

Introduction aux problèmes et méthodes de l'«École dialectométrique de Salzbourg» (avec des exemples gallo-, italo- et ibéroromans)¹

Hans Goebel

Universität Salzburg

Hans.Goebel@sbg.ac.at

Mots-clé: Dialectologie romane, atlas linguistiques romans, dialectométrie, outils cartographiques

Sommaire

Le présent article représente un voyage de découverte, richement illustré, dans le domaine de la dialectométrie (DM) telle qu'elle est pratiquée par l'«École dialectométrique de Salzbourg». Après une brève évocation des principes et concepts élémentaires de la DM, l'on passe, toujours à l'aide d'exemples illustratifs tirés d'atlas linguistiques romans bien connus (ALF, AIS, ALPI, ALDC), à la présentation des quatre étapes méthodiques suivantes: *taxation* des cartes originales d'un atlas linguistique suivi par l'établissement de la *matrice de données*, *mesuration de la similarité* entre deux vecteurs de la matrice de données et l'établissement consécutif de la *matrice de similarité*. Par la suite, il est fait mention des dépouillements heuristiques de la matrice de similarité suivants: cartes de similarités, cartes à paramètres, cartes interponctuelles (cartes à cloisons et cartes à rayons), analyse dendrographique et dialectométrie corrélative. La présentation des cartes de similarité est exemplifiée par trois séquences diaporamiques dont les différents clichés peuvent être repérés d'un côté dans le texte et de l'autre dans le web, où ces dernières peuvent être regardées sous leur forme dynamique originale. L'article se clôt par quelques mises au point cartographiques et conceptuelles très utiles en la matière.

1. La dialectométrie: une méthode inductive et exploratoire

Au début des années 70 du XX^e siècle, la dialectométrie (DM) est née sous l'impulsion de la curiosité de quelques géolinguistes² qui voulaient savoir si, derrière l'apparent désordre des structures spatiales des cartes de n'importe quel atlas linguistique, 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 main³ ou à l'aide de l'informatique alors en état de gestation,⁴ de beaucoup de cartes d'atlas. Par une vieille tradition occidentale, une telle procédure – qui part du «particulier» pour arriver au «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, allemands, anglais, hollandais et américains, il est clair que l'ordre sous-jacent («caché» ou «de profondeur») assumé d'entrée de jeu existe réellement et qu'il est l'émanation (ou la création) d'une activité particulière de l'*Homo loquens* soucieux de gérer, par tous les moyens et les niveaux de son langage basilectal, l'espace (ou les réseaux spatiaux) qu'il habite.

¹ Le présent article constitue la version écrite d'une conférence, tenue à l'Université de Lisbonne le 24-11-2011 sous la forme d'une présentation *PowerPoint* librement commentée, contenant une bonne soixantaine de diapositives couleur pourvues souvent d'éléments visuels dynamiques. Vu les modalités électroniques de cette publication, nous en avons conservé la plupart tout en y apportant quelques retouches iconiques simplificatrices et une nouvelle numérotation des figures (de 1 à 66) à laquelle il est fait constamment référence dans les lignes qui suivent.

² Je renvoie, à ce sujet, à deux articles de J. Séguéy (1971 et 1973: avec la mention du néologisme *dialectométrie*) ainsi qu'à mes propres contributions de 1971 (Goebel/Winterleitner) et 1976.

³ C'était le cas pour Jean Séguéy, décédé déjà en 1973.

⁴ J'avais le privilège de pouvoir me servir, dès le début, de l'informatique.

À Salzbourg, il est question, dans ce contexte, de la «gestion basilectale de l’espace par l’*Homo loquens*» que je considère comme une des multiples facultés innées des humains et à laquelle j’attribue le rang d’un «universal» du langage à *plein titre*.

C’est ainsi qu’en face d’un atlas linguistique (AL) à dialectométrer, nous nous sommes toujours posé les questions suivantes: (i) Quelle est la variation géolinguistique globale répertoriée sur toutes les planches de l’AL en question (ici: de l’*Atlas Linguistique de la France* [ALF])? (ii) 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?

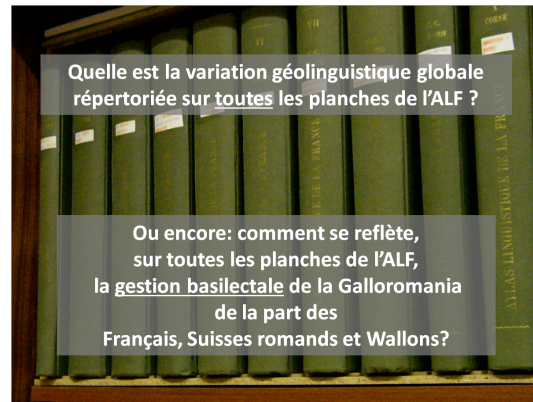


FIGURE 1. Vue d’un exemplaire de réimpression (1968-1971) de l’*Atlas Linguistique de la France* (ALF, 1902-1915) avec deux interrogations de base de l’«École dialectométrique de Salzbourg»

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 l’ALF.⁵ Ce pourcentage est pourtant suffisamment grand pour pouvoir en tirer des conclusions représentatives et, de ce fait, fiables.



FIGURE 2. Réseau d’enquête de l’*Atlas Linguistique de la France* (ALF)

⁵ Pour plus de détails sur la dialectométrisation des données de l’ALF, cf. Goebel (2002, 2003).

L'ALF comprend 638 localités où Edmond Edmont⁶ a mené 639 enquêtes. Évidemment, la pratique de la géographie linguistique – soit sous sa forme traditionnelle soit sous celle de la DM – ne peut se faire sans un recours massif à la cartographie et à la visualisation de faits géographiquement différenciés.

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 de nos jours indispensable de créer, pour le réseau de n'importe quel AL, un support cartographique capable de recevoir des informations spatiales tant qualitatives que quantitatives. Un tel support présuppose l'établissement d'un compartimentage polygonal, dit aussi «tessellation de Voronoï», dont les figures 3 et 4 montrent la genèse.

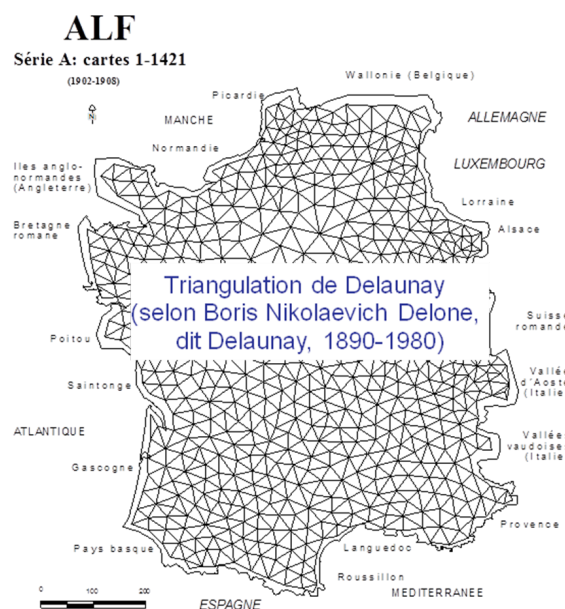


FIGURE 3. Triangulation du réseau de l'Atlas Linguistique de la France (ALF)

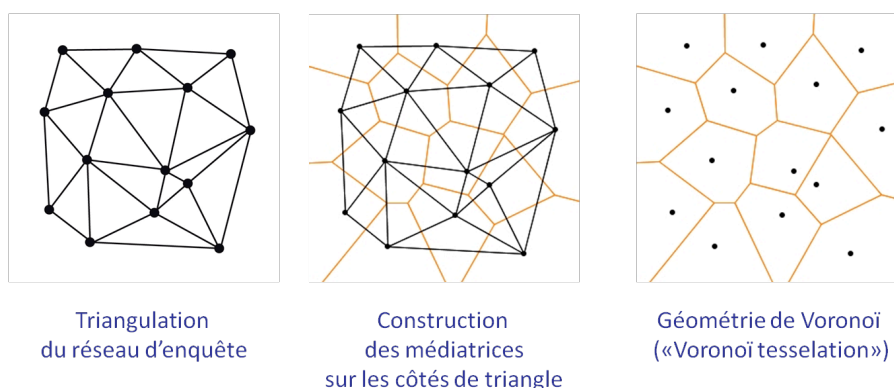


FIGURE 4. Les trois étapes de la construction de polygones de Voronoï

⁶ Rappelons qu'Edmond Edmont (1849-1926) était l'enquêteur de l'ALF.

La polygonisation passe à travers la triangulation du réseau des points d'enquête et la 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.

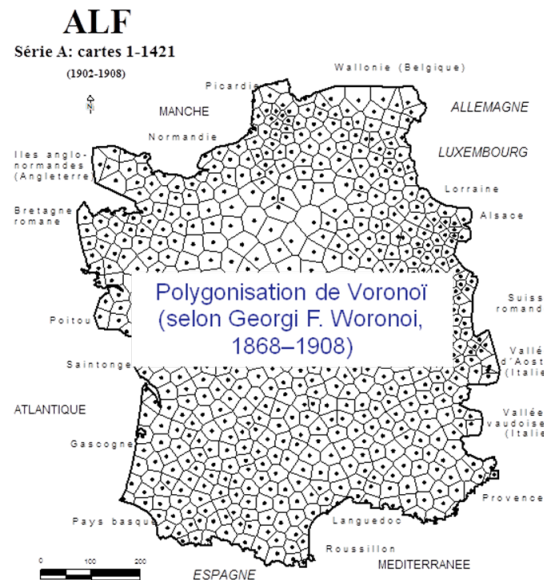


FIGURE 5. Réseau polygonisé de l'Atlas Linguistique de la France (ALF)

2. Brève présentation de la «chaîne dialectométrique»

Toute analyse-DM effectuée «à la salzbourgeoise»⁷ consiste en un certain nombre d'étapes alignées en forme de chaîne, à la fin de laquelle se situe toute une gamme de visualisations quantitatives à couleurs qui correspondent d'un côté aux vœux du géolinguiste-dialectométricien et assurent de l'autre la nature exploratoire de la démarche entière.

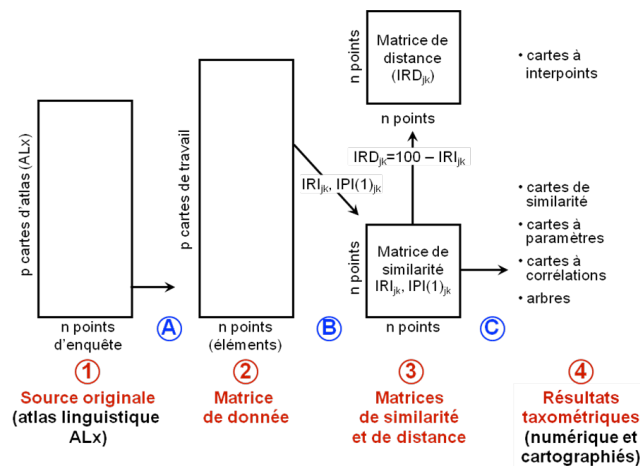


FIGURE 6. Schéma de la «chaîne dialectométrique» utilisée à Salzbourg

⁷ Cf. à ce sujet avant tout notre thèse d'habilitation de 1984.

À l'intérieur de la chaîne dialectométrique, il existe trois moments cruciaux (A, B, C) où le linguiste-dialectométricien doit intervenir avec toute sa compétence d'expert. Le moment A correspond à la transformation d'une partie bien définie de l'information contenue en une carte originale d'un AL quelconque dans une carte dite «de travail» et son insertion dans le schéma d'une matrice de données.⁸ Cette transformation – appelée à Salzbourg «taxation» – n'est absolument pas nouvelle. Elle entre pleinement dans une longue tradition d'activités similaires au cours desquelles beaucoup de linguistes (en commençant par Jules Gilliéron⁹ même) ont typisé le contenu de planches originales de l'ALF ou d'un autre AL roman. Ils y ont appliqué certains critères classificatoires judicieusement choisis au préalable tout en en confiant le résultat, avec plus ou moins de dextérité visuelle ou graphique, à une des cartes muettes mises à disposition par l'imprimeur de l'ALF.

La matrice de données une fois établie, l'étape suivante (moment B) prévoit la mesure de la similarité réciproque entre les N points d'enquête de l'AL en question. Le choix d'un indice de similarité approprié est délicat: d'un côté, il doit être fait suivant les conceptions comparatistes du linguiste et, de l'autre, selon les possibilités métrologiques qu'offre la matrice de données respective et selon les contraintes cartographiques des visualisations envisagées.

L'indice de similarité standard de la DM de Salzbourg est l'«Indice Relatif d'Identité» (IRI) auquel fait écho l'«Indice Relatif de Distance» (IRD) selon la formule $IRI + IRD = 100$. Dans la panoplie salzbourgeoise des indices de similarité, il y a, entre autres, aussi l'indice pondéré IPI (l'«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») alors que les méthodes pour la confection et la visualisation appropriée des cartes à couleurs sont fournies par la cartographie numérisée, 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.

Le moment C ouvre l'accès aux différentes espèces de visualisation que nous avons définies et mises au point uniquement suivant les nécessités de la géolinguistique traditionnelle et nullement en donnant libre cours à la richesse exubérante des méthodes numériques et visualisateurs existants.

3. La taxation des cartes d'un atlas linguistique

Les six exemples que nous allons présenter (et commenter) par la suite ont été empruntés aux atlas romans «nationaux» ALF, AIS et ALPI. Il s'agit toujours de deux espèces de taxations – d'abord *lexicale* et ensuite *phonétique* – qui ont été faites suivant des méthodes lexicologiques (ou onomasiologiques) et phonétiques «classiques», c'est-à-dire connues depuis longtemps. On trouve beaucoup d'échantillons de telles analyses dans la littérature des Études romanes des derniers cent ans.¹⁰

⁸ Cf. à ce sujet les précisions fournies dans nos contributions de 1981, 2002 et 2003.

⁹ 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), avec toute une gamme de typisations colorées de différentes planches de l'ALF.

¹⁰ Le livre de Karl Jaberg de 1908 est au début de cette longue tradition qui s'étend jusqu'à l'excellent ouvrage de Brun-Trigaud/Le Berre/Le Dû de 2005.

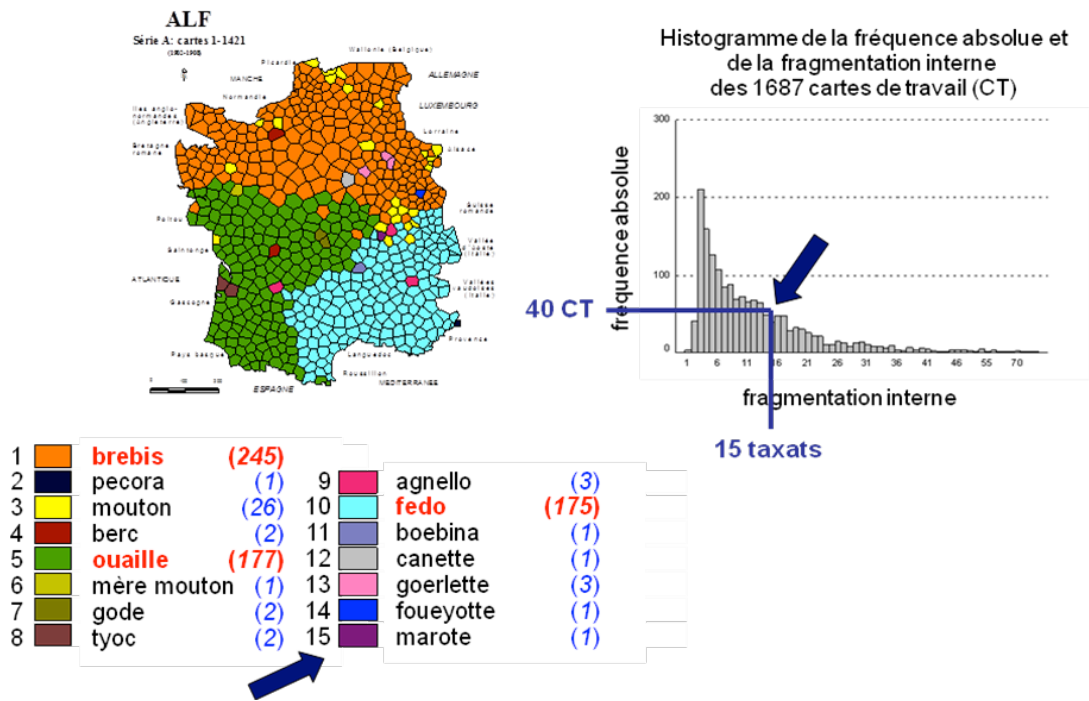


FIGURE 7. Carte de similarité de type lexical issue de la taxation de la carte 173 «la brebis» de l'Atlas Linguistique de la France (ALF)

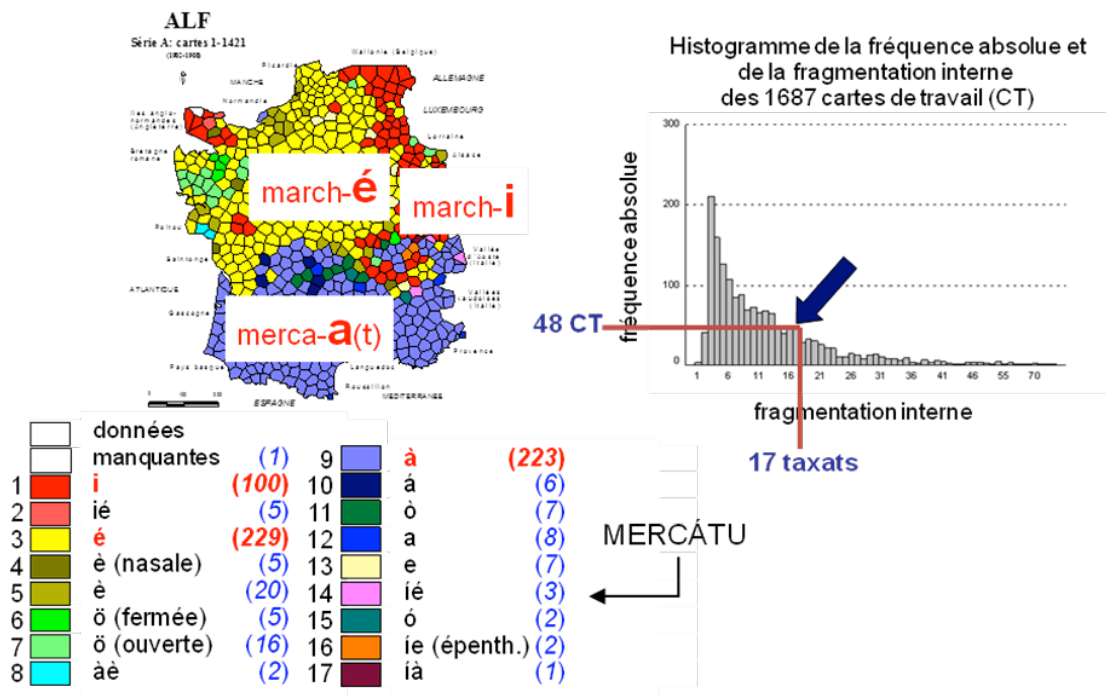


FIGURE 8. Carte de similarité de type phonétique issue de la taxation de la carte 812 «le marché» de l'Atlas Linguistique de la France (ALF)

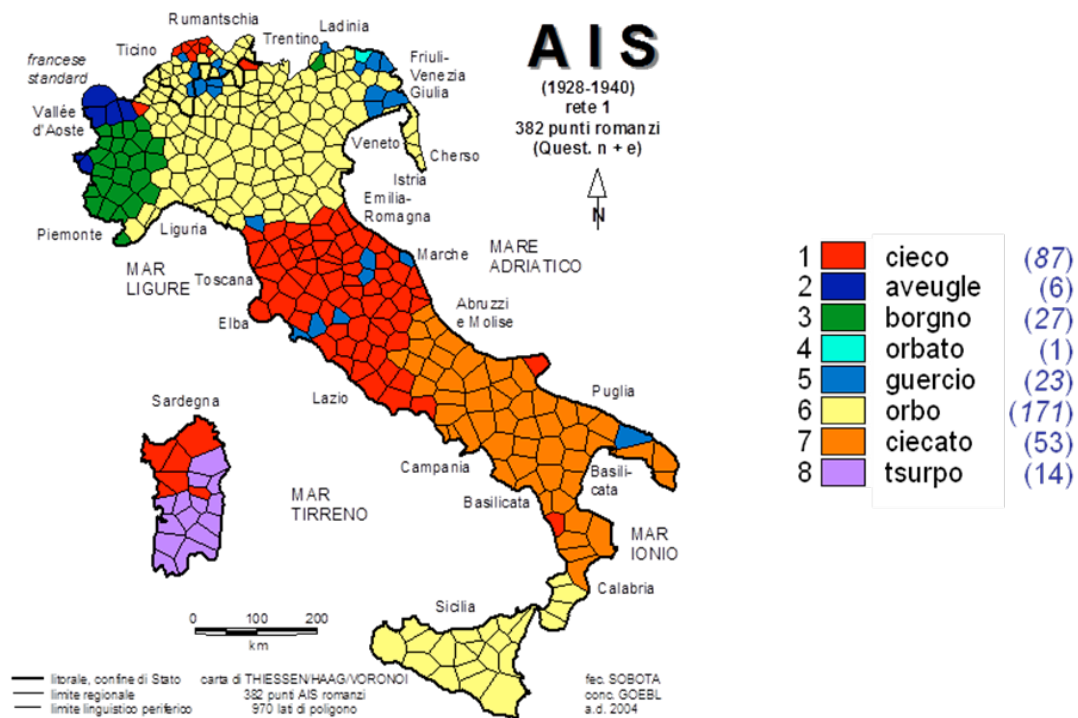


FIGURE 9. Carte de similarité de type lexical
issue de la taxation de la carte 188 «cieco» du *Sprach- und Sachatlas Italiens und der Südschweiz* (AIS)

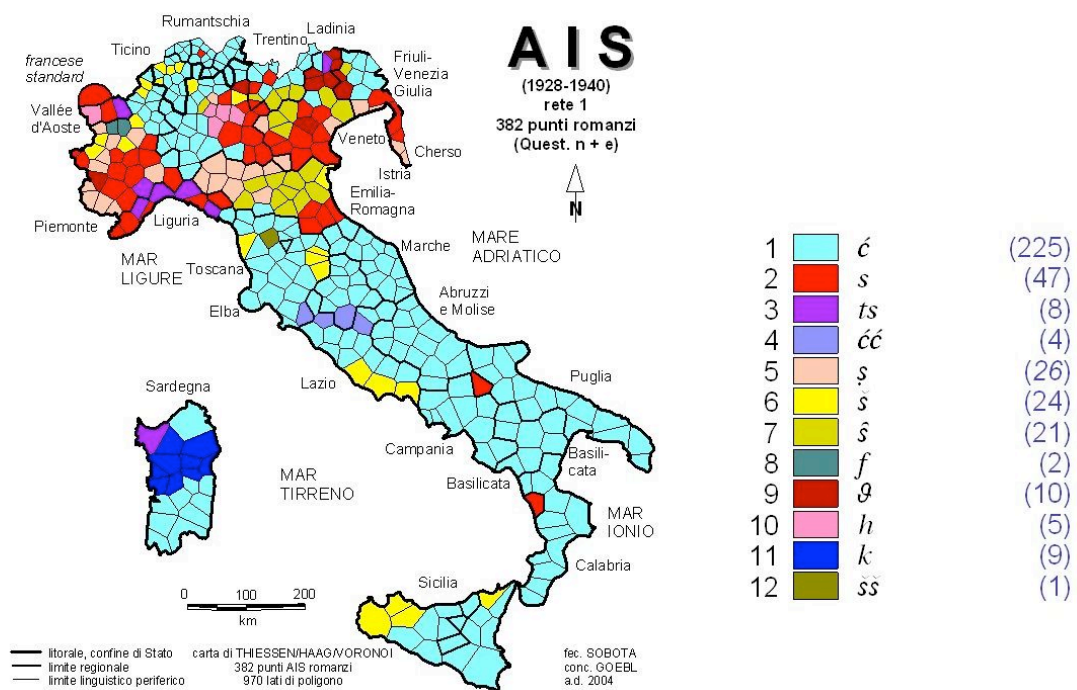
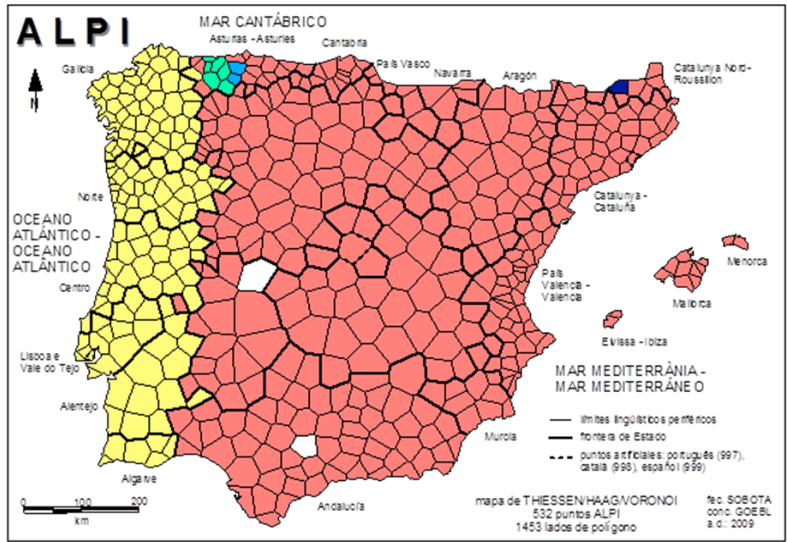
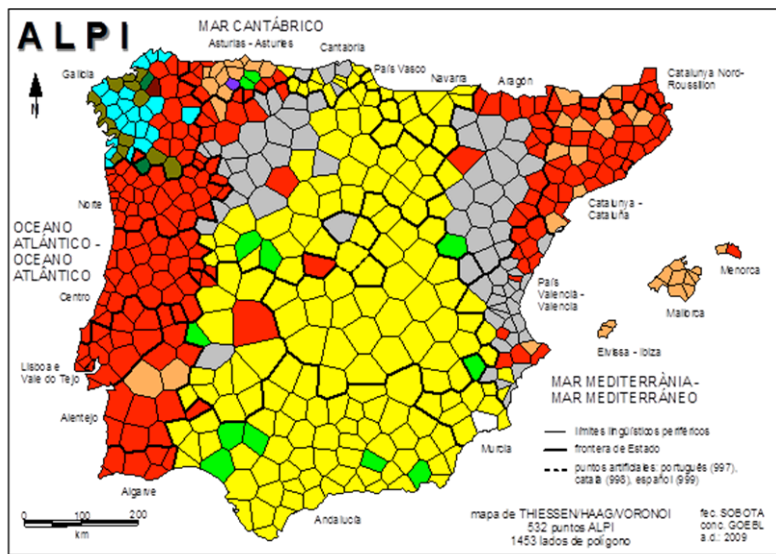


FIGURE 10. Carte de similarité de type phonétique
issue de la taxation de la carte 304 «cento» du *Sprach- und Sachatlas Italiens und der Südschweiz* (AIS)



 0: données manquantes [2]	 3: ANOITE [6]
 1: AYER [374]	 4: AMNESI [2]
 2: ONTEM [147]	 5: ZEZIA [1]

FIGURE 11. Carte de similarité de type lexical
issue de la taxation de la carte 20 «ayer» de l'Atlas Lingüístico de la Península Ibérica (ALPI)



 0: données manquantes [1]	 4: g [38]	 8: b [12]
 1: Zéro [70]	 5: ĝ [2]	 9: x [14]
 2: b [173]	 6: ħ [1]	 10: x̄ [21]
 3: g [199]	 7: ġ [1]	

FIGURE 12. Carte de similarité de type phonétique
issue de la taxation de la carte 12 «aguja» de l'Atlas Lingüístico de la Península Ibérica (ALPI)

Chaque carte de travail comprend un certain nombre de taxats et d'aires taxatoires y ayant trait. Pour l'ALF, cette variabilité se situe entre 2 et 91, pour l' AIS entre 2 et 153 et pour l'ALPI entre 2 et 105 taxats et aires taxatoires par carte de travail.

Il y a déjà quelque quarante ans que j'ai constaté qu'il existe, dans une grande masse de cartes de travail dûment taxées, une relation très curieuse entre la diversification interne et le nombre absolu des cartes de travail: voir les histogrammes à l'intérieur des figures 7 et 8. L'on y trouve une courbe exponentielle telle que j'en ai trouvé beaucoup au cours de mes recherches-DM. Le principe de base est que plus la fragmentation interne des cartes de travail est grande, moins grande se révèle leur fréquence absolue.¹¹

Au vu des figures 7-8, 9-10 et 11-12, il appert en outre que la subdivision en aires taxatoires est très «primesautière», et que ces dernières peuvent revêtir, mis à part leur taille définie par le nombre des polygones respectifs, des configurations spatiales très diversifiées, allant d'une cohérence (ou continuité) parfaite à une dislocation plus ou moins grande (ou dérégulée) dans l'espace. Toujours est-il que l'analyse et la considération d'aires représente, dans l'approche-DM de Salzbourg, la «philosophie de base». Une fois de plus, je tiens à souligner qu'avec cette orientation nous nous plaçons dans la lignée de la géolinguistique (romane) classique où la considération (et la critique) des aires – appelée aussi «aréologie» – a joué un rôle primordial.

Voici donc le bilan taxatoire et aréologique de la DM-isation des atlas ALF,¹² AIS,¹³ ALPI et ALDC:

Atlas linguistique	Σ taxats/aires taxatoires	Σ cartes de travail
ALF	20 043	1 687
AIS	43 546	3 911
ALPI	3 284	375
ALDC	15 344	1 659

Ajoutons encore que la nature spontanée des aires est la raison principale pour laquelle la coïncidence parfaite d'isoglosses est tellement rare. Mais elle explique aussi que l'enchevêtrement (ou aussi: imbrication, tuilage) d'aires de configuration très diverse constitue le principe de base de la genèse de toutes les unités dialectales à proprement parler,¹⁴ qu'elles soient grandes ou petites.

4. La mesure de similarité entre deux vecteurs de la matrice de données

Admettons qu'il s'agisse de mesurer, selon la logique algébrique de l'IRI («Indice Relatif d'Identité»), la similarité quantitative entre les points d'enquête 3 et 4 tels qu'ils figurent dans la figure 13:

¹¹ Cf. à ce sujet Altmann (1985: *passim*).

¹² Voir aussi Goebel (2002, 2003).

¹³ Voir aussi Goebel (2008).

¹⁴ Cf. le travail séminal de Karl von Ettmayer de 1924.

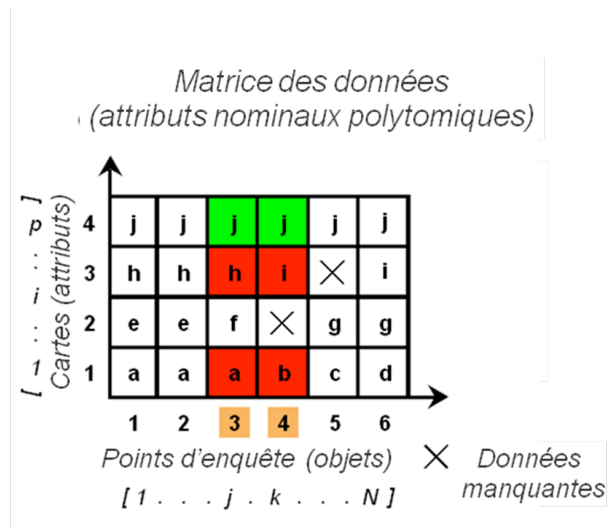


FIGURE 13. Matrice des données

L'IRI considère la fréquence absolue de «co-identités» [COI] (= co-présence de deux taxats identiques dans la même ligne de la matrice de données) et de «co-différences» [COD] au même endroit, tout en en calculant la fréquence relative en % – cf. fonctionnement algébrique de la formule de l'IRI_{jk}:

$$IRI_{3,4} = 100 \cdot \frac{\Sigma(\text{COI})}{\Sigma(\text{COI} + \text{COD})} = \frac{1}{1 + 2} = 33\%$$

Les pourcentages ainsi déterminés sont stockés par la suite dans les cellules d'une matrice de similarité qui, outre d'être carrée, est aussi parfaitement symétrique. La valeur de 33% calculée pour les localités factices 3 et 4 se retrouve donc à deux endroits de la matrice de similarité (voir les deux cercles rouges dans la figure 14):

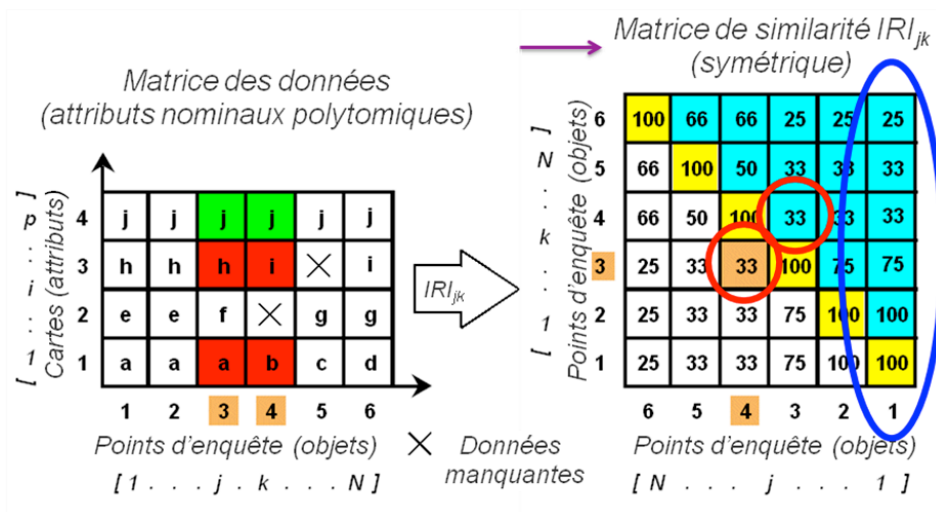


FIGURE 14. Schéma de la mensuration de la similarité linguistique entre les différents points d'un atlas linguistique

Une fois la matrice de similarité établie, il faut que le dialectométricien-classificateur se mette à la dépouiller judicieusement: voir à cela le moment C sur la figure 6. En plus, c'est le premier des nombreux pas exploratoires qui mènent des structures «de surface» aux «structures de profondeur». C'est donc le début de notre tâche exploratoire.

Le premier pas «vers l'inconnu» sera fait à l'aide de l'instrument heuristique de la «carte de similarité» (cf. figures 15-21):

1^{ère} structure de profondeur, niveau A:
carte de similarité tirée des données dialectométrisées de l'ALF

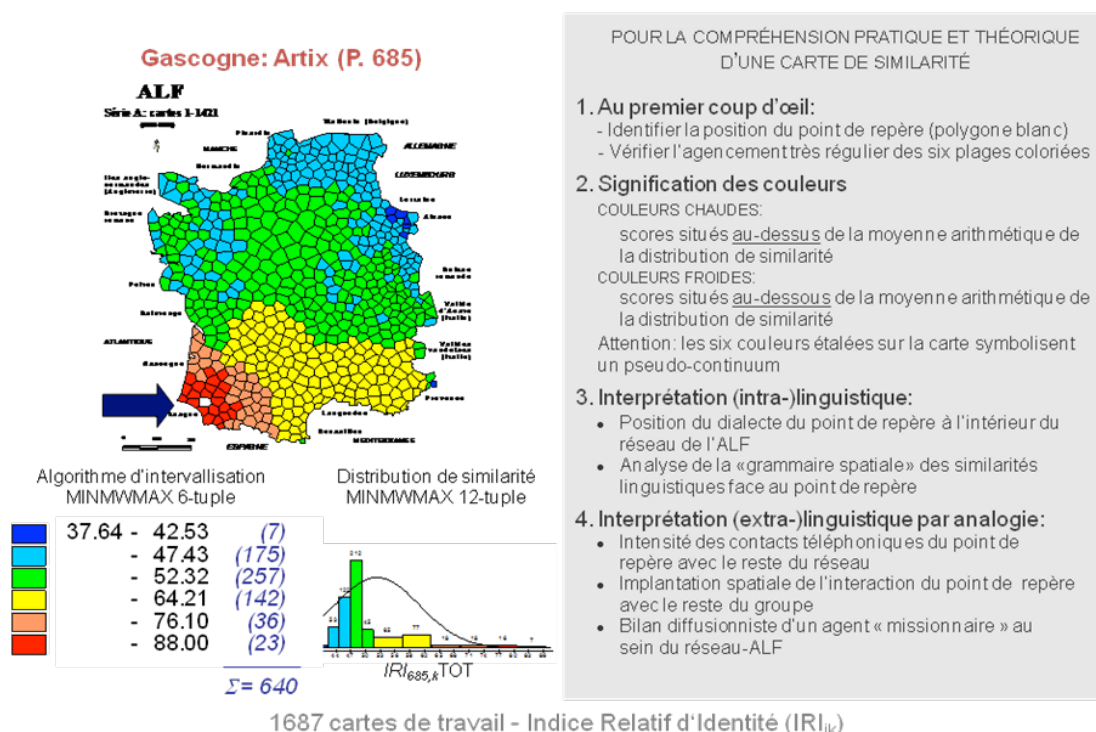


FIGURE 15. Présentation d'une carte de similarité de type *gascon* tirée de l'*Atlas Linguistique de la France* (ALF)

Le défi heuristique à relever en la matière est double:¹⁵

- du point de vue statistique: il s'agit d'extraire un des N vecteurs de la matrice de similarité, 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 valeurs de similarité extraites 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 moyenne

¹⁵Voir à cela Goebel (1981, 1984: vol. I, *passim*) et la figure 14 (ellipse bleue à la marge de droite).

arithmétique de la distribution de similarité respective, tout en définissant des intervalles numériquement toujours égaux des deux côtés de la moyenne arithmétique.¹⁶

Il en résulte des profils iconiques remarquablement stables et bien ordonnés auxquels le logiciel-DM utilisé par nous («Visual DialectoMetry» [VDM]) permet de conférer des silhouettes plus ou moins fines, allant de 2 à 18 paliers chromatiques. Les couleurs utilisées par nous 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.

Les figures 15-21 constituent sept spécimens de cartes de similarité relatives aux atlas ALF, AIS, ALPI et ALDC.

Les figures 15 et 16 montrent trois des 641 cartes de similarité calculées à partir de notre corpus intégral tiré de l'ALF (contenant 1687 cartes de travail).

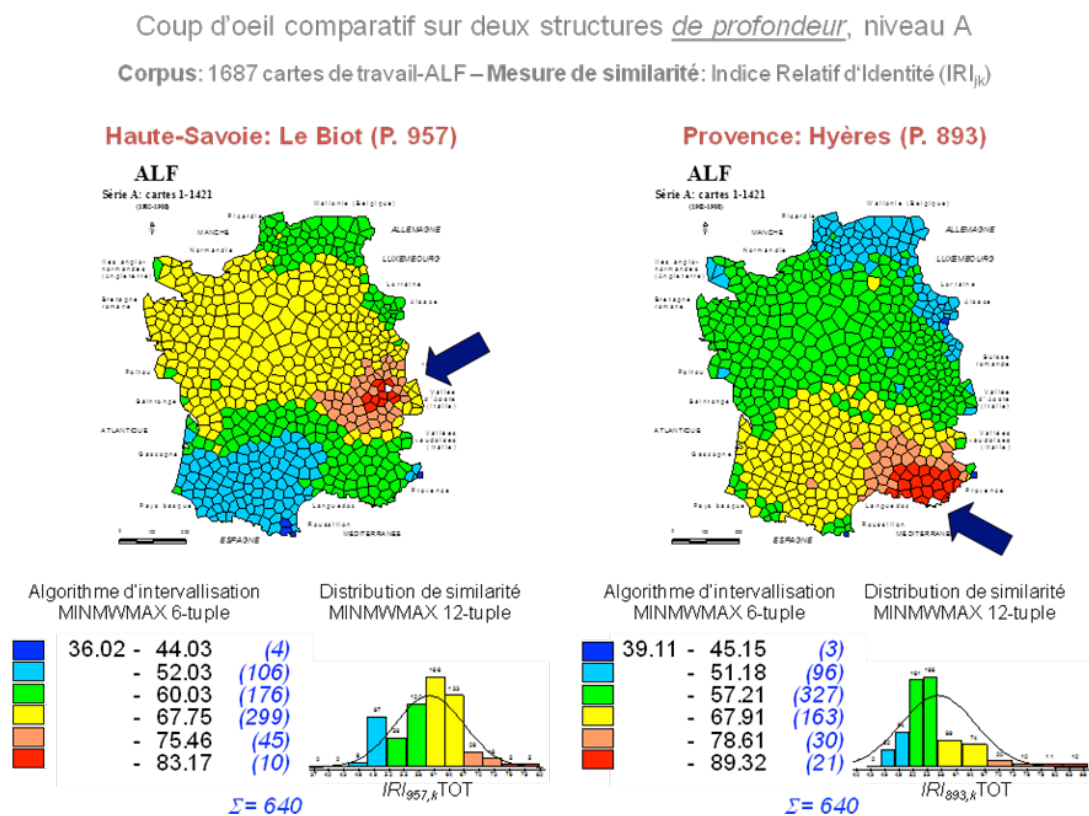


FIGURE 16. Deux cartes de similarité (de type *francoprovençal* et *provençal*) tirées de l'Atlas Linguistique de la France (ALF)

Chaque carte de similarité dispose d'un point de référence dont le polygone est toujours en blanc. Elle indique, par la structuration de son profil choroplèthe, le décalage spatial des similarités linguistiques des N-1 points d'atlas restants par rapport au point de référence préétabli. Une des grandes découvertes de la DM de Salzbourg est l'étonnante régularité avec laquelle ce décalage s'inscrit dans l'espace. Et ceci malgré l'indéniable «chaotisme» qui

¹⁶Voir à ce sujet surtout Goebel (1981, 1983, 1984).

se dégage de la considération particularisante de plus de 20 000 aires taxatoires extraites de l'ALF.¹⁷

En ligne générale et dans les conditions de visualisation déjà mentionnées, l'entourage immédiat du point de référence prend les couleurs rouge et orange qui constituent toujours une aire (quantitative) cohérente. Par contre, les couleurs bleu foncé et bleu clair 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 et vert représentent des zones-tampon (ou de transition) entre – métaphoriquement parlant – les «meilleurs amis» d'un côté et les «pires ennemis» (dialectaux) du basilecte du point de référence de l'autre.

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 figures 15 et 16 montrent des profils géotypologiques pour les domaines du gascon, du francoprovençal et du provençal, alors que la figure 17 exhibe des profils de similarité typiques des domaines piémontais et lombard.

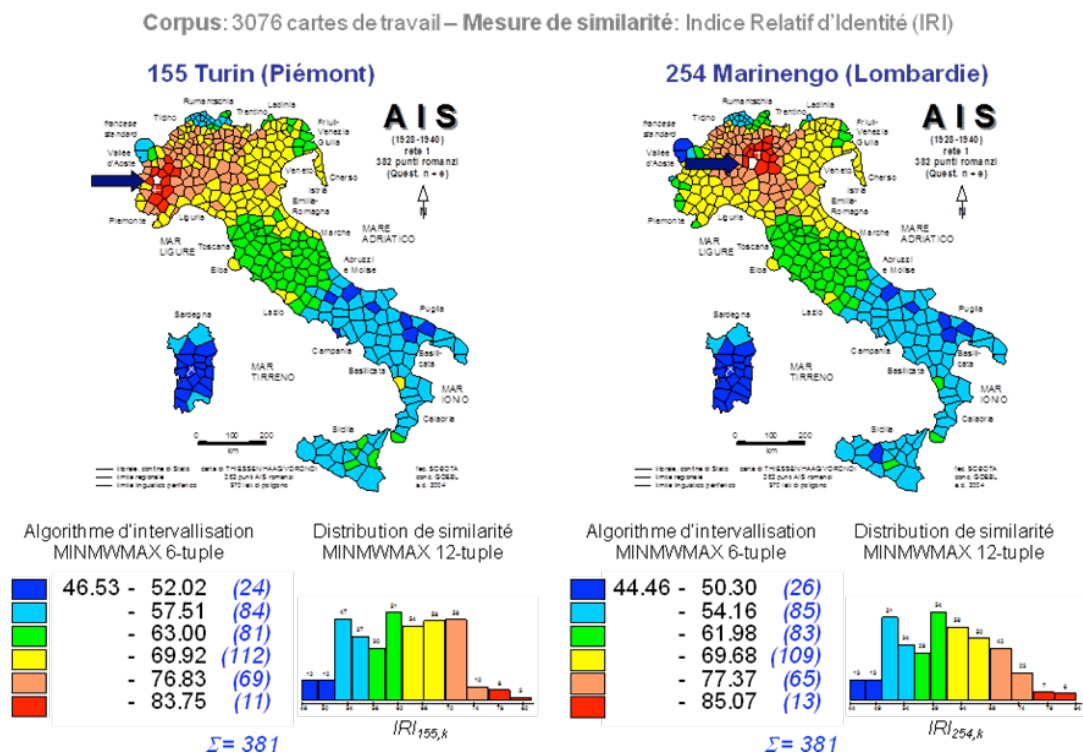


FIGURE 17. Deux cartes de similarité (de type piémontais et lombard) tirées du *Sprach- und Sachatlas Italiens und der Südschweiz* (AIS)

Le rendement heuristique des figures 18 et 19, en revanche, s'en écarte un peu. Elles permettent de démontrer la variabilité de l'insertion du dialecte alpine-occitan de l'enclave linguistique de Guardia Piemontese (P.-AIS 760) dans son entourage calabrais suivant la nature catégorielle des données analysées.¹⁸

¹⁷ Cf. Goebel (2003: 65).

¹⁸ Voir aussi les explications et visualisations présentées à ce sujet dans Goebel (2008: 61-62 et 108-110).

La position relationnelle d'un îlot linguistique occitan: Guardia Piemontese

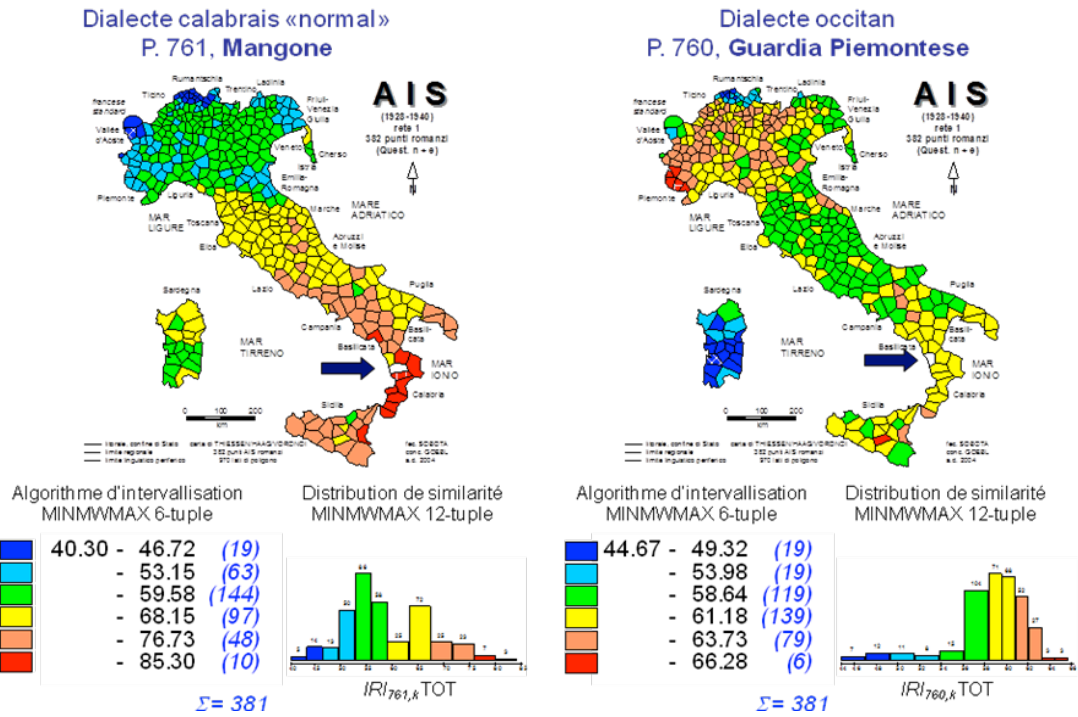


FIGURE 18. Deux cartes de similarité (de type *calabrais* et *alpine-occitan*) tirées de l' AIS

L'intégration inégale du dialecte de Guardia Piemontese dans son environnement: **phonétique** versus **lexique**

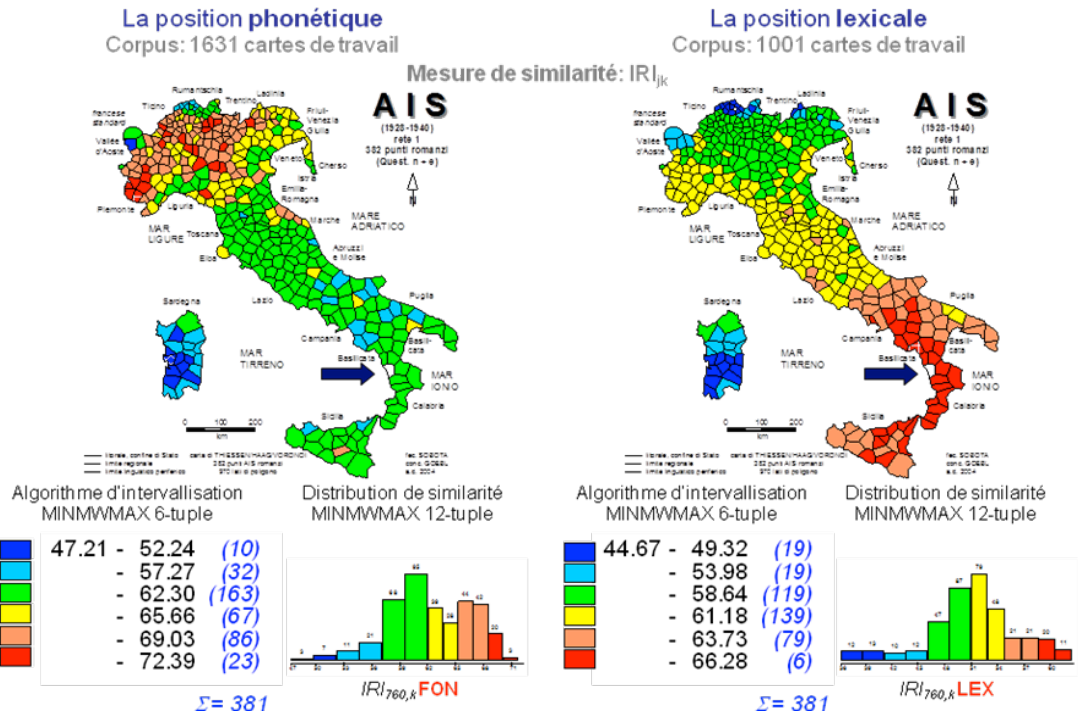


FIGURE 19. Deux cartes de similarité (de type *alpine-occitan*) tirées de l' AIS

Rappelons que la fondation de cet îlot linguistique remonte au XIV^e siècle et que, partant, le contact linguistique de ce petit groupe de Vaudois d'origine piémontaise avec son entourage dure depuis quelque sept siècles. Alors que le profil de droite de la figure 18 montre l'insertion globale du dialecte de Guardia Piemontese dans le réseau de l' AIS, les deux profils de la figure 19 fournissent des informations plus circonstanciées.

Le profil de gauche – relatif à la seule *phonétique* et basé sur quelque 1600 cartes de travail tirées de l' AIS – prouve très clairement que, de ce point de vue, les locuteurs de Guardia Piemontese ont bien conservé le caractère «nordique» (ou «alpin») de leur parler d'origine. Le profil de droite en revanche – relatif au *lexique* et basé sur 1001 cartes de travail lexicales – révèle que les mêmes locuteurs se sont largement adaptés à leur entourage calabrais du point de vue lexical.

Signalons en outre que la valeur de similarité maximale du profil de gauche de la figure 19 apparaît au point-AIS 181, Valdieri (Piémont), qui se trouve à seulement 50 km de distance de la Val Pellice d'où sont partis les colons vaudois aux temps du Dante.

La figure 20, relative à l'ALPI et, de ce fait, à une analyse-DM basée sur un nombre très restreint de cartes de travail (375), exhibe un profil de similarité nord-castillan et son homologue est-catalan. Dans les deux cas, la précision avec laquelle les domaines linguistiques du castillan et du catalan se dégagent est tout à fait remarquable.

Corpus: 375 cartes de travail – Mesure de similarité: IRI_{jk}

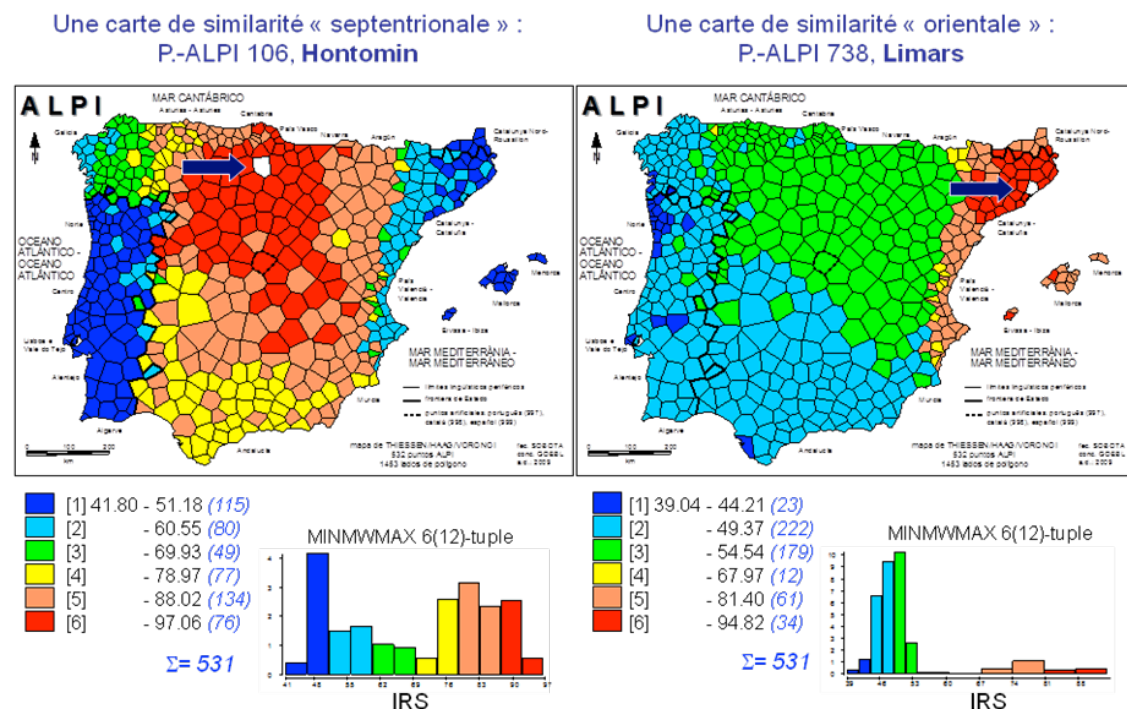
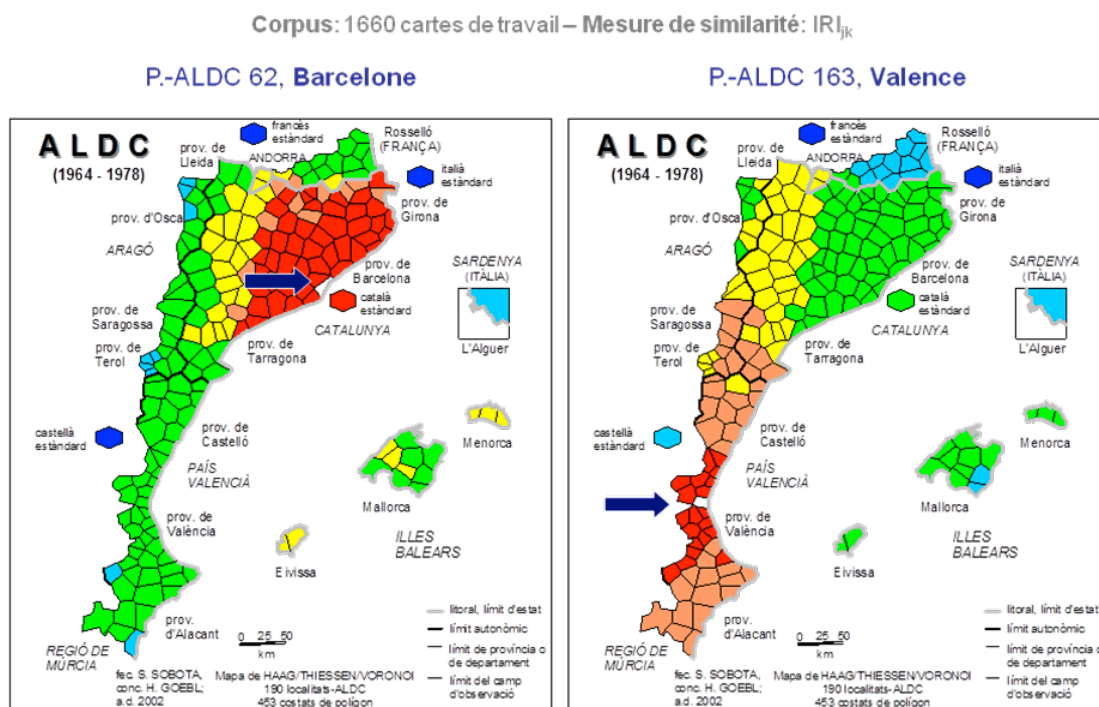


FIGURE 20. Deux cartes de similarité (de type nord-castillan et est-catalan) tirées de l'Atlas Lingüístico de la Península Ibérica (ALPI)

Non moins remarquable est la prégnance des deux profils de similarité relatifs au domaine catalan:



Ajoutons encore un détail relatif à la seule figure 21. Le réseau de l'atlas catalan ALDC comprend 190 points d'enquête. En dialectométrisant les quatre premiers volumes de cet atlas nous avons ajouté aux 190 localités-ALDC mentionnées quatre localités «factices» correspondant aux langues standards du catalan, du castillan, du français et de l'italien. Cette mesure permet d'étudier, à l'aide du réseau de l'ALDC, l'infiltration de deux langues standards ibériques dans les dialectes catalans d'Espagne et, de l'autre, de faire la même chose avec les dialectes catalans parlés en France (dans le Roussillon) et en Italie (à Alghero/L'Alguer). Des localités factices similaires ont d'ailleurs été ajoutées aussi aux réseaux de l'ALF (français et italien),¹⁹ de l'AIS (italien et français)²⁰ et de l'ALPI (portugais, castillan et catalan).

5. Au delà de la carte de similarité: les cartes à paramètres

Par «paramètre» l'on entend, en matière de statistique descriptive, une des nombreuses valeurs qui caractérisent d'une quelconque distribution de fréquence, tel que le minimum, la maximum, la moyenne arithmétique, l'écart-type, la symétrie, etc. En considérant les histogrammes au bas des figures 15-20 et surtout au bas de la figure 22, l'on constate avant tout que la symétrie des distributions de similarité peut être très variable.

¹⁹ Cf. Goebel (2002, 2003).

²⁰ Cf. Goebel (1981, 1984, 2008).

Pour gagner de profondeur analytique:
deux cartes de similarité différentes (encore: deux structures *de profondeur*, niveau A)

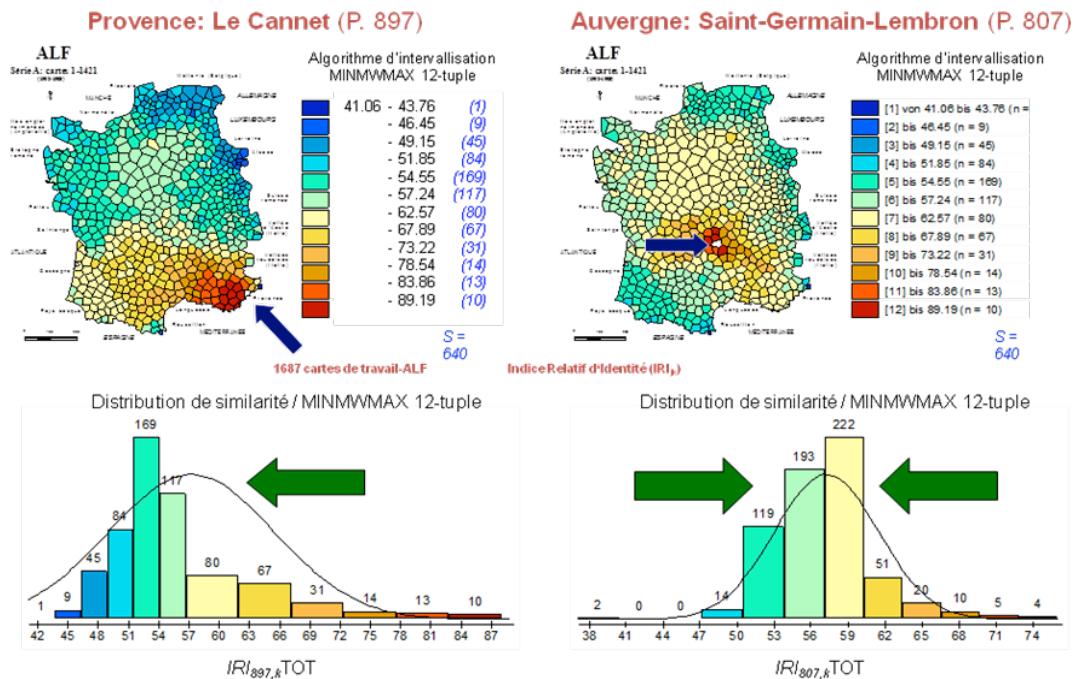


FIGURE 22. Présentation – cartographiquement détaillée – de deux cartes de similarité (de type *provençal* et *auvergnat*) tirées de l'Atlas Linguistique de la France (ALF)

À y regarder de près, cette propriété statistique revêt aussi une grande importance linguistique. Sur la figure 22, l'histogramme de gauche, relatif à un parler provençal maritime, est fortement biaisé vers la gauche. Ceci signifie que des 640 scores de similarité 425 se trouvent au-dessous de la moyenne arithmétique et seulement 215 au-dessus du même seuil.

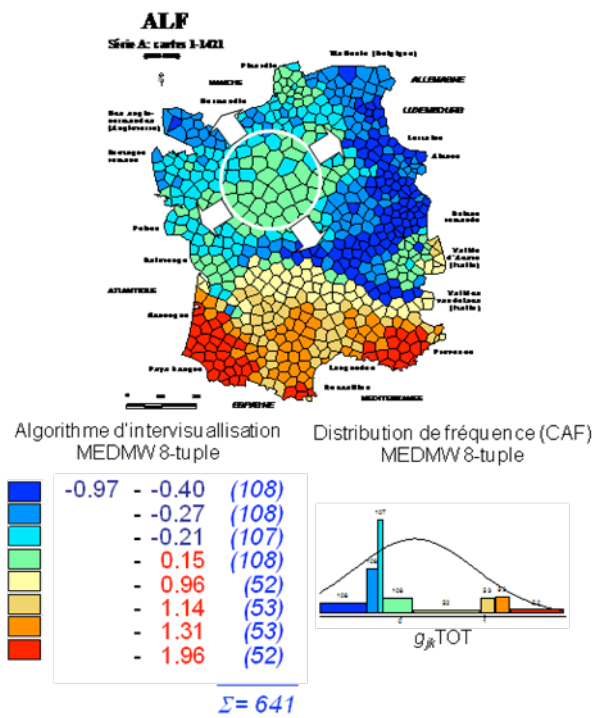
Le bilan analogue pour l'histogramme de droite se présente comme suit: valeurs situées au-dessous de la moyenne arithmétique – 328; valeurs situées au-dessus de la moyenne arithmétique – 312.

Quand on admet qu'une carte de similarité reflète – au moins en théorie – la position communicative du dialecte de la localité donnée, les valeurs mentionnées ci-dessus suggèrent que la position communicative du dialecte du point-ALF 897 (situé en Provence) est moins bonne que celle du point-ALF 807 situé en plein «Croissant» 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 la figure 23) dont la signification 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 et des autres AL examinés.

Structure *de profondeur*, niveau B:
 Carte à paramètres: synopsis des scores du «coefficient d'asymétrie de Fisher» (CAF)
 de 641 distributions de similarité



POUR LA COMPRÉHENSION PRATIQUE ET THÉORIQUE DU
 COEFFICIENT D'ASYMÉTRIE DE FISHER

1. Coefficient d'asymétrie de Fisher (CAF):
 (angl. *skewness*, all. *Schiefe*)
 une excellente mesure pour la saisie statistique de
 l'(a)symétrie d'une distribution de fréquence

2. Signification de couleurs:
 POLYGONES BLEUS (= scores *négatifs* du CAF):
 dialecticités très bien connectées avec le reste du
 réseau: compromis linguistique très intense; zones
 riches en contacts linguistiques micro- et macro-
 régionaux; zones de passage, d'interaction, de
 friction et aussi d'expansion
 POLYGONES ROUGES (= scores *positifs* du CAF):
 dialecticités très mal connectées avec le reste du
 réseau: compromis linguistique inexistant; zones
 dépourvues de contacts linguistiques micro- et
 macro-régionaux, zones résiduelles, de
 «résistance»

3. Intérêt linguistique:
 La synopsis des scores du CAF permet l'étude et
 l'analyse de la répartition spatiale du compromis
 linguistique (all. *Sprachausgleich*), c'est-à-dire de la
 plus ou moins grande intégration des lococlectes
 dans l'ensemble d'un réseau d'atlas donné; grand
 intérêt diachronique (pour l'explication de la genèse
 de l'enchevêtrement actuel des différents domaines
 linguistiques de la Galloromania); cas-modèle d'une
 structure géolinguistique «de profondeur».

Corpus: 1687 cartes de travail-ALF
 Mesure de similarité: Indice Relatif d'Identité (IRI_{ik})

FIGURE 23. Synopsis des scores du «coefficient d'asymétrie de Fisher» (CAF)
 de 641 cartes de similarité tirées de l'Atlas Linguistique de la France (ALF)

Sur la figure 23, les polygones en bleu foncé renvoient à des dialecticités bien connectées avec le reste du réseau, alors que les polygones rouges se réfèrent à des variétés isolationnistes, conservatrices et peu enclines à s'ouvrir à des interactions et échanges linguistiques. 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'un dynamisme spatial de grande envergure, les polygones rouges constituent des agglomérations ponctuelles suggérant l'idée de «réduits assiégés» ou en passe de disparition.

En effet, nous sommes en présence d'une analyse-DM très efficace dont l'utilité est avant tout diachronique. C'est ainsi que le grand agencement circulaire bleu dans le nord renvoie à l'action d'irradiation continue des dialectes de l'Ile-de-France alors que la pince entrouverte bleue autour du domaine francoprovençal renvoie aux contacts et conflits linguistiques que cet idiome entretenait tant vers le nord que vers le sud.

La figure 24 dispose également d'une structuration très claire et bien définie. Les bandes en bleu foncé qui sillonnent la plaine padane et l'Italie médiane se réfèrent à autant de zones d'échange, d'interaction et de contact alors que les polygones en rouge et orange

caractérisent des contrées où, dans le passé comme dans le présent, la stabilité linguistique était plutôt grande.

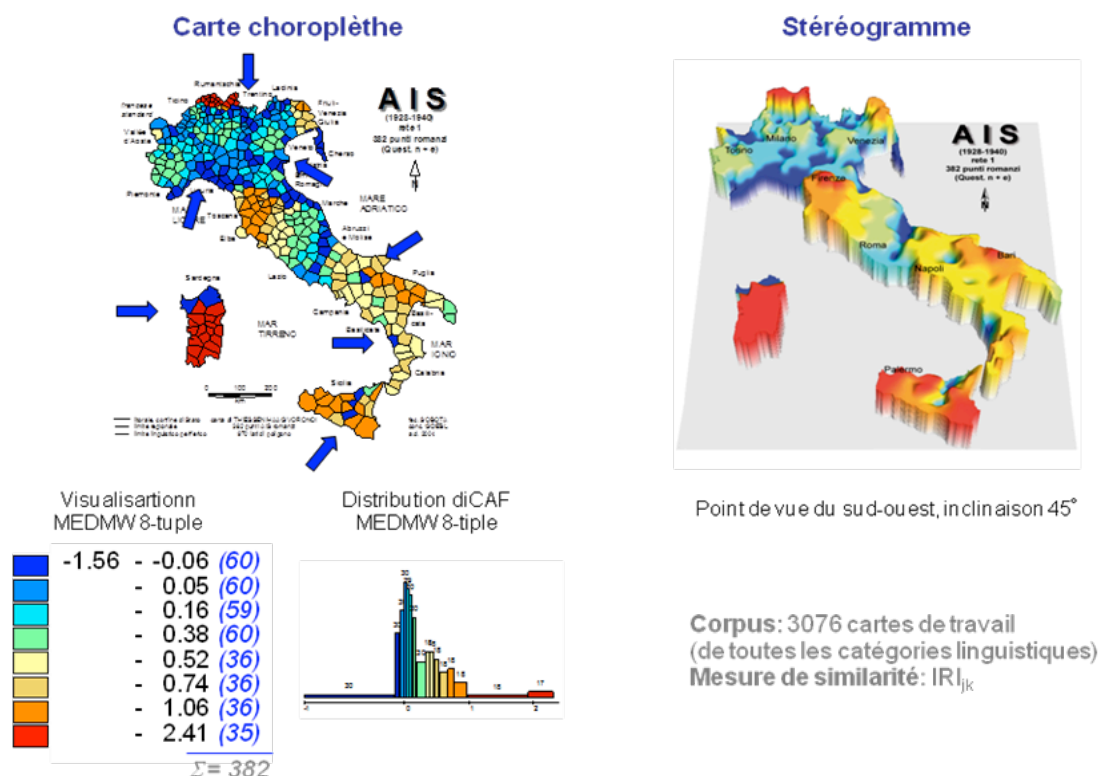


FIGURE 24. Deux visualisations différentes de la synthèse des scores du «coefficient d'asymétrie de Fisher» (CAF) de 382 cartes de similarité tirées du *Sprach- und Sachatlas Italiens und der Südschweiz* (AIS)

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 des cinq îlots linguistiques méridionaux de l'AIS (P. 715: Faeto, P. 760: Guardia Piemontese et trois points-AIS siciliens du type *gallo-italico*), qui figurent tous en bleu foncé. Aussi les parlers du nord de la Sardaigne, appartenant typologiquement au domaine corse, s'alignent sur la position-DM des îlots linguistiques à proprement parler.

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

La figure 25 en fournit deux applications très éloquentes relatives à l'Ibéroromania entière et au domaine linguistique catalan. Dans la perspective de l'ALPI la synthèse des CAF permet de reconstituer très clairement les décalages interrégionaux qui régnaient au temps de la «Reconquista» qui, elle, est caractérisée par l'avancement parallèle de trois forces linguistiques en nette concurrence entre elles: a) du bloc luso-galicien; b) du bloc castillano-léonais; c) du bloc catalano-aragonais.

Il est bien connu que l'épopée de reconquête la plus longue revenait au bloc castillano-léonais. Dans la figure 25, les polygones en bleu foncé se situent entre la Vieille- et la Nouvelle-Castille et renvoient de ce fait à des situations de forte interaction linguistique telles qu'elles s'y sont produites au milieu du XIII^e siècle.

Corpus: 375 cartes de travail

Corpus: 1660 cartes de travail

(de toutes les catégories linguistiques)

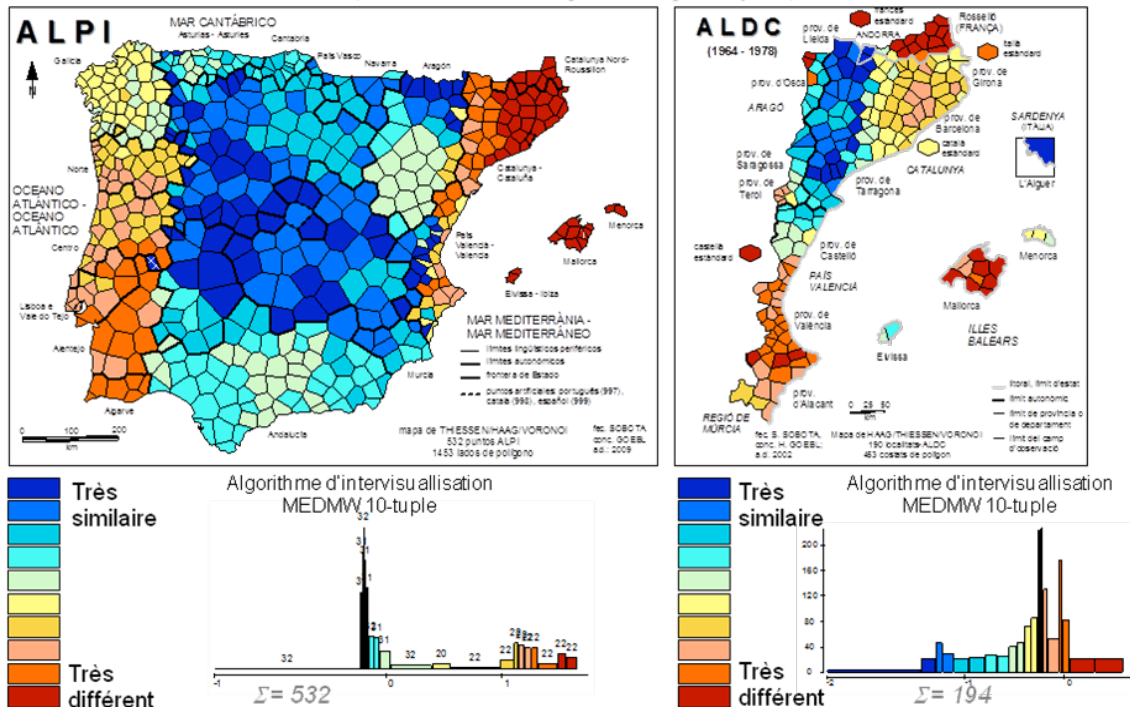


FIGURE 25. Deux synopses des scores du «coefficient d'asymétrie de Fisher» (CAF) tirées de l'Atlas Lingüístico de la Peninsula Ibérica (ALPI) et de l'Atlas Lingüístico del Domini Català (ALDC)

Le profil relatif à l'ALDC montre que l'interaction linguistique est à son comble dans la zone du catalan nord-occidental et aussi dans l'îlot linguistique d'Alghero/L'Alguer en Sardaigne.

6. L'analyse interponctuelle: carte à cloisons et carte à rayons

L'analyse dite «interponctuelle» connaît deux versants: celui de la visualisation de *distances* et celui 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.²¹ Il s'agit donc de la modélisation, faite avec les moyens de la DM, d'une technique déjà ancienne et, partant, bien connue de la géographie linguistique. La seconde visualisation, dite «carte à rayons», est, en revanche, de date récente. Du point de vue 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 régie par la formule suivante: $s + d = 100$. C'est la raison pour laquelle les silhouettes des histogrammes des figures 26 et 27 sont parfaitement symétriques. 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

²¹Voir les tableaux en appendice du travail séminal de Karl von Etmayer (1924) où se trouve un grand nombre de schémas isoglottiques exécutés à la main et contenant des bourrelets de frontière de toute sorte.

d'«aversion» qui existe entre deux lococlectes contigus, la carte à rayons représente, par l'épaisseur et le coloriage chaud des côtés de triangle, des relations d'«amitié» à très petite distance.

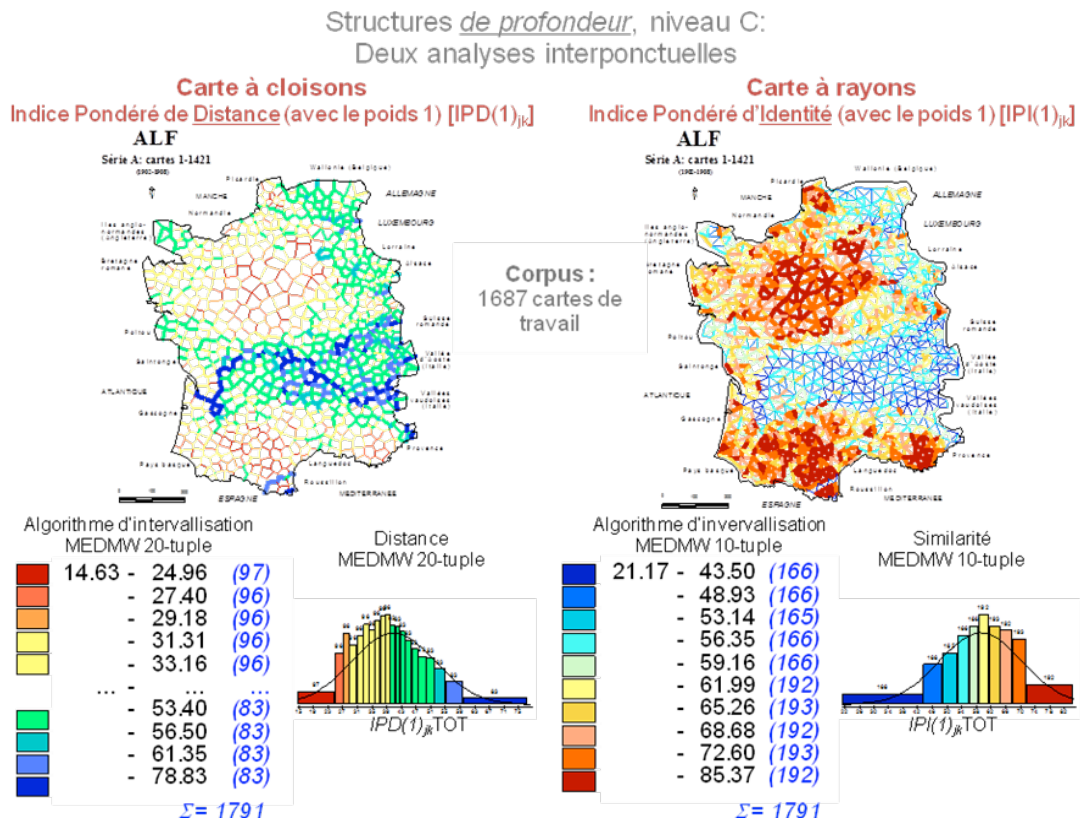


FIGURE 26. Deux cartes interponctuelles («carte à cloisons» et «carte à rayons») tirées de l'Atlas Linguistique de la France (ALF)

Le message iconique des deux cartes de la figure 26 s'appuie sur l'interaction visuelle de 1791 éléments graphiques (*i.e.* de côtés de polygone ou de triangle) et fournit une vision très nuancée des enchevêtrements interponctuels qui existent entre le nord et le sud de la France. Il n'y a cependant nulle trace d'une confrontation linéiforme entre les domaines d'Oïl et d'Oc. L'œil perçoit, par contre, une vaste zone de passage sillonnée d'effets de cloisonnements ou de réticulations finement gradués.

La même chose se manifeste aussi sur les figures 27 et 28. La carte à cloisons relative à l'AIS (ou au domaine italo-roman) montre les zones de passage bien connues situées à l'intérieur et surtout au sud de la Padanie (Italie septentrionale), appelée communément «linea La Spezia-Rimini», ainsi que la vaste bande de transition se creusant entre l'Italie centrale et méridionale («linea Roma-Ancona»). La carte à cloisons de la figure 27 montre en outre la situation interponctuelle particulière des îlots linguistiques figurant dans le réseau de l'AIS: il s'agit de Faeto (P. 715), Guardia Piemontese (P. 760) et de trois îlots gallo-italiques situés en Sicile.

Alors que les analyses interponctuelles relatives à l'ALF et à l' AIS se réfèrent à des dialecticités très anciennes leurs équivalents relatifs à l'ALPI et à l'ALDC sont pour la plupart basés sur des dialectes dits «secondaires». Il en résulte une plus grande précision et saillance des cloisonnements différentiels dont une bonne partie coïncide exactement avec le tracé de frontières étatiques actuelles (autour du Portugal et entre la France et l'Espagne).

7. Trois parcours géotypologiques (ALF, AIS et ALPI)

Les figures 29-52 servent à la mise en valeur de la vertu géotypologique d'un certain nombre de profils de similarité dont les points de référence sont alignés le long d'un bref parcours d'épreuve continu qui, lui, traverse des cloisons interponctuels très épais. Les figures 29 (pour l'ALF), 39 (pour l' AIS) et 47 (pour l'ALPI) renseignent sur la longueur et la pertinence linguistique des parcours d'épreuve respectifs.

Pour l'ALF, il s'agit d'un voyage géotypologique de la Gascogne septentrionale jusqu'aux confins septentrionaux de l'Angoumois en passant par l'unique enclave linguistique de l'ALF (P. 635, Andraut) et plusieurs points-ALF appartenant au «Croissant».

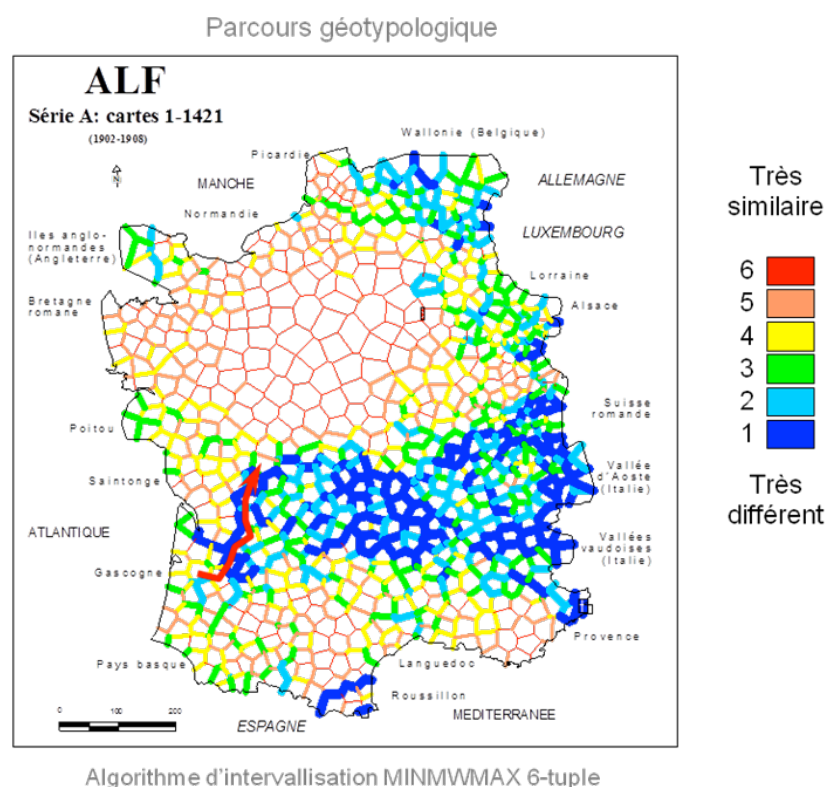
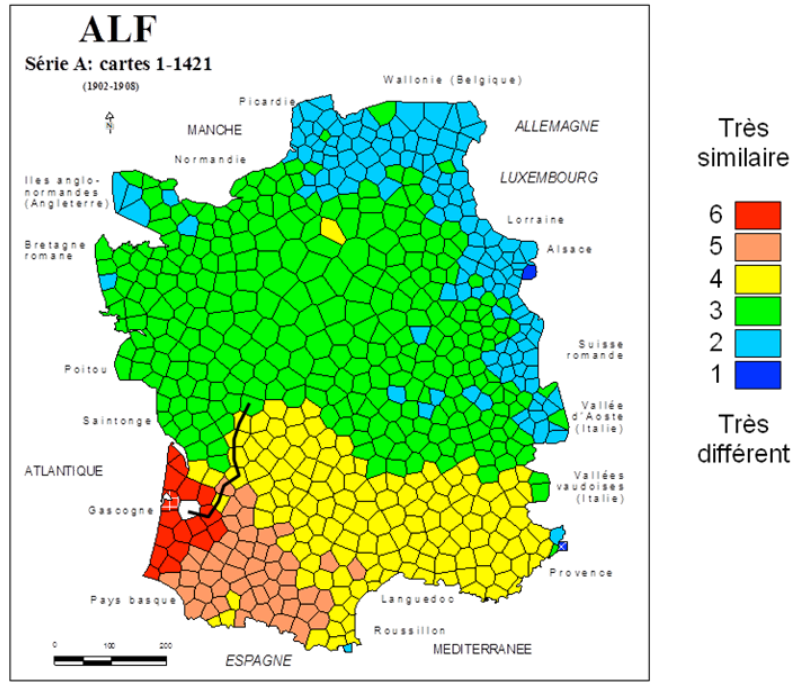


FIGURE 29. Insertion du «parcours d'épreuve» galloroman dans une carte à cloisons tirée de l'*Atlas Linguistique de la France* (ALF)

► [DIAPORAMA 1: vidéo du parcours d'épreuve galloroman](#)

Il en résulte trois profils (ou «paysages») géolinguistiques bien différenciés relatifs aux domaines d'Oc (figures 30 et 31), d'Oïl (figures 32, 37, 38) et au «Croissant» (figures 33-36).

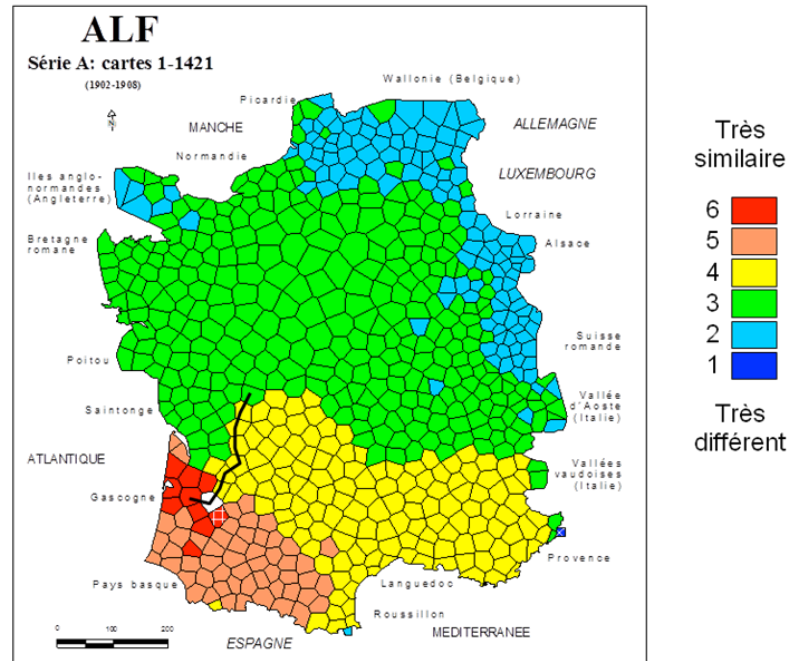
653: Hostens (gascon)



Algorithm d'intervallisation MINMWMAX 6-tuple

FIGURE 30. Première de neuf étapes du parcours d'épreuve galloroman: un profil de similarité de type *gascon* relatif au point-ALF 653: Hostens

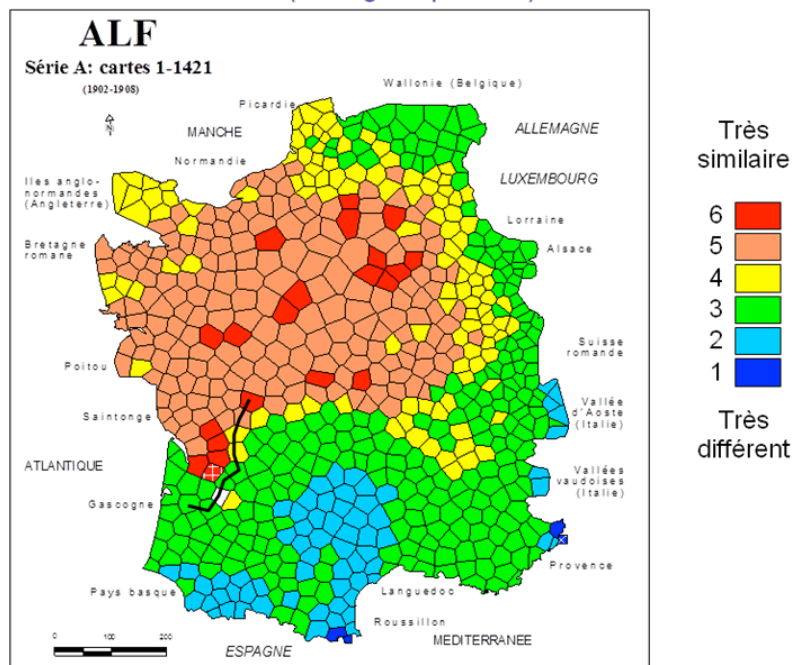
645: Saint Côme (gascon)



Algorithme d'intervallisation MINMWMAX 6-tuple

FIGURE 31. Deuxième de neuf étapes du parcours d'épreuve galloroman: un profil de similarité de type *gascon* relatif au point-ALF 645: Saint-Côme

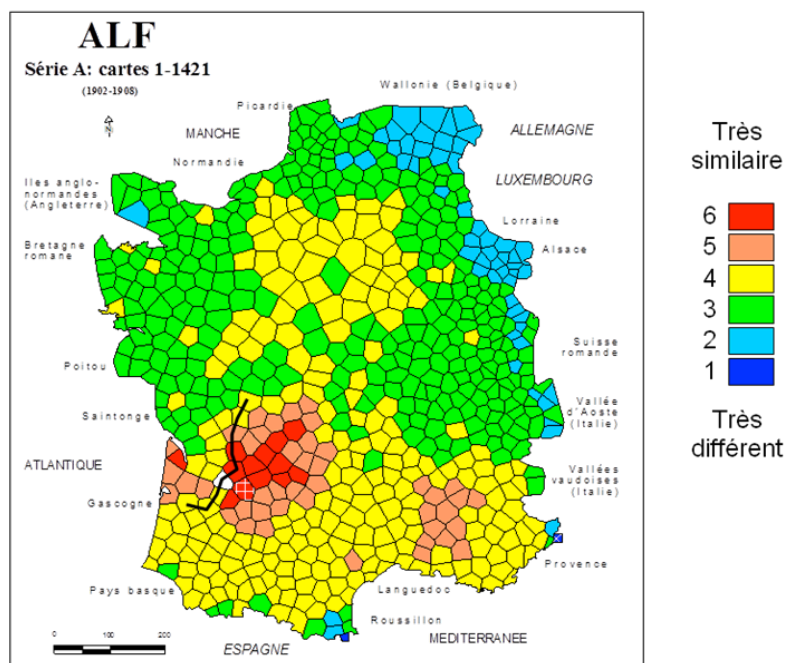
635: Andraut (îlot linguistique d'Oïl)



Algorithme d'intervallisation MINWMAX 6-tuple

FIGURE 32. Troisième de neuf étapes du parcours d'épreuve galloroman: un profil de similarité de type d'Oïl relatif au point-ALF 635: Andraut

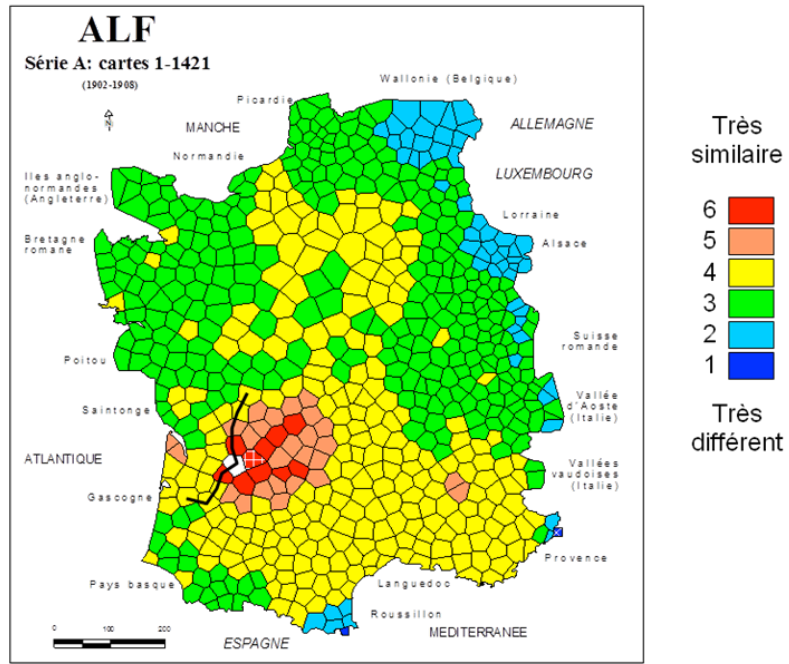
634: Vélines (croissant)



Algorithme d'intervallisation MINWMAX 6-tuple

FIGURE 33. Quatrième de neuf étapes du parcours d'épreuve galloroman: un profil de similarité de type *crescentain* relatif au point-ALF 634: Vélines

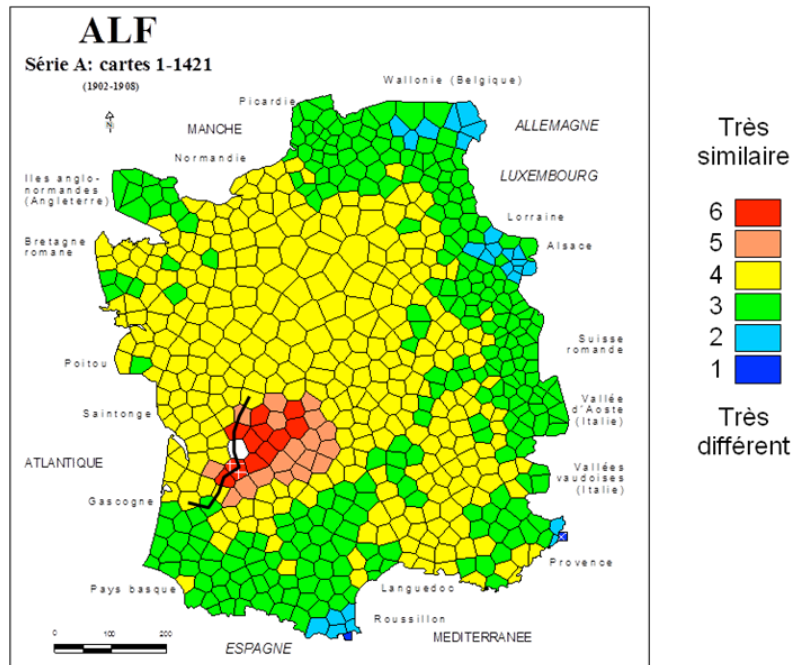
624: Bourgnac (croissant)



Algorithme d'intervallisation MINWMAX 6-tuple

FIGURE 34. Cinquième de neuf étapes du parcours d'épreuve galloroman: un profil de similarité de type *crescentain* relatif au point-ALF 624: Bourgnac

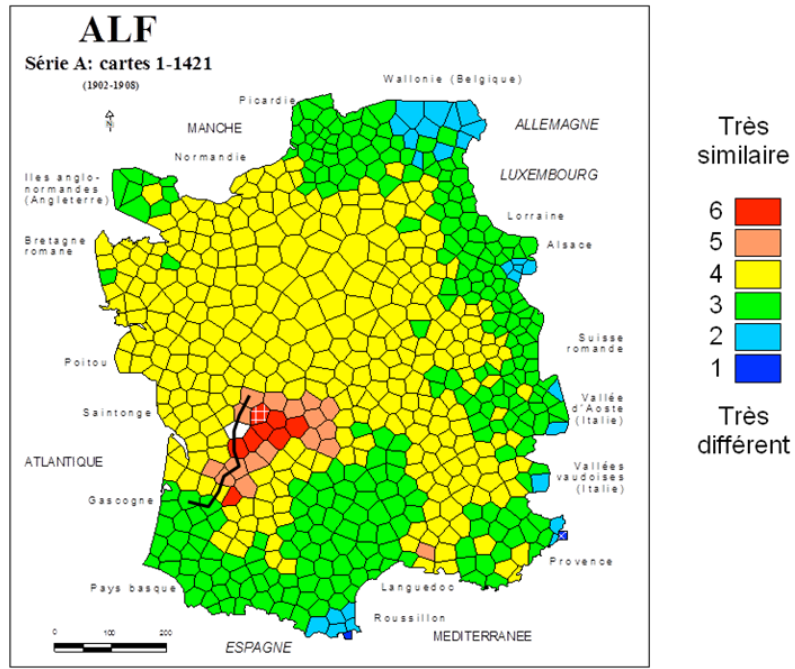
611: La Tour-Blanche (croissant)



Algorithme d'intervallisation MINWMAX 6-tuple

FIGURE 35. Sixième de neuf étapes du parcours d'épreuve galloroman: un profil de similarité de type *crescentain* relatif au point-ALF 611: La Tour-Blanche

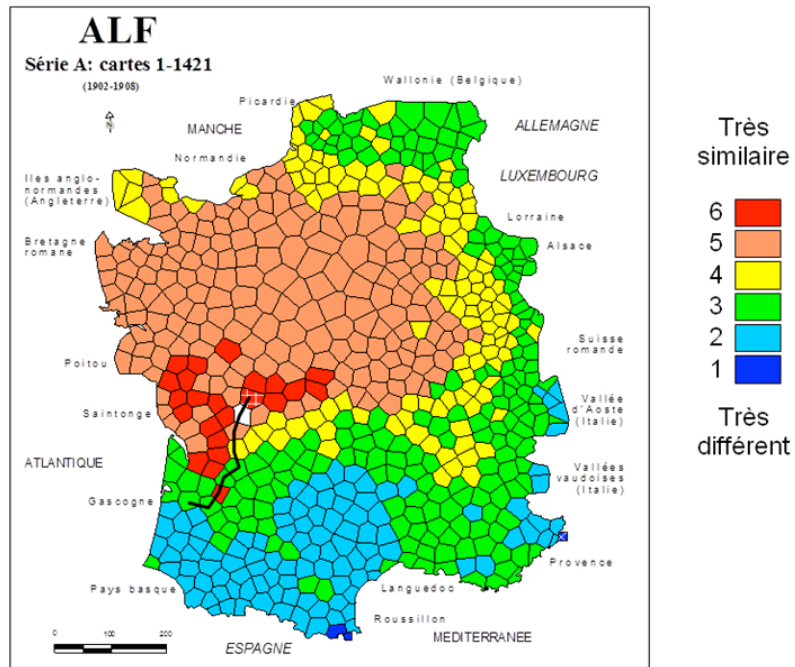
610: Chazelles (croissant)



Algorithme d'intervallisation MINWMAX 6-tuple

FIGURE 36. Septième de neuf étapes du parcours d'épreuve galloroman: un profil de similarité de type *crescentain* relatif au point-ALF 610: Chazelles

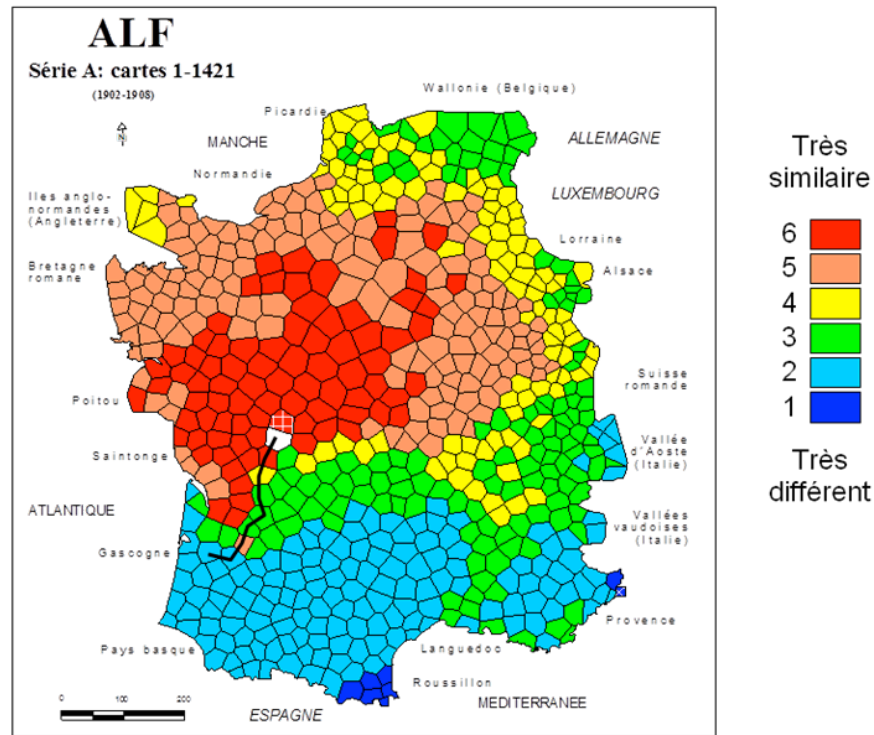
519: Saint-Claud (Oïl)



Algorithme d'intervallisation MINWMAX 6-tuple

FIGURE 37. Huitième de neuf étapes du parcours d'épreuve galloroman: un profil de similarité de type *d'Oïl* relatif au point-ALF 519: Saint-Claud

509: Millac (Oïl)



Algorithme d'intervallisation MINMWMAX 6-tuple

FIGURE 38. Dernière de neuf étapes du parcours d'épreuve galloroman: un profil de similarité de type d'Oïl relatif au point-ALF 509: Millac

Par le biais d'une projection diaporamique en fondu enchaîné (voir Diaporama 1 qui suit la figure 29), il est possible de tirer, de ces neuf étapes géotypologiques, un effet cinématographique très éloquent. L'efficacité heuristique de cet effet dépend d'ailleurs dans une large mesure de l'algorithme d'intervallisation appliqué. Les meilleurs résultats visuels s'obtiennent moyennant l'algorithme déjà bien connu de MINMWMAX (6-tuple).

Remarquons que chacun des trois paysages (ou types géolinguistiques) galloromans est relativement stable et pourvu de profils choroplèthes «typiques»: remarquons aussi que les trois parcours d'épreuve reposent tous sur le maximum des données-DM à disposition (nombre des cartes de travail: ALF: 1 687, AIS: 3 911, ALDC: 1 660).

Le voyage diaporamique italien va de Bologne jusqu'à Florence et aboutit au point factice de l'italien standard (voir Diaporama 2). Évidemment, notre parcours d'épreuve croise la «linea La Spezia-Rimini» à son endroit le plus épais si bien que le contraste géotypologique entre les figures 42 et 43 est considérable. Alors que les profils géotypologiques des figures 40 et 41 sont nettement padaniens, la figure 42 – relative au P.-AIS 466, Loiano – laisse déjà entrevoir, par la projection de polygones jaunes vers la Toscane, la nouvelle typologie linguistique.

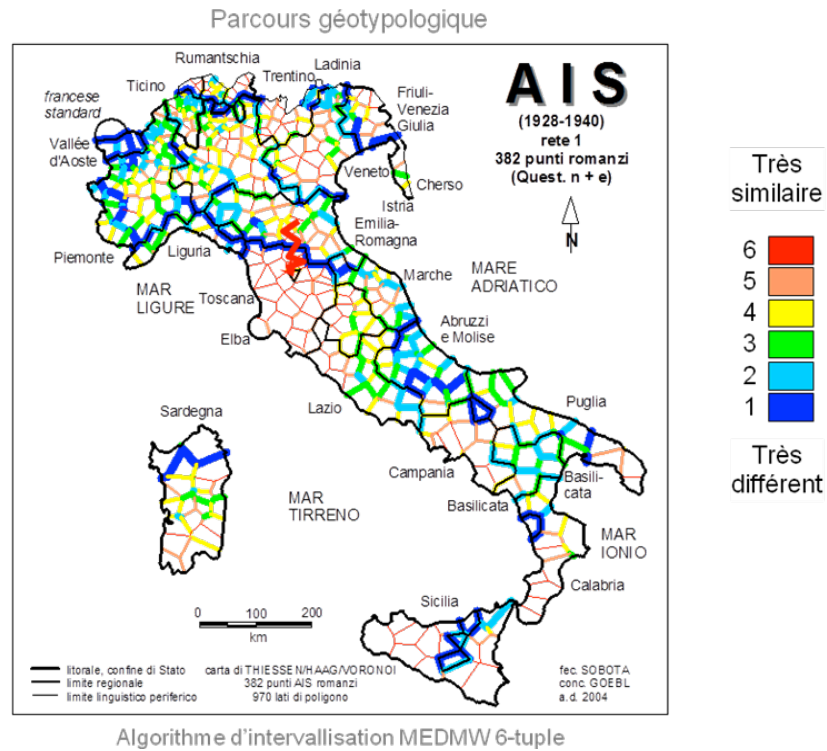


FIGURE 39. Insertion du «parcours d'épreuve» italo-roman dans une carte à cloisons tirée du *Sprach- und Sachatlas Italiens und der Südschweiz* (AIS)
 ▶ [DIAPORAMA 2: vidéo du parcours d'épreuve italo-roman](#)

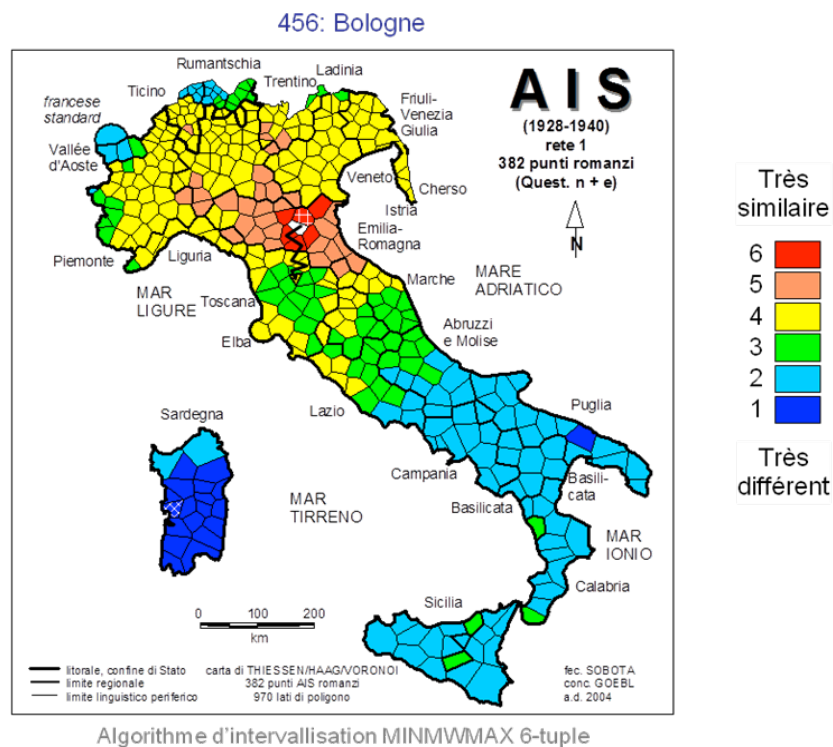
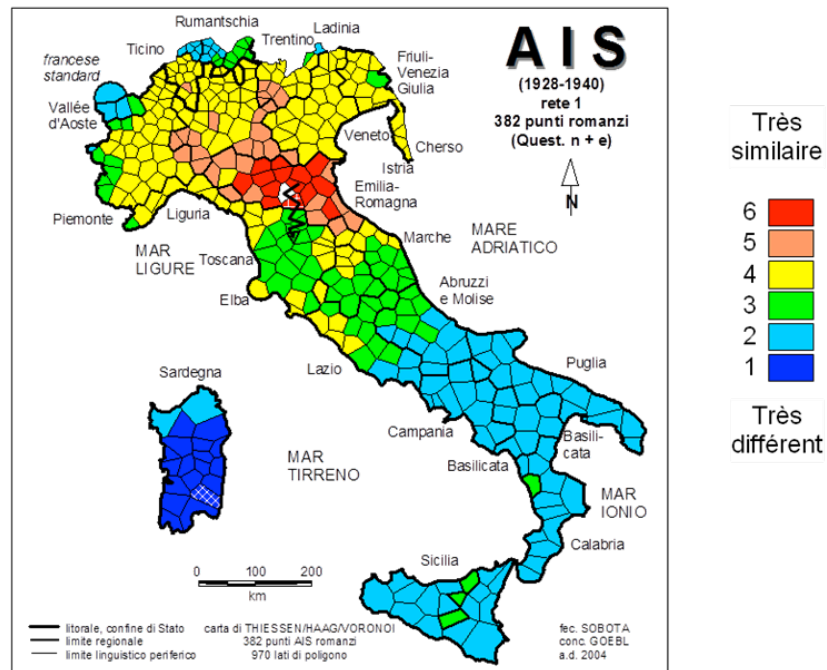


FIGURE 40. Première de sept étapes du parcours d'épreuve italo-roman: un profil de similarité de type *padanien* relatif au point-AIS 456: Bologne

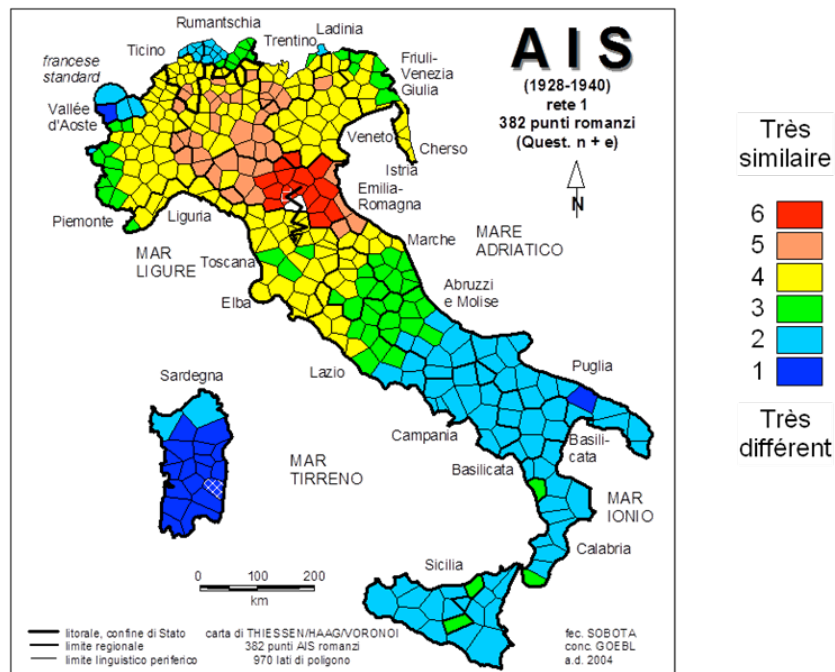
455: Savigno



Algorithme d'intervallisation MINWMAX 6-tuple

FIGURE 41. Deuxième de sept étapes du parcours d'épreuve italo-roman: un profil de similarité de type *padanien* relatif au point-AIS 455: Savigno

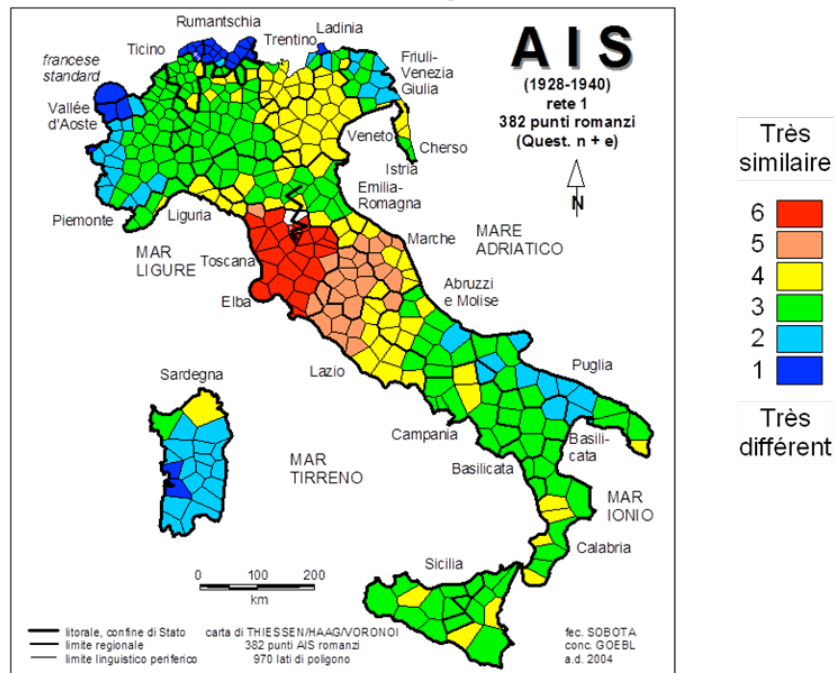
466: Loiano



Algorithme d'intervallisation MINWMAX 6-tuple

FIGURE 42. Troisième de sept étapes du parcours d'épreuve italo-roman: un profil de similarité de type *padanien* relatif au point-AIS 466: Loiano

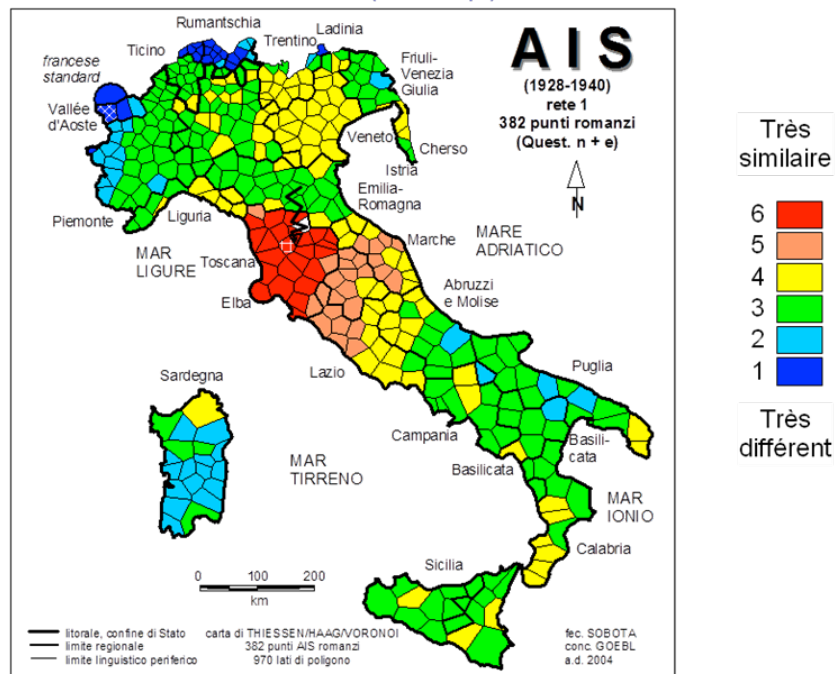
515: Barberino di Mugello



Algorithme d'intervallisation MINWMAX 6-tuple

FIGURE 43. Quatrième de sept étapes du parcours d'épreuve italo-roman: un profil de similarité de type *toscan* relatif au point-AIS 515: Barberino di Mugello

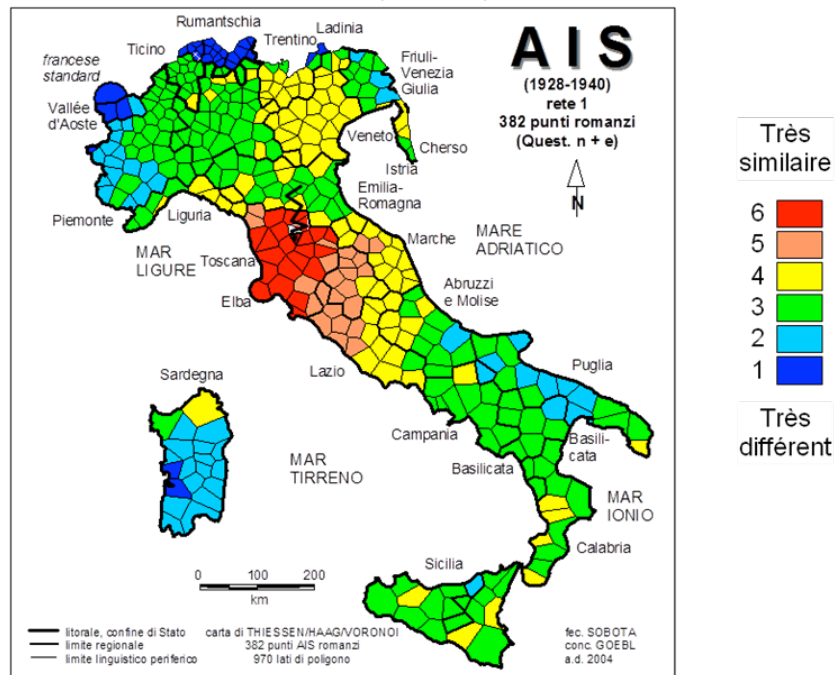
524: Florence (instr. sup.)



Algorithme d'intervallisation MINWMAX 6-tuple

FIGURE 44. Cinquième de sept étapes du parcours d'épreuve italo-roman: un profil de similarité de type *toscan* relatif au point-AIS 524: Florence (sujet avec instruction supérieure)

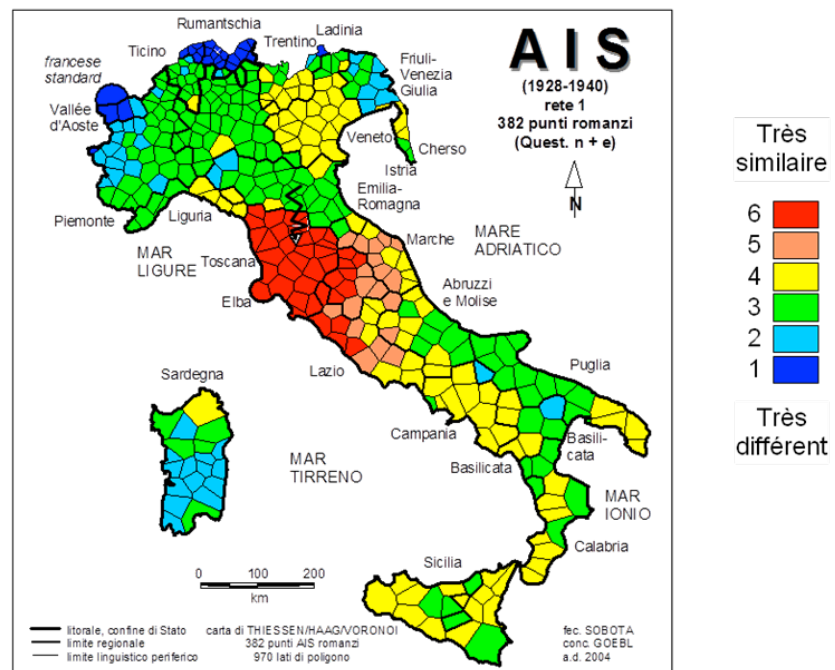
523: Florence (instr. inf.)



Algorithme d'intervallisation MINWMAX 6-tuple

FIGURE 45. Sixième de sept étapes du parcours d'épreuve italomroman: un profil de similarité de type *toscan* relatif au point-AIS 523: Florence (sujet avec instruction inférieure)

999: Italien standard



Algorithme d'intervallisation MINWMAX 6-tuple

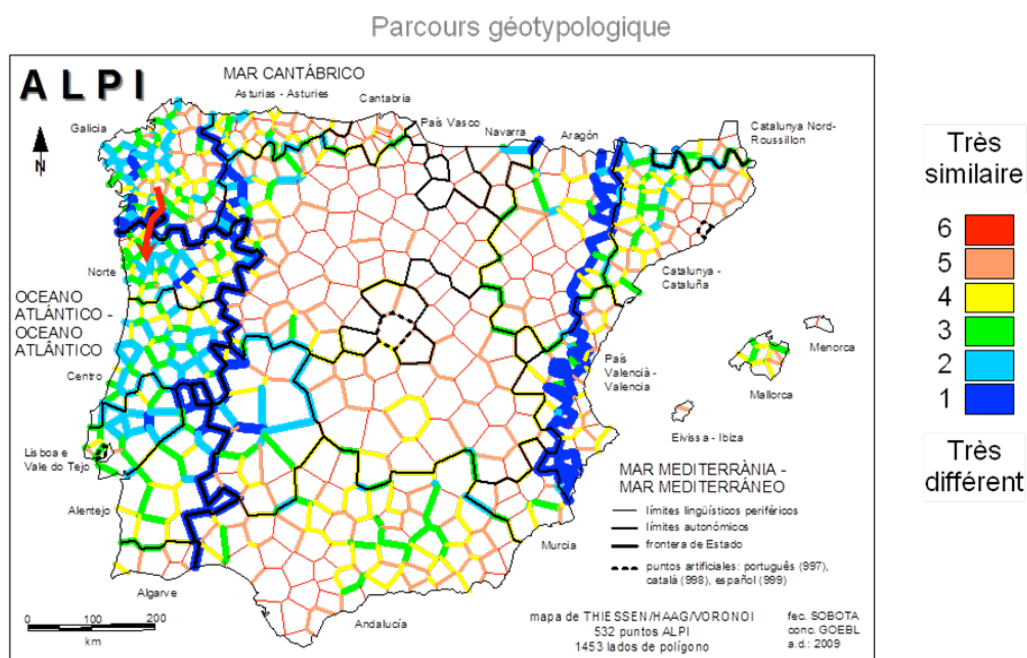
FIGURE 46. Dernière de sept étapes du parcours d'épreuve italomroman: un profil de similarité de type *toscan* relatif au point-AIS 999: italien standard

Mis à part le fait, en soi prévisible et, partant, naturel, que l'espace de la Toscane est coloré en rouge, les figures 43-46 attestent très clairement la position rapprochée du domaine vénète face au toscan en particulier et à l'Italie médiane en général. Ce fait est dû avant tout au caractère non pas *padanien*, mais *médian* voire *toscan* du phonétisme des parlers vénètes.

L'explication historique de ce phénomène curieux est difficile: deux hypothèses s'affrontent. La première veut que le caractère médian des parlers vénètes ait pris sa genèse déjà au moment de la romanisation (ca. 180 av. J. C.), soit à cause de l'origine «médiane» de la majorité des colons romains qui y ont été implantés, soit à cause de la grande parenté typologique du phonétisme des parlers paléo-vénètes avec celui du latin d'importation récente. La seconde des deux hypothèses admet une toscanisation rampante des parlers vénètes à partir de l'an 1000 apr. J. C. due surtout à l'importance de la place centrale de Venise et sa fonction comme distributrice de modes linguistiques et culturelles. Cette théorie veut que la «Serenissima» ait reçu certaines modes linguistiques du sud pour les diffuser ensuite au cours de la conquête de la «terraferma», initiée dans le premier tiers du XIV^e siècle.

On remarquera le décalage diachronique de quelque 1500 ans qui se creuse entre les deux conjectures entre lesquelles il est très difficile de trancher. Personnellement, je penche plutôt vers la première des deux solutions.²²

Le parcours géotypologique ibérique va de la Galice au Portugal (voir Diaporama 3). Il visualise non seulement la position intermédiaire de certains parlers galiciens entre les domaines castillan et portugais (voir surtout les figures 48 et 49), mais aussi le contraste géotypologique qui existe entre les domaines portugais et castillan.



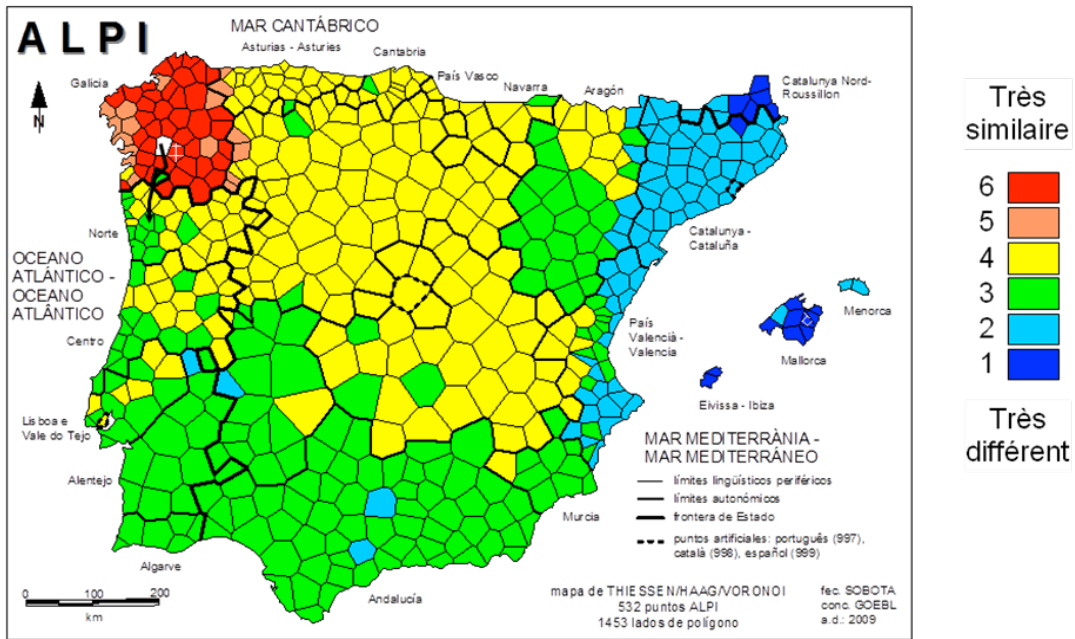
Algorithme d'intervallisation MINMWMAX 6-tuple

FIGURE 47. Insertion du «parcours d'épreuve» ibéroroman dans une carte à cloisons tirée de l' *Atlas Lingüístico de la Península Ibérica* (ALPI)

[▶ DIAPORAMA 3: vidéo du parcours d'épreuve ibéroroman](#)

²² Pour une plus ample discussion de ce phénomène, cf. Goebel (2008: 58-61).

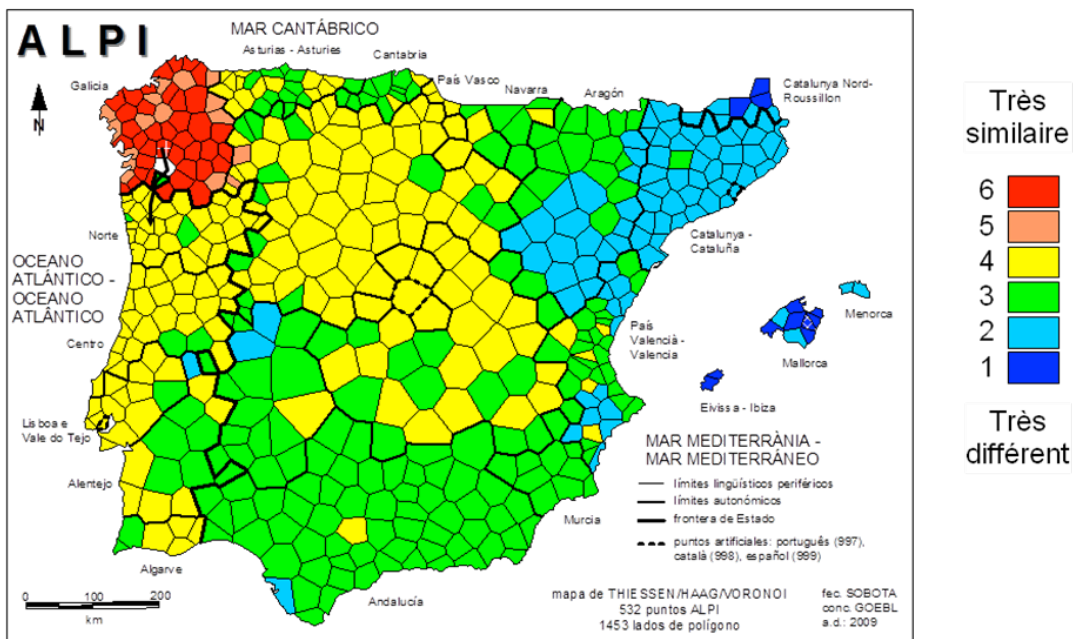
141: Irjo



Algorithme d'intervallisation MINMWMAX 6-tuple

FIGURE 48. Première de cinq étapes du parcours d'épreuve ibéroroman: un profil de similarité de type *galicien* relatif au point-ALPI 141: Irjo

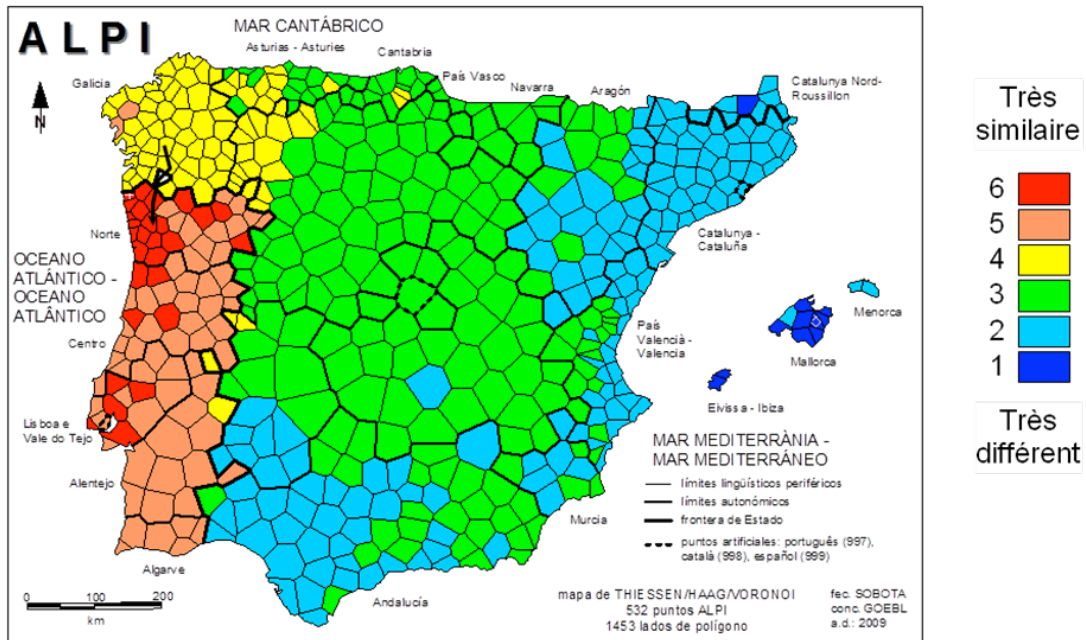
143: Arnoya



Algorithme d'intervallisation MINMWMAX 6-tuple

FIGURE 49. Deuxième de cinq étapes du parcours d'épreuve ibéroroman: un profil de similarité de type *galicien* relatif au point-ALPI 143: Arnoya

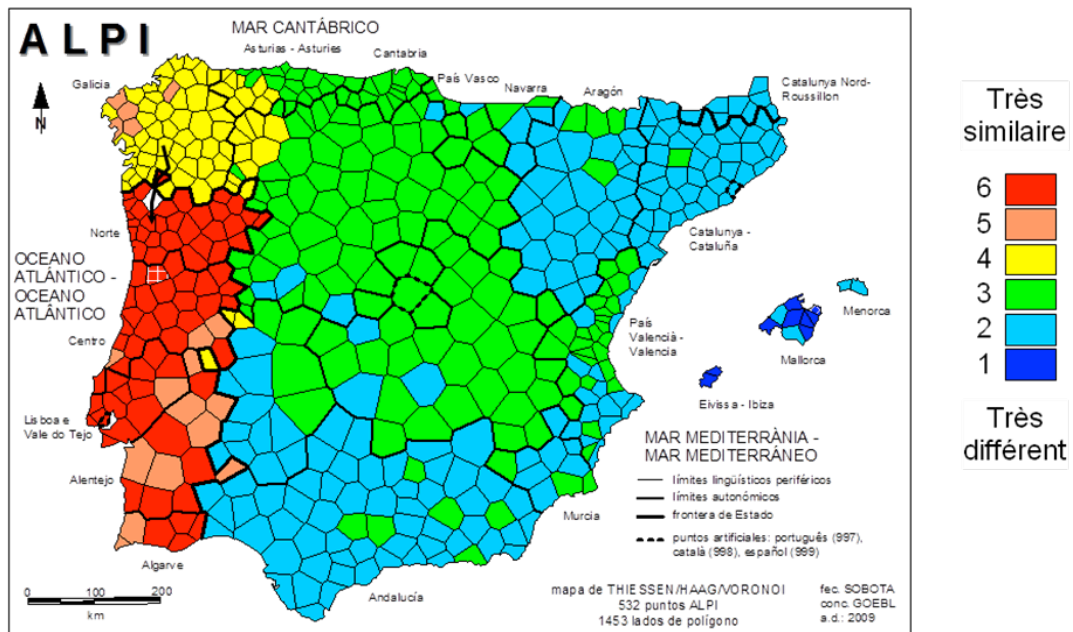
200: Paderne (Portugal)



Algorithme d'intervallisation MINMWMAX 6-tuple

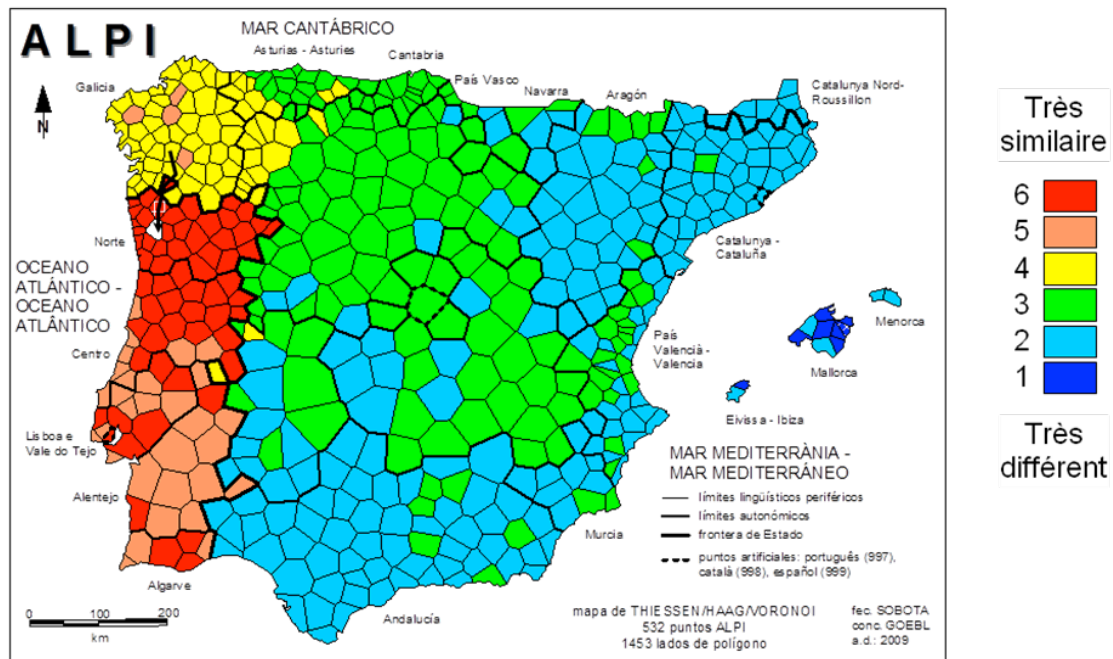
FIGURE 50. Troisième de cinq étapes du parcours d'épreuve ibéroroman: un profil de similarité de type *portugais* relatif au point-ALPI 200: Paderne

204: Tamente



Algorithme d'intervallisation MINMWMAX 6-tuple

FIGURE 51. Quatrième de cinq étapes du parcours d'épreuve ibéroroman: un profil de similarité de type *portugais* relatif au point-ALPI 204: Tamente



Algorithme d'intervallisation MINMWMAX 6-tuple

FIGURE 52. Dernière de cinq étapes du parcours d'épreuve ibéroroman: un profil de similarité de type *portugais* relatif au point-ALPI 209: Brito

8. L'analyse dendrographique

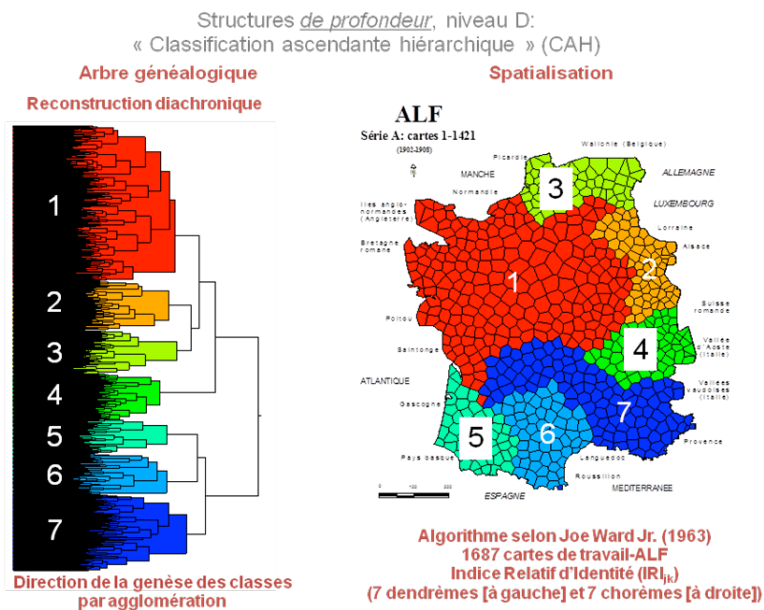


FIGURE 53. Présentation d'une analyse dendrographique (de type WARD) – avec arbre et spatialisation –, appliquée sur les données de l'Atlas Linguistique de la France (ALF)

Le niveau D de l'analyse-DM des structures de profondeur concerne la DM dendrographique ou – dans la terminologie technique de la classification numérique – la «classification ascendante hiérarchique» (CAH). La CAH est largement répandue au sein de beaucoup de sciences naturelles et humaines. Elle sert au groupement algorithmique de N objets – définis, chacun, par p attributs – en structures ramifiées binaires. 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 6).

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. Leur action agglomératrice débute par les N valeurs isolées de cette dernière: 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 répondent 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 d'un algorithme-CAH est la cohérence spatiale des différents clusters (ou «chorèmes») extraits de l'arbre calculé.

Évidemment, l'arbre en tant que configuration ramifiée ne saurait jamais suffire aux besoins d'un dialectométricien épris de schémas cartographiques et, partant, 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».

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

Ceci est dû au fait que tout arbre issu d'une CAH peut être utilisé à une reconstruction diachronique telle qu'elle est conçue et pratiquée depuis belle lurette par la lexicostatistique et la glottochronologie.²³ La logique de cette perspective veut que la proximité des différentes ramifications face à la racine (ou au tronc) de l'arbre décide de l'âge du chorème en question. Jetons à ce propos un coup d'œil sur la figure 53: le chorème 7 (en bleu foncé et situé dans le sud) est censé être linguistiquement plus ancien que le dendrème 1 (en rouge) à cause du voisinage plus grand de sa racine face au tronc central de l'arbre entier.

La figure 54 montre les résultats de l'analyse similaire appliquée aux données de l'AIS. Remarquons que la première bifurcation sépare les chorèmes 1-10 des chorèmes 11-19 de l'arbre. À y regarder de près, l'on constate que la ligne séparatrice entre ces deux mégachorèmes primaires correspond à la fameuse «linea La Spezia-Rimini». Pour découvrir d'autres subdivisions cachées dans les embranchements de l'arbre, il faut donc procéder à une analyse détaillée et toujours plus fine de toute la hiérarchie binaire calculée. Le logiciel salzbourgeois VDM («Visual Dialectometry») permet non seulement d'opérer confortablement cette décomposition, mais aussi de conférer des coloriages librement choisis aux différents dendrèmes et chorèmes.

²³ Cf. à ce sujet Goebel (2003: 84-89).

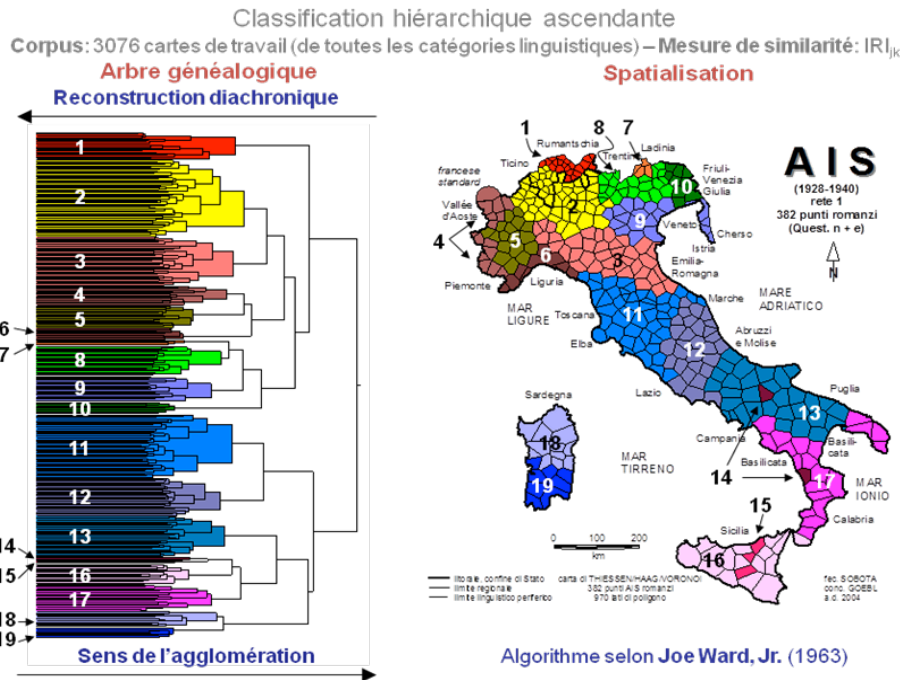


FIGURE 54. Présentation d’une analyse dendrographique (de type WARD) – avec arbre et spatialisation –, appliquée sur les données du *Sprach- und Sachatlas Italiens und der Südschweiz* (AIS)

La figure 55 montre les résultats d’une analyse dendrographique plutôt sommaire des données dialectales que nous avons tirées de l’ALPI. Ceci n’empêche que la remarquable cohérence spatiale des chorèmes ne disparaît aucunement quand on pousse la spatialisation chorématique plus à l’intérieur de l’arbre.

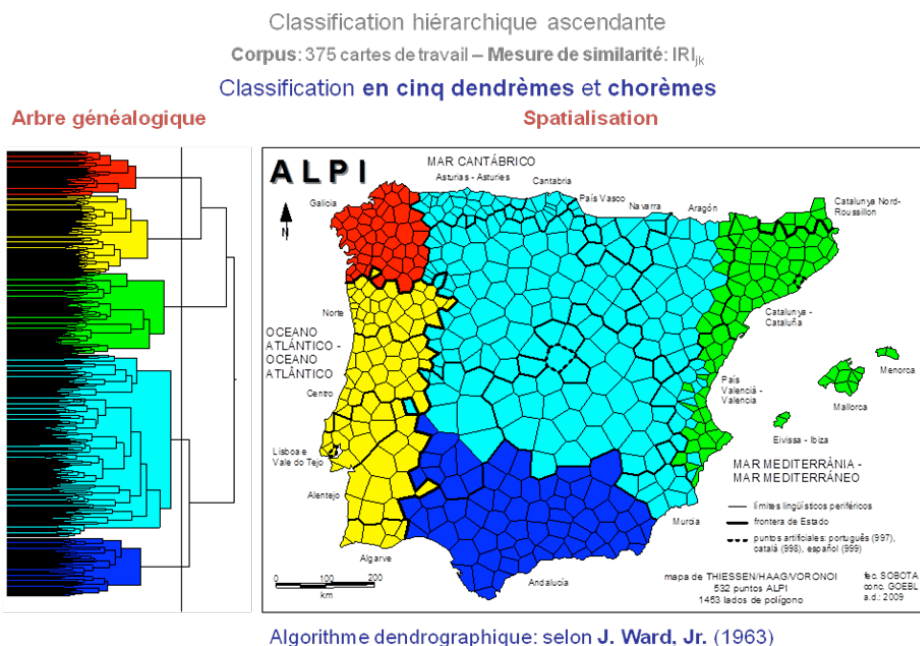


FIGURE 55. Présentation d’une analyse dendrographique (de type WARD) – avec arbre et spatialisation –, appliquée sur les données de l’*Atlas Lingüístico de la Península Ibérica* (ALPI)

9. La dialectométrie corrélative

La DM corrélative qui est une des pièces fortes de l'arsenal méthodique de l'«École dialectométrique de Salzbourg» existe depuis 2005²⁴ et a fourni 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 à dimensions extérieures égales (deux fois N^2), 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 corrélérer peut varier à loisir : l'on peut corrélérer, entre elles, différentes catégories linguistiques, les corrélérer, 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 est de ne jamais perdre de vue leur finalité *linguistique* à proprement parler.

Nous commençons la présentation du niveau E de l'analyse des structures de profondeur par la corrélation entre les *similarités linguistiques* d'un côté et les *proximités euclidiennes* de l'autre (voir les figures 56 et 57).

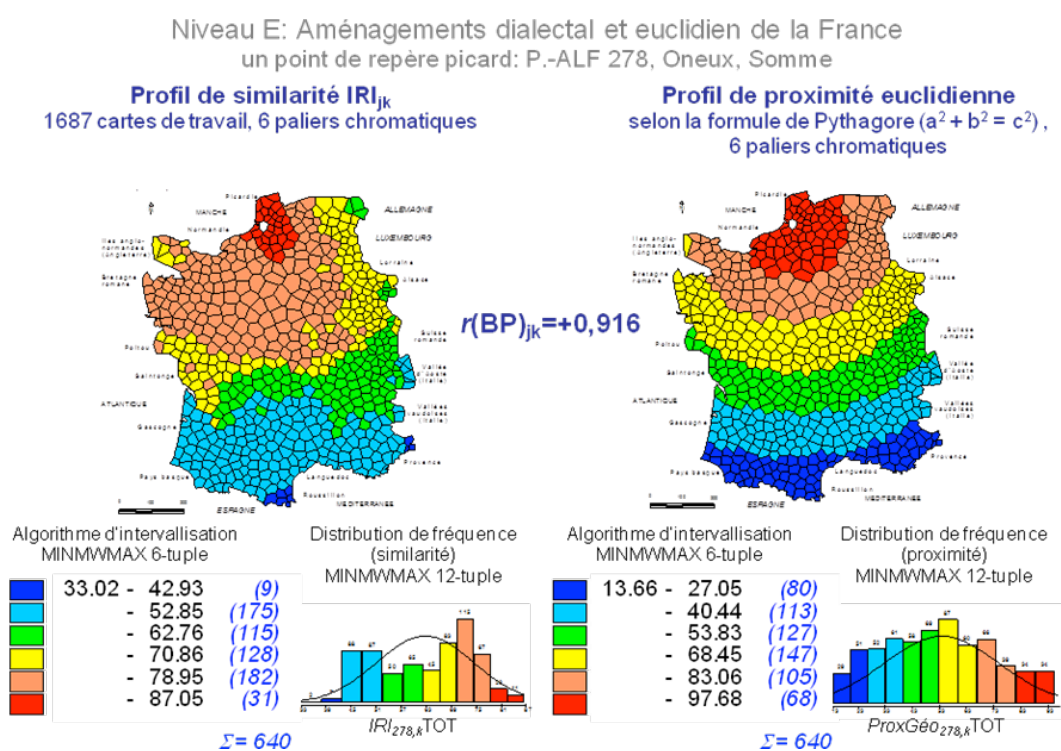


FIGURE 56. Présentation comparative d'une carte de similarité de type *picard* et d'une carte de proximité, relatives au point 278, Oneux, de l'Atlas Linguistique de la France (ALF)

Une des leçons données par les cartes de similarité était que les similarités linguistiques relatives à un locolecte de référence diminuent en fonction de l'augmentation de la distance kilométrique par rapport à ce dernier. Évidemment, les différents profils de similarité que nous venons de voir ont démontré que ce décalage peut être très variable. Toujours est-il

²⁴ Cf. Goebel (2005: *passim*).

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 de 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 des 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.

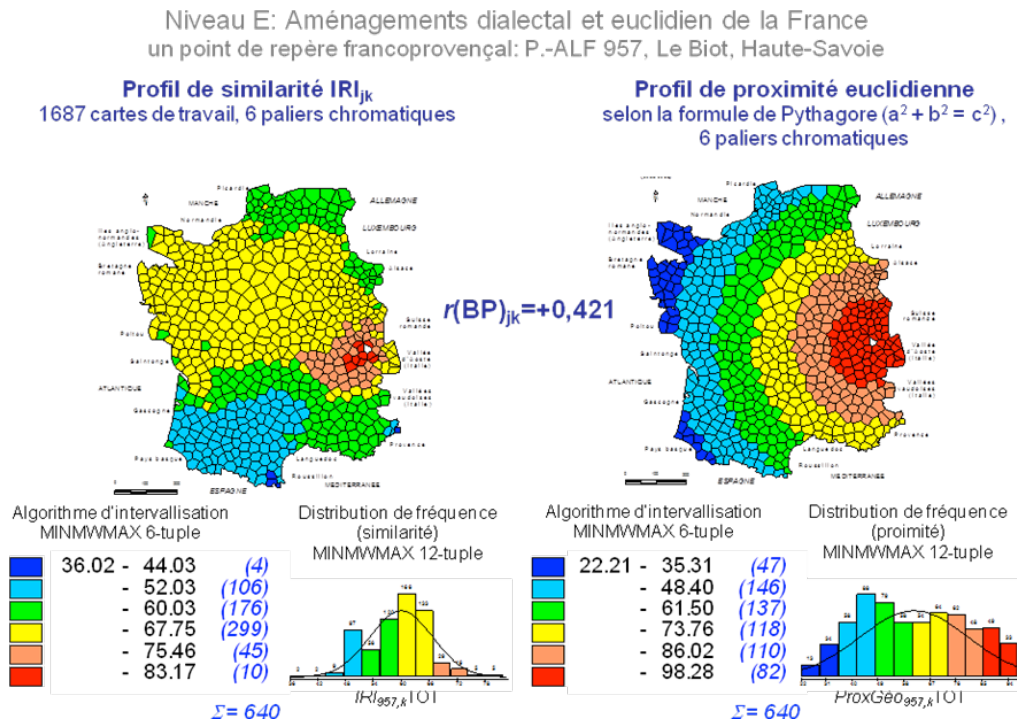


FIGURE 57. Présentation comparative d'une carte de similarité de type francoprovençal et d'une carte de proximité, relatives au point 957, Le Biot, de l'Atlas Linguistique de la France (ALF)

La figure 58 montre l'agencement de deux matrices carrées, l'organisation des calculs de corrélations et la genèse de la carte à corrélations respective. Nous utilisons, pour le calcul des corrélations, le coefficient de corrélation de (Auguste) Bravais et (Karl) Pearson [$r(BP)$] dont les scores oscillent régulièrement entre -1 et +1.

En appliquant le $r(BP)$ aux vecteurs relatifs au P.-ALF 278, contenant, d'un côté, les scores de similarité et, de l'autre, de proximité euclidienne, l'on obtient la valeur +0,916 (voir la figure 56). Des mêmes calculs effectués pour les deux vecteurs du P.-ALF 957, nous obtenons la valeur +0,421 (voir la figure 57). Les deux valeurs de corrélation, prises comme telles, ne nous disent rien en tant que linguistes. Il faut les faire parler aux oreilles – mieux: aux yeux – des linguistes en les visualisant: voir à cet effet la figure 59.

Schéma du calcul de la carte à corrélations
à l'aide de deux matrices (carrées) de similarité (ou de proximité)
avec N vecteurs d'objets

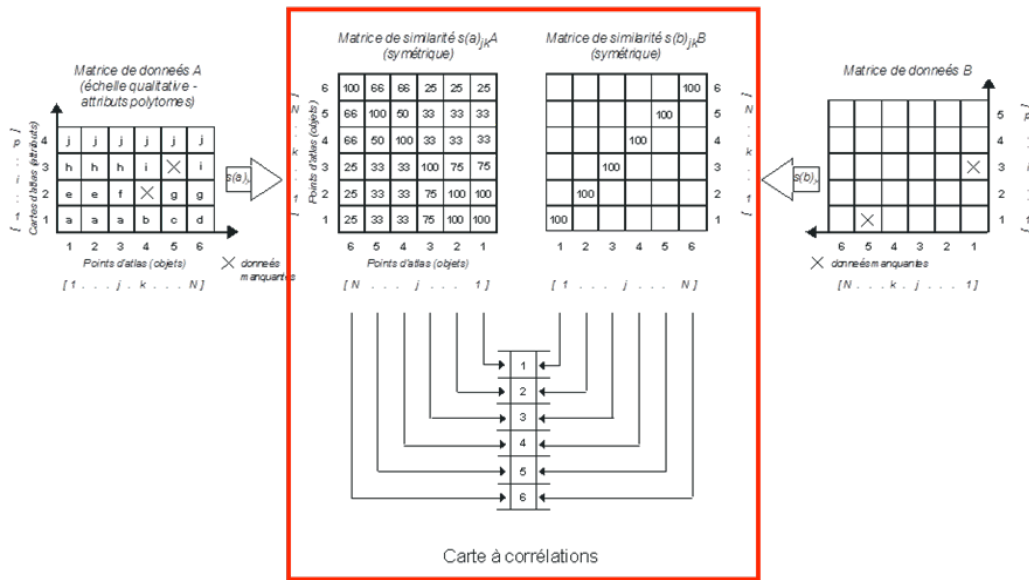


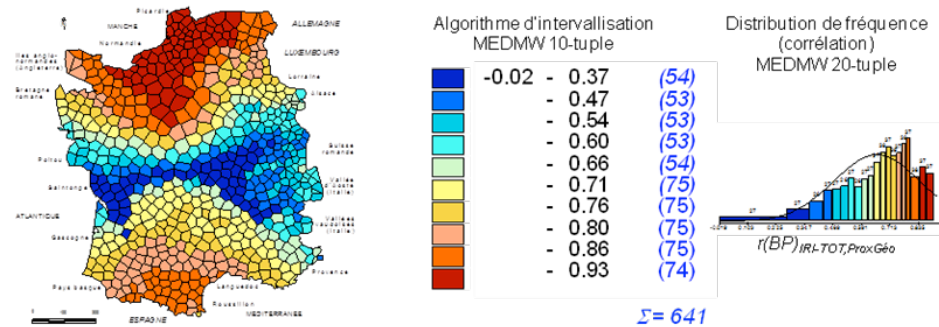
FIGURE 58. Présentation du schéma du calcul d'une carte à corrélations

Carte choroplèthe de la corrélation spatiale
entre les aménagements dialectal et euclidien de la Galloromania

Mensuration de la similarité dialectale: IRI_{jk}

Mensuration de la proximité euclidienne: formule de Pythagore ($a^2 + b^2 = c^2$)

Coefficient de corrélation: $r(BP)_{jk}$



Teintes rouges (etc.):
grande convergence ("harmonie") entre les aménagements dialectal et euclidien de l'espace galloroman: > évolution spatio-linguistique lente et continue

Teintes bleues (etc.):
grande divergence ("disharmonie") entre les aménagements dialectal et euclidien de l'espace galloroman: > évolution spatio-linguistique abrupte et discontinue

FIGURE 59. Présentation d'une carte à corrélations
entre les aménagements dialectal et euclidien de la Galloromania

Le profil choroplèthe de la figure 59 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 dynamismes. Évidemment, de telles disharmonies sont dues à tout ce qui peut influencer de dehors sur l'évolution «naturelle», de la langue dans l'espace, soit en l'accéléralant, soit en la retardant. Les facteurs en question sont bien entendu d'ordre historique, politique, 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, du Massif Central vers les 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 d'un îlot linguistique 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. Il s'agit du P.-ALF 635, Andraut, faisant partie de la micro-région de la «Petite Gavacherie» qui avait été fondée au XIV^e 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. Soit dit en passant, le score minimal de la figure 59 se trouve exactement à l'endroit de P.-ALF 635.

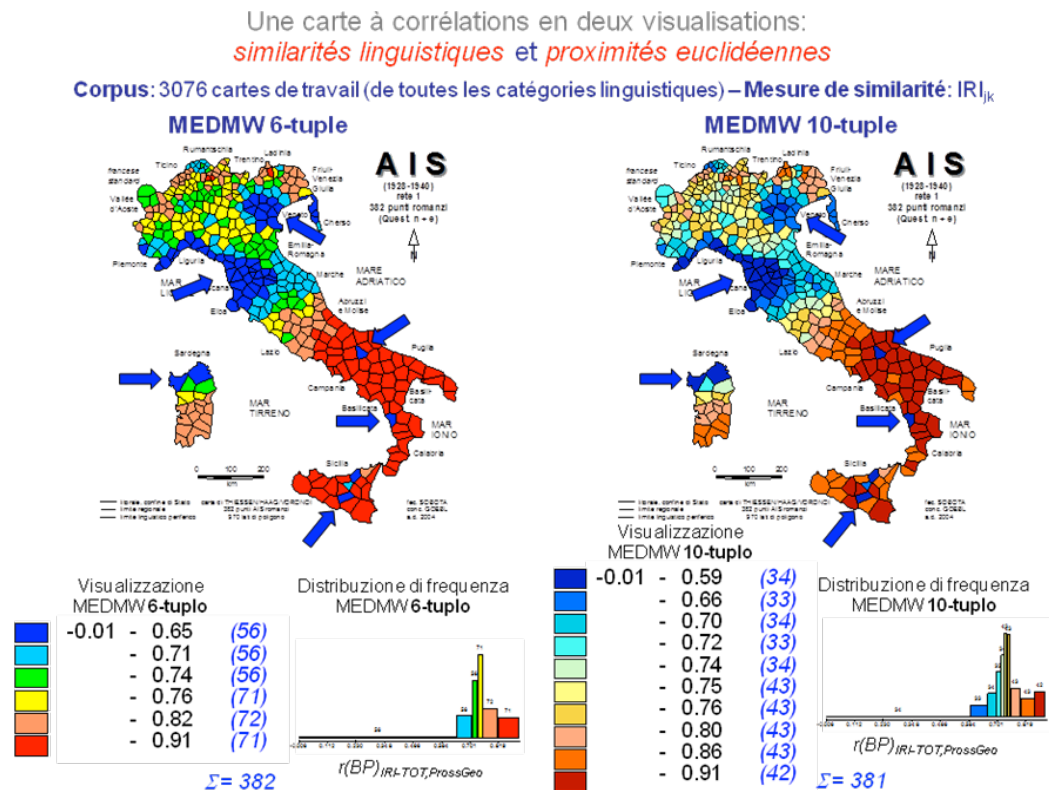


FIGURE 60. Présentation comparative de deux visualisations d'une carte à corrélations entre les aménagements dialectal et euclidien de l'Italoromania

Les deux profils de la figure 60, basés sur les mêmes valeurs numériques mais ayant subi deux visualisations légèrement différentes l'une de l'autre, montrent que l'Italie méridionale

est marquée par une grande harmonie entre la langue et l'espace alors que l'Italie centrale (Toscane) et septentrionale (Ligurie et Vénétie) accusent la présence d'intenses divergences entre ces deux dimensions dont la genèse a dû se produire dans un passé plus ou moins éloigné. L'«inquiétude» linguistique de la Toscane, de la Ligurie et de la Vénétie aurait pu être causée par l'action *rayonnante* de la Toscane d'un côté et sa contrepartie *réceptrice* de la Ligurie et de la Vénétie de l'autre. Signalons en outre que les îlots linguistiques de l' AIS (francoprovençal, occitan alpin, dialectes gallo-italiques de la Sicile et parlers corses au nord de la Sardaigne) se comportent exactement comme l'unique enclave linguistique de l'ALF (P. 635, Andraut) sur la figure 59.

Évidemment, il est possible de répéter le même exploit-DM pour la *phonétique* et l'*espace* d'un côté, ainsi que pour le *lexique* et l'*espace* de l'autre. La figure 61 en montre les résultats.

La corrélation entre l'espace et deux catégories linguistiques différentes
(*phonétique* et *lexique*)

Mesure de similarité: IRI_{jk} – Mesure de proximité: théorème de Pythagore

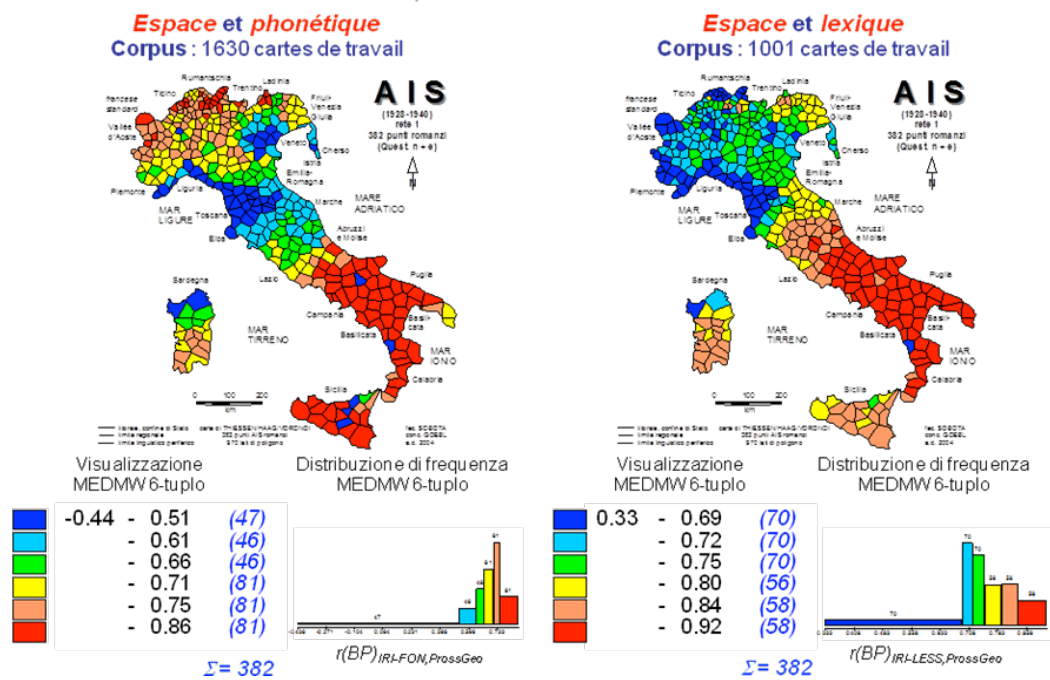


FIGURE 61. Présentation comparative des cartes à corrélations relatives aux aménagements euclidiens et phonétique (à gauche) et euclidiens et lexicaux (à droite) de l'Italoromania

À nouveau, l'Italie méridionale exhibe une situation harmonieuse alors que pour l'Italie septentrionale il s'annonce des situations conflictuelles en partie différentes. Ces différences regardent surtout les domaines du lombard, du piémontais et du romanche des Grisons. Remarquons en outre que pour la corrélation entre le *lexique* et l'*espace* (voir la moitié droite de la figure 61), la position fortement conflictuelle des îlots linguistiques a été considérablement réduite. Ceci s'explique par la facilité avec laquelle, en ligne générale, le lexique d'une enclave linguistique s'adapte à celui de son entourage.

La corrélation entre deux catégories linguistiques différentes

Mesure de similarité: IRI_{jk}

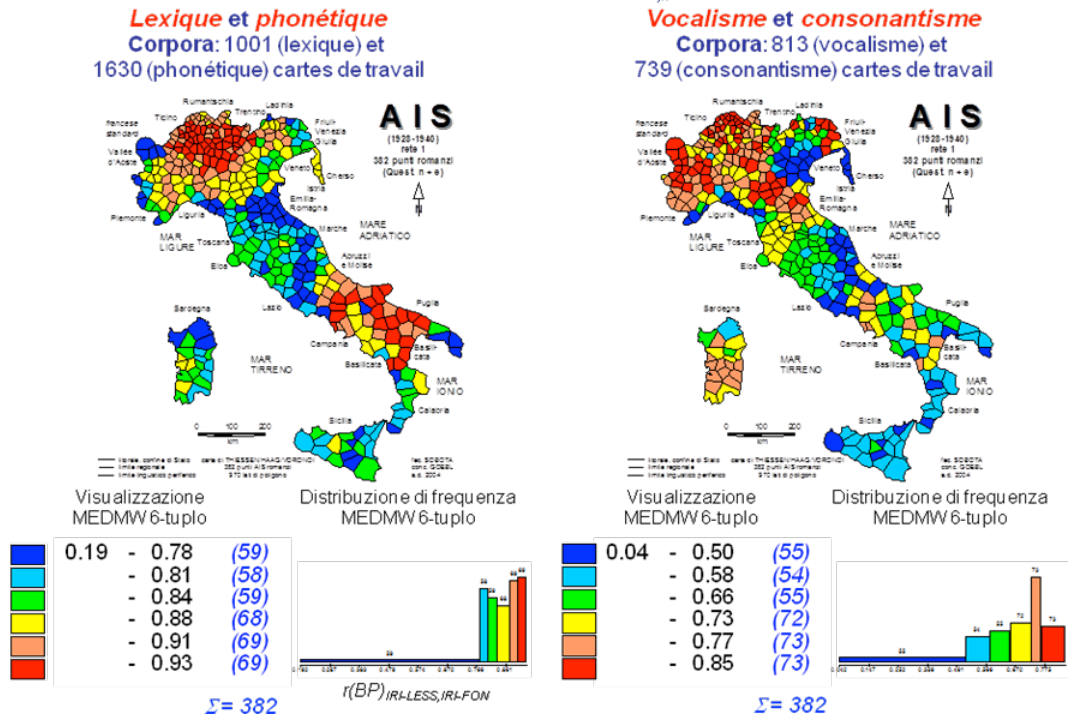


FIGURE 62. Présentation comparative des cartes à corrélations relatives aux aménagements *lexical et phonétique* (à gauche) et *vocalique et consonantique* (à droite) de l'Italoromania

Les deux clichés de la figure 62 se réfèrent à un problème éminemment grammatical. 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 des profils choroplèthes très clairs. Alors que les gestions *lexicale et phonétique* divergent le long des «lignes séparatrices» allant de La Spezia à Rimini et de Rome à Ancône, et qu'une grande convergence des deux gestions catégorielles se manifeste tant au sud qu'au nord de la péninsule, le cas de la corrélation spatiale entre le *vocalisme* et le *consonantisme* est plus complexe. C'est que surtout la Vénétie et la zone située entre Rome et Ancône exhibent les plus grandes dissonances entre la gestion *vocalique* et *consonantique* de l'espace auxquelles participent également la Calabre et la Sicile.

Les deux clichés ibériques sont à nouveau d'une clarté exemplaire (voir la figure 63). Le cliché de droite retrace avec une netteté parfaite les deux couloirs conflictuels (coloriés en bleu plus ou moins foncé) le long desquels s'est effectuée la «Reconquista» entre les VIII^e et XV^e siècles. Le couloir occidental (en bleu foncé) symbolise la totalité des frictions qui devaient exister entre l'action colonisatrice *luso-galicienne* d'un côté et sa contrepartie *castillano-léonaise* de l'autre. Le couloir oriental, par contre, également en bleu plus ou moins foncé, se réfère aux divergences qui accompagnaient l'avancement parallèle vers le sud des colonisations *castillano-léonaise* (au centre) et *catalano-aragonaise* (à l'est).

Deux cartes à corrélations: ALPI et ALDC

Mesure de similarité: IR_{ijk}

Mesure de la corrélation: coefficient de corrélation de Bravais et Pearson $[r(BP)]$

Corpus: 375 cartes de travail

Corpus: 1660 cartes de travail

(toutes les catégories linguistiques)

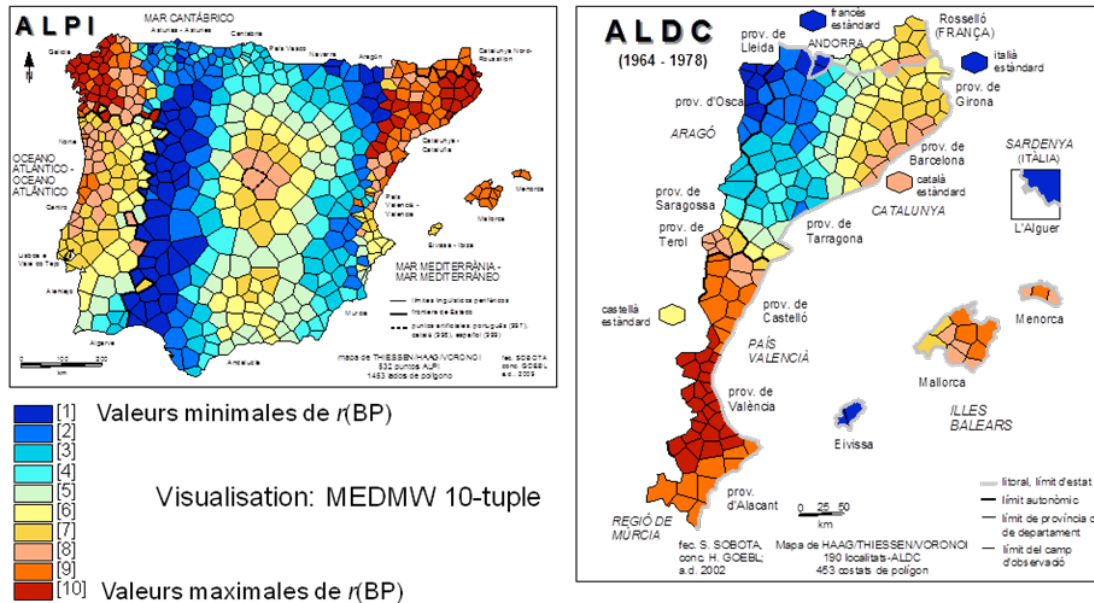


FIGURE 63. Présentation de deux cartes à corrélations entre les aménagements *dialectal* et *euclidien* de l'Ibéroromania et du domaine catalan, tirées de l'Atlas Lingüístico de la Península Ibérica (ALPI) et de l'Atlas Lingüístico del Domini Català

Le cliché de droite, relatif au seul domaine catalan, montre que la plus grande divergence conflictuelle concernait non seulement la zone du catalan nord-occidental mais aussi l'îlot linguistique d'Alghero/L'Alguer situé en Sardaigne ainsi que deux des points-ALDC factices relatifs au français et à l'italien standard.

De tous ces clichés, il se dégage la grande utilité linguistique qu'offrent les cartes-DM à corrélations pour l'élargissement de nos connaissances des mécanismes profonds du fonctionnement de réseaux géolinguistiques tout court.

10. Quelques mises au point cartographiques et conceptuelles

En général, le langage des cartes est bien loin d'être simple ou univoque. Souvent, il est aussi complexe et difficile que celui des chiffres. C'est pourquoi il me semble indiqué de mettre en valeur, en guise de conclusion de cet article, quelques différenciations sémiotiques très importantes.

En matière de classification géolinguistique et partant spatiale, il faut distinguer deux perspectives: la perspective purement *qualitative* ou *monothétique* (voir la figure 64) et sa contrepartie *quantitative* ou *polythétique* (voir la figure 65).

Perspective classificatoire monothétique

ALF 173 brebis (type FEDO)

ALF 812 le marché (résultat é dans -ÂTU)

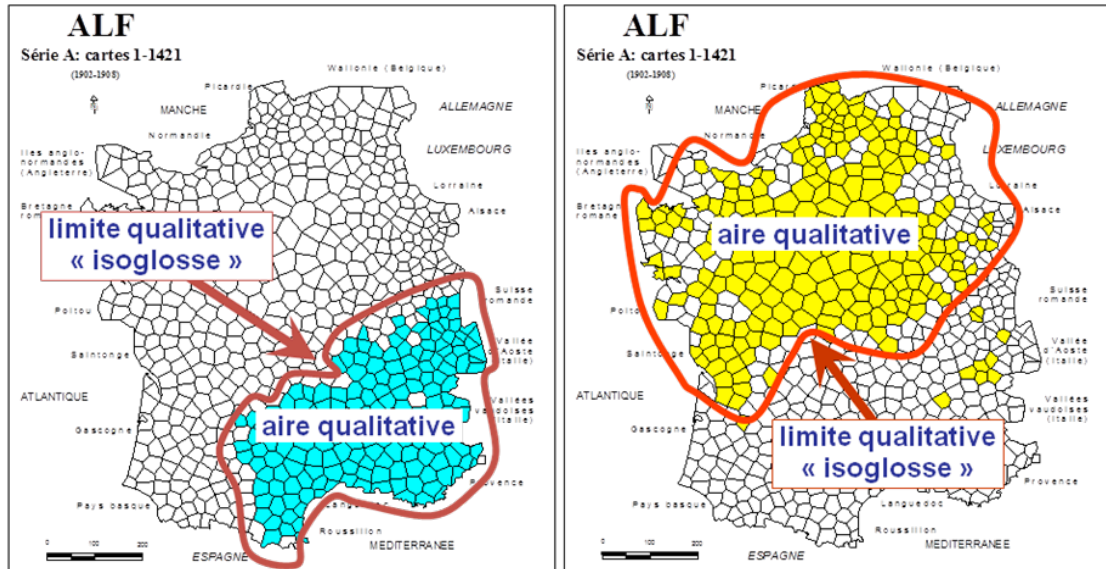


FIGURE 64. Présentation comparative de deux aires taxatoires tirées de l'Atlas Linguistique de la France (ALF)

La figure 64 montre deux phénomènes de type *monothétique*: à gauche, l'aire bleue visualise la répartition du taxat *fédo* pour désigner la brebis (voir aussi la figure 7). L'on y perçoit d'abord l'étendue relativement cohérente de l'aire et ensuite le tracé polygonal de la «frontière» qui l'entoure. Cette frontière s'appelle communément «isoglosse». À droite, l'aire jaune, relative à un phénomène phonétique (voir à cela la figure 8), est déjà moins cohérente quant à sa superficie. Elle est en outre entourée d'une frontière polygonale extérieure très effilochée à laquelle il est convenu d'attribuer également le qualificatif d'«isoglosse» malgré la perte de la cohérence linéaire de son tracé.

Il est d'ailleurs dangereux d'attribuer à ces deux aires des qualificatifs géographiques préétablis (tels que *provençal*, *picard*, *wallon*, etc.) et absolument inadmissible de croire en la coïncidence spatiale entre l'étendue d'aires taxatoires isolées et de régions historiques particulières. Précisons, à ce propos, que du point de vue épistémologique, l'aire de répartition d'un taxat linguistique quelconque se situe au niveau du «particulier» alors que l'aire d'une région historique occupe celui du «général». Ces deux configurations spatiales évoluent donc à des niveaux logiques différents.

La figure 65 montre la mise en carte de classifications spatiales-DM *quantitatives* ou *polythétiques*. L'on y distingue, à gauche, la perspective «isarithmique» (ou linéiforme) et, à droite, la perspective «choroplèthe» (ou chorique) dont le profil à trois paliers chromatiques a été extrait de la carte choroplèthe de gauche de la figure 17. Dans les deux cas, il s'agit, en dernière analyse, de visualisations tridimensionnelles où le *quantitatif* – à savoir le cloisonnement interponctuel d'un côté et la similarité linguistique par rapport au dialecte de Turin de l'autre – évolue dans la troisième dimension. Pour bien comprendre ce qui se passe dans les deux cas, il est indispensable de toujours cerner de près le *sujet thématique* de la visualisation en question.

Perspective classificatoire polythétique

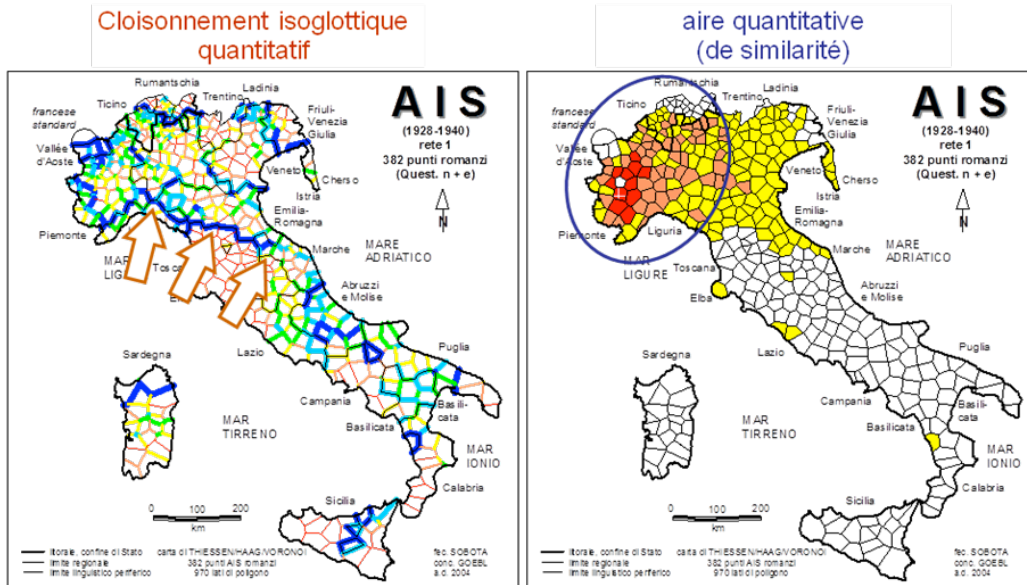


FIGURE 65. Présentation comparative de deux visualisations quantitatives, de type *isarithmique* (à gauche) et de type *choroplèthe* (à droite), tirées du *Sprach- und Sachatlas Italiens und der Südschweiz* (AIS)

Parmi les visualisations présentées dans cet article, le cas de la perspective *multidimensionnelle* est le plus compliqué.

Perspective classificatoire quantitative et multivariée

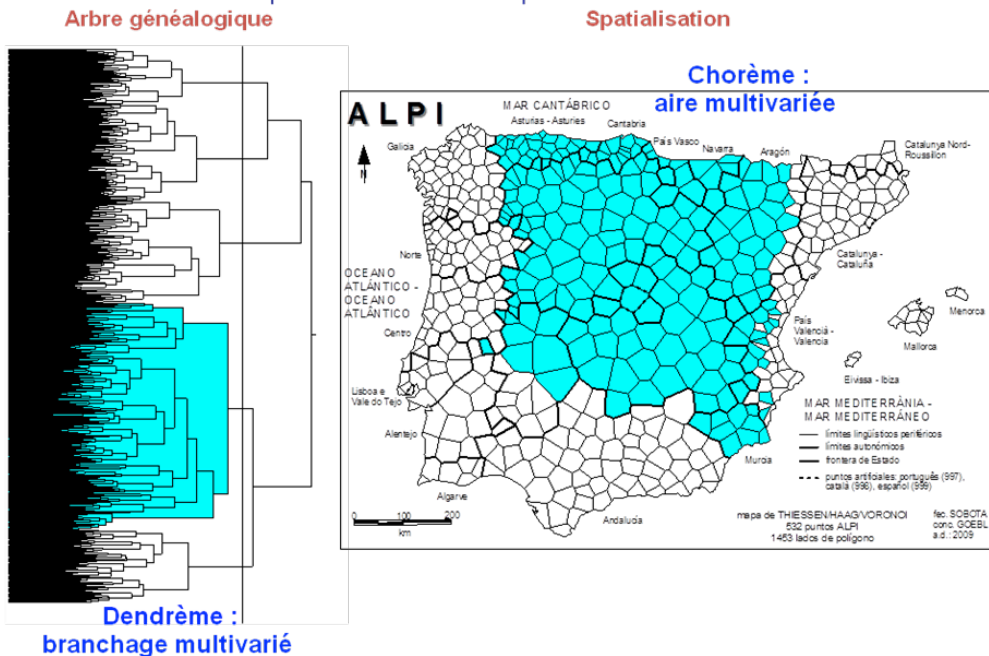


FIGURE 66. Présentation comparative d'un dendrome (à gauche) et du chorème y ayant trait (à droite), tirés d'une analyse dendrographique multivariée (de type WARD) appliquée sur les données de l'*Atlas Lingüístico de la Península Ibérica* (ALPI)

Le chorème bleu situé dans la moitié de droite de la figure 66 ne signale nullement la présence, graduée ou non, d'un phénomène linguistique quelconque, mais la projection spatiale *sommaire* – et partant extrêmement *simplifiée* – d'une information classificatoire finement articulée stockée dans une partie de l'arbre généalogique calculé.

L'effet classificatoire qui s'exprime à travers l'individualité du dendrème bleu à l'intérieur de l'arbre repose en réalité sur un enchevêtrement hiérarchisé de beaucoup de micro-dendrèmes dont la variabilité quantitative interne croît avec l'augmentation de la distance des différentes racines par rapport au niveau des «feuilles» de l'arbre. En même temps, la variabilité interne est régie par les impératifs algébriques de l'algorithme-CAH utilisé. Or, une bonne compréhension de la signification linguistique de cet état des choses ne peut se passer d'une connaissance approfondie des engrenages mathématiques de l'algorithme en question. Ceci n'empêche que la visualisation parallèle de dendrèmes et de chorèmes est toujours très suggestive voire inspiratrice, aussi pour un public non averti.

11. En guise de conclusion: quelques observations finales

En guise de conclusion

- L'importance du concept de la «gestion basilectale de l'espace par l'*Homo loquens*» [GBEsp]
- La GBEsp se manifeste sur tous les niveaux de la faculté langagière de l'*Homo loquens*
- Les atlas linguistiques [AL]: un instrument idéal pour l'étude de la GBEsp
- Les AL → données de masse → analyse par des moyens quantitatifs
- La DM: découverte de structures globales enfouies dans la profondeur des données d'un AL
- «Cosi fan tutte [le scienze]»: la nature interdisciplinaire de la DM (de la géographie humaine à la génétique des populations)

FIGURE 67. Brève évocation du bilan final de la présentation *PowerPoint* de Lisbonne (novembre 2011)

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 – et, partant, saisissables par le biais d'un atlas linguistique de type ALF, AIS, ALPI ou 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 en outre de savoir qu'il y a, en dehors de la géolinguistique, de nombreuses sciences étroitement liées avec l'espace dont les méthodes géo-classificatoires sont très proches 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.²⁵

²⁵ Cf. la contribution interdisciplinaire de Scapoli/Goebel/Sobota/Mamolini/Barrai de 2005.

Remerciements

Établissement technique de la présentation *PowerPoint* de Lisbonne: Werner Goebel (Vienne); préparation et confection définitive des clichés de cet article: Werner Goebel (Vienne), Pavel Smečka (Salzbourg); contrôle linguistique de mon texte français: Lily Ditz-Fuhrich (Salzbourg). Que Mme Ditz-Fuhrich, mon fils Werner et M. Smečka reçoivent ici l'expression de ma sincère et profonde gratitude.

Abréviations

AL:	Atlas linguistique.
CAF:	Coefficient d'Asymétrie de Fischer.
DM:	Dialectométrie, relative à la dialectométrie, dialectométrie.
IRD:	Indice Relatif de Distance.
IRI:	Indice Relatif d'Identité.
MEDMW:	Algorithme d'intervallisation (centré sur la moyenne arithmétique, avec la formation de classes égales en éléments des deux côtés de la moyenne arithmétique).
MINMWMAX:	Algorithme d'intervallisation (centré sur la moyenne arithmétique, avec la formation de n classes, des deux côtés de la moyenne arithmétique, dont la largeur numérique est égale).
VDM:	Visual DialectoMetry (logiciel dialectométrique utilisé à Salzbourg; librement disponible sur demande).

Références

- AIS: Jaberg, Karl & Jakob Jud (1928-1940) *Sprach- und Sachatlas Italiens und der Südschweiz*. 8 vols. Zofingen: Ringier (réimpression: Nendeln: Kraus 1971).
- ALDC: Veny, Joan & Lídia Pons i Griera (2001-2009) *Atles Lingüístic del Domini Català*. 4 vols. Barcelona: Institut d'Estudis Catalans.
- ALF: Gilliéron, Jules & Edmond Edmont (1902-1910) *Atlas Linguistique de la France*. 10 vols. Paris: Champion (réimpression: Bologne: Forni 1968).
- ALPI: (1962) *Atlas Lingüístico de la Península Ibérica*. Vol. I: *Fonética*. Madrid: CSIC.
- Altmann, Gabriel (1985) "Die Entstehung diatopischer Varianten. Ein stochastisches Modell". *Zeitschrift für Sprachwissenschaft* 4. 139-155.
- Brun-Trigaud, Guylaine, Yves Le Berre & Jean Le Dû (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.
- 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 tableaux).
- Gilliéron, Jules (1918) *Généalogie des mots qui désignent l'abeille d'après l'Atlas Linguistique de la France*. Paris: Champion.

- Goebel, Hans (1976) "La dialectométrie appliquée à L'ALF (Normandie)". In Alberto Vàrvaro (ed.) *Atti del XVI Congresso Internazionale di Linguistica e Filologia Romanza* (Naples 1974). Vol. 2. Naples, Amsterdam: Marchiaroli, Benjamins. 165-195.
- Goebel, Hans (1981) "Éléments d'analyse dialectométrique (avec application à l'ALS)". *Revue de Linguistique Romane* 45. 349-420.
- Goebel, Hans (1983) "Parquet polygonal et treillis triangulaire. Les deux versants de la dialectométrie interponctuelle". *Revue de Linguistique Romane* 47. 353-412.
- Goebel, Hans (1984) *Dialektometrische Studien. Anhand italoromanischer, rätoromanischer und galloromanischer Sprachmaterialien aus ALS und ALF*. 3 vols. Tübingen: Niemeyer.
- Goebel, Hans (2002) "Analyse dialectométrique des structures de profondeur de l'ALF". *Revue de Linguistique Romane* 66. 5-63.
- Goebel, 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.
- Goebel, Hans (2005) "La dialectométrie corrélatrice: un nouvel outil pour l'étude de l'aménagement dialectal de l'espace par l'homme". *Revue de Linguistique Romane* 69. 321-367.
- Goebel, Hans (2008) "La dialettometrizzazione integrale dell'ALS. Presentazione dei primi risultati". *Revue de Linguistique Romane* 72. 25-113.
- Goebel, Hans & Norbert Winterleitner (1971) "Projekt einer sprachstatistischen Auswertung von in Sprachatlanten gespeicherter linguistischer Information mit Hilfe elektronischer Rechenanlagen". *Linguistische Berichte* 14. 60-61.
- 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»*. Traduction de A. Llorente et M. Alvar. Granada: Universidad de Granada 1959).
- Scapoli, Chiara, Hans Goebel, Slawomir Sobota, Elisabetta Mamolini, Alvaro Rodriguez-Larralde & Italo Barra (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". *Revue de Linguistique Romane* 35. 335-357.
- Séguy, Jean (1973) "La dialectométrie dans l'Atlas Linguistique de la Gascogne". *Revue de Linguistique Romane* 37. 1-24.