**Zeitschrift:** Revue de linguistique romane

**Herausgeber:** Société de Linguistique Romane

**Band:** 69 (2005)

**Heft:** 275-276

**Artikel:** La dialectométrie corrélative : un nouvel outil pour l'étude de l'aménagement dialectal de l'espace par l'homme

**Autor:** Goebl, Hans

**DOI:** https://doi.org/10.5169/seals-400100

**Nutzungsbedingungen**

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

**Conditions d'utilisation**

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

**Terms of use**

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

**Download PDF:** 11.07.2025

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

# LA DIALECTOMÉTRIE CORRÉLATIVE: UN NOUVEL OUTIL POUR L'ÉTUDE DE L'AMÉNAGEMENT DIALECTAL DE L'ESPACE PAR L'HOMME

## 1. Remarques préliminaires

### *1.1. La dialectométrie corrélative*

Dans les 32 ans de son existence formelle (1973-2005), la dialectométrie (DM) a fait plusieurs fois peau neuve. La mue la plus importante (et lourde de conséquences scientifiques) a été l'implémentation, réalisée dans les années 1999-2000 par mon collaborateur Edgar Haimerl, de toutes les méthodes-DM développées jusqu'alors, dans un logiciel à la fois très puissant et confortable appelé 'Visual DialectoMetry' (VDM). Or, le logiciel VDM a été enrichi, l'an dernier (2004), par un nouveau module dédié entièrement à la DM corrrélative. Dans cet article, il ne sera pas seulement question des multiples possibilités computationnelles et cartographiques offertes par la DM corrélative, mais aussi des nouvelles perspectives scientifiques qu'entraîne cette innovation technologique en matière de linguistique. Néanmoins, vu l'intérêt surtout méthodique de cet article, nos commentaires sur les nouvelles possibilités linguistiques de la DM corrélative seront forcément provisoires et, de ce fait, plutôt sommaires. Il va de soi qu'ils devraient être approfondis dans un autre contexte scientifique.

### *1.2. L'aménagement dialectal de l'espace par l'homme*

C'est à Jean Séguy, créateur du terme de 'dialectométrie'(1), que nous devons la constatation (d'ordre empirique et théorique à la fois) que la fonction des dialectes est double: qu'ils servent tant à la communication qu'à la démarcation symbolique de leurs locuteurs et peuvent, de ce fait, être analysés à partir de ces deux points de vue(2). N'oublions pas non

(1) Cf. Séguy 1973a, 1.
(2) Cf. Séguy 1973b, passim.

plus que les activités en question – c'est-à-dire la communication et la démarcation – émanent d'un comportement conscient (mais aussi subconscient) des locuteurs, lequel, inévitablement, dépend de la condition spatiale (ou géographique) de l'homme. Les humains sont donc, métaphoriquement parlant, 'condamnés' à tout jamais à 'avoir maille à partir' avec l'espace. Qu'ils le veuillent ou non: ils doivent aménager l'espace à l'aide de leur faculté langagière: d'où le concept scientifique de l''aménagement dialectal de l'espace par l'homme (en particulier de l'HOMO LOQUENS)' que nous utilisons depuis un certain temps pour la caractérisation de l'impostation théorique centrale de nos recherches-DM.

Le même concept permet en outre d'établir un contexte de recherche interdisciplinaire étant donné que l'étude et l'analyse de l'aménagement de l'espace en tant que tel représentent un défi théorique et empirique aussi pour d'autres sciences, tant humaines que naturelles, de l'anthropologie et l'ethnologie jusqu'à la géographie humaine et la génétique des populations. Il serait donc hautement souhaitable que s'instaure, sous le signe de *l'aménagement multiple de l'espace par l'homme*, une coopération convergente de toutes les disciplines qui, de loin ou de près, s'intéressent à l'interaction entre l'homme et l'espace(3).

### *1.3. Le caractère foncièrement typologique de l'aménagement dialectal de l'espace par l'homme*

L'on sait que la typologie, éminente sous-branche de la linguistique générale, s'occupe de l'analyse des traits retenus comme 'typiques' ou 'constitutifs' d'une partie plus ou moins grande des quelque 4000 à 6000 langues de notre planète. Jusqu'à maintenant, les recherches typologiques à proprement parler se sont limitées à des études comparatives sur le plan de la substance intra-linguistique des langues (phonétique, morphologie, syntaxe, etc.) tout en négligeant, hormis quelques louables exceptions (cf. Nichols 1992 et le WALS), l'étude des mécanismes et de la logique des imbrications spatiales des traits intra-linguistiques respectifs des langues analysées. Comme, en ligne générale, nulle langue ne peut être parlée par des locuteurs a-spatiaux et que chaque langue dispose, inévitablement, d'une grande richesse de traits langagiers qu'elle partage aussi avec d'autres langues, l'étude de leur enchevêtrement spatial devrait faire, à plus forte raison, partie intégrante de la recherche typologique en tant que telle. Il est donc grand temps que la branche aréale de la typologie linguistique

(3) A ce propos nous renvoyons, en guise d'exemple, au travail interdisciplinaire de Scapoli et al. 2005.

moderne[4] s'en rende compte et, partant, jette un coup d'œil intéressé aussi sur les prérequis théoriques et les acquis méthodologiques de la DM, que celle-ci soit corrélative ou non.

*1.4. «Extra mappas nulla salus dialectometrica sive geolinguistica»*

Cet adage latin, forgé sur le modèle d'un autre longtemps utilisé («Nulla salus extra ecclesiam»), souligne une nécessité que je considère comme absolue en matière de géographie linguistique et, partant, aussi de dialectométrie: la conception, la confection, l'usage et l'interprétation de cartes. Les phénomènes dont la géographie linguistique s'occupe depuis plus de cent ans, ne peuvent se manifester et dérouler que dans l'espace. Or, vu les conditions matérielles et les contraintes psychologiques qui sont les nôtres, il est de toute première importance que les géolinguistes reportent les faits observés par eux sur des schémas iconiques appropriés et qu'ils s'y réfèrent, en permanence, dans leurs réflexions et discussions. Qui, en tant que géolinguiste, renonce à l'usage de cartes ou finit même par les répudier[5], anéantit de ce fait sa propre discipline. En bonne logique, le lien indissoluble qui existe entre la géolinguistique et les cartes concerne aussi toutes les réflexions et démarches, tant théoriques que pratiques, qui sous-tendent la conception et la fabrication de cartes géolinguistiques. À l'instar des géographes (humains ou physiques), des géologues ou d'autres tenants de disciplines attachées à l'étude de l'espace, les géolinguistes à proprement parler doivent donc vouer une attention toute particulière à la cartographie, pour l'application et la pratique de laquelle l'informatique moderne offre d'ailleurs de multiples possibilités.

Soulignons que le recours aux cartes tel qu'il est pratiqué en matière de DM remplace, très souvent, l'usage de formules mathématiques et de tests statistiques. C'est qu'une carte thématique quantitative – dont 24 accompagnent cet article – remplit plusieurs fonctions épistémologiques, allant de la théorisation initiale jusqu'à la vérification d'hypothèses ou de suppositions probabilistes. Si Jean Séguy, dans son article stimulant de 1971 sur la relation entre les distances linguistiques et spatiales, n'avait pas dû, faute de moyens, se limiter au tracement de courbes et à la présentation de simples tableaux numériques et qu'il eût donné une expression vraiment cartographique à ses idées, l'impact de sa pensée innovatrice aurait été beaucoup plus grand, tant à l'intérieur qu'à l'extérieur des études romanes à proprement parler.

---

(4) Cf. Dahl 2001, passim. Nous renvoyons ici expressément au WALS ('World Atlas of Language Structures') récemment paru (été 2005). Les données du WALS constituent (entre autres) une matrice de données très grande susceptible d'être dialectométrisée en bonne et due forme.

(5) Voir à ce sujet Grassi 2001, 226-230.

Rappelons à ce propos que, dans sa contribution capitale de 1971, il présente plusieurs tableaux numériques relatifs à un certain nombre d'atlas linguistiques romans – parmi lesquels figure aussi l'AIS – dont les chiffres se réfèrent à l'augmentation de la distance linguistique le long d'une ligne droite tracée arbitrairement à travers les réseaux de plusieurs atlas linguistiques. Le tableau relatif à l'AIS contient les valeurs relevées le long d'une ligne droite allant des confins occidentaux de la Ligurie jusqu'au Frioul. En visualisant les scores collectés par Séguy tout en les comparant avec les nôtres relevés au cours de la dialectométrisation, sur la base des volumes I, II et IV, de la partie septentrionale de l'AIS, l'on obtient un graphique très simple qui montre deux courbes dont la valeur heuristique – pour être sincère – est d'ailleurs très limitée: voir la figure 1. Ajoutons encore qu'en établissant les 'courbes' de la figure 1, nous avons remplacé les concepts séguyens des *distances* linguistique et géographique par leurs compléments logiques, c'est-à-dire par les *proximités* linguistique et géographique.

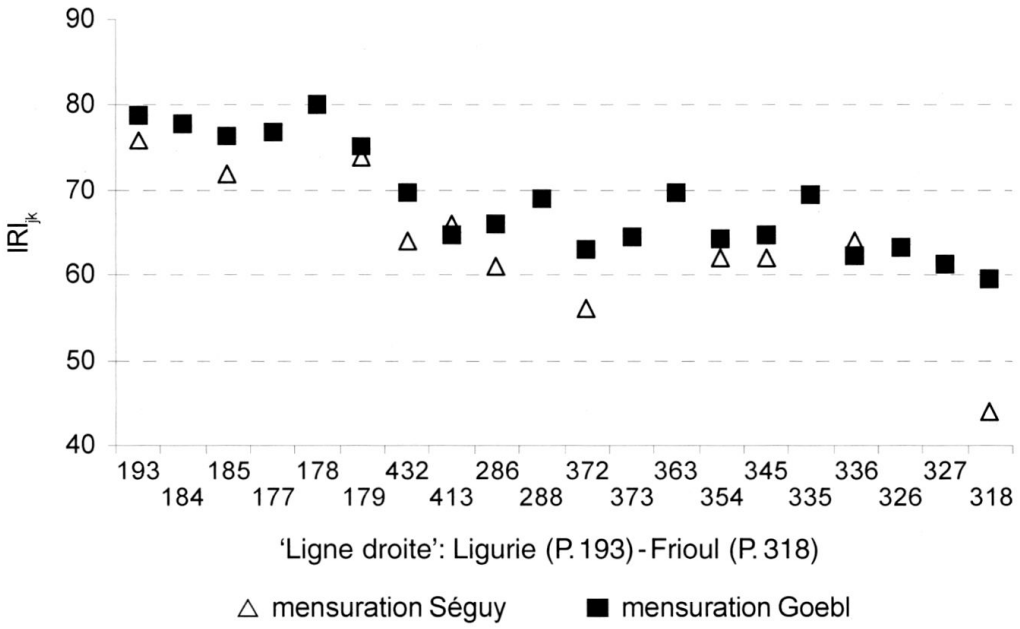



Figure 1: La relation entre les proximités linguistique et géographique le long d'un parcours d'épreuve ('ligne droite') tracé au sein du réseau de l'AIS entre la Ligurie [AIS-P. 193] et le Frioul [AIS-P. 318] (d'après Séguy 1971, 344, et Goebl 1984 II, 284-293).

Certes, il peut être intéressant, du point de vue mathématique, de chercher une courbe théorique qui s'adapte aussi bien que possible à la distribution empirique présentée ci-dessus. Ce travail a d'ailleurs été fait

à plusieurs reprises par Henri Guiter(6). Étant donné que les valeurs empiriques relatives à une quelconque ligne droite tracée arbitrairement à travers un domaine linguistique donné ne sauraient jamais dépasser le rang d'un échantillon fortuit, j'ai depuis toujours nourri de sérieux doutes au sujet de cette procédure.

Ceci n'empêche qu'il faut reconnaître qu'il y a tout de même un grain de vérité. Mais, pour passer d'un grain de vérité isolé à la vérité tout entière, il faut essayer de transporter, dans un graphique à double entrée (sur l'axe x: distances/proximités kilométriques, sur l'axe y: distances/proximités linguistiques), non pas les quelques valeurs d'une ligne droite choisie au petit bonheur, mais le contenu d'une matrice de similarité tout entière: voir la figure 2.

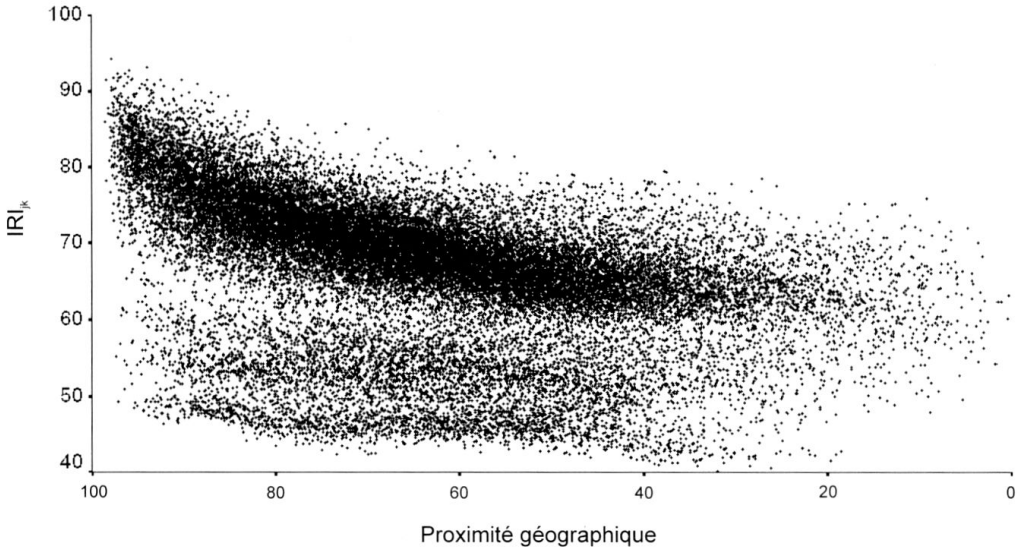



Figure 2: Diagramme de dispersion ('scattergram') de 31 375 valeurs de similarité linguistique (d'après l'IRI$_{jk}$) [en ordonnée] et de proximité géographique [en abscisse] (d'après Goebl 1984 I, 114 ss.).

Commentaire: Les minces agglomérations nébuleuses au-dessous du nuage principal se réfèrent principalement aux dialectes rhéto-romans et francoprovençaux du réseau de l'AIS.

(6) Cf. Guiter 1984. Dans un compte rendu (publié in: *Zeitschrift für romanische Philologie* 102, 1986, 499) de l'article en question, le linguiste suisse Wulf Müller s'exprime comme suit: «Henri Guiter [...] zieht mit dem Lineal Linien durch Europa, um dann die lexikalische Distanz zwischen den Hochsprachen der durchquerten Länder zu bestimmen. Kann die Fragestellung noch als vernünftig bezeichnet werden? Oder wollte der Verfasser damit die Séguy'sche Dialektometrie ad absurdum führen?»

C'est ainsi que l'on passe, du point de vue heuristique, de la perception d'une modeste courbe à celle d'un (ou même de plusieurs) nuage compact dont la configuration laisse entrevoir, dans leur plénitude, les dimensions dans lesquelles se déploient les deux espèces de proximité (ou de distance), celle des dialectes et celle de l'espace euclidien. Un premier coup d'œil sur la configuration compacte de l'agglomération nébuleuse centrale de la figure 2 nous dit qu'il doit exister, dans la totalité des données analysées en la matière, entre les proximités géographiques et linguistiques, un rapport tout autre qu'aléatoire. Il va de soi que l'établissement de tels graphiques ne peut être que l'étape initiale d'une recherche relative à l'interdépendance de ces deux variables.

Evidemment, au vu de telles configurations nébuleuses, le comportement de l'observateur sera différent suivant ses propres intérêts scientifiques. Alors que le statisticien tentera, très vraisemblablement, de calculer, pour le nuage de la figure 2, une courbe de régression, le géolinguiste se penchera, par contre, sur la question de savoir dans quelle mesure la structuration toute somme faite bien marquée de la figure 2, se reflète dans l'espace respectif et essayera donc de reporter dans l'espace (naturel ou géographique) l'information corrélative de la figure 2.

Or, tel était notre propos depuis le début de nos propres travaux dialectométriques. C'est de là que résulte notre perpétuel effort de mettre non pas en courbes (et autre schémas non-cartographiques) mais en cartes les résultats de nos mensurations dialectométriques.

### *1.5. L'intérêt général de l'étude des corrélations spatiales*

Les géolinguistes se sont penchés plus d'une fois sur la question de savoir dans quelle mesure les aires de traits linguistiques appartenant à des catégories grammaticales différentes se ressemblent ou se distinguent les unes des autres(7). Ces questions ne touchaient pas seulement les traits phonétiques par rapport aux traits lexicaux mais aussi l'intercorrélation qui existe entre la distance (mieux: similarité) linguistique d'un côté et la distance (mieux: proximité) géographique (ou kilométrique) de l'autre (cf. Séguy 1971, passim). Cette dernière question intrigue d'ailleurs, depuis longtemps, également les généticiens des populations (humaines ou autres).

---

(7) Voir à ce sujet plusieurs passages très pertinents dans l'ouvrage classique de Karl von Ettmayer (de 1924) sur l'essence de la formation des dialectes («Über das Wesen der Dialektbildung»): 12-13, 44-47 et le tableau («Tafel») VII.

Il est donc naturel que l'usage qui pourrait être fait de la nouvelle DM corrélative soit multiple (ou carrément interdisciplinaire) et que cette dernière puisse fournir des réponses claires et innovatrices à des questions traditionnelles. Il est cependant non moins évident que la DM corrélative soulève aussi des questions franchement nouvelles et complètement inconcevables jusqu'à une date récente auxquelles il est cependant difficile de fournir des réponses adéquates aujourd'hui même.

## 2. De la carte de similarité à la carte à corrélations

Il est temps de se pencher sur le côté purement technique de la DM corrélative: voir la figure 3.

L'on y discerne, dans la partie supérieure, deux configurations métrologiques parfaitement symétriques l'une face à l'autre, dont chacune comprend le schéma d'une petite matrice de données (avec 6 points d'atlas) et la matrice de similarité (aux dimensions 6 fois 6) respective. Le contenu de la configuration métrologique de gauche (caractérisée par la lettre A) correspond exactement à celle que nous utilisons, évidemment à des fins purement explicatives, dans tous nos écrits dialectométriques publiés depuis 1984. L'indice de similarité appliqué $[s(a)_{jk}]$ correspond à l'Indice Relatif d'Identité ($IRI_{jk}$). C'est donc par l'application de l'$IRI_{jk}$ que l'on obtient les scores répertoriés dans les cellules de la matrice de similarité A.

La configuration métrologique de droite (caractérisée par la lettre B) doit disposer, pour permettre le calcul des corrélations entre les 6 vecteurs d'objets (ou de points d'atlas) des deux matrices de similarité en question (A et B), également de 6 (= N) points d'atlas alors que le nombre des attributs (ou cartes d'atlas) retenus dans la matrice de données respective (B) peut varier par rapport à celle de la matrice de données A. Voilà pourquoi la matrice de données de droite (B) dispose de 5 au lieu de seulement 4 attributs.

C'est ainsi que l'on obtient, après l'établissement des deux matrices de similarité aux dimensions extérieures strictement identiques (N fois N), la possibilité de comparer, par voie numérique, N paires $[s(a)_j/s(b)_j]$ de scores de similarité. Or, la réalisation quantitative d'une telle comparaison numérique par paires peut se prévaloir, depuis longtemps, des acquis de la statistique corrélative.

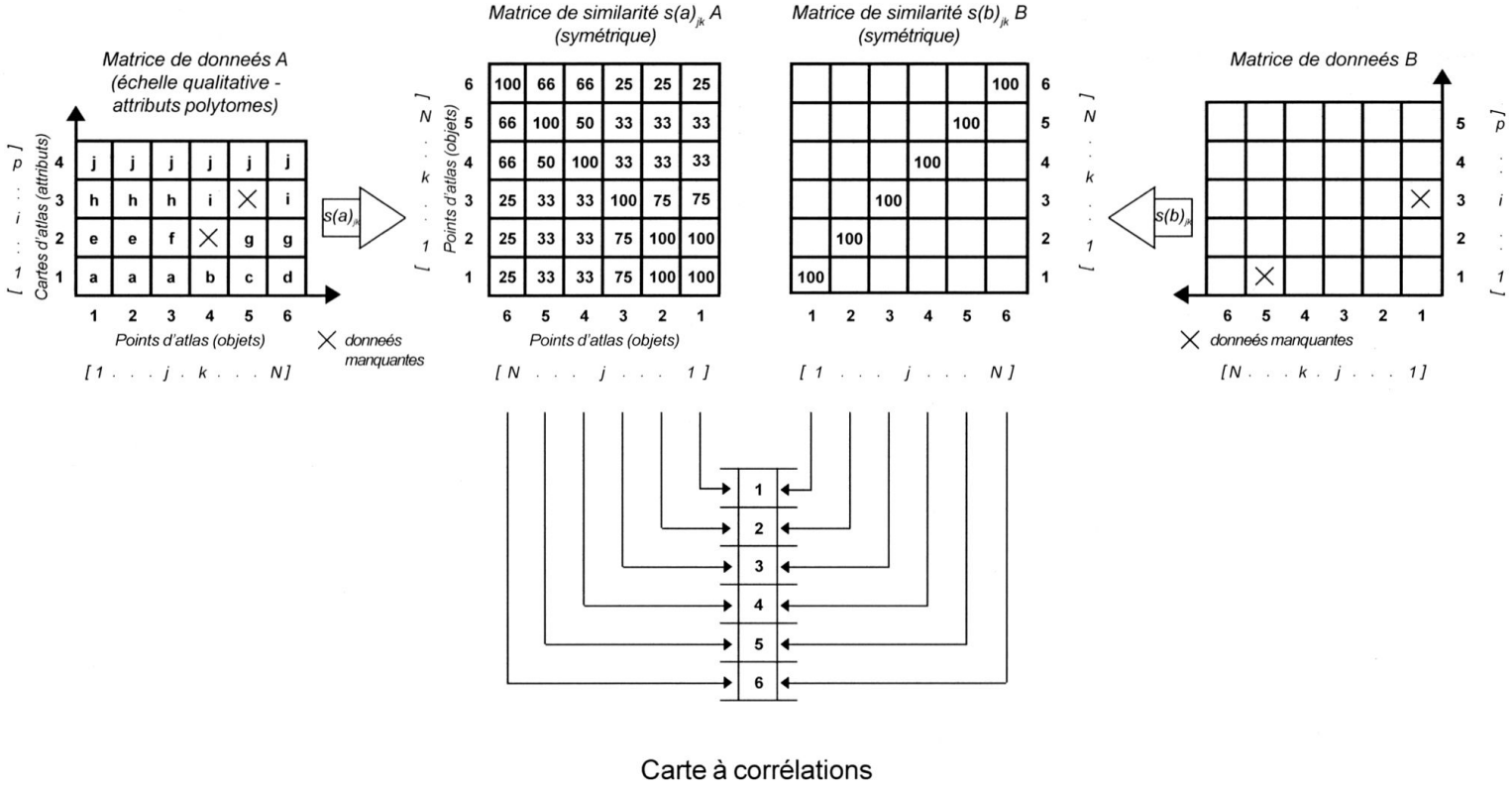



Figure 3: Schéma du calcul de la carte à corrélations à l'aide de deux matrices (carrées) de similarité (ou de proximité, de distance, etc.) disposant de N vecteurs d'objets.

L'indice qui avait été choisi pour le calcul de la corrélation (numérique et linéaire) entre les N paires de valeurs tirées de deux matrices carrées prises en considération, est le 'coefficient de corrélation de Bravais[8] et Pearson[9]' [r(BP)]. En voici la formule[10]:

$$r(BP) = \frac{\sum_{i=1}^{N} (x_{jk} - \bar{x}_j)(y_{jk} - \bar{y}_k)}{\sqrt{\frac{1}{N}\sum_{i=1}^{N} (x_{jk} - \bar{x}_j)^2 \; \frac{1}{N}\sum_{i=1}^{N} (y_{jk} - \bar{y}_k)^2}}$$

Signification des symboles utilisés:

| | |
|---|---|
| $j$ | index du vecteur du point de référence $j$ de la matrice de similarité A |
| $k$ | index du vecteur du point de référence $k$ de la matrice de similarité B |
| $x_{jk}$ | score de similarité entre les vecteurs $j$ et $k$ de la matrice de similarité A |
| $y_{jk}$ | score de similarité entre les vecteurs $j$ et $k$ de la matrice de similarité B |
| $\bar{x}_j$ | moyenne arithmétique des scores de similarité du vecteur $j$ de la matrice de similarité A |
| $\bar{y}_k$ | moyenne arithmétique des scores de similarité du vecteur $k$ de la matrice de similarité B |
| N | nombre des points d'atlas (objets) disponibles |
| $i$ | un des N vecteurs des deux matrices de similarité (A et B) |

Les valeurs du r(BP) oscillent entre -1 (corrélation négative), 0 (corrélation inexistante) et +1 (corrélation positive ou parfaite). Ceci signifie que des scores-r(BP) proches de la valeur +1 renvoient à une certaine interdépendance des deux variables intercorrélées dont l'intérêt scientifique respectif devra d'ailleurs être analysé de plus près par le chercheur concerné. Plus les scores-r(BP) s'approchent de la valeur 0, moins l'inter-

(8) Auguste Bravais (1811-1863), naturaliste et cristallographe français.

(9) Karl Pearson (1857-1936), mathématicien anglais cultivant des intérêts scientifiques très variés.

(10) Pour la genèse, l'explication et la fonction de cette formule voir aussi Bahrenberg/Giese 1975, 144-151 (en allemand), Chandon/Pinson 1981, 65 (en français) et Legendre/Legendre 1998, 139-144 (en anglais).

dépendance postulée est probable. Elle est carrément nulle pour r(BP) = 0. Quant aux scores-r(BP) situés entre 0 et -1, ils signalent une corrélation *inverse* des deux variables examinées. Ceci signifie, entre autres, qu'à l'augmentation des scores de la variable x correspond, d'une façon plus ou moins systématique, une diminution parallèle des scores de la variable y.

Or, le géolinguiste-dialectométricien, bien loin de s'arrêter aux calculs des N scores de corrélation, vise à la mise en carte de ces derniers. Toute carte à corrélations constitue donc la visualisation de N scores de corrélation: voir à ce sujet la partie inférieure de la figure 3.

Comme nous l'avons déjà précisé, ces N valeurs de corrélation sont les résultats de N calculs de corrélation, effectués par paires de vecteurs, à l'aide des N vecteurs de deux matrices de similarité (ou de proximité, de distance, etc.). Il en appert que les deux matrices en question doivent disposer exactement des mêmes dimensions (N). Il est également nécessaire que les scores numériques qu'elles contiennent doivent osciller à l'intérieur d'une marge quantitative bien définie et suivant des modalités statistiques similaires. Si tel n'est pas le cas, il est indiqué de soumettre les scores des deux matrices (carrées) à une transformation numérique appropriée avant de procéder au calcul des N corrélations(11). Or, la nature numérique des données utilisées dans cet article est telle que cette mesure correctrice devient superflue.

Jetons, pour mieux comprendre cet état des choses, un coup d'œil sur la figure 4. L'on y discerne deux histogrammes dont chacun repose sur 50 colonnes de longueur inégale, et qui se réfèrent aux données de l'ALF utilisées dans cet article. L'histogramme de gauche constitue la visualisation de tous les 205 120 scores de similarité stockés dans la matrice de similarité (calculée à l'aide de l'indice $IRI_{jk}$) relative à l'ALF alors que l'histogramme de droite sert à la visualisation des 205 120 valeurs de proximité géographique (GeoProx) répertoriées dans la matrice de proximité. Ces deux histogrammes constituent donc la somme de 641 histogrammes particuliers dont les cartes 1, 2, 15 et 16 montrent des échantillons isolés.

Au premier abord, l'on perçoit que la nature mathématique (ou statistique) des deux histogrammes (ou courbes) est fort différente et que l'apparente régularité de la courbe de droite contraste avec l'allure quelque peu inégale de la courbe de gauche qui, elle, est caractérisée surtout

(11) Cf. Legendre/Legendre 1998, 33-47.

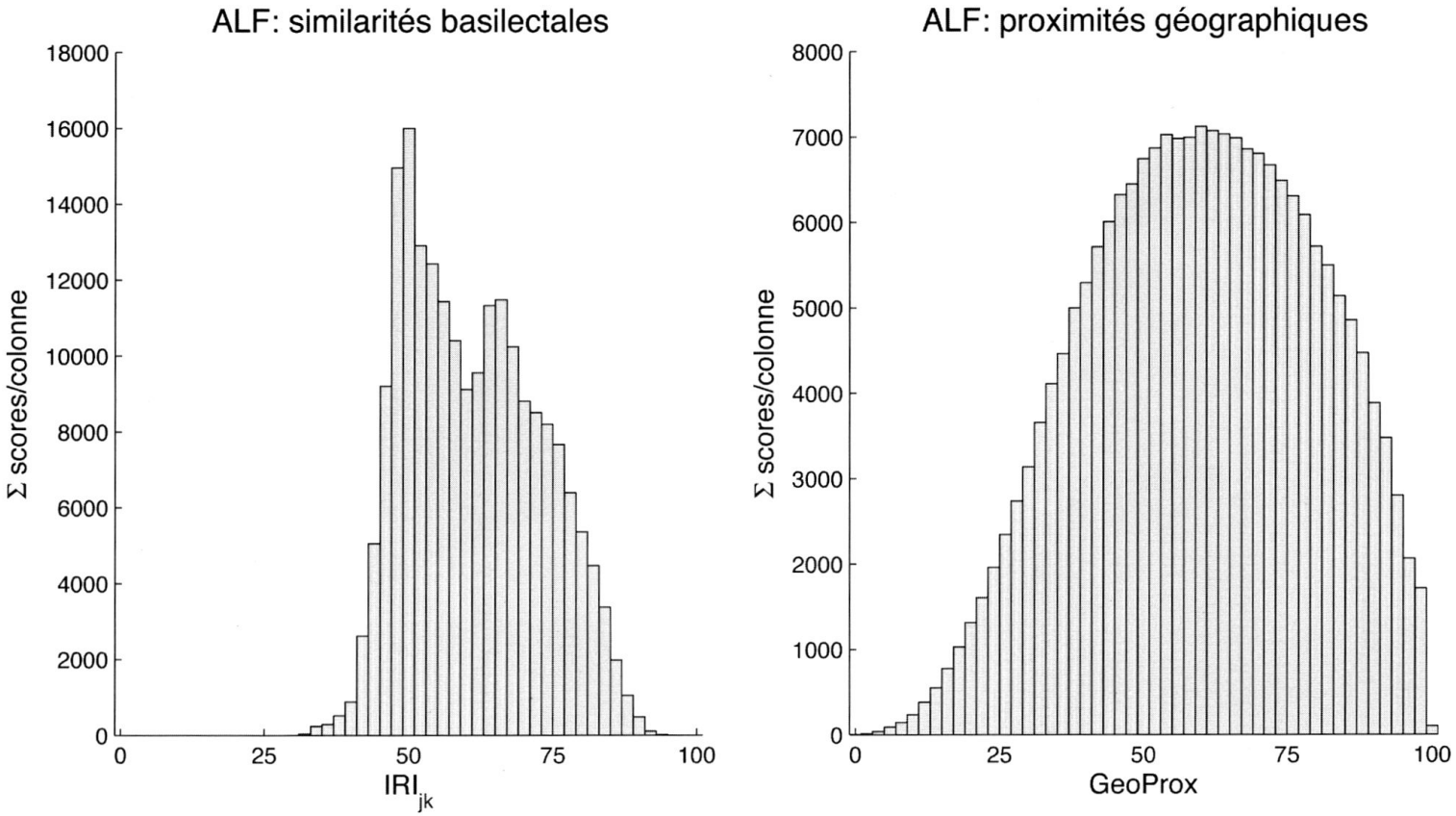



Figure 4: Histogrammes des 205 120 scores de la similarité basilectale (selon l'$IRI_{jk}$) (à gauche) et de la proximité euclidienne (selon le théorème de Pythagore) (à doite).

par une bimodalité (c.-à-d. disposant de deux 'modes' ou pics) clairement marquée. Cette bimodalité s'explique par la coexistence, au sein du réseau de l'ALF, de deux (géo)types très divergents correspondant aux domaines d'Oïl et d'Oc.

Il serait donc intéressant de croiser ces deux dimensions relationnelles, relatives aux aménagements basilectal et euclidien de l'espace de l'ALF, et d'en tirer, si possible, une leçon cartographique adéquate. Examinons, à ce propos, encore une fois la figure 3.

La figure 3 montre donc le schéma de la mise en carte de N (ici: 6) valeurs de corrélation, dont chacune dispose cependant d'une appartenance géographique particulière. Le souci du dialectométricien sera donc avant tout de redonner à ces N scores de corrélation géographiquement 'déracinés' durant l'exécution du calcul des corrélations, leur contexte spatial original, et d'en tirer une carte choroplèthe dûment visualisée. Ce faisant, nous écartons d'autres possibilités (statistiques ou mathématiques) de traiter les N scores de corrélation calculés. Celles-ci consisteraient, entre autres, dans l'application de N tests de signification isolés qui du reste ne pourraient jamais porter sur les structures choroplèthes globales à notre disposition. C'est pourquoi ces tests ne sauraient nous fournir que des informations particularisantes(12) et, qui plus est, ne toucheraient pas la nature cartographique de nos cartes à corrélations. L'œil du linguiste-dialectométricien est, par contre, désireux de capter des structures globales, synthétiques et, de surcroît, mises en cartes, et d'en tirer des leçons d'ordre (géo)linguistique.

## 3. La corrélation entre les aménagements dialectal et euclidien de l'espace

Nous présenterons par la suite quatre applications de la DM corrélative tirées toujours d'un autre atlas linguistique ou d'une collection de

(12) Un test classique en la matière est le test de Mantel: cf. Legendre/Legendre 1998, 552-557. Ajoutons que pour l'application correcte d'un test statistique inférentiel, il est également nécessaire de définir, au préalable, une hypothèse nulle appropriée pour le rejet ou l'acceptation de laquelle le test en question sera utilisé. Or, la formulation d'une telle hypothèse nulle sera, dans notre contexte de recherche, très difficile. C'est qu'elle devrait porter sur la signification statistique (et autre) de structures (angl. *patterns*, all. *Muster*) géographiques entières, en l'occurrence sur celles de nos cartes à corrélations.

données géolinguistiques similaire. En particulier, il s'agit des recueils atlantographiques suivants: ALF, AIS, ALD-I et CLAE (SED). Chaque application sera illustrée par trois cartes choroplèthes: deux des trois cartes choroplèthes en question sont des cartes de similarité ou de proximité (disposant toujours du même point de référence) dont le but est de montrer aux lecteurs dans quelle mesure les deux espèces d'aménagement du territoire convergent ou s'écartent l'une de l'autre. Comme il serait possible de présenter, au moins théoriquement, pour chacune des quatre applications mentionnées ci-dessus, N paires de cartes choroplèthes (de similarité et de proximité), il se vérifie que le message iconique émanant d'une seule des N paires de cartes est très limité. Ceci n'empêche que l'œil (et l'esprit) du dialectométricien ne serait jamais en mesure de parvenir, même par la considération comparative de N paires de cartes, à une vision d'ensemble scientifiquement valable. Il est donc tout à fait légitime de se limiter ici à la présentation exemplaire de quelques échantillons (choisis plus ou moins de façon aléatoire) et de confier le processus de synthèse à la cartographie statistique, c'est-à-dire au jeu combiné de la statistique et de la mise en carte judicieuse des scores respectifs.

Ajoutons encore que la carte de proximité et les scores de la proximité euclidienne sur lesquels elle repose font fonction, dans la démarche scientifique qui est la nôtre, d'un modèle théorique avec lequel la réalité empirique des similarités linguistiques relevées sera confrontée. On verra par la suite que la comparaison de ces deux niveaux épistémiques fournit des résultats très intéressants et, qui plus est, complètement inédits en matière de géolinguistique.

Par la suite, chacun des quatre recueils atlantographiques utilisés, ainsi que les modalités de leur traitement dialectométrique préalable, seront présentés sous la forme de tableaux standardisés avec trois entrées principales: la *première* se réfère à la dénomination officielle de la source utilisée, la *deuxième* aux modalités du traitement dialectométrique respectif (avec l'indication du nom du chercheur responsable et des références bibliographiques les plus importantes) et la *troisième* aux dimensions (nombre des points d'enquête du réseau utilisé, nombre des cartes de travail synthétisées) de la matrice de données respective, et aux prérequis métrologiques (ou taxatoires) utilisés pour l'établissement de la matrice de données en question. C'est ainsi que les lecteurs sont à même de confronter les différents projets dialectométriques et d'en repérer rapidement les coordonnées quantitatives et qualitatives.

### *3.1. L'exemple de l'ALF*

Voir les cartes 1, 2 et 9.

| | | |
|---|---|---|
| 1. | Source originale | Atlas linguistique de la France (ALF) |
| 2. | Dialectométrisation | |
| | Auteur responsable | H. Goebl |
| | Références bibliographiques | Goebl 2000, 2002, 2003 et 2004a |
| 3. | Matrice de données | |
| | Critères de taxation | variations phonétiques, morphologiques et lexicales |
| | Nombre des points du réseau dialectométrisé | 638 + 3 (points artificiels) |
| | Nombre des cartes de travail | 1687 (corpus intégral) |

Les cartes 1 et 2 sont des cartes de similarité qui indiquent le décalage spatial des proximités linguistiques et géographiques par rapport à un point de repère préalablement sélectionné. Nous avons choisi comme tel un point-ALF situé en position centrale (ALF-P. 1, Marcigny, Département Nièvre). Du point de vue iconique, les cartes 1 et 2 reposent sur les mêmes principes cartographiques (algorithme d'intervallisation MINMW-MAX 6-tuple, arrangement des six teintes chromatiques choisies suivant la logique du spectre solaire(13)). La carte 1 visualise 640 scores de similarité saisis à l'aide de l''Indice Relatif d'Identité' ($IRI_{jk}$)(14). La carte 2 en fait autant pour 640 scores de proximité qui ont été calculés à l'aide du théorème bien connu de Pythagore ($a^2 + b^2 = c^2$)(15).

(13) Pour toutes les conventions cartographiques utilisées cf. Goebl 1981, 365-368; 1982, 25-30 et 1984 I, 93-98.

(14) Pour une explication détaillée de l'$IRI_{jk}$ cf. Goebl 1981, 357-361; 1982, 23-24 et 1984 I, 74-78.

(15) La relation entre les concepts de proximité (prox) et de distance (dist) est la suivante: prox + dist = 100. Comme les coordonnées cartésiennes (en x et y) de tous les points d'un réseau d'atlas considéré sont à la disposition du logiciel VDM, le calcul des distances (euclidiennes) entre les points d'atlas est vraiment chose facile. Les distances ainsi déterminées seront converties, par la suite, en valeurs de proximité (selon la formule indiquée ci-dessus) et, en outre, standardisées à l'intérieur d'une marge numérique située entre 0 et 100 pour pouvoir servir à des comparaisons directes avec les similarités linguistiques respectives qui, elles, oscillent également entre 0 et 100.

Précisons – quant au profil choroplèthe de la carte 1 – qu'il s'agit d'un profil de similarité typique des régions du centre du domaine d'Oïl: voir à ce sujet aussi la carte 2 dans Goebl 2002, 41 et la figure 1 dans Goebl 2000, 224. Le profil en question montre la position géotypologique d'un dialecte d'Oïl central, situé au milieu du réseau de l'ALF. Evidemment, l'implantation diatopique des similarités dialectales les plus importantes sur la carte (voir les polygones en rouge, jaune et orange) tire vers le nord. Il est non moins évident que les dialecticités les plus divergentes (c'est-à-dire les moins similaires) de celle du point de repère 1, se trouvent dans le sud du domaine d'Oc (voir les polygones en bleu foncé, en bleu clair et en vert). A noter également la position excentrique (voir les polygones verts) des domaines wallons et francoprovençal ainsi que le couloir jaune qui s'insinue entre le francoprovençal et l'auvergnat. Ce curieux phénomène est dû à l'infiltration d'éléments oïliques dans la zone de contact entre les parlers francoprovençaux et auvergnats où les locuteurs, le plus souvent tiraillés entre trois registres linguistiques en lice (francoprovençal, auvergnat et français) donnent, dans un accès d'insécurité linguistique, leur préférence au registre local le plus prestigieux, à savoir au français régional respectif: voir aussi les explications détaillées dans Goebl 2003, 79.

Evidemment, la structure de la carte 1 est due à l'action conjointe d'innombrables locuteurs de tous les idiomes parlés, depuis belle lurette, sur le sol de la Galloromania entière (à commencer par les langues de substrat pré-romaines jusqu'aux dialectes romans de l'époque de l'ALF), alors que celle de la carte 2 repose uniquement sur les impératifs de la géométrie euclidienne bidimensionnelle.

Au premier coup d'oeil, l'on est en mesure de capter les grandes différences (surtout géo-structurales) qui se creusent entre les deux profils choroplèthes sans d'ailleurs pouvoir porter un jugement quantitatif précis sur l'état des choses considéré. Ce n'est qu'en consultant les fichiers du logiciel VDM que l'on apprend que le coefficient de corrélation r(BP) entre les deux séries de 640[16] valeurs qui sous-tendent les profils iconiques de deux cartes est de +0,433. Evidemment, la taille numérique de cette valeur prise isolément ne nous dit rien. Ce n'est que dans un contexte majeur – et qui ne saurait être que de nature géo- ou cartogra-

(16) Le calcul de corrélation ne prend pas en considération les valeurs situées le long de la diagonale des matrices de similarité (etc.) où se trouvent répertoriés les scores réflexifs ($s_{ii}$) et, de ce fait, toujours munis de la valeur 100.

phique – que la valeur en question adopte une certaine importance et signification (géo)linguistiques.

La mise en carte des 641 scores de corrélation telle qu'elle ressort de la carte 9 fournit des résultats très intéressants et, de surcroît, complètement inédits. L'architecture de la carte – mis à part le fait qu'elle reflète une parfaite symétrie et accuse même une beauté iconique franchement étonnante – nous dit que dans le nord et le sud de la Galloromania les aménagements linguistique et euclidien de l'espace sont (numériquement) très convergents, alors que, dans une zone intermédiaire qui s'étend de l'Atlantique aux Alpes et englobe le domaine francoprovençal pratiquement dans sa totalité, l'impact humain sur la gestion dialectale de l'espace s'écarte (ou a dû s'écarter) considérablement de la texture euclidienne du même espace.

Quelle est donc la leçon à proprement parler linguistique de cette carte choroplèthe(17)? Tout d'abord, l'agencement de la structure iconique de la carte 9 ne peut pas être le produit du hasard: au linguiste, il rappelle avant tout la fameuse bipartition linguistique de la Galloromania en Oïl et Oc. Il est donc admis d'interpréter cette carte dans les termes de l'histoire antagoniste des types linguistiques d'Oïl et d'Oc et d'y voir par conséquent le bilan d'affrontements (c'est-à-dire de mouvements d'expansion et de recul) et de brassages linguistiques majeurs qui se sont déroulés à plus ou moins grande distance géographique et qui, en tout cas, sont pourvus d'une longue durée historique.

La situation géographique des zones coloriées en rouge et en orange, et leur adossement soit à la Manche dans le nord (où se trouve le berceau de la langue d'Oïl) soit aux Pyrénées dans le sud (c'est-à-dire non loin des foyers originaux de la langue d'Oc), nous suggèrent qu'il y subsiste les derniers restes d'une 'convergence primitive' entre la gestion dialectale de l'espace et ses dimensions euclidiennes dont la genèse doit d'ailleurs remonter loin dans l'histoire des deux langues en question et, de ce fait, se situer bien avant l'an mil.

Précisons que par 'convergence primitive' nous entendons un état quasiment 'naturel' où les similarités glotto-diatopiques se développent et

(17) Précisons que la leçon de la carte 9 peut être lue aussi en d'autres termes que ceux de la linguistique. Pour un chercheur dont les intérêts sont tournés vers la statistique, la structuration spatiale de cette carte est de loin moins intéressante que le fait que l'écart entre les 641 scores de corrélation mis en carte se situe entre –0,15 et +0,98 (voir à ce sujet la légende de la carte 9).

s'échelonnent dans l'espace uniquement en dépendance directe de la structuration euclidienne de ce dernier. Il en résulte, par rapport à un point de référence donné, une corrélation hautement positive (c'est-à-dire près de la valeur +1) entre la distance (ou proximité) linguistique et la distance (ou proximité) kilométrique. Evidemment, il est possible que cet étalement 'harmonieux' des similarités glotto-diatopiques dans l'espace soit déformé, sous l'influence de facteurs anthropiques externes (relevant de l'histoire ou de la politique), ou bien de manière homogénéisante (par l'émancipation de la similarité linguistique des contraintes de la proximité géographique) ou bien de manière discrétisante voire atomisante (par une fragmentation dialectale excessive de l'espace).

Les zones vertes et bleues, par contre, nous signalent les effets de la dissociation progressive entre l'emprise linguistique des humains sur l'espace et la structure euclidienne de ce dernier, due à l'intensité croissante de contacts, d'influences et de mélanges linguistiques survenus le plus souvent en provenance du nord et dirigés vers le sud. L'étalement des zones vertes et bleus sur la carte est très éloquent: il symbolise les zones de contact entre les domaines d'Oïl et d'Oc avec inclusion du domaine francoprovençal.

Evidemment, toutes ces interactions linguistiques reviennent à des effets de refoulement (ou d'expansion) linguistiques dont la durée diachronique ne devrait pas être sous-estimée. L'analyse corrélative entre les dimensions de la langue et de l'espace revêt donc une importance diachronique toute particulière.

Que nos suppositions contiennent une bonne dose de vérité peut être vérifié, sur la carte 9, par l'examen de la position d'une localité-témoin. Il s'agit du point-ALF 635 (Andraut, Département de la Gironde), situé au sud-est de Bordeaux, dont le polygone en bleu foncé est facilement repérable, au sein de la texture choroplèthe de la carte 9, par sa position détachée du reste du domaine d'Oc environnant. Or, Andraut fait partie de l'îlot linguistique de la Petite Gavacherie[18] qui fut fondée, en terre linguistiquement occitane, peu après la fin de la Guerre de Cent Ans (en 1453) par des colons saintongeais. Il est bien évident que, pour un îlot linguistique, une éventuelle 'convergence primitive' entre l'homme et l'espace n'a plus aucune raison d'être et que, de ce fait, le taux de corrélation respectif représente le minimum de la carte entière: r(BP) = -0,149.

(18) Cet îlot saintongeais en territoire gascon est souvent appelé, d'après son chef-lieu, 'Gavacherie de Montségur'.

Toute somme faite, nous pouvons postuler la présence et l'action d'une véritable 'grammaire spatiale'(19) qui régit les relations entre les dimensions psycho-sociales de la faculté langagière de l'homme et les dimensions euclidiennes de l'espace à l'intérieur duquel celle-ci se déploie. Si cette 'grammaire' a la force d'une véritable loi (ou d'un fait universel)(20), elle devrait se manifester aussi dans d'autres réseaux atlantographiques. Or, nous verrons par la suite que, en effet, tel est le cas.

### *3.2. L'exemple de l'AIS*

Voir les cartes 3, 4 et 10.

| | | |
|---|---|---|
| 1. | Source originale | Sprach- und Sachatlas Italiens und der Südschweiz (AIS) |
| 2. | Dialectométrisation | |
| | Auteur responsable | H. Goebl |
| | Références bibliographiques | Goebl 1981, 1982 et 1984 |
| 3. | Matrice de données | |
| | Critères de taxation | variations lexicales et morphologiques |
| | Nombre des points du réseau dialectométrisé | 247 + 4 (points artificiels) |
| | Nombre des cartes de travail | 696 (corpus intégral) |

(19) Rappelons à ce propos les termes de 'Raumgrammatik' et 'raumgrammatische Methode' forgés par le germaniste Heinz Rosenkranz dès 1938 au cours de ses études géolinguistiques sur les dialectes franconiens parlés des deux côtés de la frontière entre la Bavière et la Thuringie: cf. Rosenkranz 1938, 18-20.

(20) Précisons que les régularités glotto-diatopiques découvertes par la DM constituent le pendant spatial des fameuses 'lois linguistiques' découvertes par le comparatisme indo-européen du XIX^e^ siècle. Les lois linguistiques se réfèrent à des régularités observables dans la dimension du *temps* alors qu'ici il est question de la dimension de l'*espace*. Précisons en outre que le savoir sur l'intercorrélation étroite entre le temps et l'espace constitue, depuis l'antiquité gréco-romaine, un des acquis fondamentaux de la science occidentale tout court. Un des auteurs à qui nous devons des définitions classiques en la matière est le philosophe allemand Emmanuel Kant (1724-1804) qui, des dimensions du temps et de l'espace, a fait deux de ses fameux *"a priori"*.

A nouveau, il s'agit de données dialectométrisées par nous-même. Le point de repère, situé, une fois de plus, en position centrale, engendre deux profils choroplèthes encore plus divergents qu'auparavant dont le taux de corrélation r(BP) est de +0,313. La carte 3 constitue un profil de similarité 'typiquement' lombard[21] alors que la texture choroplèthe de la carte 4 reflète uniquement l'étalement spatial des proximités euclidiennes par rapport au point de référence 236 (Branzi, province de Bergame, Lombardie).

La carte 10 par contre, issue de la corrélation entre les proximités linguistiques et les proximités euclidiennes, dispose d'une structuration très claire qui laisse entrevoir, à l'instar de la carte 9, la double action de mouvements spatiaux carrément 'expansifs' (en provenance du sud et du sud-est) d'un côté, et 'récessifs' (situés sur les marges septentrionales du Piémont, de la Lombardie et du Trentin nord-occidental) de l'autre. La zone (coloriée en bleu foncé) où se manifeste le plus grand dérèglement de la 'convergence (glotto-diatopique) primitive', jouxte les foyers des parlers romanches des Grisons qui, à cause de leur position quelque peu détachée, évoquent – métaphoriquement parlant – l'idée d'une place forte assiégée en passe de se rendre.

A remarquer en outre que ni en France ni dans l'Italie du Nord, la configuration orographique naturelle du terrain ne semble avoir influé directement sur le profil des corrélations. Ceci est vrai avant tout pour la fameuse barrière des Apennins dont on ne trouve, sur la carte 10, aucune trace. L'impression qui se dégage d'une considération globale de la carte 10 est que son profil reflète les conséquences d'une poussée 'harmonisante' énergique en provenance de la Toscane et de la Padanie orientale dont la force finit par tarir complètement aux confins des Grisons[22]. Ajoutons qu'une supposition analogue peut être avancée au regard de la carte 11, également relative aux domaines italo- et rhéto-romans.

---

(21) Pour d'autres profils de similarité lombards voir Goebl 1984 III, 29, 45 et 175 ss.

(22) Précisons que cette constatation est en plein accord avec une vision nouvelle de l'histoire linguistique de la Padanie selon laquelle le faciès (géo)linguistique de l'Italie septentrionale, à l'origine encore fortement galloromanisant, a été modifié progressivement par une lente et irrésistible méridionalisation toscanisante à partir de la fin du premier millénaire: cf. Videsott 2001 passim.

### *3.3. L'exemple de l'ALD-I*

Voir les cartes 5, 6 et 11.

| | | |
|---|---|---|
| 1. | Source originale | Atlant linguistich dl ladin dolomitich y di dialec vejins, 1ª pert (ALD-I) |
| 2. | Dialectométrisation | |
| | Auteur responsable | R. Bauer |
| | Références bibliographiques | Bauer 2003 (a et b) et 2004 |
| 3. | Matrice de données | |
| | Critères de taxation | variations phonétiques, morphologiques et lexicales |
| | Nombre des points du réseau dialectométrisé | 217 + 1 (point artificiel) |
| | Nombre des cartes de travail | 3 899 |

L'ALD-I a été dialectométrisé, peu après sa parution en 1998, par notre collaborateur salzbourgeois Roland Bauer dans le cadre de sa thèse d'Etat: cf. Bauer 2003b. La quantité des cartes de travail tirées de l'analyse des quelque 900 cartes originales de l'ALD-I est impressionnante: il s'agit de 3 899 cartes de travail qui, évidemment, confèrent à cette recherche une fiabilité toute particulière.

La carte 5, relative à un point de repère géographiquement central et situé à l'extrémité septentrionale du Val de Non (all. Nonsberg), est du type trentin: le profil choroplèthe en fait ressortir, par l'agencement des polygones en rouge et orange, la silhouette du Trentin presque dans sa totalité, alors que les parages rhéto-romans de l'Engadine, de la Ladinie dolomitique et, dans une moindre mesure, aussi du Frioul, occupent les antipodes du point de repère choisi: voir les polygones respectifs en bleu foncé, bleu clair et vert. Au premier coup d'œil, le profil (euclidien) de la carte 6 s'en écarte considérablement[23]. Cette impression subjective est confirmée par la valeur relativement basse du coefficient de corrélation: +0,316.

En examinant la légende de la carte à corrélations (carte 11: en bas, à gauche) l'on constate, d'entrée de jeu, que l'écart des scores de corrélation qui varient entre +0,25 (valeur minimale) et +0,86 (valeur maximale)

(23) Voir aussi Bauer 2004, 220-221 où l'on trouve, sur la base des données de l'ALD-I, deux paires de cartes de similarité, relatives à la proximité linguistique d'un côté et à la proximité euclidienne de l'autre.

ressemble de très près à celui que nous avons relevé pour les données de l'AIS (entre +0,24, valeur minimale, et +0,75, valeur maximale): voir la légende de la carte 10 (en bas, à gauche).

Quant à son profil géographique, la leçon linguistique de la carte 11 est très claire: une fois de plus, les zones coloriées en bleu foncé (où se manifeste la part de lion des divergences entre les dimensions du basilecte et de l'espace naturel) forment, dans la partie supérieure du réseau de l'ALD, une espèce de ceinture-tampon située entre l'engadinais et quelques parlers ladins d'un côté et, dans la partie inférieure du même réseau, une grande plage vénéto-lombarde (coloriée en rouge et orange) de l'autre qui semble s'élancer vers le nord sous la forme d'un plateau cohérent.

À nouveau il se dégage l'impression de la coexistence, devenue quelque peu précaire, de zones 'expansives' (Lombardie, Trentin sud-oriental, Vénétie) situées à la marge méridionale de la carte, et 'récessives' (Engadine, Val Gardena et les parties septentrionales du Val de Fassa et du Val Badia). Toujours est-il que l'allure harmonieuse du gradient spatial de la carte 11 suggère, une fois de plus, la présence et l'action d'une 'grammaire spatiale' latente.

*3.4. L'exemple du CLAE («Computer Developed Linguistic Atlas of England»)*

Voir les cartes 7, 8 et 12.

| | | |
|---|---|---|
| 1. | Source originale | Computer Developed Linguistic Atlas of England (CLAE I et II) |
| 2. | Dialectométrisation | |
| | Auteurs responsables | W. Viereck et H. Ramisch |
| | Références bibliographiques | CLAE (I et II), Goebl 1997, Goebl/Schiltz 1997 |
| 3. | Matrice de données | |
| | Critères de taxation | variations lexicales et morphologiques |
| | Nombre des points du réseau dialectométrisé | 313+1 (point artificiel) |
| | Nombre des cartes de travail | 597 (corpus-MACRO: 321, corpus-MICRO: 276) |

Le souci des auteurs – Wolfgang Viereck et Henrich Ramisch – du 'Computer Developed Linguistic Atlas of England' (CLAE), n'était pas de dialectométriser une partie des données du trésor géolinguistique de l'Angleterre appelé 'Survey of English Dialects' (SED), mais bien plutôt d'en visualiser la distribution spatiale moyennant la confection de cartes à symboles (carré, triangle, cercle, losange, etc.) qui, eux, ont été sélectionnés en fonction de la variabilité lexicale et morphologique des cartes-SED choisies pour l'analyse visualisatrice. Les planches thématiques des deux volumes du CLAE s'appuient, en dernière analyse, sur 321 cartes originales tirées des quatre volumes du SED.

Précisons en outre que les données (transcrites) du SED sont présentées non pas sous la forme de cartes ou de planches (géolinguistiques) mais sous celle de colonnes et tableaux, ce qui les rend difficilement lisibles voire intelligibles du point de vue géolinguistique à proprement parler. C'est pourquoi les dialectologues anglais ont très tôt commencé à en extraire un certain nombre de 'cartes' et à les visualiser suivant leurs propres besoins: cf. Kolb/Glauser/Elmer/Stamm 1979, Orton/Sanderson/Widdowson 1978 et Orton/Wright 1974.

Au premier abord, le nombre des cartes de travail indiqué dans l'encadré ci-dessus (597 unités) semble être relativement haut. Ce chiffre demande cependant une précision. Il s'explique par la combinaison de deux analyses taxatoires (MICRO et MACRO), l'une plus particularisante (MICRO: 276 cartes de travail), l'autre plus enveloppante (MACRO: 321 cartes de travail). Ces deux types d'analyse taxatoire figurent déjà sur les légendes des planches respectives des deux volumes du CLAE. Il suffit de les feuilleter pour s'en convaincre.

Ajoutons encore que l'analyse taxatoire que W. Viereck et H. Ramisch ont appliquée aux données brutes du SED ne reposait, d'entrée de jeu, sur aucune intention dialectométrique globalisante à proprement parler. Le même travail, effectué sur les mêmes données de départ mais dans un autre esprit théorique, aurait abouti, très vraisemblablement, à des résultats taxatoires quelque peu différents.

Comme néanmoins il nous a paru utile de tenter, à des fins purement méthodologiques, une analyse dialectométrique des données des deux volumes du CLAE, nous avons demandé, au début des années 90 du siècle dernier, à W. Viereck de nous les confier – sous forme numérisée –

pour une analyse dialectométrique d'essai. Pour l'historique et les résultats de cette expérience nous renvoyons à Goebl 1997 et Goebl/Schiltz 1997.

La carte 7 montre l'implantation géotypologique d'un dialecte anglais oriental au sein du réseau du SED (ou du CLAE) et peut être considérée, de ce fait, comme étant 'anthropo-gène'. Tel n'est pas le cas pour la carte 8 qui, en tant que carte de proximité, mériterait, par contre, le qualificatif de 'géomètro-gène'.

La valeur de corrélation [selon r(BP)] entre les scores de similarité et de proximité sur lesquels reposent les cartes 7 et 8 est de +0,631 et, partant, relativement 'bonne'. L'appréciation visuelle de la corrélation entre les deux profils choroplèthes entraînerait un jugement similaire. La légende de la carte à corrélations (n° 12) nous enseigne toutefois que cette valeur est carrément exiguë par rapport aux autres étant donné le fait que l'écrasante majorité des scores de corrélation se situe entre +0,50 et +0,89.

Ceci n'empêche que l'allure générale du profil choroplèthe de la carte 12 laisse entrevoir, une fois de plus, le choc de deux dynamismes spatio-linguistiques antagonistes apparemment très anciens. En fouillant dans l'histoire de la langue anglaise il nous semble indiqué de mentionner, à ce propos, l'antagonisme viscéral qui régnait jadis, au sein de l''heptarchie saxonne' entre les anciens royaumes anglo-saxons 'sudistes' de Kent, Wessex, Sussex, Essex, Est-Anglie et Mercie d'une part, et le royaume 'nordiste' de Northumbrie de l'autre(24).

Précisons à ce propos que la faille bleue et verte de la carte 12 traverse l'Angleterre le long d'une bande s'étendant de l'estuaire du Mersey (à l'ouest) à celui du Humber et du Wash (à l'est). Il s'agit là d'une zone de frontière chère aux dialectologues anglicistes. Une fois de plus, il s'avère que la mise en carte des scores de la corrélation spatio-linguistique est de toute première importance pour la saisie de tensions et antagonismes diatopiques très anciens.

---

(24) Voir les cartes respectives dans les deux excellents manuels portant sur tous les aspects de la langue anglaise publiés d'une part par les soins de W. Viereck, K. Viereck et H. Ramisch (2002, 54), et par ceux de D. Crystal (1995, 28-29) de l'autre.

## 4. La corrélation entre les aménagements patronymique et euclidien de l'espace

Voir les cartes 13, 14 et 21.

| | | |
|---|---|---|
| 1. | Source originale | CD-ROM téléphonique italien (diffusé en 1996 par l'entreprise Topware) contenant les noms de 24 millions d'abonnés du téléphone italiens |
| 2. | Dialectométrisation | |
| | Auteurs responsables | I. Barrai, A. Rodrigez-Larralde, E. Mamolini et C. Scapoli |
| | Références bibliographiques | Barrai et al. 1999 |
| 3. | Matrice de données | |
| | Critères de taxation | variation patronymique |
| | Nombre des points du réseau dialectométrisé | 123 villes italiennes (de taille moyenne ou grande) |
| | Nombre des cartes de travail | 215 623 patronymes linguistiquement italiens ('types') appartenant à 5 043 580 abonnés du téléphone ('tokens') |

Les trois cartes en question émanent d'une recherche interdisciplinaire effectuée en coopération avec Italo Barrai, professeur émérite de biologie à l'Université de Ferrare (Italie), et ses collaborateurs. Comme tant d'autres généticiens, le professeur Barrai s'intéresse depuis longtemps à la géographie des patronymes ('noms de famille', pris dans un sens général) étant donné que la transmission des patronymes d'un côté et de certains allèles génétiques pour le chromosome Y de l'autre s'effectue uniquement dans la lignée paternelle. Une comparaison diatopique des deux espèces de diffusion – l'une de nature sociale et l'autre sujette aux mécanismes de la biologie – permet aux généticiens de mieux saisir certains aspects du phénomène de l'intermariage (angl. *interbreeding*). L'intérêt des généticiens pour la géographie des noms de famille a abouti à un grand nombre d'études portant le plus souvent sur des États entiers(25).

(25) Comme les contributions respectives des généticiens sont largement ignorées par les linguistes, nous renvoyons ici à l'adresse électronique de la bibliographie la plus importante en la matière («PubMed»): http://www.ncbi.nlm.nih.gov/entrez/query.fcgi

C'est ainsi que le professeur Barrai a étudié, entre autres, la géographie patronymique de l'Allemagne[26], de l'Autriche[27], de l'Italie[28] et aussi de la France[29]. Pour ce faire, il s'est toujours servi de CD-ROM de téléphone, de nos jours universellement disponibles, dont il a extrait, pour un nombre discret d'agglomérations urbaines d'une certaine importance démographique, plusieurs centaines de milliers de noms de famille différents ('types') appartenant à plusieurs millions d'abonnés de téléphone ('tokens').

Or, il est évident que, pour la saisie comparative de la répartition géographique des noms de famille, il faut considérer, pour chaque ville prise en considération, les trois grandeurs suivantes:

le total des habitants ('tokens') de la ville en question

le nombre des types patronymiques y ayant cours ('types')

le nombre des individus porteurs d'un patronyme donné.

Le mérite d'en avoir tiré une formule mathématique efficace et simple à la fois revient à l'anthropologue américain Gabriel W. Lasker[30]. À notre demande expresse, I. Barrai a mis à notre disposition les matrices de *distance*-Lasker pour l'Allemagne, l'Autriche, l'Italie et la France que nous avons converties ensuite, par l'application de la formule dist + sim (prox) = 100, en matrices de *similarité*. Après la confection d'un fond de carte approprié pour le réseau des 123 villes choisies par Italo Barrai pour la seule Italie péninsulaire, nous étions à même d'intégrer les données patronymiques d'I. Barrai dans notre logiciel VDM et d'en appliquer tous les modules analytiques, de la carte de similarité jusqu'aux dendrogrammes et les cartes à corrélations.

Les cartes 13, 14 et 21 n'en constituent que des échantillons illustratifs. Alors que la carte 13 montre – en tant que mensuration relationnelle de la similarité par rapport à un point de référence préalablement choisi – la position patronymique de la ville de Pérouse (ital. Perugia) à l'intérieur de l'Italie (définie comme ensemble de 123 agglomérations urbaines), la carte 14 en fait autant du point de vue euclidien. Au premier coup d'œil, les profils iconiques des deux cartes choroplèthes se ressemblent

(26) Cf. Rodriguez-Larralde/Barrai/Nesti/Mamolini/Scapoli 1998.

(27) Cf. Barrai/Rodriguez-Larralde/Mamolini/Scapoli 2000.

(28) Cf. Barrai/Rodriguez-Larralde/Mamolini/Scapoli 1999.

(29) Cf. Scapoli/Goebl/Sobota/Mamolini/Rodriguez-Larralde/Barrai 2005. Nous nous sommes d'ailleurs associé à cette recherche moyennant les données issues de la dialectométrisation de l'ALF (cf. Goebl 2000, 2002, 2003 et 2004a).

(30) Cf. ses études de 1977 (passim) et 1985 (22-24).

beaucoup. La valeur du coefficient de corrélation entre les 123 scores des cartes 13 et 14, est de +0,599 qui, prise isolément, a seulement une utilité heuristique fort limitée. Ce n'est que par la considération de la carte 21 qui synthétise 123 valeurs de corrélation tout en fournissant, une de fois de plus, une structuration bien accusée et très claire, que nous sommes à même d'avancer une interprétation globale de la covariance qui existe entre la diffusion des noms de famille d'un côté et les dimensions de l'espace de l'autre.

Ici encore, les polygones rouges renvoient à des zones où la 'convergence primitive' entre les aménagements euclidien et sémiotique (ici: patronymique) de l'espace semble avoir conservé son importance originale alors que de vastes zones de l'Italie du nord (situées surtout dans le Piémont) sont ou bien restées à l'écart de cette 'convergence primitive' ou bien l'ont perdue après coup à la suite de divergences spatio-linguistiques ultérieures.

Evidemment, l'interprétation de la carte 21 est plus difficile que celle des autres cartes à corrélations. Ceci s'explique par le fait que les données patronymiques brutes (reposant sur une liste de 215 623 patronymes italiens différents) ne nous sont pas connues directement et que la nature géolinguistique des patronymes, dont la lente genèse et diffusion ultérieure remonte au haut Moyen Âge(31), est quelque peu différente de celle des dialectes à proprement parler.

Il est vrai d'ailleurs que les autres analyses corrélatives que nous avons faites à l'aide de données patronymiques relatives à l'Autriche, l'Allemagne et la France fournissent également des résultats non moins clairs et éloquents si bien que nous sommes en droit de dire qu'il s'agit là, de toute évidence, d'une piste de recherche très prometteuse du point de vue interdisciplinaire. C'est pourquoi il serait souhaitable de multiplier de telles expériences interdisciplinaires pour aiguiser et raffiner nos connaissances analytiques en la matière(32).

### 5. La corrélation entre deux mensurations de la similarité

Voir les cartes 15, 16 et 22.

Evidemment, l'analyse corrélative peut être appliquée aussi à deux matrices de similarité dont chacune repose sur une autre espèce de men-

(31) Cf. Bagola 1988 (passim), Kunze 1998, 61, Kremer 1996 (passim) et les deux ouvrages d'E. De Felice (1980 et 1982).

(32) Voir aussi notre étude de 1996 qui porte sur la convergence entre les aménagements dialectal et génétique de l'espace de l'Italie du Nord.

suration de la similarité linguistique (ou dialectale). Or, l'utilité d'appliquer, le cas échéant, plus d'un seul indice de similarité aux mêmes stocks de données brutes, a été démontrée à plusieurs reprises (cf. Goebl 1983 et 1987). C'est ainsi que nous avons élargi la panoplie des indices de similarité à notre disposition qui, elle, ne comprenait d'abord que l'Indice Relatif d'Identité ($IRI_{jk}$), par l'Indice Pondéré d'Identité ($IPI_{jk}$)(33). L'$IPI_{jk}$ peut adopter 99 différents degrés (numériques) de pondération. L'utilité taxométrique et typodiagnostique d'un indice pondéré réside dans sa capacité de mieux faire ressortir, dans la texture du profil choroplèthe d'une carte de similarité, les plus grandes et les plus petites valeurs de similarité: voir à ce sujet le contraste iconique entre les cartes 15 (réalisée avec l'indice standard $IRI_{jk}$) et 16 (réalisée avec l'indice pondéré $IPI(1)_{jk}$). La corrélation numérique entre les deux cartes est d'ailleurs très grande: r(BP) = +0,969.

Alors que la carte 15 montre la position d'un dialecte wallon au sein du réseau de l'ALF dans la perspective d'une mensuration 'isocratique' (non pondérée), la carte 16 en fait autant du point de vue d'une mensuration 'anisocratique' (pondérée). A noter, sur la carte 16, le rétrécissement de la zone orange par rapport à celle de la carte 15, et aussi la prolifération des polygones en bleu foncé dans le sud de la Galloromania. L'allure générale du profil choroplèthe de la carte 16 est donc plus serrée et plus compacte que celle de la carte 15, plus équilibrée et, partant, plus 'plate'.

Comme les plus grandes divergences entre l'$IRI_{jk}$ et l'$IPI_{jk}$ reposent, en dernière analyse, sur la prise en compte (quantitative) différentielle de l'étendue spatiale des taxats répertoriés dans la matrice de données, l'établissement d'une carte à corrélations entre les deux philosophies métrologiques semblait être très prometteur, surtout dans la perspective de détecter ainsi la distribution spatiale de l'impact quantitatif des deux espèces de mensuration.

C'est qu'en bonne logique, la mise en carte de tous les 641 scores de corrélation devrait fournir une sorte de calibrage spatial des deux indices de similarité, susceptible de mettre en évidence les zones où l'emprise numérique de la mensuration pondérée (ou anisocratique) s'écarte le plus de la mensuration 'normale', c'est-à-dire non pondérée (ou isocratique).

En fait, la carte 22 est très éloquente à ce sujet. Alors que, pour les zones rouges, l'effet pondérateur n'a qu'une importance très limitée et

(33) Pour une explication détaillée de l'$IPI_{jk}$ cf. Goebl 1983 (en allemand) et 1987 (en français).

que, de ce fait, les différences quantitatives entre les deux sortes de mensuration sont minimes, le même effet est lourd de conséquences dans les zones bleues dont l'étalement sur la carte rappelle la fameuse ceinture-tampon curviforme entre les domaines d'Oïl et d'Oc. A remarquer que cette zone-frontière englobe aussi le domaine francoprovençal dont la nature hybride est connue depuis longtemps.

L'impact taxométrique de l'$IPI_{jk}$ repose sur deux effets: a) sur une mise en vedette numérique de taxats de moindre diffusion spatiale(34) ('effet augmentatif') et b) sur une dévalorisation numérique de taxats spatialement très répandus ('effet réducteur'). Or, l'impact numérique des deux effets dépend du pourcentage qui existe, au sein du vecteur d'attributs d'une localité (ou d'un pont d'atlas) donnée, entre les taxats à faible et à forte distribution spatiale. C'est que les zones bleues renvoient, en dernière analyse, à des dialectes dont les vecteurs respectifs disposent, dans l'ensemble, de traits linguistiques munis d'une diffusion spatiale très inégale. Ces dialectes (et, partant, leurs vecteurs d'attributs) sont donc caractérisés par la présence combinée d'un nombre plutôt réduit de taxats disposant d'aires grandes et d'un nombre beaucoup plus important de taxats ayant des aires moyennes et très petites: considérer, pour une bonne intelligence de ces faits, la matrice de données dans la moitié gauche de la figure 3. Or, la genèse d'aires linguistiques mineures est souvent l'apanage de zones de contacts riches en contaminations et innovations dialectales de toute sorte.

A remarquer en outre la configuration circulaire en rouge et orange, située au sein du domaine d'Oïl, qui rappelle de très près celle que nous avons constatée, à plusieurs reprises (voir surtout Goebl 2002, 54 [carte 15] et 2003, 112-113 [cartes 15 et 16]), en appliquant la synthèse des 'coefficients d'asymétrie de Fisher' (CAF). Les zones coloriées en jaune et ocre (répertoriées dans les intervalles 5 et 6), situées des deux côtés du cercle rouge et orange, donnent donc, à la mensuration pondérée, une prise plus grande que ne le font les zones rouges. Ceci s'explique de deux manières: à l'intérieur de la configuration circulaire en rouge et orange, par l'impact de l''effet réducteur' (mentionné ci-dessus)(35) alors que, à l'extérieur de la même configuration circulaire (et, partant, à la périphérie

(34) Rappelons qu'à chaque taxat correspond, à l'intérieur du réseau d'atlas examiné, une aire bien déterminée.

(35) La zone en question est, vu son voisinage avec les centres irradiateurs du domaine d'Oïl, caractérisée par beaucoup de taxats à grande diffusion spatiale ('méga-chorique').

du réseau d'enquête), c'est l''effet augmentatif'[36] qui crée l'écart entre les mensurations anisocratique et isocratique.

Soulignons encore le fait que la carte 22 repose sur un écart numérique très réduit (corrélation minimale: +0,78, corrélation maximale: +0,99). Il est d'autant plus remarquable que, malgré cela, la mise en carte de ces valeurs numériquement si peu distantes aboutisse à une configuration spatiale aussi précise et éloquente.

### 6. La corrélation entre les aménagements *phonétique* (etc.) et *lexical* (etc.) de l'espace

Voir les cartes 17, 18 et 23 ainsi que 19, 20 et 24.

Étant donné la grande utilité heuristique du calcul de corrélation, il semble logique de l'appliquer aussi aux différentes sous-catégories linguistiques telle que la *phonétique* face au *lexique* ou le *vocalisme* face au *consonantisme*. Ce procédé est d'autant plus intéressant qu'il s'agit là de questions de recherche très anciennes qui avaient été posées déjà par les pionniers de la géographie linguistique peu après la parution de l'ALF et dans le premier quart du 20e siècle (cf. Ettmayer 1924, 12-13, 44-47). Il va de soi que les comparaisons inter-catégorielles qui ont été faites alors devaient se limiter à la confection de cartes isoglottiques (très souvent mal réalisées) et à leur confrontation visuelle successive. Avec la DM corrélative, nous disposons, de nos jours, de bien d'autres moyens.

Notre présentation respective s'appuie sur deux volets: les cartes 17, 18 et 23 se réfèrent à la corrélation entre le sous-corpus des cartes de travail *phonétiques* (avec 1117 unités) et celui des cartes de travail *lexicales* (avec 471 unités) d'un côté, alors que les cartes 19, 20 et 24 visualisent la corrélation entre les cartes de travail *vocaliques* (612) et *consonantiques* (505) de l'autre. Les quatre sous-corpus tirés de l'ALF sont suffisamment grands pour nous permettre d'aboutir à des conclusions fiables.

Déjà la considération comparative des deux paires de cartes de similarité (17 et 18, 19 et 20) est très révélatrice. Elle nous montre dans quelle mesure divergent, toujours à partir du même point de repère de l'ALF (P. 1, Marcigny, Département Nièvre), les différentes espèces de l'aménagement linguistique du territoire de l'ALF. Mises à part certaines différences à la périphérie du réseau, les quatre profils présentés se recouvrent

(36) Les zones périphériques du domaine d'Oïl sont caractérisées par un surplus de taxats dont l'étendue est plutôt réduite ('micro-chorique').

étonnamment bien. Il n'est donc pas étonnant que les taux de corrélations respectifs soient relativement grands: cartes 17 et 18: +0,907; cartes 19 et 20: +0,869.

Face à ces valeurs numériques, une mise en garde est d'ailleurs indiquée: prises isolément et pour elles seules, ces valeurs ne nous disent pratiquement rien et ne contribuent en aucune manière à l'augmentation de nos connaissances comparatives sur les deux paires de catégories linguistiques en question. Ce n'est que par la mise en carte des valeurs en question que les 'chiffres commencent à parler' et que les brouillards se dissipent sur le coup: voir les cartes 23 et 24 dont les profils choroplèthes sont très éloquents et se ressemblent beaucoup.

Une fois de plus, le réseau de l'ALF est traversé d'une faille (en bleu foncé) qui s'étend de l'Atlantique aux Alpes et dont le fond court au sud du domaine francoprovençal. Le coloriage en rouge et orange symbolise que, dans les zones respectives, l'aménagement phonétique et lexical (ou vocalique et consonantique) de l'espace est largement convergent (ou 'en phase') alors que le coloriage bleu (c'est-à-dire en bleu foncé et bleu clair) renvoie à des régions où tel n'est pas le cas, et où, de ce fait, l'impact des deux catégories corrélées sur l'espace n'est plus 'au pas cadencé' et, partant, 'hors phase'. En jaugeant les profils choroplèthes des cartes 23 et 24 du nord au sud – p. ex. le long d'une ligne allant de la Picardie au Roussillon – l'on constate un mouvement de pendule harmonieux allant d'une coïncidence quasiment parfaite des deux 'phases' à un déphasage complet (mais spatialement bien circonscrit), pour aboutir, dans le Languedoc, à un nouveau recouvrement des deux phases en question.

La même constatation peut être faite à l'aide d'autres sous-corpus catégoriels, p. ex. en corrélant le vocalisme *pré*-tonique avec le vocalisme *post*-tonique ou tonique, etc. Il en appert que l'imbrication bien ordonnée des différentes catégories linguistiques que beaucoup de courants de la linguistique moderne admettent comme une évidence valable pour chacune des 4000 à 6000 langues de ce bas monde, se répercute, dans la perspective de la dialectométrie, également dans l'espace tout en engendrant de nouvelles régularités dont personne n'aurait soupçonné l'existence il y a encore quelques années. Sous le coup de cette évidence, pour ne pas dire: découverte, nous soulignons, une fois de plus, notre conviction que l'étude dialectométrique de la répartition spatiale des différents traits grammaticaux et lexicaux d'un nombre aussi grand que possible de langues (ou de dialectes), devrait faire partie intégrante de la typologie classique en général et de la typologie aréale en particulier.

## 7. Epilogue

En effet, la DM corrélative offre un grand nombre de nouvelles approches exploratoires en matière de géolinguistique et d'autres sciences engagées dans l'analyse de l'espace. C'est qu'elle permet d'analyser de plus près les conséquences spatiales de l'aménagement combiné d'un espace donné par l'HOMO LOQUENS d'un côté, et par la géométrie euclidienne de l'autre. La plateforme théorique et heuristique de la DM corrélative sont les cartes à corrélations. Comme, du reste, le principe des cartes à corrélations peut s'appliquer également à différentes catégories linguistiques (*phonétique* versus *lexique,* etc.) et à des données spatiales d'origine scientifique diverse (patronymes, génétique des populations, etc.), les résultats y ayant trait dépassent rapidement le cadre relativement étroit de la géolinguistique à proprement parler, et acquièrent, de ce fait, une importance franchement interdisciplinaire. Ceci concerne des disciplines situées tant à l'intérieur (typologie générale, aréologie typologique, etc.) qu'à l'extérieur (recherches patronymiques au sein de la génétique des populations, etc.) de la linguistique en tant que telle.

Bien que reposant sur des calculs statistiques en apparence relativement complexes au premier coup d'oeil, la DM corrélative ne tire aucunement son utilité scientifique d'un amas hétéroclite de chiffres, mais bien plutôt des résultats iconiques d'une visualisation adéquate de ces derniers, accompagnée, bien sûr, d'une lecture intelligente des cartes respectives de la part des (géo)linguistes.

C'est donc par le truchement de structures spatiales – le plus souvent très éloquentes au premier abord et ressortissant de cartes choroplèthes polychromes dûment confectionnées – que la DM corrélative parle aux yeux et à l'esprit des (géo)linguistes. Face à cette nouvelle perspective méthodologique, le défi le plus important que doivent relever les (géo)linguistes consiste d'un côté à s'approprier la maîtrise du langage iconique des cartes, et de l'autre à s'initier aux moyens techniques pour en produire en grand nombre et – surtout – d'excellente qualité. C'est exactement dans un tel dessein qu'a été conçu le logiciel VDM.

En tant que module récent du logiciel VDM, la DM corrélative est d'un maniement relativement simple, à condition toutefois que celui qui l'utilise soit au courant de 'tous les secrets' de VDM. Précisons que, de nos jours, l'existence et la pratique courante de la méthode de la DM dépendent non seulement du savoir-faire personnel d'un chercheur désireux de s'en servir, mais aussi du bon fonctionnement d'un contexte scientifique (et administratif) majeur qui embrasse la disponibilité de tous les moyens informatiques nécessaires, les moyens de payer les licences indis-

pensables pour l'usage d'un certain nombre de logiciels de grande diffusion, et aussi l'expertise et la bonne volonté de spécialistes sachant naviguer avec aisance sur la mer devenue très houleuse de logiciels plus ou moins complexes.

Bien que Edgar Haimerl, le créateur de VDM, ait toujours tenu à le rendre aussi souple et simple que possible, il subsiste néanmoins, quant à son maniement courant, un certain seuil de 'familiarité naturelle' avec l'informatique que les adeptes de la DM(37) en général et les futurs utilisateurs de VDM en particulier doivent franchir avant de pouvoir s'aventurer dans le domaine captivant de la détection de structures et de lois géolinguistiques jusqu'alors inconnues.

Ajoutons, en guise de point final, que le logiciel VDM est à la libre disposition de tous ceux qui en feront la demande(38). Ceci vaut également pour quelques leçons introductives à effectuer dans le laboratoire-DM à Salzbourg.

Université de Salzbourg Hans GOEBL

Avec la participation de: Italo BARRAI, Werner GOEBL, Edgar HAIMERL, Elisabetta MAMOLINI, Alvaro RODRIGUEZ-LARRALDE, Chiara SCAPOLI et Slawomir SOBOTA.

Remerciements:

Le présent travail est le fruit de la coopération patiente et assidue de plusieurs spécialistes à qui nous exprimons ici notre vive reconnaissance.

Création et mise à jour régulière du logiciel VDM: Edgar HAIMERL (Blaustein, Seattle)

Confection des 24 cartes et des 3 figures de cet article: Slawomir SOBOTA (Salzbourg) et Werner GOEBL (Vienne)

Mise à disposition des données dialectométrisées relatives à l'ALD-I: Roland BAUER (Salzbourg)

Mise à disposition des données patronymiques de l'Italie: Italo BARRAI (Ferrare)

Mise à disposition des données géolinguistiques du CLAE: Wolfgang VIERECK (Bamberg)

Toilettage stylistique de mon texte français: Lily DITZ-FUHRICH (Salzbourg).

Nous devons également une chaleureuse parole de remerciement à l'organisme de recherche autrichien FWF («Fonds zur Förderung der wissenschaftlichen For-

(37) Soit dit en passant, il existe, pour l'intelligence adéquate de la DM, un seuil similaire qui se réfère à la bonne compréhension du quantitatif en général et de la linguistique quantitative en particulier.

(38) Nous renvoyons nos lecteurs désireux de s'initier au logiciel VDM à notre contribution de 2004 (Goebl 2004b) où ils trouveront une introduction graduée et munie d'un grand nombre de "screen-shots" dûment commentés.

schung») pour le soutien financier qu'il a prêté à la dialectométrisation de l'ALF, réalisée au cours des projets de recherche avec les numéros 12 414 et 13 349.

## 9. Références bibliographiques et abréviations

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

ALD-I = Goebl, Hans/Bauer, Roland/Haimerl, Edgar et alii (éd.), 1998. *Atlant linguistich dl ladin dolomitic y di dialec vejins, 1ª pert. Atlante linguistico del ladino dolomitico e dei dialetti limitrofi, 1ª parte. Sprachatlas des Dolomitenladinischen und angrenzender Dialekte, 1. Teil,* Wiesbaden, L. Reichert, 7 vol.; accompagné de 3 CD-ROM (Salzbourg, 1999-2000) avec le module parlant de l'ALD-I, d'un DVD (Salzbourg, 2002sq.) et aussi d'une version spéciale de l'atlas parlant sur internet: http://ald.sbg.ac.at/ald/ald-i.

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

Bagola, Holger, 1988. «Italienisch: Anthroponomastik», *in*: *LRL* 4, 419-431.

Bahrenberg, Gerhard/Giese, Ernst 1975. *Statistische Methoden und ihre Anwendung in der Geographie*, Stuttgart, Teubner.

Barrai, Italo/Rodriguez-Larralde, Alvaro/Mamolini, Elisabetta/Scapoli, Chiara,1999. «Isonymy and Isolation by Distance in Italy», *in*: *Human Biology* 71, 947-961.

Barrai, Italo/Rodriguez-Larralde, Alvaro/Mamolini, Elisabetta/Scapoli, Chiara. 2000: «Elements of the Surname Structure of Austria», *in*: *Annals of Human Biology* 27, 607-622.

Bauer, Roland, 2003. «Sguardo dialettometrico su alcune zone di transizione dell'Italia nord-orientale (Lombardo vs. Trentino vs. Veneto)», *in*: Bombi, Raffaella/Fusco, Fabiana (éd.). Parallela 10. Sguardi reciproci. Vicende linguistiche e culturali dell'area italofona e germanofona, Udine, Forum, 93-119.

Bauer, Roland, 2003b: *Dialektometrische Analyse des Sprachatlasses des Dolomitenladinischen und angrenzender Dialekte (ALD-I)*, Salzburg, Institut für Romanistik, 2 vol.

Bauer, Roland, 2004: «Dialekte-Dialektmerkmale-dialektale Spannungen. Von 'Cliquen', 'Störenfrieden' und 'Sündendböcken' im Netz des dolomitenladinischen Sprachatlasses ALD-I», *in*: *Ladinia* 28, 201-242.

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

CLAE = Viereck, Wolfgang/Ramisch, Heinrich (éd.), 1991 et 1997. *The Computer Developed Linguistic Atlas of England*, Tübingen, Niemeyer, 2 vol.

Crystal, David, 1995. *The Cambridge Encyclopedia of the English Language,* Cambridge, Cambridge University Press, (2nd edition).

Dahl, Östen, 2001. «Principles of Areal Typology», *in*: Haspelmath, Martin/König, Ekkehard/Österreicher, Wulf/Raible, Wolfgang (éd.): *Language Typology and Language Universals. Sprachtypologie und sprachliche Universalien. La typologie des langues et les universaux linguistiques. An International Handbook. Ein internationales Handbuch. Manuel international.* Berlin, New York, de Gruyter, Handbücher der Sprach- und Kommunikationswissenschaft 20, vol. 2, 1456-1470.

De Felice, Emidio, 1980. *I cognomi italiani. Rilevamenti quantitativi dagli elenchi telefonici. Informazioni socioeconomiche e culturali, onomastiche e linguistiche*, Bologne, Mulino.

De Felice, Emidio, 1982: *I nomi italiani. Informazioni onomastiche e linguistiche, socioeconomiche e religiose. Rilevamenti quantitativi dei nomi personali dagli elenchi telefonici*, Rome, Venise, Sarin, Marsilio.

DM = *dialectométrie, dialectométrique.*

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

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

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

Goebl, Hans, 1983. «'Stammbaum' und 'Welle'. Vergleichende Betrachtungen aus numerisch-taxonomischer Sicht», *in*: *Zeitschrift für Sprachwissenschaft* 2,1, 3-44.

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

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

Goebl, Hans, 1996. «La convergence entre les fragmentations géo-linguistique et géogénétique de l'Italie du Nord», *in*: *Revue de linguistique romane* 60, 25-49.

Goebl, Hans, 1997. «Some Dendrographic Classifications of the Data of CLAE I and CLAE II», *in*: *CLAE* 2, 23-32.

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

Goebl, Hans, 2002. «Analyse dialectométrique des structures de profondeur de l'ALF», *in*: *Revue de linguistique romane* 66, 5-62 (avec 24 cartes dialectométriques).

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

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

Goebl, Hans, 2004b. «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 (éd.): *Romanistik und neue Medien. Romanistisches Kolloquium XVI*, Tübingen, Narr, 209-241.

Goebl, Hans/Schiltz, Guillaume, 1997. «A Dialectometrical Compilation of CLAE I and CLAE II. Isoglosses and Dialect Integration», *in*: *CLAE* 2, 13-21.

Grassi, Corrado, 2001: «La geografia linguistica», *in*: *LRL* I,1, 207-235.

Guiter, Henri: «L'Europe en ligne droite», in: Goebl, Hans (ed.): Dialectology, Bochum 1984 (Quantitative Linguistics 21), 121-135

Kolb, Eduard/Glauser, Beat/Elmer, Willy/Stamm, Renate, 1979. *Atlas of English Sounds*, Berne, Francke.

Kremer, Dieter, 1996. «Morphologie und Wortbildung der Familiennamen: Romanisch», *in*: Eichler, Ernst/Hilty, Gerold/Löffler, Heinrich/Steger, Hugo/Zgusta, Ladislav (eds): *Namenforschung. Name Studies. Les noms propres. Ein internationales Handbuch zur Onomastik. An International Handbook of Onomastics. Manuel international d'onomastique*, Berlin, New York, de Gruyter, Handbücher der Sprach- und Kommunikationswissenschaft 11, vol. 2, 1263-1275.

Kunze, Konrad, 1998. *dtv-Atlas Namenkunde. Vor- und Familiennamen im deutschen Sprachgebiet*, Munich, dtv.

Lasker, Gabriel W., 1977. «A Coefficient of Relationship by Isonymy: a Method for Estimating the Genetic Relationship between Populations», *in*: *Human Biology* 49, 489-493.

Lasker, Gabriel W., 1985. *Surnames and Genetic Structure,* Cambridge, Londres, New York, New Rochelle, Melbourne, Sydney, Cambridge University Press.

Legendre, Pierre/Legendre, Louis, 1998. *Numerical Ecology*. Second English Edition, Amsterdam, Lausanne, New York, Oxford, Shannon, Singapore, Tokyo, Elsevier.

LRL = Holtus, Günter/Metzeltin, Michael/Schmitt, Christian (éd.), 1988 sq. *Lexikon der Romanistischen Linguistik*, Tübingen, Niemeyer.

Nichols, Johanna, 1992. *Linguistic Diversity in Space and Time*, Chicago, Londres, University of Chicago Press.

Orton, Harold/Sanderson, Stewart/Widdowson, John David Allison, 1978. *The Linguistic Atlas of England*, Londres, Croom Helm.

Orton, Harold/Wright, Nathalia, 1974. *A Word Geography of England*, Londres, Croom Helm.

Rodriguez-Larralde, Alvaro/Barrai, Italo/Nesti, Carlo/Mamolini, Elisabetta/Scapoli, Chiara, 1998. «Isonymy and Isolation by Distance in Germany», *in*: *Human Biology* 70, 1041-1056.

Rosenkranz, Heinz, 1938. *Mundart und Siedlung im Gebiet der oberen Saale und des nördlichen Frankenwaldes*, Jena, Fischer.

Scapoli, Chiara/Goebl, Hans/Sobota, Slawomir/Mamolini, Elisabetta/Rodriguez-Larralde, Alvaro/Barrai, Italo, 2005. «Surnames and Dialects in France. Population Structure and Cultural Evolution», *in*: *Journal of Theoretical Biology* (sous presse).

SED = Orton, Harold/Halliday, Wilfried J. (éd.), 1962-1972. *Survey of English Dialects*, Leeds, Arnold, 4 vol.

Séguy, Jean, 1971. «La corrélation entre la distance spatiale et la distance lexicale», *in*: *Revue de linguistique romane* 35, 335-357.

Séguy, Jean, 1973a. «La dialectométrie dans l'Atlas linguistique de la Gascogne», *in*: *Revue de linguistique romane* 37, 1-24.

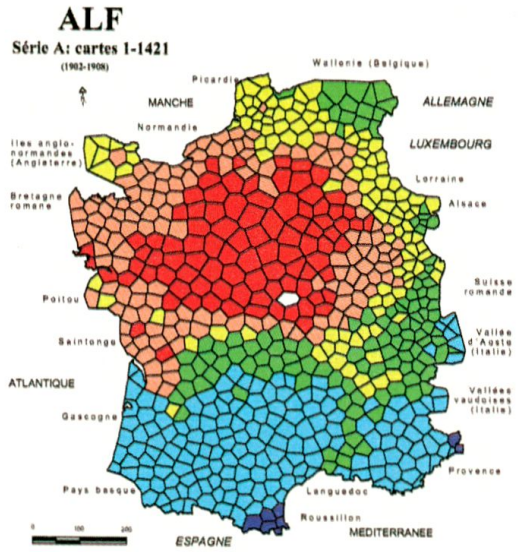
Séguy, Jean, 1973b. «La fonction minimale du dialecte», in: Straka, Georges/Gardette, Pierre (éd.): *Les dialectes romans de France à la lumière des atlas régionaux (Colloque de Strasbourg 1971)*, Paris, 27-37.

VDM = *Visual DialectoMetry*, logiciel dialectométrique créé par Edgar Haimerl, cf. le paragraphe 1.

Videsott, Paul, 2001. «La palatalizzazione di CA e GA nell'arco alpino orientale. Un contributo alla delimitazione dei confini dell'Italia linguistica nell'anno 1000», *in*: *Vox Romanica* 60, 25-50.

Viereck, Wolfgang/Viereck, Karin/Ramisch, Heinrich, 2002. *dtv-Atlas Englische Sprache*, Munich, dtv.

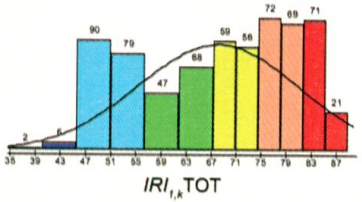
WALS = Haspelmath, Martin/Dryer, Matthews S./Gil, David/Comrie, Bernard (éd), 2005. *The World Atlas of Language Structures*, Oxford, Oxford University Press.



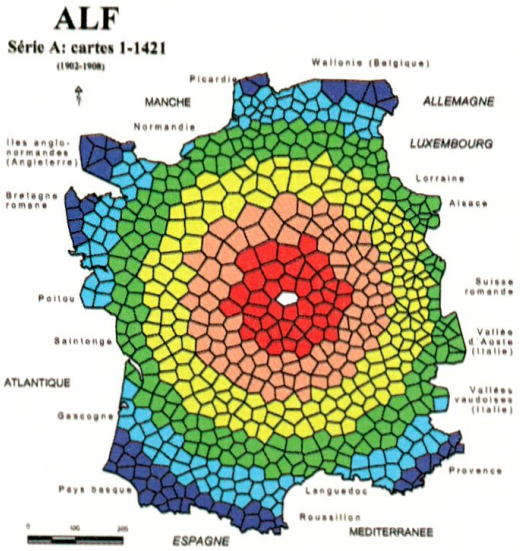

Algorithme d'intervallisation
MINMWMAX 6-tuple

| | | | |
|---|---|---|---|
| 1 | 34.61 - | 45.76 | (8) |
| 2 | - | 56.92 | (169) |
| 3 | - | 68.07 | (115) |
| 4 | - | 75.46 | (115) |
| 5 | - | 82.84 | (141) |
| 6 | - | 90.22 | (92) |
| | | | Σ = 640 |

Distribution de fréquence (similarité)
MINMWMAX 12-tuple



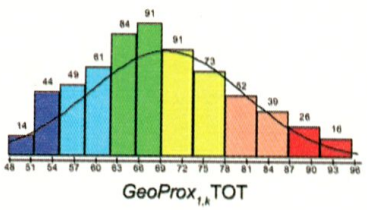

Carte 1: Carte choroplèthe de la distribution de similarité ($IRI_{1,k}$) relative au point de référence 1 (Marcigny, Département Nièvre).
Corpus: ALF, 1687 cartes de travail. Voir aussi le paragraphe 3.1.



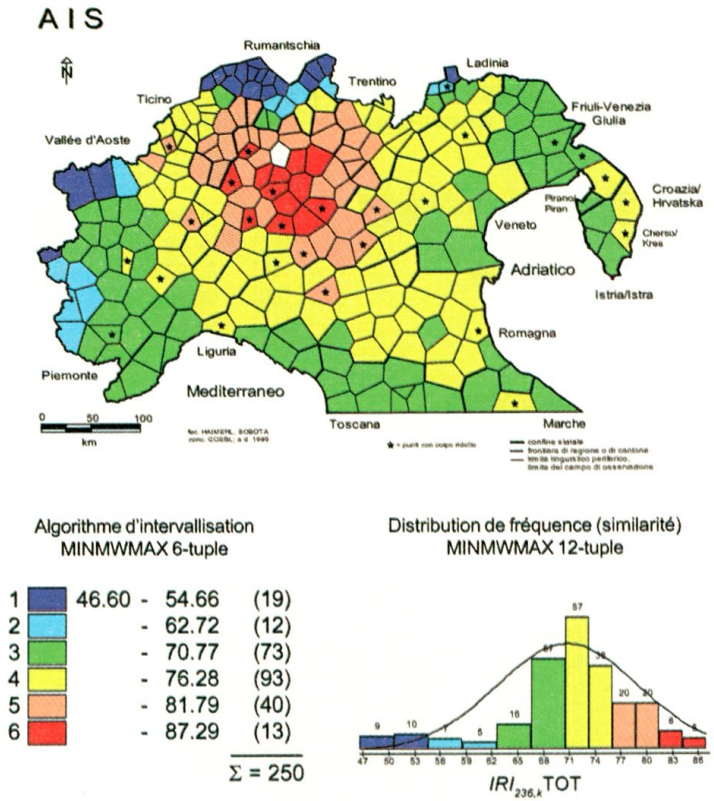

Algorithme d'intervallisation
MINMWMAX 6-tuple

| | | | |
|---|---|---|---|
| 1 | 47.85 - | 55.15 | (58) |
| 2 | - | 62.46 | (110) |
| 3 | - | 69.76 | (175) |
| 4 | - | 78.80 | (164) |
| 5 | - | 87.84 | (91) |
| 6 | - | 96.88 | (42) |
| | | | Σ = 640 |

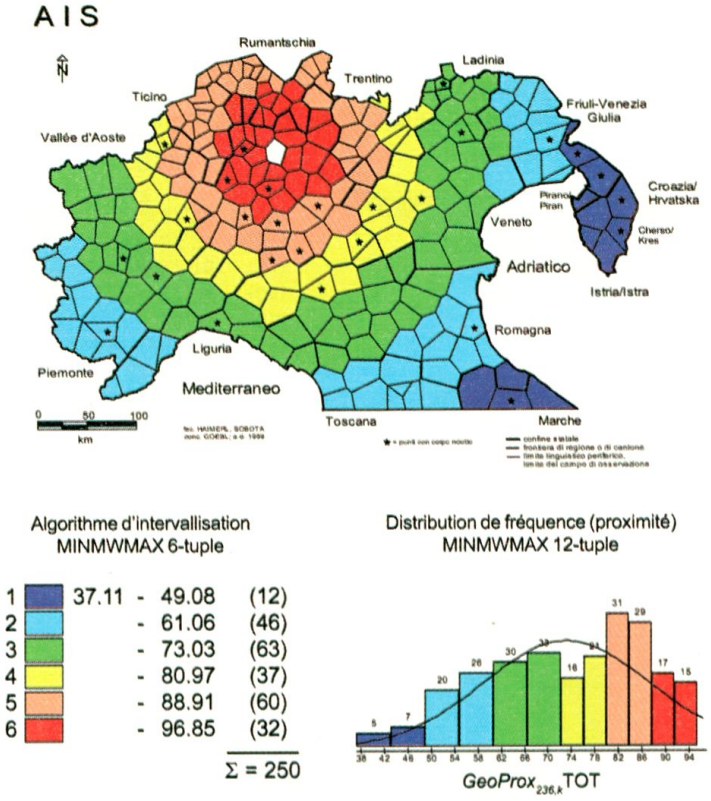
Distribution de fréquence (proximité)
MINMWMAX 12-tuple



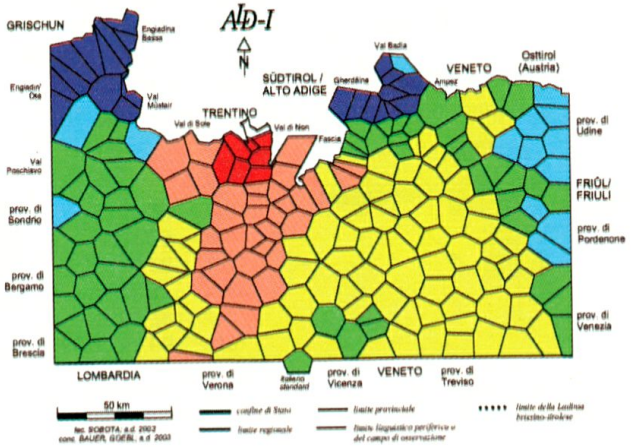

Carte 2: Carte choroplèthe de la proximité euclidienne relative au point de référence 1 (Marcigny, Département Nièvre).
Voir aussi le paragraphe 3.1.




Carte 3: Carte choroplèthe de la distribution de similarité ($IRI_{236,k}$) relative au point de référence 236 (Branzi, province de Bergame, Lombardie).
Corpus: AIS (vol. I, II et IV), 696 cartes de travail.
Voir aussi le paragraphe 3.2.



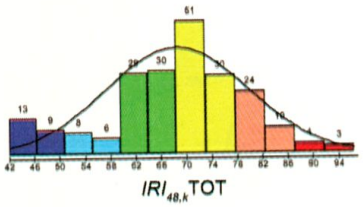

Carte 4: Carte choroplèthe de la proximité euclidienne relative au point de référence 236 (Branzi, province de Bergame, Lombardie).
Voir aussi le paragraphe 3.2.



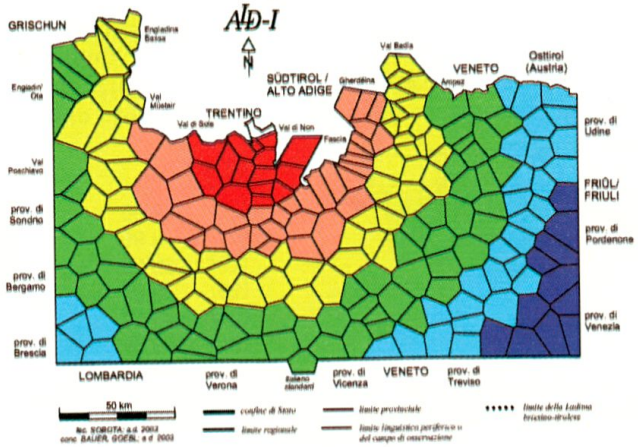

Algorithme d'intervallisation
MINMWMAX 6-tuple

| | | | |
|---|---|---|---|
| 1 | 41.95 - | 50.82 | (22) |
| 2 | - | 59.68 | (14) |
| 3 | - | 68.55 | (59) |
| 4 | - | 78.02 | (81) |
| 5 | - | 87.48 | (34) |
| 6 | - | 96.95 | (7) |
| | | Σ = | 217 |

Distribution de fréquence (similarité)
MINMWMAX 12-tuple



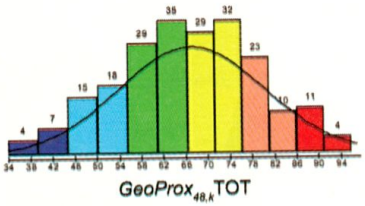

Carte 5: Carte choroplèthe de la distribution de similarité ($IRI_{48,k}$) relative au point de référence 48 (Castelfondo, Trentin).
Corpus: ALD-I, 3899 cartes de travail.
Voir aussi le paragraphe 3.3.



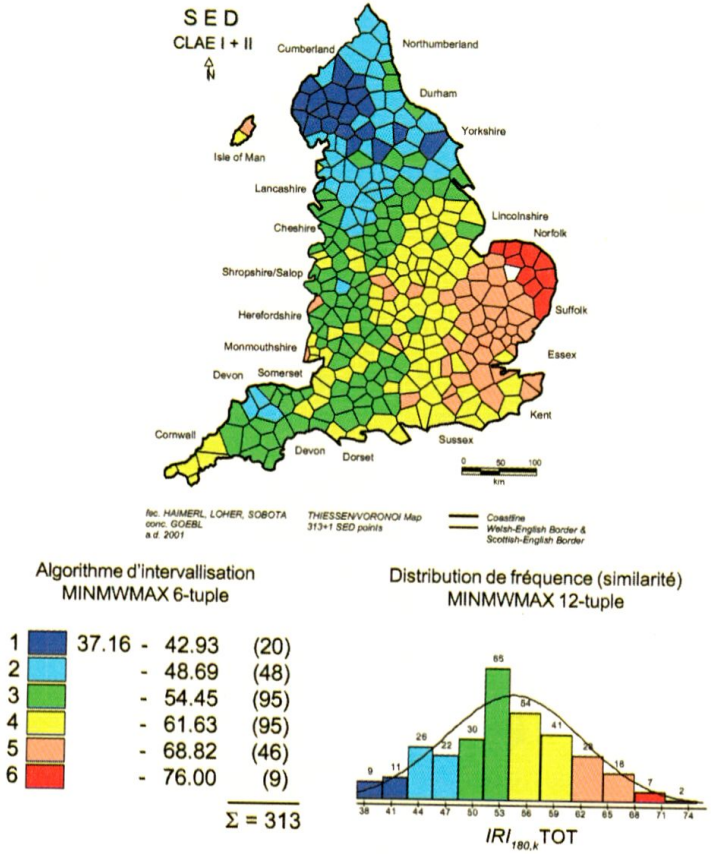

Algorithme d'intervallisation
MINMWMAX 6-tuple

| | | | |
|---|---|---|---|
| 1 | 33.86 - | 44.92 | (11) |
| 2 | - | 55.98 | (33) |
| 3 | - | 67.04 | (64) |
| 4 | - | 77.16 | (61) |
| 5 | - | 87.27 | (33) |
| 6 | - | 97.38 | (15) |
| | | Σ = | 217 |

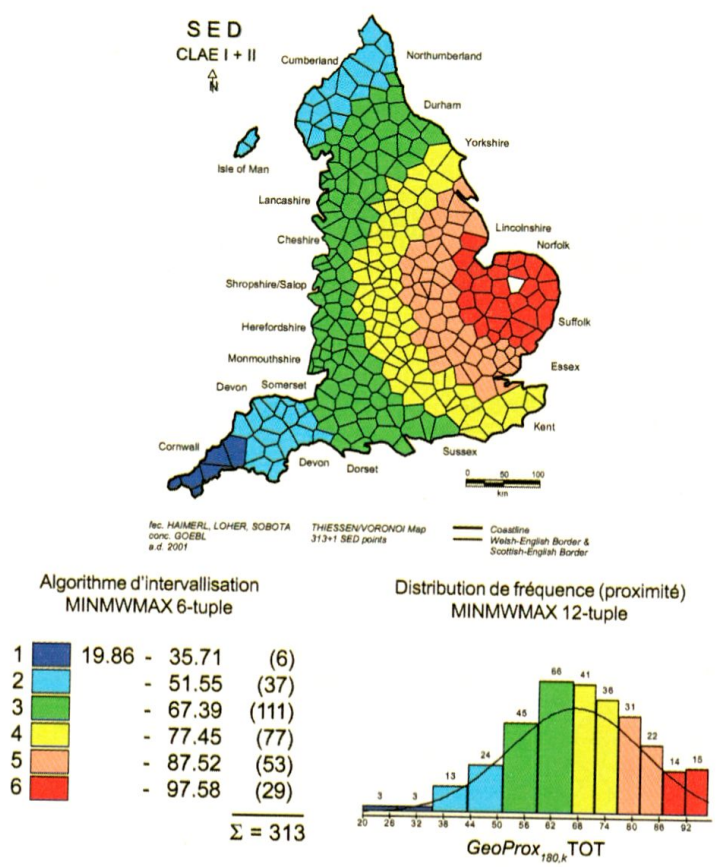
Distribution de fréquence (proximité)
MINMWMAX 12-tuple



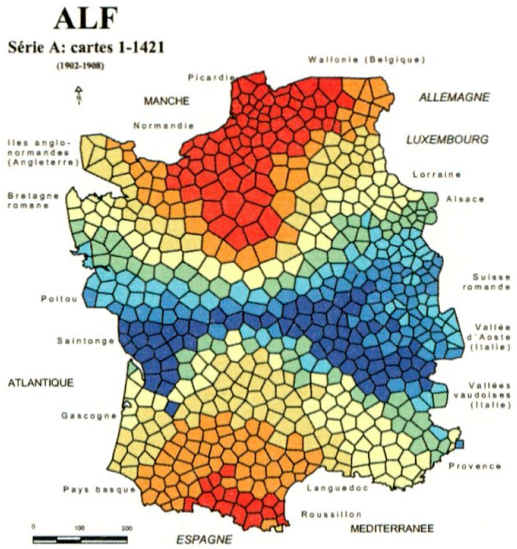

Carte 6: Carte choroplèthe de la proximité euclidienne relative au point de référence 48 (Castelfondo, Trentin).
Voir aussi le paragraphe 3.3.

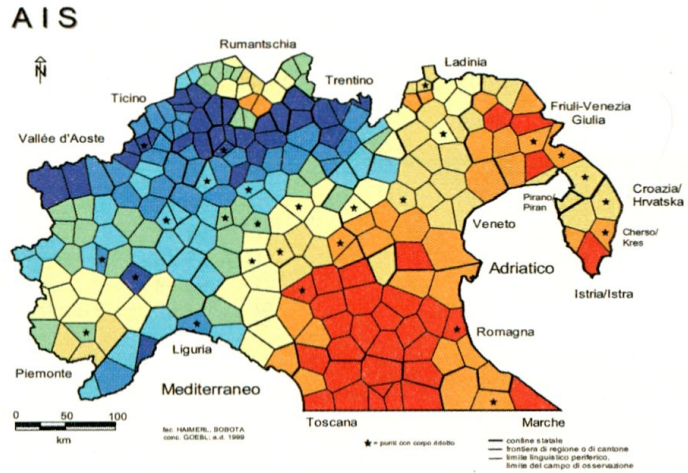



Carte 7: Carte choroplèthe de la distribution de similarité ($IRI_{Nf\ 9,k}$) relative au point de référence Nf 9 (Shipdham, Norfolk). Corpus: CLAE I et II, 597 cartes de travail. Voir aussi le paragraphe 3.4.




Carte 8: Carte choroplèthe de la proximité euclidienne relative au point de référence Nf 9 (Shipdham, Norfolk). Voir aussi le paragraphe 3.4.

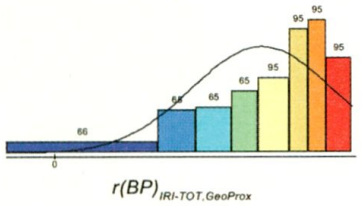



Algorithme d'intervallisation
MEDMW 8-tuple

| | | | | |
|---|---|---|---|---|
| 1 | -0.15 | - | 0.32 | (66) |
| 2 | | - | 0.44 | (65) |
| 3 | | - | 0.55 | (65) |
| 4 | | - | 0.63 | (65) |
| 5 | | - | 0.73 | (95) |
| 6 | | - | 0.79 | (95) |
| 7 | | - | 0.85 | (95) |
| 8 | | - | 0.93 | (95) |
| | | | | Σ = 641 |

Distribution de fréquence
MEDMW 8-tuple

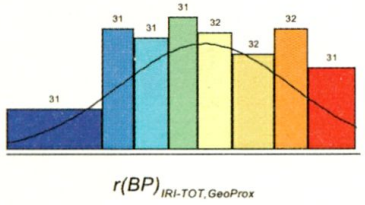



Carte 9: Carte choroplèthe de la corrélation spatiale ($r(BP)_{IRI,\ GeoProx}$) entre les aménagements dialectal et euclidien de la Galloromania (d'après l'ALF).
Coefficient de corrélation: $r(BP)_{jk}$. Voir aussi le paragraphe 3.1.

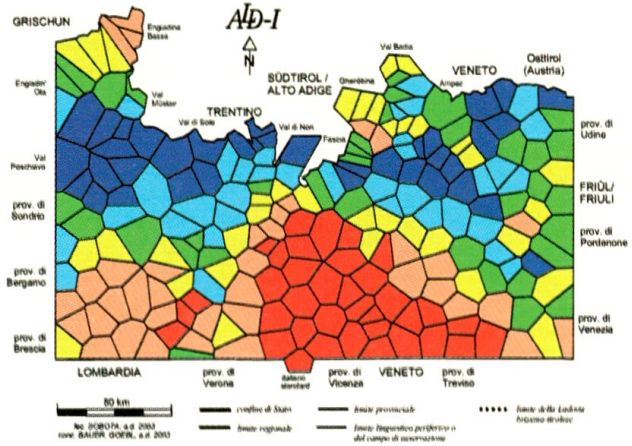
Algorithme d'intervallisation
MEDMW 8-tuple

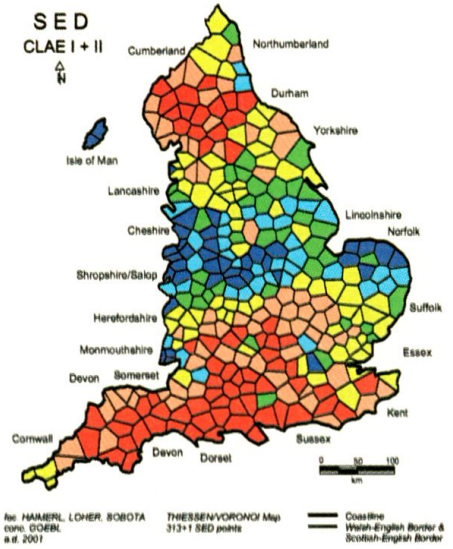
| | | | | |
|---|---|---|---|---|
| 1 | 0.24 | - | 0.38 | (31) |
| 2 | | - | 0.43 | (31) |
| 3 | | - | 0.48 | (31) |
| 4 | | - | 0.52 | (31) |
| 5 | | - | 0.57 | (32) |
| 6 | | - | 0.63 | (32) |
| 7 | | - | 0.68 | (32) |
| 8 | | - | 0.75 | (31) |
| | | | | Σ = 251 |

Distribution de fréquence
MEDMW 8-tuple




Carte 10: Carte choroplèthe de la corrélation spatiale ($r(BP)_{IRI,\ GeoProx}$) entre les aménagements dialectal et euclidien de l'Italie septentrionale (d'après l'AIS).
Coefficient de corrélation: $r(BP)_{jk}$. Voir aussi le paragraphe 3.2.

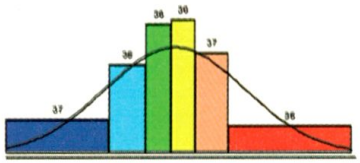



Algorithme d'intervallisation
MEDMW 6-tuple

| | | | | |
|---|---|---|---|---|
| 1 | 0.25 | - | 0.43 | (37) |
| 2 | | - | 0.50 | (36) |
| 3 | | - | 0.54 | (36) |
| 4 | | - | 0.59 | (36) |
| 5 | | - | 0.65 | (37) |
| 6 | | - | 0.86 | (36) |
| | | | | Σ = 218 |

Distribution de fréquence
MEDMW 6-tuple

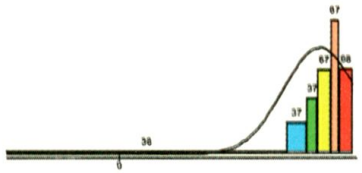



$r(BP)_{IRP\text{-}TOT,\,GeoProx}$

Algorithme d'intervallisation
MEDMW 6-tuple

| | | | | |
|---|---|---|---|---|
| 1 | -0.43 | - | 0.63 | (38) |
| 2 | | - | 0.71 | (37) |
| 3 | | - | 0.75 | (37) |
| 4 | | - | 0.80 | (67) |
| 5 | | - | 0.84 | (67) |
| 6 | | - | 0.89 | (68) |
| | | | | Σ = 314 |

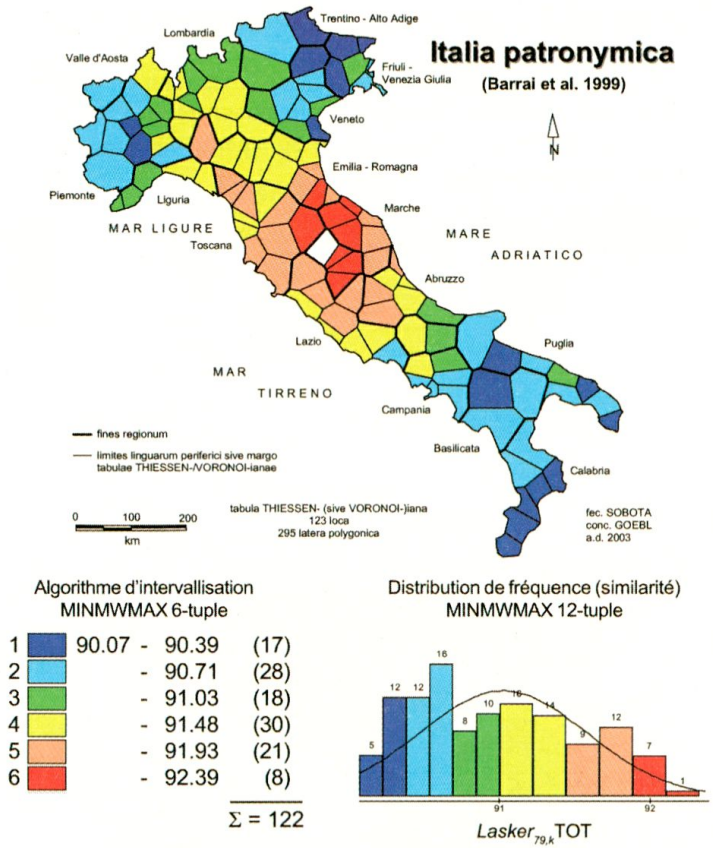
Distribution de fréquence
MEDMW 6-tuple



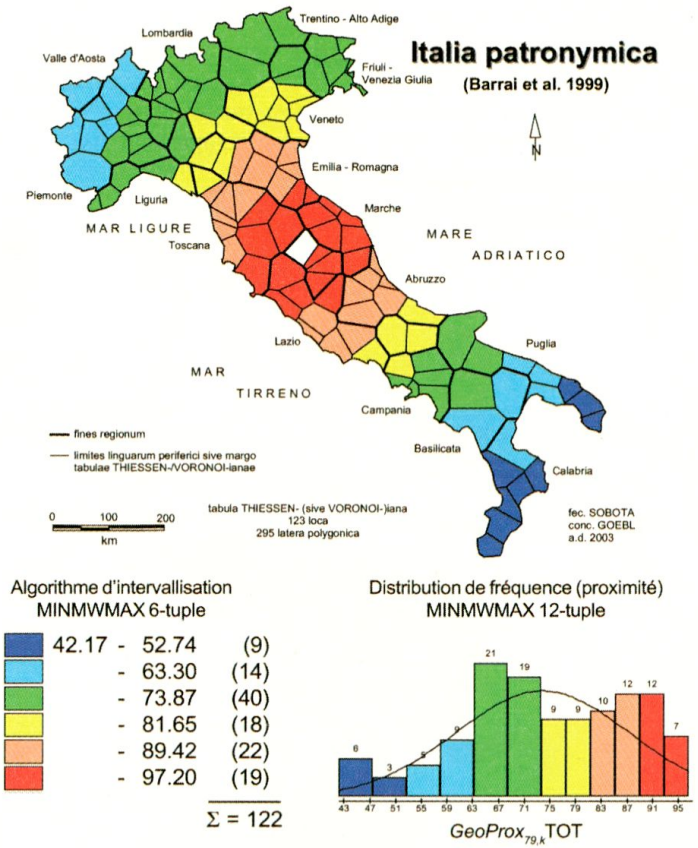

$r(BP)_{IRI\text{-}TOT,\,GeoProx}$

Carte 11: Carte choroplèthe de la corrélation spatiale ($r(BP)_{IRI,\,GeoProx}$) entre les aménagements dialectal et euclidien de l'Italie septentrionale du nord-est et de la Suisse sud-orientale (d'après l'ALD-I).
Coefficient de corrélation: $r(BP)_{jk}$. Voir aussi le paragraphe 3.3.

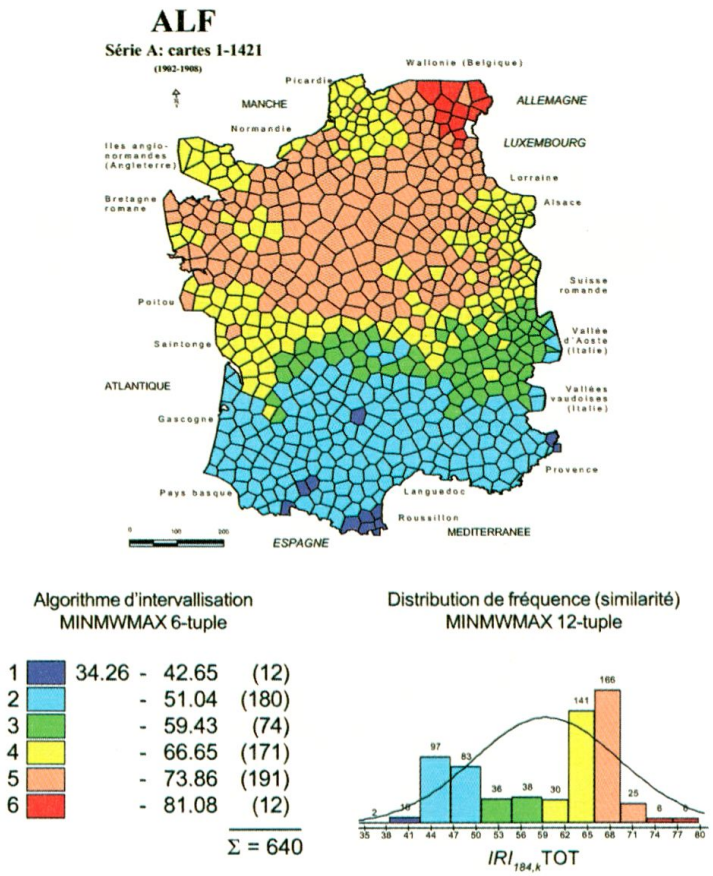
Carte 12: Carte choroplèthe de la corrélation spatiale ($r(BP)_{IRI,\,GeoProx}$) entre les aménagements dialectal et euclidien de l'Angleterre (d'après CLAE I et II).
Coefficient de corrélation: $r(BP)_{jk}$. Voir aussi le paragraphe 3.4.



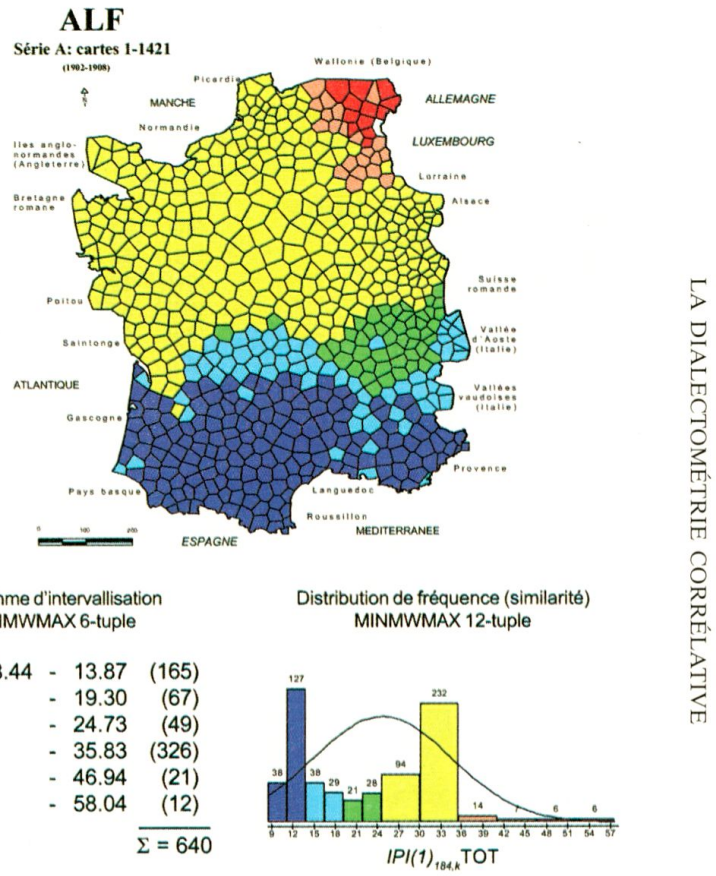

Carte 13: Carte choroplèthe de la distribution de similarité ($Lasker_{79,k}$) relative au point de référence 79, Pérouse (Ombrie).
Corpus: 215 623 patronymes de 5 043 580 abonnés de téléphone (d'après l'étude de Barrai et al. 1999).
Voir aussi le paragraphe 4.



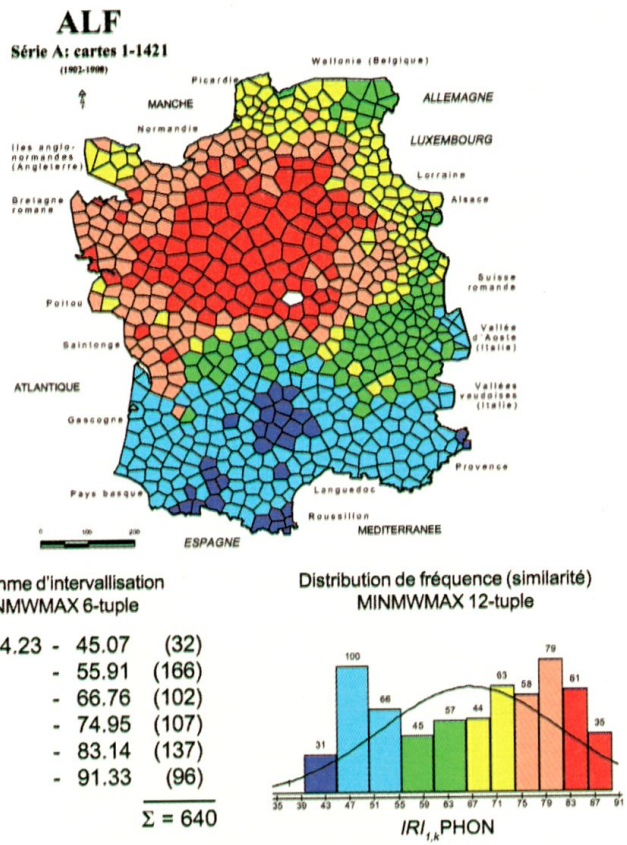

Carte 14: Carte choroplèthe de la proximité euclidienne relative au point de référence 79, Pérouse (Ombrie).
Voir aussi le paragraphe 4.



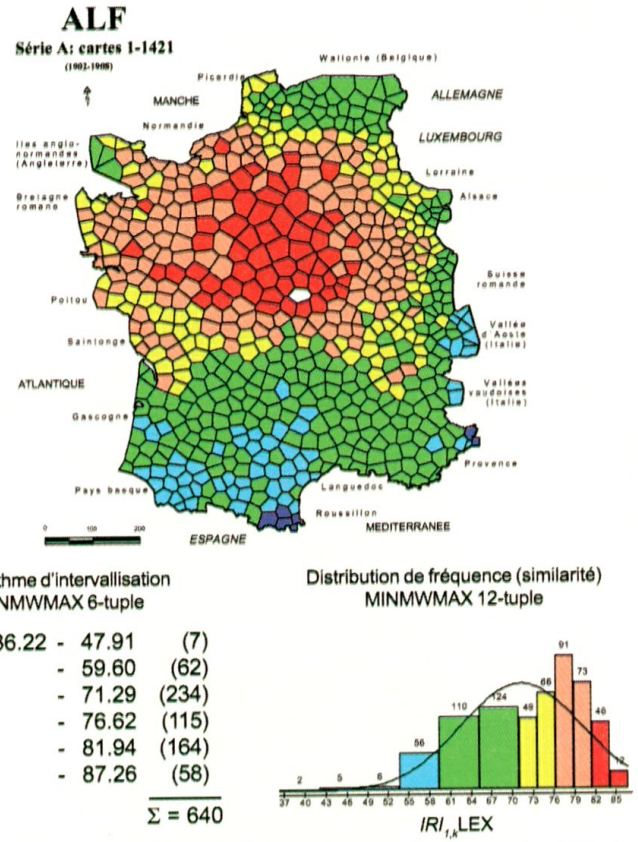

Carte 15: Carte choroplèthe de la distribution de similarité ($IRI_{184,k}$) relative au point de référence 184 (Bastogne, province de Luxembourg, Belgique).
Corpus: ALF, 1687 cartes de travail. Voir aussi le paragraphe 5.



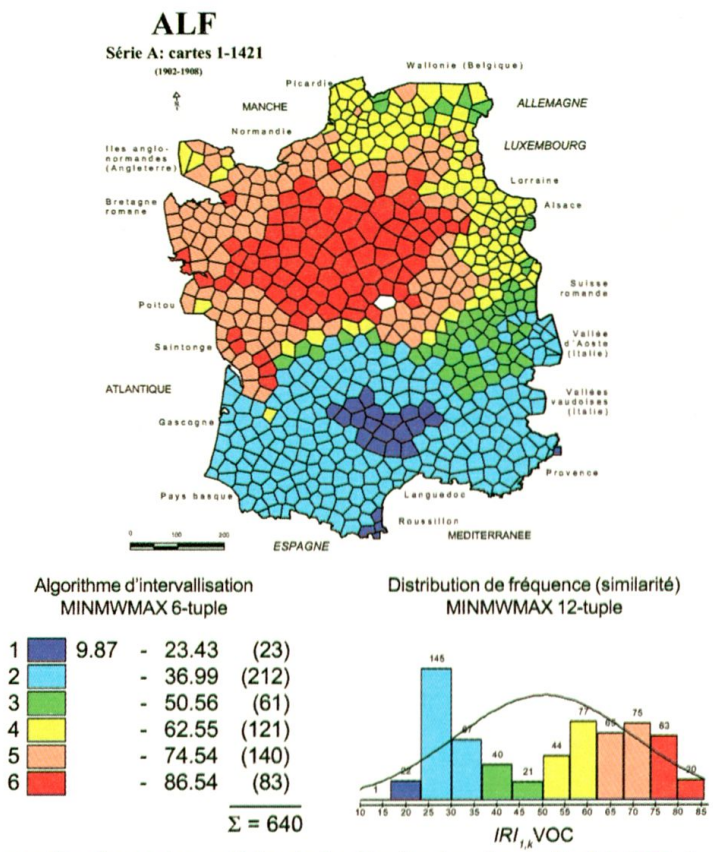

Carte 16: Carte choroplèthe de la distribution de similarité ($IPI(1)_{184,k}$) relative au point de référence 184 (Bastogne, province de Luxembourg, Belgique).
Corpus: ALF, 1687 cartes de travail. Voir aussi le paragraphe 5.



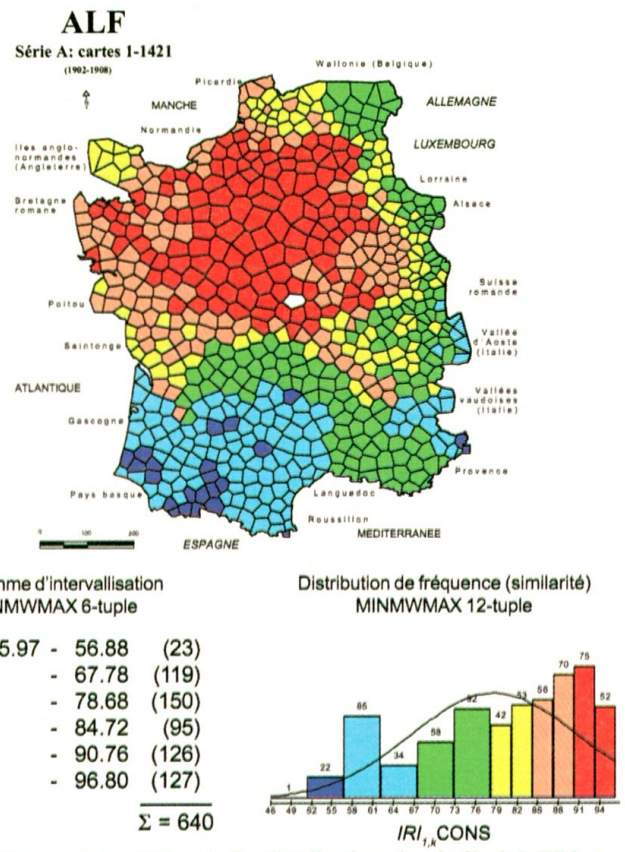

Carte 17: Carte choroplèthe de la distribution de similarité ($IRI_{1,k}$) relative au point de référence 1 (Marcigny, Département Nièvre).
Corpus: ALF, 1117 cartes de travail (données phonétiques).
Voir aussi le paragraphe 6.



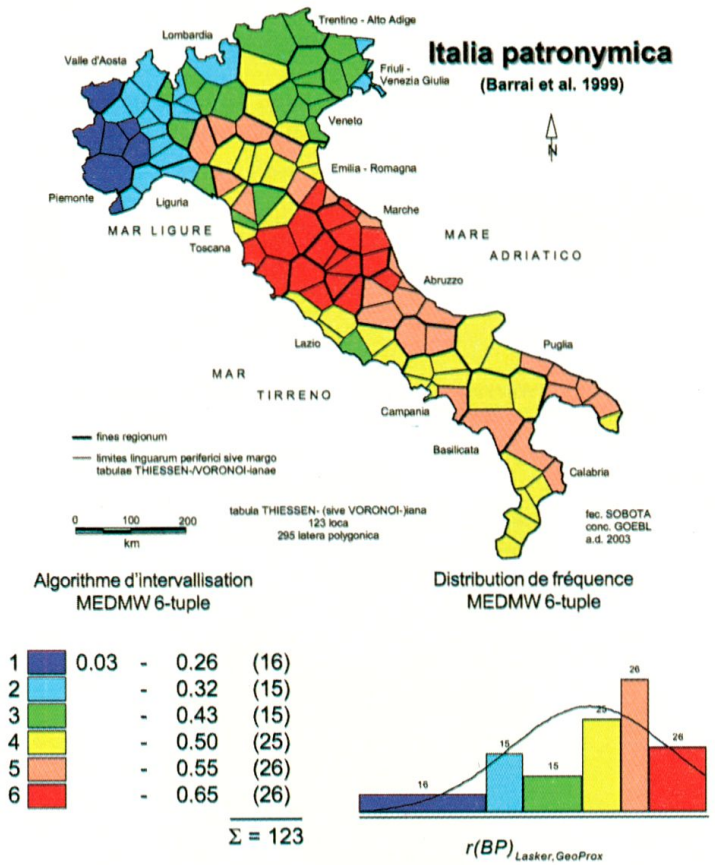

Carte 18: Carte choroplèthe de la distribution de similarité ($IRI_{1,k}$) relative au point de référence 1 (Marcigny, Département Nièvre).
Corpus: ALF, 471 cartes de travail (données lexicales).
Voir aussi le paragraphe 6.



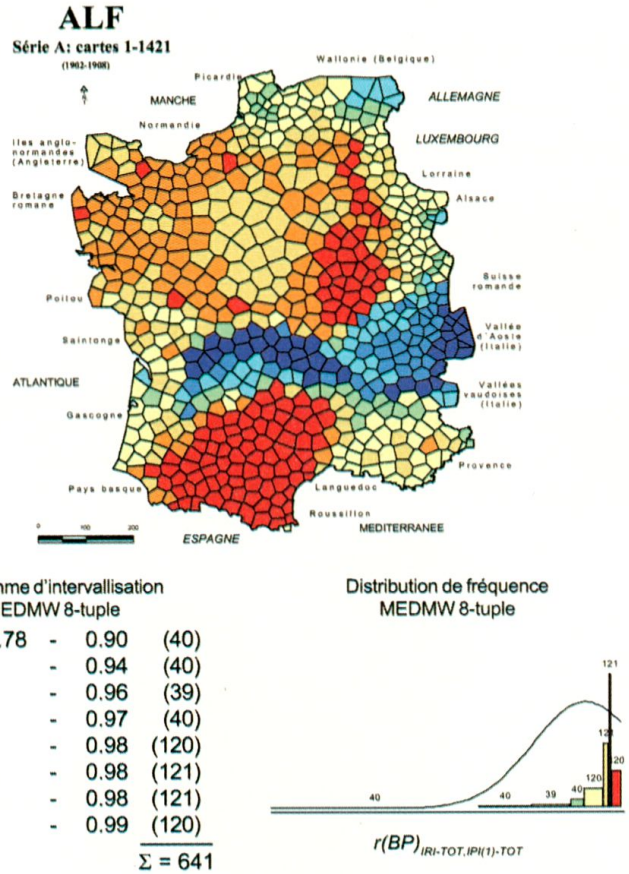

Carte 19: Carte choroplèthe de la distribution de similarité ($IRI_{1,k}$) relative au point de référence 1 (Marcigny, Département Nièvre).
Corpus: ALF, 612 cartes de travail (données vocaliques).
Voir aussi le paragraphe 6.



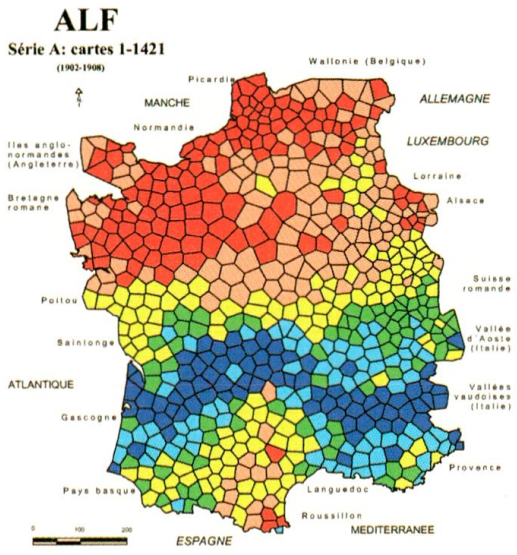

Carte 20: Carte choroplèthe de la distribution de similarité ($IRI_{1,k}$) relative au point de référence 1 (Marcigny, Département Nièvre).
Corpus: ALF, 505 cartes de travail (données consonantiques).
Voir aussi le paragraphe 6.




Carte 21: Carte choroplèthe de la corrélation spatiale ($r(BP)_{Lasker,\ GeoProx}$) entre les aménagements patronymique et euclidien de l'Italie (d'après l'étude de Barrai et al. 1999). Coefficient de corrélation: $r(BP)_{jk}$. Voir aussi le paragraphe 4.



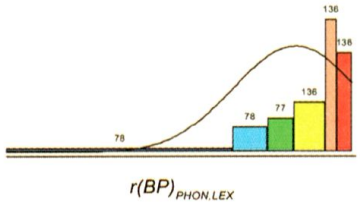

Carte 22: Carte choroplèthe de la corrélation spatiale ($r(BP)_{IRI,\ IPI(1)}$) entre deux mensurations de la similarité (selon IRI et IPI(1)) appliquées à la Galloromania.
Coefficient de corrélation: $r(BP)_{jk}$. Voir aussi le paragraphe 5.



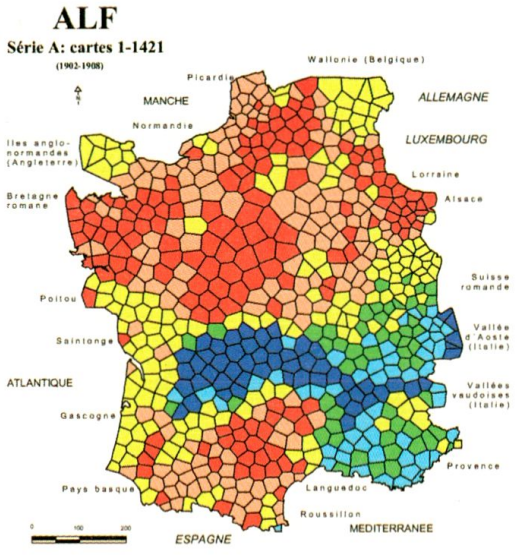

Algorithme d'intervallisation
MEDMW 6-tuple

| | | | | |
|---|---|---|---|---|
| 1 | 0.40 | - | 0.75 | (78) |
| 2 | | - | 0.80 | (78) |
| 3 | | - | 0.84 | (77) |
| 4 | | - | 0.89 | (136) |
| 5 | | - | 0.91 | (136) |
| 6 | | - | 0.93 | (136) |
| | | | | Σ = 641 |

Distribution de fréquence
MEDMW 6-tuple



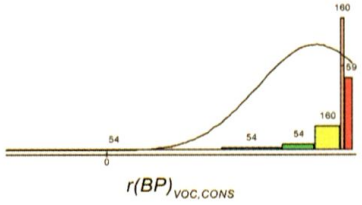

Carte 23: Carte choroplèthe de la corrélation spatiale ($r(BP)_{Phon, Lex}$) entre les aménagements phonétique et lexical de la Galloromania (d'après l'ALF). Coefficient de corrélation: $r(BP)_{jk}$. Voir aussi le paragraphe 6.




Algorithme d'intervallisation
MEDMW 6-tuple

| | | | | |
|---|---|---|---|---|
| 1 | -0.38 | - | 0.43 | (54) |
| 2 | | - | 0.66 | (54) |
| 3 | | - | 0.78 | (54) |
| 4 | | - | 0.88 | (160) |
| 5 | | - | 0.90 | (160) |
| 6 | | - | 0.93 | (159) |
| | | | | Σ = 641 |

Distribution de fréquence
MEDMW 6-tuple




Carte 24: Carte choroplèthe de la corrélation spatiale ($r(BP)_{Voc, Cons}$) entre les aménagements vocalique et consonantique de la Galloromania (d'après l'ALF). Coefficient de corrélation: $r(BP)_{jk}$. Voir aussi le paragraphe 6.