinsèque* du déploiement des similarités linguistiques et des proximités euclidiennes dans l'espace, alors que les teintes *froides* signalent de fortes *frictions* et *divergences* entre ces deux « énergies » spatiales. Evidemment, de telles disharmonies sont dues à tout ce qui peut influer de l'extérieur sur l'évolution « naturelle » de la langue dans l'espace, soit en l'accélérant, soit en la retardant. Les facteurs en question sont bien entendu d'ordre *historique*, *politique* et *social*.

Dans cette perspective, la Galloromania se présente comme le théâtre d'une grande confrontation entre le Nord et le Sud où le premier a fini par prendre le dessus tout en déstabilisant graduellement le Sud, en passant par le Massif Central pour arriver aux Pyrénées. La ceinture médiane de la Galloromania, allant de la Gironde à l'ouest jusqu'à l'éventail du francoprovençal à l'est, constitue la zone des affrontements les plus violents.

Une fois de plus, la position de l'îlot linguistique du P.-ALF 635 (Andraut) – où l'on trouve le score r(BP) minimal – nous fournit la preuve du bien fondé de notre supposition selon laquelle la divergence de fond entre les gestions *basilectale* et *euclidienne* de l'espace crée les scores de corrélation les plus bas et, partant, des coloriages

en bleu foncé. Rappelons que la « Petite Gavacherie » dont Andraut (= P.-ALF 635) fait partie, avait été fondée au XVe siècle en territoire d'oc comme colonie linguistique d'oïl d'origine saintongeaise. Il va de soi que la fondation d'un îlot linguistique en terre linguistiquement étrangère constitue une violation éclatante de la relation « naturelle » entre les gestions *basilectale* et *euclidienne* de l'espace.

Évidemment, il est possible de répéter le même exploit-DM pour la *phonétique* et l'*espace* d'un côté (voir la carte 16), ainsi que pour le *lexique* et l'*espace* de l'autre (voir la carte 17). Les plus grandes divergences entre les profils choroplèthes des deux cartes à corrélations se situent en Provence, alors que pour le reste de la Galloromania, la distribution spatiale des corrélations entre les gestions linguistique et euclidienne de l'espace reste sensiblement la même.

Selon toute vraisemblance, l'allure éminemment régulière des cartes 15-17 est due, entre autres, à la grande ancienneté des divergences qui se sont manifestées, au cours des siècles, entre les gestions *linguistique* et *euclidienne* de la Galloromania.

La dernière des quatre cartes corrélatives que nous venons de présenter, se réfère à un problème éminemment *grammatical* : voir la carte 18. Il s'agit de savoir dans quelle mesure la gestion de l'espace moyennant deux *catégories linguistiques différentes* s'annonce comme convergente ou divergente.

Les mensurations et les visualisations y ayant trait fournissent un profil choroplèthe très clair : les gestions *lexicale* et *phonétique* divergent considérablement le long d'un couloir « delphino-forme » allant de la Gironde aux Alpes alors qu'elles convergent tant au nord qu'au sud de la Galloromania. À remarquer, une fois de plus, l'allure très régulière du profil choroplèthe dans son ensemble.

Ajoutons encore que le couloir *delphino*-forme mentionné ci-dessus a fait son apparition dans beaucoup d'autres analyses-DM relatives à l'ALF. Il est cependant bon de ne pas le confondre avec le tracé du « croissant » de mémoire ronjatienne. Ce dernier se situe plus au nord et dispose d'un contour beaucoup plus mince[52].

Précisons en outre que l'application de la DM corrélative à d'autres domaines linguistiques – romans ou autres – a fourni également des résultats très clairs et bien structurés : cf. à ce propos Goebl (2005, passim, et 2012, 155 sq.).

12.Structure de profondeur E : l'analyse dendrographique

Voir les cartes 19-20.

Le niveau E de l'analyse-DM des structures de profondeur concerne la DM dendrographique ou – dans la terminologie technique de la classification numérique (ou taxométrie) – la « classification ascendante hiérarchique » (CAH). La CAH est largement employée au sein de beaucoup de sciences naturelles et humaines. Elle sert au

[52] Pour l'historique du *croissant*, voir le travail séminal de Guylaine Brun-Trigaud de 1990 : pour les contours géographiques traditionnels du « croissant » tel qu'il avait été défini en 1913 par Jules Ronjat (1864-1925), voir la carte à la page 274 (ibidem).

groupement algorithmique de N objets – définis, chacun, par p attributs – en structures hiérarchiques pourvues de ramifications binaires. Du point de vue statistique, il s'agit d'une analyse dite « multivariée ». Dans la chaîne-DM, la CAH ne représente cependant qu'une seule des nombreuses possibilités de dépouillement de la matrice de similarité (voir la figure 3). Vu la longue tradition de la *pensée arborescente* au sein de pratiquement toutes les branches de la linguistique, ce genre d'analyse offre un certain attrait « naturel » aussi pour les besoins de la géolinguistique.

Les algorithmes à utiliser (ou utilisables) dans le cadre de la CAH sont très variés. Ils opèrent, tout en exécutant des impératifs mathématiques préalablement définis, des fusions (ou agglomérations) binaires entre les N objets de la matrice de similarité en question. En général, leur action agglomératrice débute par les différences *minimales* qui existent entre les N^2 valeurs de la matrice de similarité pour en rejoindre les différences *maximales* : d'où le qualificatif « ascendant » de la CAH. Quant au choix de l'algorithme-CAH le plus approprié aux besoins de la DM, il n'y a pas de solution fixe ou « passe-partout ». C'est pourquoi le dialectométricien doit sélectionner, dans l'arsenal algorithmique de la CAH, la procédure dont les résultats correspondent le plus à ses propres besoins et visées.

Un des algorithmes-CAH qui répond le mieux à ces impératifs est celui qui a été proposé, en 1963, par le statisticien américain Joe Ward Jr. Il fournit, en ligne générale, d'excellents résultats quand il est appliqué à des données de nature géolinguistique. Un des critères pour déterminer l'utilité *linguistique* d'un algorithme-CAH est la cohérence spatiale des différents clusters (= « chorèmes ») extraits de l'arbre calculé.

Évidemment, l'arbre en tant que configuration ramifiée ne saurait jamais suffire aux besoins d'un *géolinguiste* épris de schémas *cartographiques* et, par conséquent, *spatiaux*. Il faut donc projeter le rendement classificatoire de l'arbre dans l'espace. J'appelle ce transfert « spatialisation ». C'est ainsi que chacun des branchages de l'arbre (appelés par nous « dendrèmes ») retenus comme linguistiquement « significatifs », est transposé dans le secteur approprié du réseau en question qui, lui, est appelé à Salzbourg « chorème ». Inutile d'ajouter que le coloriage des dendrèmes correspond parfaitement à celui des chorèmes.

Le jeu alternatif des différents chorèmes disséminés dans l'espace établit donc une excellente classification des locolectes examinés, alors que le message hiérarchique concomitant de l'arbre permet d'y ajouter des nuances quantitatives – et aussi diachroniques – supplémentaires.

Ceci est dû au fait que tout arbre issu d'une CAH peut être utilisé également à une *reconstruction diachronique* telle qu'elle est conçue et pratiquée depuis belle lurette par la *lexicostatistique* et la *glottochronologie*[53]. La logique de cette perspective veut que la ramification croissante à partir de la racine de l'arbre corresponde au déploiement diachronique de la fragmentation linguistique du territoire en question.

[53] Cf. à ce sujet Goebl (2003, 84-89).

Jetons à ce propos un coup d'œil sur la carte 19 : les deux embranchements de la première bifurcation correspondent, en effet, aux domaines d'oc (= dendrèmes/chorèmes 1-4) et d'oïl (= dendrèmes/chorèmes 5-8). La bifurcation suivante – toujours selon la logique de l'algorithme de Ward – engendre la séparation entre le francoprovençal (= dendrème/chorème 5) et le reste du domaine d'oïl (= dendrèmes/chorèmes6-8). Evidemment, il est possible de continuer ce genre d'analyse jusqu'au niveau des feuilles. VDM en offre les possibilités techniques nécessaires.

Une autre particularité – d'ordre statistique – mérite également d'être mentionnée. C'est que les différents dendrèmes (et chorèmes) doivent leur genèse à la considération serrée de leur *variabilité quantitative interne*. C'est ainsi que naît une certaine hiérarchisation des dendrèmes/chorèmes en fonction de leur hétérogénéité quantitative interne (et aussi réciproque) : plus un dendrème est proche de la racine de l'arbre, plus grande en est l'hétérogénéité interne. Et vice versa.

Ceci signifie que, sur la carte 19, la variabilité numérique interne du dendrème/chorème relatif au domaine d'oïl (= dendrèmes/chorèmes 5-8) est moins grande que celle du dendrème/chorème relatif au domaine d'oc (= dendrèmes/chorèmes 1-4)[54].

Soulignons encore que les cartes 19 et 20 se distinguent par leurs corpus de base : la carte 19 repose sur notre sous-corpus *phonétique* (avec 1 117 CT) alors que la carte 20 émane d'une matrice de données *lexicale* (avec 465 CT). Il va de soi que les deux arborescences et leurs spatialisations ne sont pas complètement identiques. Ceci vaut surtout pour l'agencement (relationnel) des huit dendrèmes à l'*intérieur* des arbres correspondants.

Pour en faciliter la comparaison, nous avons conféré, sur les cartes 19 et 20, le même coloriage (et numérotation) aux chorèmes placés dans une situation géographique comparable, d'où il résulte cependant une numérotation non plus continue (c'est-à dire allant de 1 à 8) mais discontinue des différents dendrèmes de la carte 20. L'on peut en déduire que la première bifurcation aboutit, sur les deux cartes, aux mêmes résultats (= nette séparation entre Oc et Oïl), certes, mais aussi que la hiérarchisation des huit dendrèmes/chorèmes colorés à l'intérieur de l'arbre de la carte 20 se distingue quelque peu de la situation analogue sur la carte 19. Toujours est-il que la cohérence spatiale des chorèmes telle qu'elle se manifeste près de la racine, persiste aussi au *milieu* de l'arbre et près des *feuilles*. Ceci prouve le haut degré de « classificabilité » de nos données.

13. Epilogue

Le fondement théorique de la DM de Salzbourg est l'hypothèse de l'existence de la gestion basilectale de l'espace par l'homo loquens dont les manifestations réelles – toutes saisissables par le biais d'un atlas linguistique de type ALF, AIS, ALPI ou

[54] La raison en est que la racine du méga-dendrème 5-8 est plus proche du tronc de l'arbre que celle du méga-dendrème 1-4.

ALDC – obéissent à certaines régularités voire lois générales qui ne peuvent être découvertes qu'à travers un certain nombre d'opérations *inductives* ou de *synthèse*. Il importe que les résultats de ces opérations de synthèse soient dûment visualisés afin qu'ils puissent déclencher les opérations mentales voulues.

Il est en outre très important de ne pas oublier qu'il y a, en dehors de la géolinguistique à proprement parler, de nombreuses sciences étroitement liées avec l'espace dont les méthodes géo-classificatoires sont très poches de celle de la DM, et avec lesquelles une coopération interdisciplinaire s'impose. Nous n'en citons ici que la *géographie humaine* et la *génétique des populations*[55]. Dans la rétrospective, il est bien évident que de telles disponibilités – ou faudrait-il dire « velléités » – interdisciplinaires existaient déjà du temps de Coquebert de Montbret père et fils.

Il est cependant non moins évident que – du point de vue *empirique* et *méthodique* – un fil conducteur rouge s'étend de leurs activités à la DM actuelle.

Université de Salzbourg Hans GOEBL

Remerciements

Préparation et confection de tous les clichés de cet article : Werner Goebl (Vienne)

Contrôle linguistique de mon texte français : Stéphanie Lescure (Marbourg, Hesse)

Que mon fils Werner et Mme Lescure reçoivent ici l'expression de ma sincère et profonde gratitude.

Abréviations souvent utilisées

AL	atlas linguistique
CAF	coefficient d'asymétrie de [Ronald A.] Fisher
CAH	classification ascendante hiérarchique
CT	carte de travail
DM	dialectométrie, relative à la dialectométrie, dialectométrique
EDMS	Ecole dialectométrique de Salzbourg
IRD_{jk}	Indice Relatif de Distance (entre les points d'atlas j et k)
IRI_{jk}	Indice Relatif d'Identité (entre les points d'atlas j et k)
MA	moyenne arithmétique

[55] Citons, à titre d'exemple, la contribution interdisciplinaire de Scapoli / Goebl / Sobota / Mamo lini / Barrai de 2005.

MEDMW	algorithme d'intervallisation (centré sur la MA, avec la formation, des deux côtés de la MA, de *n* classes égales en nombre de points d'atlas)
MINMWMAX	algorithme d'intervallisation (centré sur la MA, avec la formation, des deux côtés de la MA, de *n* classes dont la largeur numérique est égale)
r(BP)	coefficient de corrélation de [Auguste] Bravais et [Karl] Pearson
VDM	Visual DialectoMetry (logiciel dialectométrique utilisé à Salzbourg ; librement disponible sur demande)

Références bibliographiques

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

ALDC = Veny, Joan / Pons i Griera, Lídia, 2001-2013. *Atles Lingüístic del Domini Català*, 6 vol., Barcelona, Institut d'Estudis Catalans.

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

ALPI = *Atlas lingüístico de la Península Ibérica*, 1962, vol. I : Fonética, Madrid, CSIC.

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

Baehr, Rudolf, 1958. « Zu den romanischen Wochentagsnamen », in : Lausberg, Heinrich / Weinrich, Harald (ed.), *Romanica. Festschrift für Gerhard Rohlfs*, Halle, Niemeyer, 26-56.

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.

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

Bord, Jean-Paul, 2012. *L'univers des cartes. La carte et le cartographe*, Paris, Belin.

Bourguet, Marie-Noëlle, 1988. *Déchiffrer la France. La statistique départementale à l'époque napoléonienne*, Paris, Editions des Archives Contemporaines.

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

Brun-Trigaud, Guylaine, 1990. *Le croissant : le concept et le mot. Contribution 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.

Chartier, Roger, 1992. « La ligne Saint-Malo – Genève », in : Nora, Pierre (ed.), *Les lieux de mémoire, vol. 3 : Les France, tome 1 : Conflits et partages*, Paris, Gallimard, 738-775.

Ettmayer, Karl von, 1924. « Das Wesen der Dialektbildung, erläutert an den Dialekten Frankreichs », *Denkschriften* [= Mémoires] *der Akademie der Wissenschaften in Wien*, philosophisch-historische Klasse, 66/3, 1-56, 7 Tafeln [= tableaux].

FEW = Wartburg, Walther von (1922/1928 sq.). *Französisches etymologisches Wörterbuch*, Bonn, Bouvier.

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

Gilliéron, Jules / Roques, Mario, 1912. *Etudes de géographie linguistique d'après l'Atlas linguistique de la France*, Paris, Champion.

Goebl, Hans,1976. « La dialectométrie appliquée à L'ALF (Normandie) », in : Vàrvaro, Alberto (ed.), *Atti del XVI Congresso Internazionale di Linguistica e Filologia Romanza* (Naples 1974), Naples, Marchiaroli/Amsterdam, Benjamins, vol. 2, 165-195.

Goebl, Hans, 1981. « Eléments d'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*, Tübingen, Niemeyer, 3 vol.

Goebl, Hans, 2000. « La dialectométrisation de l'ALF : présentation des premiers résultats », *Linguistica* 40, 209-236.

Goebl, Hans, 2002a. « Analyse dialectométrique des structures de profondeur de l'ALF » *RLiR* 66, 5-63.

Goebl Hans, 2002b. « Sprachatlanten : woher? womit? wozu? Einige buntgemischte Reflexionen am Gartenzaun zwischen Romanistik und Germanistik », in : Anreiter, Peter / Ernst, Peter / Hausner, Isolde / Kalb, Helmut (ed.), *Namen, Sprachen und Kulturen. Imena, Jeziki in Kulture. Festschrift für Heinz Dieter Pohl zum 60. Geburtstag*, Vienne, Edition Präsens, 257-274.

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, 59-96.

Goebl, Hans, 2004. « VDM – Visual DialectoMetry. Vorstellung eines dialektometrischen Software-Pakets auf CD-ROM (mit Beispielen zu ALF und Dees 1980) », in : Dahmen, Wolfgang / Holtus, Günter / Kramer, Johannes / Metzeltin, Michael / Schweickard, Wolfgang / Winkelmann, Otto (ed.), *Romanistik und neue Medien*. Romanistisches Kolloquium XVI (Rauischholzhausen, 2000), Tübingen, Narr, 209-241.

Goebl, Hans, 2005. « La dialectométrie corrélative : un nouvel outil pour l'étude de l'aménagement dialectal de l'espace par l'homme », *RLiR* 69, 321-367.

Goebl, Hans, 2006. « 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 (Göttingen, 2004), Tübingen, Narr, 291-317.

Goebl, Hans, 2008. « La dialettometrizzazione integrale dell'AIS. Presentazione dei primi risultati », *RLiR* 72, 25-113.

Goebl, Hans, 2011. « Quo vadis, atlas linguistice? Einige wissenschaftshistorische und zeitgeistkritische Reflexionen zur atlasgestützten Geolinguistik », in : Schlaak, Claudia / Busse, Lena

(ed.),*Sprachkontakte, Sprachvariation und Sprachwandel. Festschrift für Thomas Stehl zum 60. Geburtstag*, Tübingen, Narr, 5-27.

Goebl, Hans, 2012. « Introduction aux problèmes et méthodes de l' ‹Ecole dialectométrique de Salzbourg› (avec des exemples gallo-, italo- et ibéroromans) », in : Álvarez Pérez, Xosé Afonso / Carrilho, Ernestina / Magro, Catarina (ed.), *Proceedings of the International Symposium on Limits and Areas in Dialectology (LimiAr)*, Lisbonne, Centro de Linguística da Universidade de Lisboa, 117-166 (dans le web : http ://limiar.clul.ul.pt/proceedings/general_session/7_goebl.pdf).

Goebl, Hans / Scherrer, Yves / Smečka, Pavel, 2013. « Kurzbericht über die Dialektometrisierung des Gesamtnetzes des ‹Sprachatlasses der deutschen Schweiz› (SDS) », in : Schneider-Wiejowski, Karina / Kellermeier-Rehbein, Birte / Haselhuber, Jakob (ed.), *Vielfalt, Variation und Stellung der deutschen Sprache* [Mélanges pour le 70e anniversaire de Ulrich Ammon], Berlin/Boston, Walter de Gruyter, 153-175.

Goebl, Hans / Schiltz, Guillaume, 1997. « A Dialectometrical Compilation of CLAE 1 and CLAE 2. Isoglosses and Dialect Integration », in : Viereck, Wolfgang / Ramisch, Heinrich (ed.), *The Computer Developed Linguistic Atlas of England*, vol. 2, Tübingen, Niemeyer, 13-21.

Goebl, Hans / Winterleitner, Norbert, 1971. « Projekt einer sprachstatistischen Auswertung von in Sprachatlanten gespeicherter linguistischer Information mit Hilfe elektronischer Rechenanlagen », *Linguistische Berichte* 14, 60-61.

Herrgen, Joachim, 2001. « Die Dialektologie des Deutschen », in : Auroux, Sylvain / Koerner, E. F. Konrad / Niederehe, Hans-Josef / Versteegh, Kees (ed.), *History of the Language Sciences / Geschichte der Sprachwissenschaften / Histoire des sciences du langage. An International Handbook on the Evolution of the Study of Language from the Beginning to the Present / Ein internationales Handbuch zur Entwicklung der Sprachforschung von den Anfängen bis zur Gegenwart / Manuel international sur l'évolution de l'étude du langage des origines à nos jours*, Berlin/New York, de Gruyter, vol. 2, 1513-1535.

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

Jochnowitz, George, 1973. *Dialect Boundaries and the Question of Franco-Provençal*, La Haye, Mouton.

Ködel, Sven, 2013. *L'enquête Coquebert de Montbret (1806-1812) sur les langues et dialectes de France et la représentation de l'espace linguistique français sous le Premier Empire*, Bamberg, Paris, thèse doctorale.

Laboulais-Lesage, Isabelle, 1999. *Lectures et pratiques de l'espace. L'Itinéraire de Coquebert de Montbret, savant et grand commis d'Etat (1755-1831)*, Paris, Champion.

Ozouf-Marignier, Marie-Vic, 1989. *La formation des départements : la représentation du territoire français à la fin du 18ème siècle*, Paris, Editions de l'Ecole des Hautes Etudes en Sciences Sociales.

Palsky, Gilles, 1996. *Des chiffres et des cartes. Naissance et développement de la cartographie quantitative française au XIXe siècle*, Paris, CTHS.

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

Ravier, Xavier, 1976. « Jean Séguy et la traversée du langage gascon. Réflexions sur une topogenèse géolinguistique », *RLiR* 40, 389-402.

Rohlfs, Gerhard, 1949. « Les noms des jours de la semaine dans les langues romanes », in : *Mis-*

celânea de filologia, literatura e história cultural, à memória de Francisco Adolfo Coelho (1847-1919), Lisbonne, Centro de Estudos Filológicos, 88-94.

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).

Scapoli, Chiara / Goebl, Hans / Sobota, Slawomir / Mamolini, Elisabetta / Rodriguez-Larralde, Alvaro / Barrai, Italo, 2005. « Surnames and Dialects in France : Population structure and cultural evolution », *Journal of Theoretical Biology* 237, 75-86.

Séguy, Jean, 1971. « La relation entre la distance spatiale et la distance lexicale ». *RLiR* 35, 335-357.

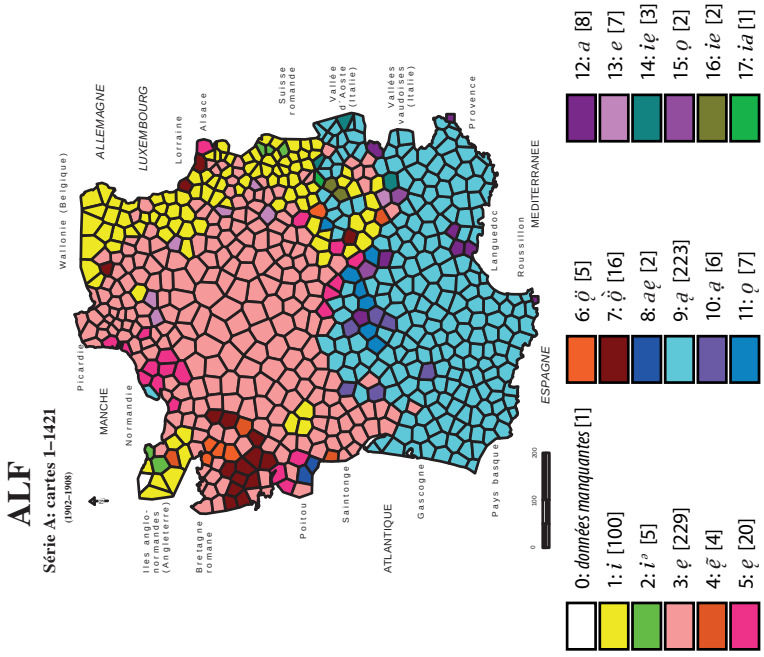
Séguy, Jean, 1973. « La dialectométrie dans l'Atlas linguistique de la Gascogne », *RLiR* 37, 1-24.

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

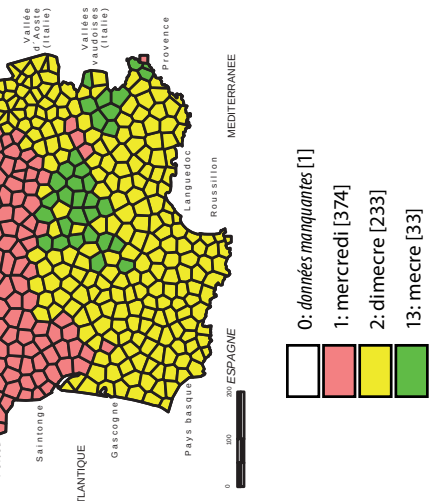
Storost, Jürgen, 1994. «Noch einmal zur Diez-Stiftung», *RJb* 45, 74-84.

Tuaillon, Gaston, 1983. « Les atlas linguistiques régionaux de France », *BALI* III Serie, Dispensa n. 7, 68-89.

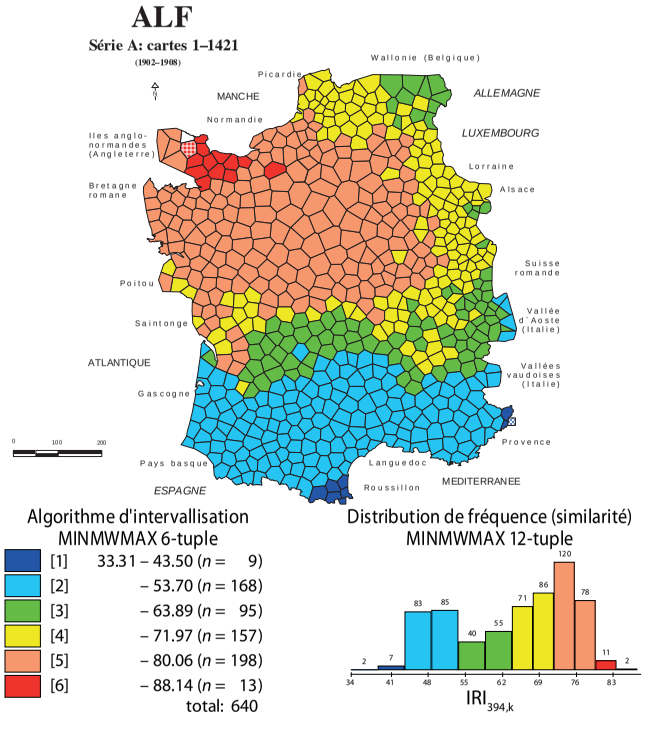
Wartburg,Walther von, 1949. «Los nombres de los días de la semana», *RFE* 33, 1-14.



Carte 1: Exemple d'une carte de travail 17-nyme de nature *phonétique* (selon ALF 812 *marché*) : les succédanés du nexus latin -ÁTU



Carte 2: Exemple d'une carte de travail 3-nyme de nature *lexicale* (selon ALF 839 *mercredi*) : les réalisations lexicales du concept « mercredi »

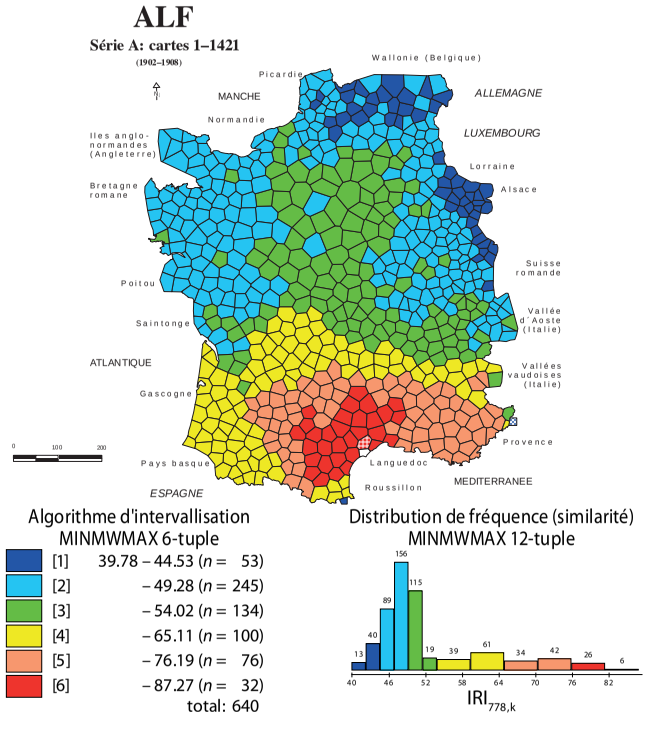


Carte 3: Profil choroplèthe d'une carte de similarité relative au point-ALF 394 (Auderville, Manche)

Corpus: corpus intégral (1681 cartes de travail)

Indice de similarité: IRI_{jk}

Algorithme d'intervallisation: MINMWMAX 6-tuple

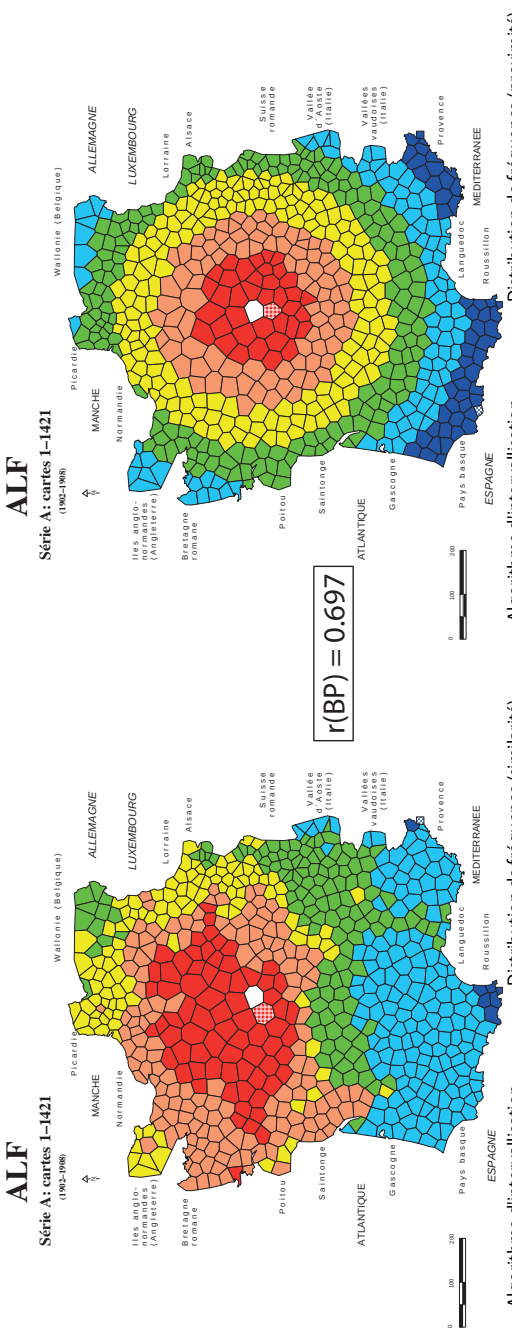


Carte 4: Profil choroplèthe d'une carte de similarité relative au point-ALF 778 (Agde, Hérault)

Corpus: corpus intégral (1681 cartes de travail)

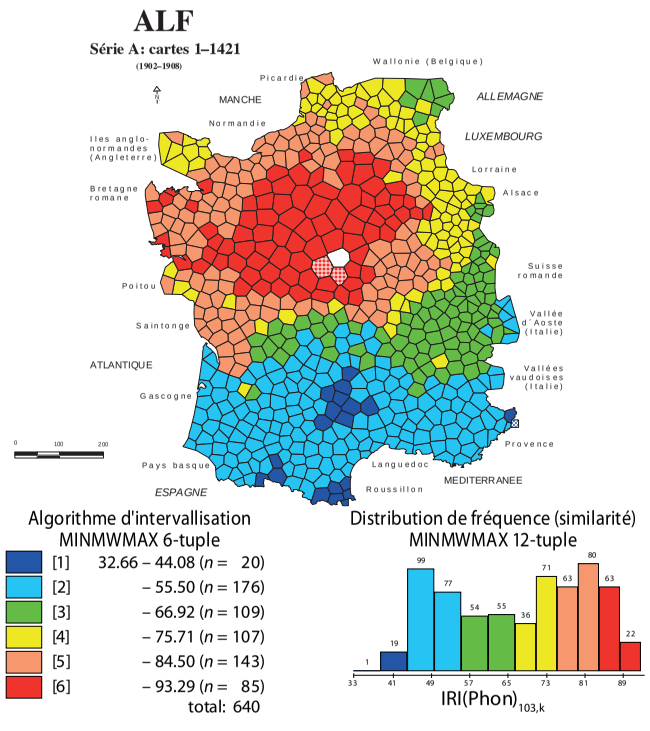
Indice de similarité: IRI_{jk}

Algorithme d'intervallisation: MINMWMAX 6-tuple



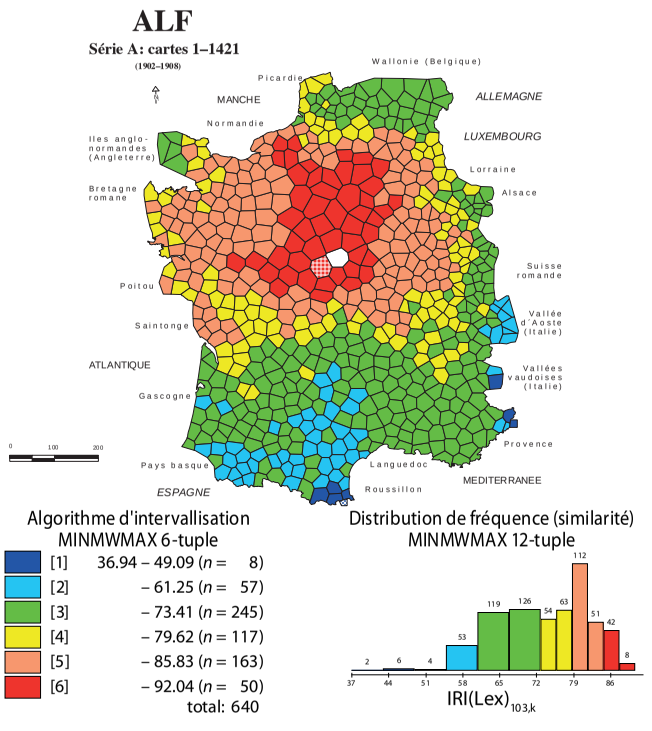
Carte 5: Profil choroplèthe d'une carte de similarité relative au point-ALF 103 (Bué, Sancerre, Cher)
Corpus: corpus intégral (1681 cartes de travail)
Indice de similarité: IRIjk
Algorithme d'intervallisation: MINMWMAX 6-tuple

Carte 6: Profil choroplèthe d'une carte de proximité relative au point-ALF 103 (Bué, Sancerre, Cher)
Indice de proximité: théorème de Pythagore
Algorithme d'intervallisation: MINMWMAX 6-tuple



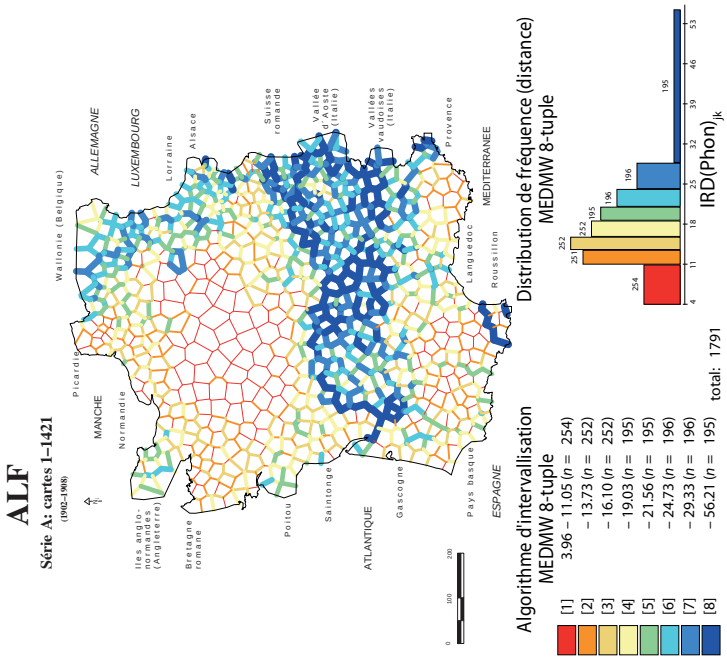
Carte 7: Profil choroplèthe d'une carte de similarité relative au point-ALF 103 (Bué, Sancerre, Cher)

Corpus: sous-corpus *phonétique* (1117 cartes de travail)
Indice de similarité: IRI_{jk}
Algorithme d'intervallisation: MINMWMAX 6-tuple

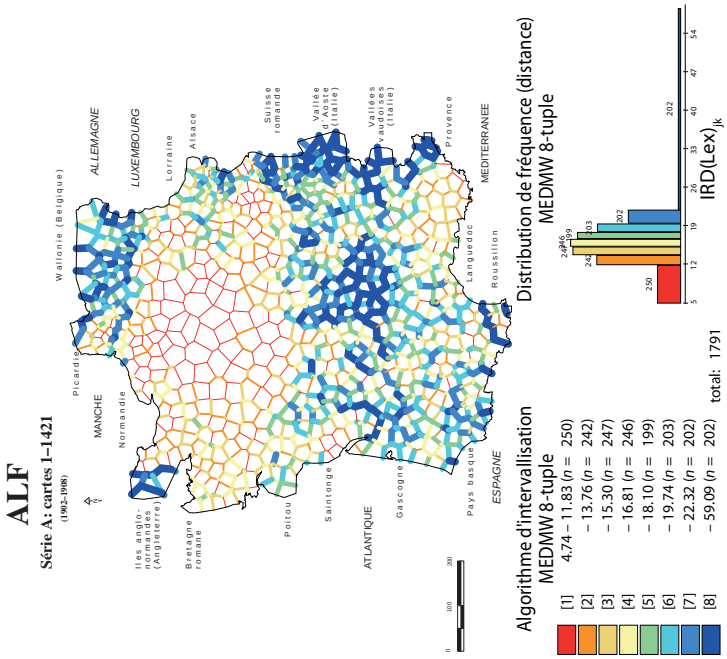


Carte 8: Profil choroplèthe d'une carte de similarité relative au point-ALF 103 (Bué, Sancerre, Cher)

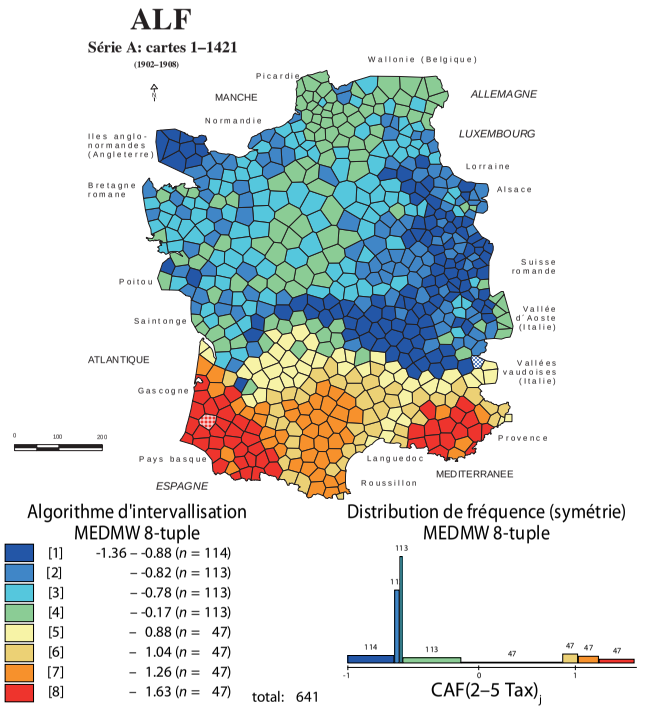
Corpus: sous-corpus *lexical* (465 cartes de travail)
Indice de similarité: IRI_{jk}
Algorithme d'intervallisation: MINMWMAX 6-tuple



Carte 9: Analyse interponctuelle (carte à cloisons)
Corpus: sous-corpus *phonétique* (1117 cartes de travail)
Indice de distance: IRDjk
Algorithme d'intervallisation: MEDMW 8-tuple



Carte 10: Analyse interponctuelle (carte à cloisons)
Corpus: sous-corpus *lexical* (465 cartes de travail)
Indice de distance: IRDjk
Algorithme d'intervallisation: MEDMW 8-tuple

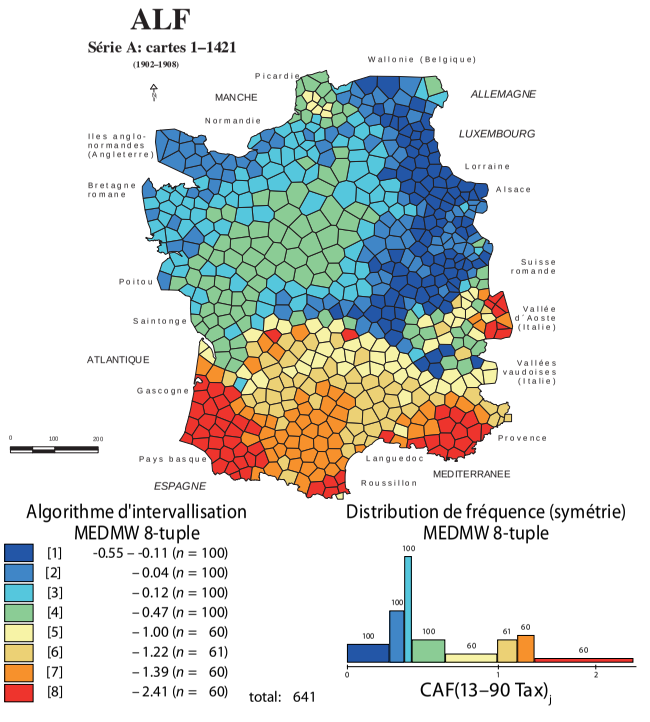


Carte 11: Synopse des scores du « coefficient d'asymétrie de Fisher » (CAF)

Corpus: sous-corpus A (entre 2 et 5 taxats par carte de travail): 536 cartes de travail

Indice de similarité: IRI_{jk}

Algorithme d'intervallisation: MEDMW 8-tuple

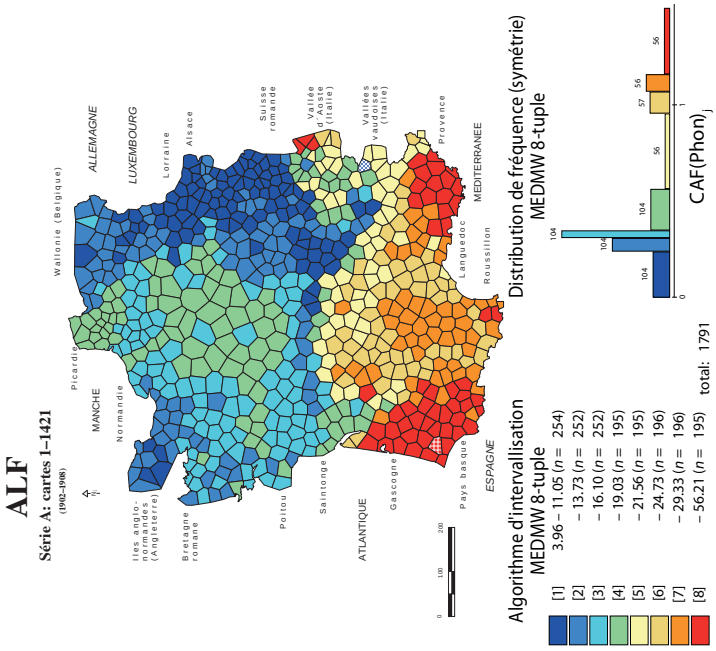


Carte 12: Synopse des scores du « coefficient d'asymétrie de Fisher » (CAF)

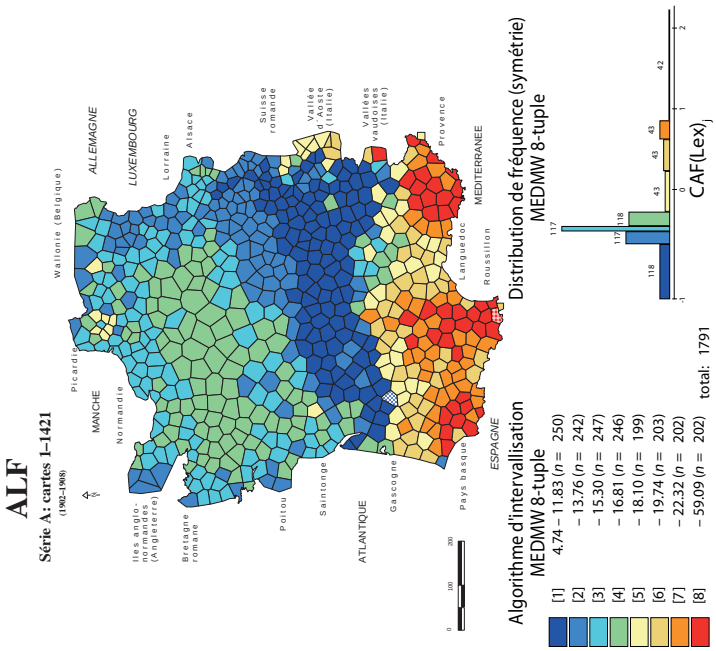
Corpus: sous-corpus C (entre 13 et 91 taxats par carte de travail): 588 cartes de travail

Indice de similarité: IRI_{jk}

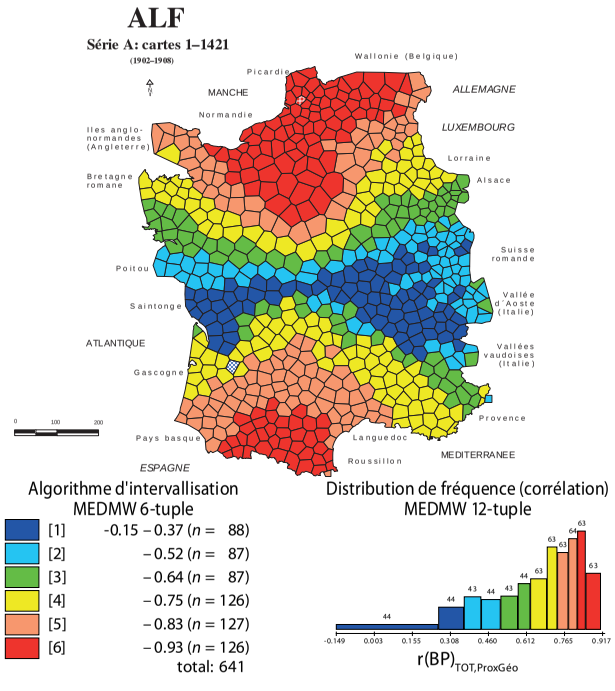
Algorithme d'intervallisation: MEDMW 8-tuple



Carte 13: Synopse des scores du « coefficient d'asymétrie de Fisher » (CAF)
Corpus: sous-corpus *phonétique* (1117 cartes de travail)
Indice de similarité: IRIjk
Algorithme d'intervallisation: MEDMW 8-tuple



Carte 14: Synopse des scores du « coefficient d'asymétrie de Fisher » (CAF)
Corpus: sous-corpus *lexical* (465 cartes de travail)
Indice de similarité: IRIjk
Algorithme d'intervallisation: MEDMW 8-tuple



Carte 15: Profil choroplèthe d'une carte à corrélations

Matrices carrées corrélées:

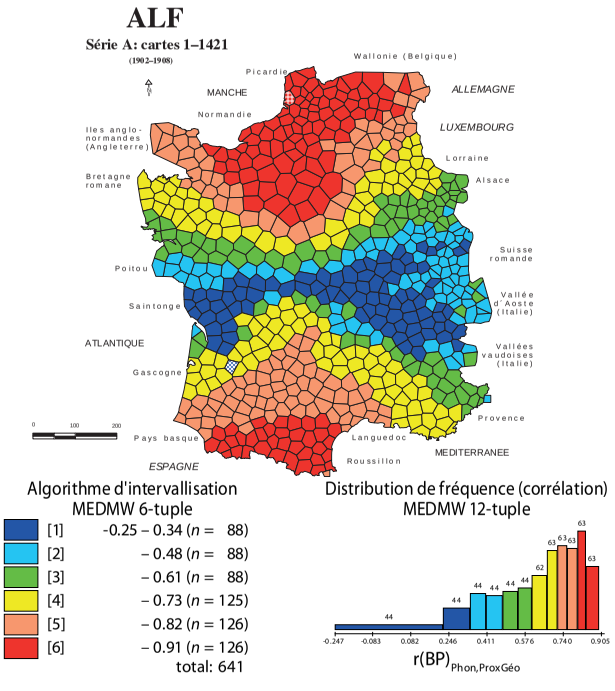
matrice de similarité (selon IRI_{jk})

- corpus *intégral* (1681 cartes de travail)

matrice de proximité (selon le théorème de Pythagore)

-

Algorithme d'intervallisation: MEDMW 6-tuple



Carte 16: Profil choroplèthe d'une carte à corrélations

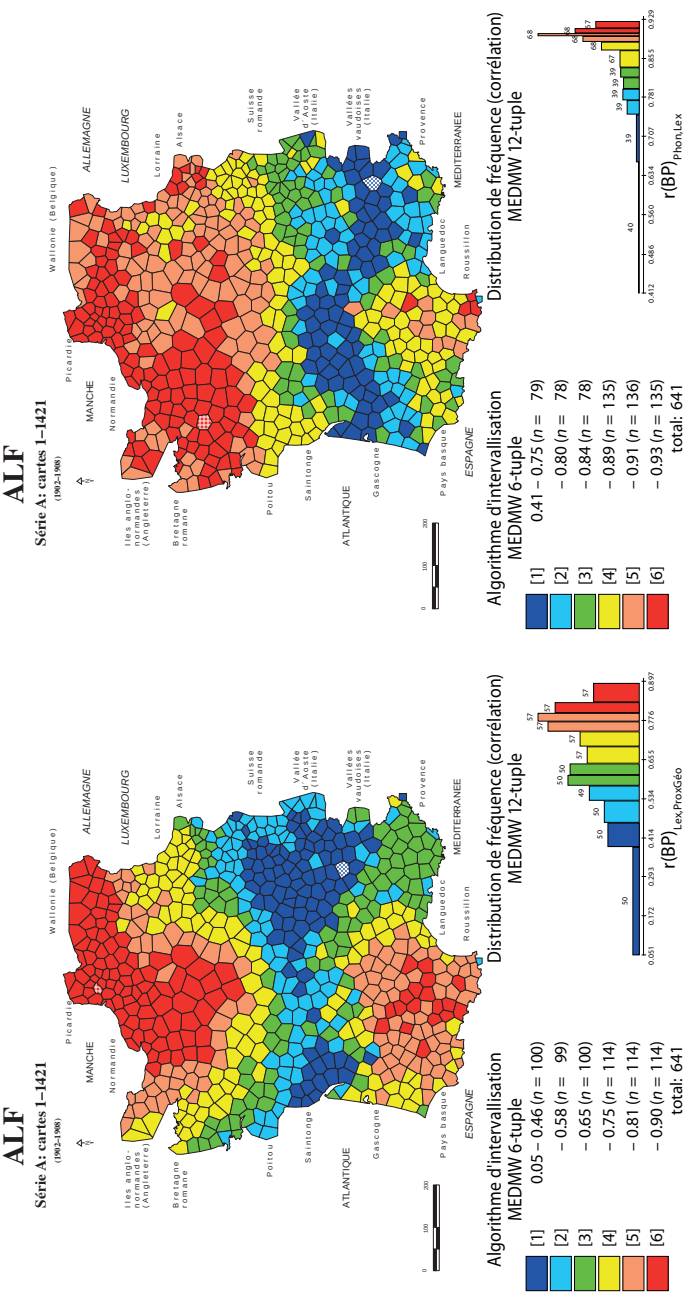
Matrices carrées corrélées:

- matrice de similarité (selon IRI_{jk})

sous-corpus *phonétique* (1117 cartes de travail)

- matrice de proximité (selon le théorème de Pythagore)

Algorithme d'intervallisation: MEDMW 6-tuple

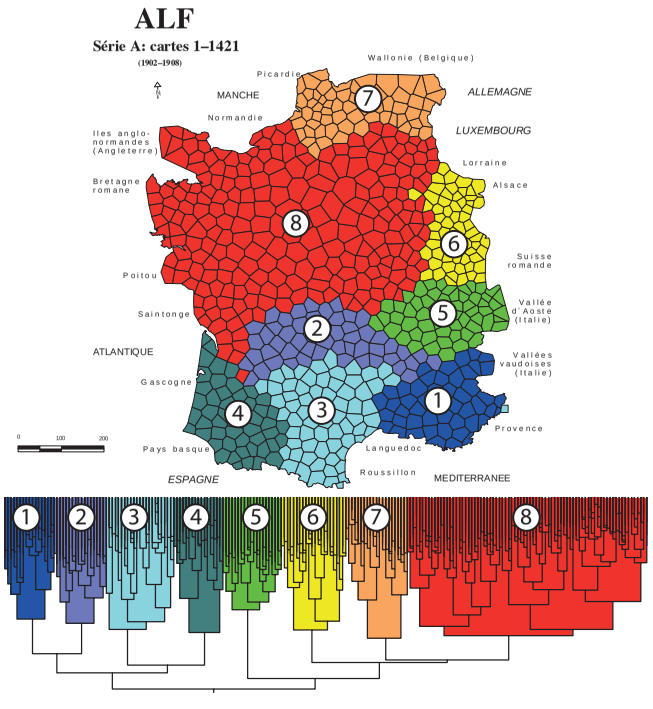


Carte 17: Profil choroplèthe d'une carte à corrélations
Matrices carrées corrélées:
- matrice de similarité (selon IRI_{jk})
  sous-corpus *lexical* (465 cartes de travail)
- matrice de proximité (selon le théorème de Pythagore)

Algorithme d'intervallisation: MEDMW 6-tuple

Carte 18: Profil choroplèthe d'une carte à corrélations
Matrices de similarité (selon IRI_{jk}) corrélées:
- sous-corpus *phonétique* (1117 cartes de travail)
- sous-corpus *lexical* (465 cartes de travail)

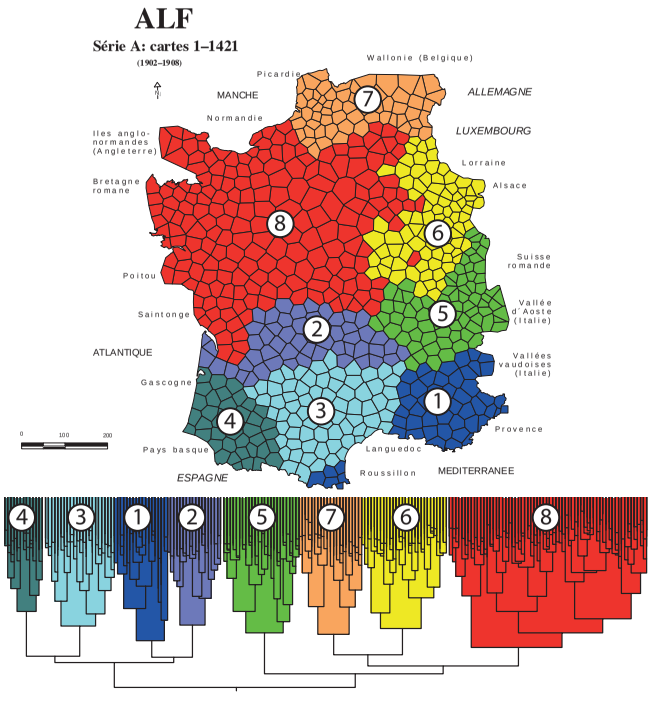
Algorithme d'intervallisation: MEDMW 6-tuple



Carte 19: Analyse dendrographique (selon l'algorithme de Joe Ward Jr.)

- En bas: arbre (classification en 8 dendrèmes)
- En haut: spatialisation de l'arbre (classification en 8 chorèmes)

Indice de similarité: IRI_{jk}
Corpus: sous-corpus *phonétique* (1117 cartes de travail)



Carte 20: Analyse dendrographique (selon l'algorithme de Joe Ward Jr.)

- En bas: arbre (classification en 8 dendrèmes)
- En haut: spatialisation de l'arbre (classification en 8 chorèmes)

Indice de similarité: IRI_{jk}
Corpus: sous-corpus *lexical* (465 cartes de travail)