

Received 8 February 2024

Accepted 21 May 2024

Published November 2024

DOI: 10.1344/DIALECTOLOGIA2024.2024C.1

DU NOUVEAU DE L'« ÉCOLE DIALECTOMÉTRIQUE DE SALZBOURG » : BRÈVE PRÉSENTATION DU MODE « BÉTA » DE LA DIALECTOMÉTRIE DE SALZBOURG

Hans GOEBL *

Université de Salzbourg

hans.goebel@plus.ac.at

Résumé

La présente contribution est avant tout d'inspiration méthodique. Il s'agit d'augmenter les étapes taxométriques utilisées dans le cadre de l'École dialectométrique de Salzbourg. Jusqu'à présent, cette dernière a travaillé avec quatre étapes : 1) données originales de l'atlas linguistique analysé → 2) matrice de données en dérivée → 3) matrice de similarité → 4) visualisation du contenu de la matrice de similarité (selon six modalités taxométriques différentes). La nouvelle « dialectométrie *béta* » présentée ici dispose de cinq étapes : 1) données originales de l'atlas linguistique analysé → 2) matrice de données en dérivée → 3) matrice de similarité → 4) traitement ultérieur des données de la matrice de similarité au moyen du Coefficient de corrélation de Bravais-Pearson → 5) visualisation (selon les modalités taxométriques habituelles) du contenu de la nouvelle matrice de corrélation. Il en résulte des effets de lissage très intéressants des points de vue *géolinguistique* et *classificatoire*. Pour la démonstration, nous avons utilisé les données de la taxation du contenu de l'ALPI, faite à Salzbourg il y a une dizaine d'années. Les résultats de l'analyse quadripartite (= dialectométrie-*alpha*) et ceux de l'analyse en cinq parties (= dialectométrie-*béta*) sont présentés de manière comparative à l'aide de 12 planches en couleurs, toujours divisées en deux parties.

Mots clés : dialectométrie, dialectométrie-alpha, atlas linguistique, dialectométrie-béta, matrice de corrélation, matrice de similarité, lissage

* Département des Études Romanes, Erzabt-Klotz-Strasse 1, A-5020 Salzbourg.

© Author(s)



NOTÍCIES SOBRE L'«ESCOLA DIALECTOMÈTRICA DE SALZBURG»: BREU PRESENTACIÓ DEL MODE «BETA» DE LA DIALECTOMETRIA DE SALZBURG

Resum

Aquesta contribució és, abans de res, d'inspiració metodològica. Es tracta d'augmentar els passos taxomètrics utilitzats en el marc de l'«Escola dialectomètrica de Salzburg». Fins ara, aquesta ha treballat en quatre etapes: 1) dades originals de l'atles lingüístic analitzat → 2) matriu de dades derivada → 3) matriu de similitud → 4) visualització del contingut de la matriu de similitud (segons sis modalitats taxomètriques diferents). La nova “dialectometria *beta*” presentada aquí disposa de cinc passos: 1) dades originals de l'atles lingüístic analitzat → 2) matriu de dades derivada → 3) matriu de similitud → 4) tractament ulterior de les dades de la matriu de similitud mitjançant el Coeficient de correlació de Bravais-Pearson → 5) visualització (segons les sis modalitats taxomètriques habituals) del contingut de la nova matriu de correlació. Els resultats, a efectes de suavització, molt interessants des dels punts de vista *geolingüístic* i *classificatori*. Per a la demostració, hem utilitzat les dades de la taxació del contingut de l'ALPI, feta a Salzburg fa una desena d'anys. Els resultats de l'anàlisi quadripartida (= dialectometria-*alfa*) i els de l'anàlisi en cinc parts (= dialectometria-*beta*) es presenten de manera comparativa mitjançant 12 mapes en colors, sempre dividits en dues parts.

Paraules clau: dialectometria, dialectometria-alpha, atlas lingüístic, dialectometria-beta, matriu de correlació, matriu de similitud, allisatge

NEWS FROM THE “SALZBURG SCHOOL OF DIALECTOMETRY”: A BRIEF PRESENTATION OF THE “BETA” MODE OF SALZBURG DIALECTOMETRY

Abstract

This contribution is primarily methodological in nature. The aim is to extend the taxometric steps used in the Salzburg School of Dialectometry. Until now, the Salzburg School has worked with four steps: 1) original data from the analysed linguistic atlas → 2) derivative data matrix → 3) similarity matrix → 4) visualisation of the content of the similarity matrix (according to six different taxometric modalities). The new “*beta* dialectometry” presented here has five steps: 1) original data from the analysed linguistic atlas → 2) derivative data matrix → 3) similarity matrix → 4) further processing of the similarity matrix data by means of the Bravais-Pearson Correlation Coefficient, → 5) visualisation (according to the six usual taxometric modalities) of the content of the new correlation matrix. The result consists of smoothing effects that are very interesting from a *geolinguistic* and *classificatory* point of view. To demonstrate this, we used data from the typification of the 70 maps of the unique volume of ALPI carried out in Salzburg around ten years ago. The results of the quadripartite analysis (= *alpha* dialectometry) and those of the five-part analysis (= *beta* dialectometry) are presented comparatively using 12 colour plates, always divided into two parts.

Keywords: dialectometry, alpha-dialectometry, linguistic atlas, beta-dialectometry, correlation matrix, similarity matrix, smoothing

1. Introduction

Dans cet article, nous nous proposons de présenter – moyennant trois figures (en noir-et-blanc) et douze planches (en couleurs) – les capacités classificatoires et visualisatrices d’un nouveau module du logiciel dialectométrique VDM (« Visual DialectoMetry ») tel qu’il avait été développé ces dernières années dans le cadre de l’« École Dialectométrique de Salzburg » (EDMS).¹

Ce nouveau module n’est rien d’autre qu’un prolongement de la chaîne méthodique traditionnelle de l’EDMS telle qu’elle ressort de la Figure 1.

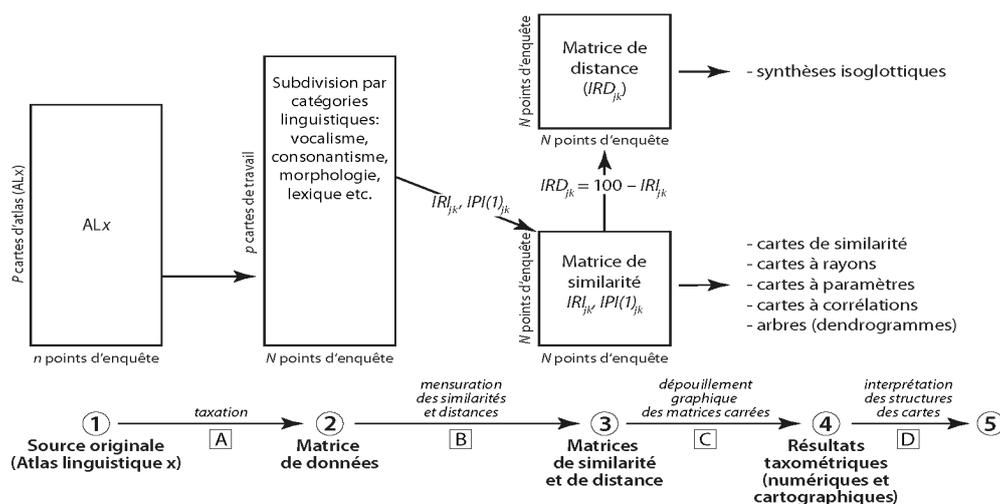


Figure 1. Alignement standard (mode-DM « alpha ») des démarches méthodiques de l’« École Dialectométrique de Salzburg »

¹ Pour une introduction aux méthodes de l’EDMS il y a plusieurs « filières » linguistiques: voir à ce sujet mes travaux suivants : en allemand : 1984, en français : 1981, 2000 et 2002 ; en anglais : 2009b, 2018b. Comme, de nos jours, le qualificatif de *dialectométrie* (ou de *dialectometry*, *Dialektometrie* etc.) est devenu l’apanage de plusieurs méthodes quantitatives dont le fond commun est en effet très réduit, la mise en vedette de *Salzburg* pour la caractérisation exclusive des méthodes présentées ici, est très utile, voire indispensable.

Dans sa configuration classique, cette dernière est composée de deux sections majeures :

- de la section *empirique* : celle-ci s'étend de l'*atlas linguistique* examiné [ALx] soumis à une analyse-DM [voir la position A-1], jusqu'à la *matrice de données* (MD) [voir la position A-2]. (Cette dernière est toujours créée par un processus de *typisation* (c.-à.-d. de « taxation ») des données originales de l'ALx.),
- et de la section *taxométrique* à proprement parler qui s'étend de la mesure de la similarité entre les N vecteurs d'attributs de la MD [voir la position B-3] jusqu'à la visualisation [voir la position C-4] des données ainsi calculées et leur interprétation linguistique à la lumière des six volets taxométriques traditionnels de l'EDMS [voir la position D-5].

La section *taxométrique* forme un bloc cohérent centré sur la matrice de *similarité* (MSim), calculée à l'aide d'un des nombreux indices de similarité de l'EDMS, et sur la matrice de *distance* (MDist), qui en est dérivée par l'application de la formule suivante: $dist = 1$ (ou: 100) – sim .

Au cours de différentes expériences-DM effectuées surtout dans le cadre de recherches de type scriptométrique² nous avons découvert l'utilité de l'application supplémentaire d'indices de *corrélation* à proprement parler, parmi lesquels l'indice de corrélation $r(BP)$ – proposé, il y a belle lurette, par le physicien français Auguste Bravais (1811-1863) et le mathématicien anglais Karl Pearson (1857-1936) – est le plus universel et le mieux connu.³

De là est née l'idée de soumettre les scores de similarité stockés dans la MSim de base (voir la position 3 sur les Figures 1 et 2) à un traitement statistique supplémentaire par l'application de l'indice $r(BP)$ ⁴ aux données (de nature quantitative

² Par « scriptométrie » nous entendons l'application des méthodes-DM de l'EDMS aux données (et problèmes) de la scriptologie (médiévale): voir, à titre d'exemple, Dees 1980, nos contributions de 2008a, 2016a et 2017 (les deux dernières contributions en coopération avec P. Smečka) et Videsott 2009.

³ Pour la formule du $r(BP)$ voir Chandon/Pinson (1981 : 65) (→ la formule S4).

⁴ Les scores du $r(BP)$ oscillent entre -1 et +1. Ils signalent la corrélation linéaire entre deux variables métriques.

ou « métrique ») de la MSim déjà existante, et de dédoubler ainsi la partie *taxométrique* de la chaîne-DM traditionnelle : voir la position 4 sur la Figure 2.

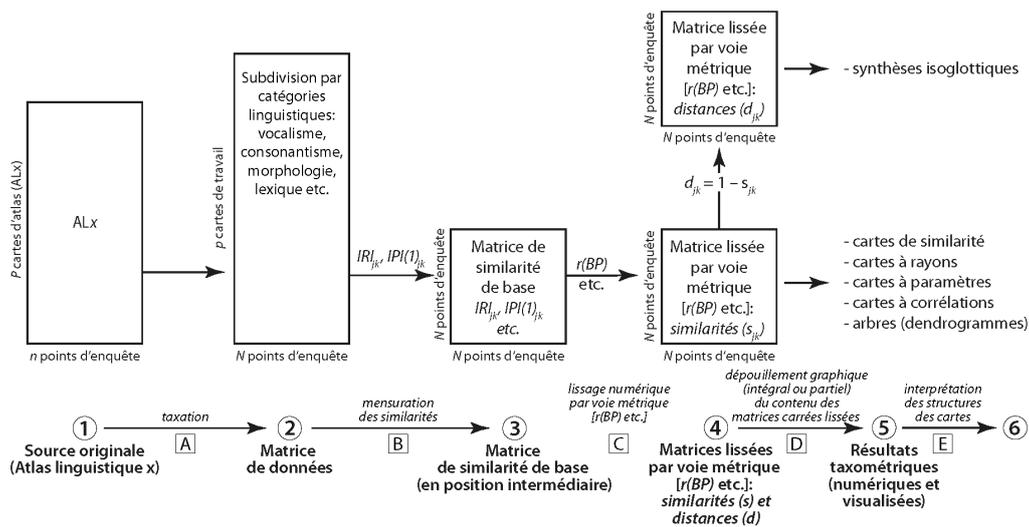


Figure 2. Alignement élargi (mode-DM « bêta ») des démarches méthodiques de l'« École Dialectométrique de Salzbourg »

Pour bien distinguer les deux procédures statistiques, nous les avons baptisées « mode-DM *alpha* » (mode ancien ou traditionnel) et « mode-DM *bêta* » (mode nouveau).

Évidemment, le traitement visualisateur final des scores de *similarité* (ou de *distance*) issus des calculs du type nouveau *bêta*, reste le même.

2. Plan du présent article

Par la suite, nous présenterons à l'aide de douze planches – issues respectivement de l'application des modes-DM *alpha* et *bêta* aux six volets taxométriques traditionnels de l'EDMS – les différences heuristiques qui se creusent entre les deux méthodes. Chacune de ces planches contient deux cartes, calculées

selon les deux modes-DM mentionnés : la carte de gauche selon le mode traditionnel α , la carte de droite selon le mode innovateur β .

Les données de base ont été tirées du premier (et unique) volume de l'ALPI qui avait été dépouillé, peu avant le tournant du millénaire, par une équipe de jeunes romanistes salzbourgeois,⁵ parallèlement au développement du logiciel VDM par les soins de mon ami Edgar Haimerl.

3. Présentation du corpus-ALPI utilisé

Voir la fig. 3.

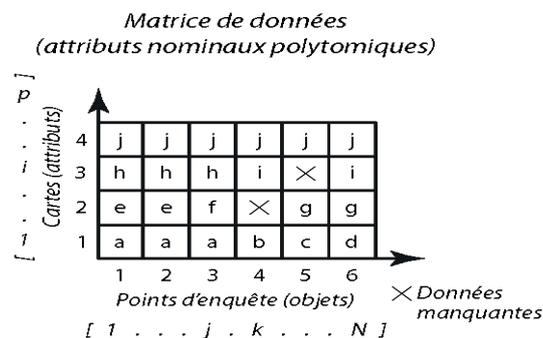


Figure 3. Schéma d'une matrice de données tirée par voie taxatoire d'un atlas linguistique

La génération de la MD relative à l'ALPI pouvait se prévaloir de nos expériences pratiques faites précédemment au cours des dialectométrisations de l'ALF⁶ et de l'ALS.⁷ Ceci n'empêche que la dialectométrisation de l'ALPI a toujours souffert d'une carence

⁵ En voici leurs noms : Barbara Aigner, Irmgard Dautermann, Hildegund Eder, Susanne Oleinek et Annette Schatzmann. Au cours des deux dernières décennies, notre gratitude envers elles n'a cessé de croître. Voir Goebel (2013) pour un premier bilan de ces recherches.

⁶ Voir Goebel 2000, 2002 et 2003.

⁷ Voir Goebel 2008b.

capitale, à savoir de l'exiguïté des données originales de l'ALPI. Des 75 cartes originales seulement 64 ont pu être taxées en bonne et due forme si bien que la MD à proprement parler ne comprenait que 375 cartes de travail (CT). Ce chiffre est minime par rapport à nos projets-DM antérieurs, relatifs aux atlas linguistiques ALF (→ 1681 CT) et AIS (→ 3911 CT).

C'est pourquoi la valeur linguistique des résultats présentés par la suite est purement expérimentale. Néanmoins, je nourris l'espoir qu'il sera possible un jour d'augmenter considérablement le nombre des CT déjà disponibles à l'aide de la nouvelle version digitale de l'ALPI et grâce à l'engagement personnel de quelques jeunes néophytes ibériques de la DM en général et du programme VDM en particulier.⁸

En établissant nos différentes matrices de données (MD) de base nous avons suivi, en dernière analyse, l'exemple que Jules Gilliéron (1854-1926) avait donné dans sa fameuse étude onomasiologique sur l'action de *scier* dans le Sud-Est de la France (Gilliéron 1905). Nous avons donc commencé par typiser (« taxer ») les données originales de l'atlas examiné suivant plusieurs critères linguistiques (phonétique, morpho-syntaxe et lexique) pour enfourner ensuite les données d'atlas ainsi taxées dans les lignes et colonnes de la MD, au lieu de les visualiser tout de bon sur une ou plusieurs cartes muettes (CM) comme le fit jadis Jules Gilliéron.

Notre travail de « taxation » se distinguait par deux particularités de celui de Gilliéron :

- par le principe de synthétiser un nombre aussi grand que possible de cartes d'atlas alors que Gilliéron se limitait toujours à l'analyse de cartes-ALF isolées,
- et par le souci d'analyser (ou: de « taxer ») les cartes originales (CO) de l'atlas en question selon toutes les catégories linguistiques et non seulement, à l'instar de Gilliéron, selon le *lexique*, et d'inclure ainsi aussi la *phonétique* et la *morpho-syntaxe* dans notre travail taxatoire.

⁸ Voir le site officiel de l'ALPI: <http://www.alpi.csic.es/>.

C'est ainsi que nous avons pu tirer, moyennant l'analyse taxatoire de seulement 64 des 75 *cartes originales* (CO) de l'ALPI (toutes appartenant au premier et unique volume de l'ALPI) 375 *cartes « de travail »* (CT) dont la bigarrure linguistique interne (*polynymie*) est très inégale. Elle va de deux types (ou « taxats ») par CT jusqu'à 104 types/taxats par CT.⁹ Chacun de ces taxats¹⁰ occupe d'ailleurs une aire particulière qui, depuis belle lurette, est appelée à Salzbourg « aire taxatoire » (AT).¹¹

Voici le bilan de notre travail taxatoire:

Catégorie linguistique	Σ cartes de travail (CT)
Vocalisme	189
Consonantisme	142
Lexique	44

Encore quatre remarques explicatives :

- Quant à la qualité intrinsèque des CO dépouillées : nous avons dû écarter, pour des raisons statistiques, toutes les CO compromises par un surplus de données manquantes (< 5% du réseau intégral de 529 points d'enquête).

- Quant aux éléments (= points) de notre réseau de recherche : nous avons ajouté, aux 529 points originaux de l'ALPI,¹² trois points factices¹³ qui correspondent aux langues standard du *portugais*, de l'*espagnol* et du *catalan*. Il en résulte un total (= N) de 532 points-ALPI (correspondant à d'autant de polygones).

- Quant au support cartographique de base : une des actions innovatrices de Jules Gilliéron était le recours systématique aux *cartes muettes* (CM) que l'imprimerie Georges Protat & Frères (à Mâcon, Saône-et-Loire), chargée de l'impression de l'ALF,

⁹ Pour une analyse du rôle de la polynymie cf. Goebel 2014.

¹⁰ Dans notre analyse de l'ALPI, le nombre total de ces taxats (et aussi des aires taxatoires y ayant trait) est de 2 945.

¹¹ Les « cartes de travail » (CT) et les « aires taxatoires » (AT) constituent les piliers empiriques centraux de l'EDMS: cf. Goebel / Smečka 2016b.

¹² Rappelons que le réseau de l'ALF comprend 638 *points d'enquête* et 639 *enquêtes*. Ce petit décalage s'explique par le fait que l'enquêteur Edmond Edmont a fait deux *enquêtes* au P.-ALF 284, Saint-Pol-sur-Ternoise, son village natal.

¹³ Voici les correspondances : *espagnol* = P. 999 (situé à l'emplacement de Madrid), *catalan* = P. 998 (à l'emplacement de Barcelone), *portugais* = P. 997 (à l'emplacement de Lisbonne).

avait mises à la disposition des romanistes peu après 1902.¹⁴ Ce faisant il a inauguré ainsi une espèce de « navette typisante » entre les données originales de l'ALF et leurs métabolismes cartographiques simplifiés. En effet, la géographie linguistique romane à proprement parler est née de la visualisation systématique de la variabilité des faits linguistiques sur des vecteurs en papier. Depuis le début du XX^e siècle, le maniement courant de CM est donc de rigueur au sein des études romanes.¹⁵ Dans le cadre de l'EDMS, la préférence est donnée aux CM *polygonisées* qui ont l'avantage de créer un lien sémiotique bien défini entre la mise en carte de faits linguistiques par *plages*¹⁶ d'un côté, et par *lignes*¹⁷ de l'autre. Rappelons aussi que la découverte de ce lien est le mérite d'un germaniste allemand (Carl Haag 1860-1946) qui en avait parlé la première fois en 1898 dans une thèse sur les dialectes de Souabe (Allemagne du Sud-Ouest) moyennant des échantillons cartographiques remarquables. Malheureusement, la clairvoyance pionnière de C. Haag a longtemps été méconnue à l'intérieur et à l'extérieur des études germaniques.

- Quant à la structure interne de la MD : dans tous les atlas linguistiques les superficies (et leurs pourtours) des AT sont très changeantes et bien loin d'être « prévisibles » ou « prédictibles ». Il s'agit là d'un fait universel ! La coïncidence exacte des AT ne se réalise pratiquement jamais, même quand deux AT se réfèrent à la même origine étymologique. Ce phénomène représente le pendant spatial de la fameuse incohérence des « lois » phonétiques qui a fait couler tant d'encre parmi les linguistes à partir du dernier quart du XIX^e siècle (et au-delà).

C'est ainsi qu'à l'intérieur d'une MD, il en résulte un « désordre » curieux pour lequel nos prédécesseurs ont créé le slogan « Chaque mot a son histoire » (cf. Jaberg 1908 et Christmann 1971), et que nous appelons, depuis plusieurs années, « enchevêtrement particulier » [EP] (des AT). Toutes les AT extraites d'un atlas

¹⁴ C'est Karl Jaberg (1877-1958) qui a présenté le premier, en 1906, les nouvelles CM aux romanistes du temps. Voir aussi notre bilan historique de 2018a.

¹⁵ Il y a là une grande différence par rapport aux *germanistes* (et *anglicistes*) dont les atlas linguistiques ne contiennent pas de données brutes (c.-à-d. des transcriptions originales), mais des données préalablement typisées (ou : « taxées ») par les éditeurs respectifs.

¹⁶ Terme technique de la cartographie « officielle » : cartes « choroplèthes ».

¹⁷ La cartographie les appelle cartes « isarithmiques ».

linguistique sont donc enchevêtrées entre elles d'une façon fort complexe ce qui n'empêche nullement leur analyse combinatoire par voie statistique : c'est là le véritable « sel » de la DM en général et de l'EDMS en particulier.

Le tableau suivant est une synopse comparative entre le contenu de la figure 3 et notre corpus-ALPI de base :

	Nombre des points d'enquête	Nombre des cartes de travail (CT)	Nombre des taxats (= aires taxatoires, AT)
figure 3	6	4	10
notre corpus-ALPI	532	64	2 945

4. Présentation des planches 1-12

Chacune des 12 planches de l'annexe contient deux cartes multicolores. Toutes les cartes de *gauche* sont issues d'un traitement selon le mode-DM *alpha* traditionnel (voir la fig. 1) alors que les cartes de *droite* se réfèrent au mode-DM *béta* de date récente (voir la fig. 2) : elles sont donc inédites et nettement innovatrices.¹⁸

Néanmoins leur visualisation suit les ornières bien rodées de l'EDMS. C'est ainsi que le coloriage des cartes *choroplèthes* et *isarithmiques* de l'annexe a été réalisé à l'aide de deux algorithmes de visualisation (MINMWMAX et MEDMW) qui depuis longtemps font partie intégrante de toute la philosophie visualisatrice de l'EDMS et du logiciel VDM (« Visual DialectoMetry »).

Les deux algorithmes réunissent les scores calculés (et stockés ensuite dans les MSim et MDist) en six intervalles numériques préalablement définis¹⁹ tout en leur

¹⁸ Quant à nos travaux-DM antérieurs – réalisés, en mode-DM *alpha*, avec les données de l'ALF – nous ne renvoyons ici qu'à nos contributions, toutes rédigées en langue française, de 2000, 2002, 2003 et 2009a. Les lecteurs intéressés y trouveront un grand nombre d'exemples, tous dûment présentés, expliqués et interprétés. Pour une bibliographie complète de nos travaux-DM, toujours tenue à jour, voir notre site personnel : https://www.sbg.ac.at/rom/people/prof/goebl/dm_publici.htm.

¹⁹ Dans VDM, le nombre des intervalles chromatiques peut varier entre 2 et 20. Malheureusement la perception de l'œil humain n'est pas à la hauteur de ces possibilités techniques. En règle générale, la meilleure perception de la structure de cartes choroplèthes multicolores s'obtient moyennant 6 ou 8 classes chromatiques.

attribuant des couleurs arrangées selon la logique de l'arc-en-ciel (ou du spectre solaire).

C'est ainsi que les couleurs *froides* (= bleu foncé, bleu clair et vert) correspondent aux scores situés *au-dessous* de la moyenne arithmétique (MA) de la distribution de fréquence qui sous-tend la carte en question, alors que les couleurs *chaudes* (= rouge, orange et jaune) se réfèrent aux valeurs situées *au-dessus* de la même MA.

L'algorithme MINMWMAX confère aux intervalles des *largeurs* numériques égales, alors que l'algorithme MEDMW crée des intervalles *égaux* quant au nombre de *leurs éléments* (= points d'enquête ou polygones).

Précisons encore que le recours systématique aux couleurs de l'arc-en-ciel offre de grands avantages perceptifs à l'œil humain.

En règle générale, l'algorithme MINMWMAX engendre des profils choroplèthes plutôt nivelés alors que l'algorithme MEDMW crée des profils plus accidentés. Avec VDM, l'on peut facilement sauter entre MINMWMAX et MEDMW tout en variant librement le nombre des intervalles chromatiques entre 2 et 20. Le coloriage algorithmique, à la fois multiple et nuancé, des scores-DM stockés dans les colonnes des deux matrices métriques carrées (MSim et MDist) facilite énormément la comparaison visuelle entre les deux séries de cartes, calculées selon les modes α et β .

4.1 Les cartes de similarité (CSim): voir les planches 1-6²⁰

Chacune des CSim dispose d'un point *de référence* [ou : point *de repère*] (PR) précis qui sert d'étalon « de comparaison » avec le reste du réseau. Chaque CSim comprend donc 531 (= $N - 1$; $N = 532$) polygones coloriés alors que le polygone du PR reste en blanc.

Les PR que nous avons extraits du réseau de l'ALPI à titre d'exemple, concernent le nord-ouest (zone Eo-Navia → planche 1), la Galice (→ planche 2), l'ouest avec

²⁰ Sur les cartes choroplèthes des planches 1-10, deux des 531 ou 532 polygones respectifs sont marqués par des hachures blanches. Ce graphisme renvoie soit à la valeur *maximale* (*blanc sur rouge*), soit à la valeur *minimale* (*blanc sur bleu foncé*) de la distribution de fréquence visualisée.

Lisbonne (→ planche 3), le centre avec Madrid (→ planche 4), le nord-est avec Barcelone (→ planche 5) et l'île de Majorque (→ Planche 6).

En général, le contenu et la structure des CSim dépendent largement de l'indice de similarité utilisé pour leur calcul. Or, cet indice devrait correspondre – au moins en théorie – à une conception préétablie de la similarité linguistique (en termes quantitatifs) à être relevée entre les différents idiomes. Comme la statistique et la taxométrie offrent une gamme très vaste d'indices de *similarité* (et aussi de *distance*) qui tous répondent à des finalités scientifiques très différentes, les dialectométriciens doivent régler leur choix sur les présupposés méthodiques de la géolinguistique.

Parmi les indices adoptés par l'EDMS sous ces auspices, l'« Indice Relatif d'Identité » (IRI)²¹ occupe une place de choix. Il peut être considéré comme indice standard de l'EDMS.

4.1.1 Présentation et interprétation des deux cartes de la planche 1

Le PR des deux cartes de la planche 1 se trouve aux confins entre la Galice et l'Asturies, magnifiquement illustrés par l'atlas linguistique ETLEN, publié en 2017 par les soins de Ramón de Andrés Díaz et son équipe.

Alors que la carte de gauche reflète bien la position intermédiaire du P.-ALPI 301 (Campas), la structuration beaucoup plus « compacte » de la carte de droite (établie en mode β) témoigne (par le biais d'un puissant effet de lissage) de la vertu « égalisatrice » de ce module-VDM. Évidemment, sur les deux cartes apparaissent également – avec différents degrés de netteté – les trois grands domaines linguistiques de l'Ibéro-Romania (i. e. le portugais, le castillan et le catalan) dont la composante centrale a littéralement « engouffré », sur la carte de droite (en mode β), le P.-ALPI 301 et ses alentours.

²¹ La dénomination originale allemande est *Relativer Identitätswert*, l'équivalent anglais est *Relative Identity Value*. Pour une explication détaillée de l'IRI cf. Goebel (1981, 2000 et 2002).

Signalons que l'effet de lissage du mode-VDM *béta* s'est manifesté puissamment dans toutes nos analyses-DM majeures. Il semble pourtant que l'exigüité du nombre des CT tirées de l'ALPI atténue considérablement cet effet.²²

4.1.2 Présentation et interprétation des deux cartes de la planche 2

Cette fois-ci, le PR (= P.-ALPI 108, Corcubióm) se situe sur la pointe occidentale de la Galice. Il est donc normal que le profil choroplèthe de gauche (exécuté en mode α) en reflète les filiations « galiciennes » au sein de l'Ibéro-Romania, alors que profil de droite (réalisé en mode β) en donne une vision beaucoup plus grossière et compacte. Ceci n'empêche que, sur cette carte, à la lisière entre la Galice, les Asturies et la Vieille Castille, pointe, en jaune, un couloir oblong de transition.

4.1.3 Présentation et interprétation des deux cartes de la planche 3

Le PR (997, situé à l'endroit de Lisbonne), correspond au portugais standard. C'est ainsi que les deux cartes exhibent fort bien l'excentricité géographique de la latinité lusitane par rapport au reste de l'Ibéro-Romania. Une fois de plus, le profil de gauche (calculé en mode α) en offre une vision détaillée, alors que la carte de droite (créée en mode β) dispose d'une structuration plus ramassée.

4.1.4 Présentation et interprétation des deux cartes de la planche 4

Le PR (= P.-ALPI 999), équivalent à l'espagnol standard, se situe aux environs de Madrid. Il s'agit – come sur les planches 3 et 5 – d'un point d'atlas artificiel qui n'appartient pas au réseau original de l'ALPI. Sur les deux cartes, les larges plages en rouge symbolisent l'extension géographique de la mouvance castillane, qui évidemment, ressort avec plus de netteté à droite (mode β) qu'à gauche (mode α).

²² Voir à ce sujet les graphiques présentés dans Goebel (2022), tous relatifs à l'ALF.

4.1.5 Présentation et interprétation des deux cartes de la planche 5

Sur cette planche, le PR, situé à la place de Barcelone, correspond au catalan standard (= P. 998). Les deux profils choroplèthes – qui se ressemblent d'ailleurs beaucoup – illustrent bien l'altérité linguistique du catalan par rapport au reste de l'Ibéro-Romania. Alors que la fameuse bipartition interne (ouest-est) du domaine catalan ressort bien du profil de gauche (en mode α) – voir le contraste entre les plages en rouge et en orange –, celle-ci a complètement disparu sur le profil choroplèthe de droite sous le coup de l'effet égalisateur du mode-DM *béta*.

4.1.6 Présentation et interprétation des deux cartes de la planche 6

Du point de vue linguistique, le PR (P. 797) est toujours de type catalan : il se trouve à Campos en Majorque. En dernière analyse, les deux profils choroplèthes accusent les mêmes structures et propriétés qu'auparavant (voir la planche 5).

4.2 L'analyse interponctuelle en fonction discriminatoire (« cartes isoglottiques ») et en fonction communicative (« cartes à rayons ») : voir les planches 7 et 8.

Le terme et le concept de l'*interpoint* ont été créés, à partir de 1949, par l'abbé Théobald Lalanne (1880-1952) qui était un linguiste-amateur très inventif. Ce concept (de nature géométrique) présuppose l'existence d'un réseau de recherche avec N points d'enquête, tous répartis plus ou moins équitablement dans l'espace, entre lesquels s'effectuent, l'une après l'autre, deux opérations géométriques: d'abord le traçage de *lignes droites* entre les points avoisinants du réseau (\rightarrow triangulation de Delaunay),²³ et ensuite la construction des verticales (ou: perpendiculaires) respectives

²³ Mathématicien russe (B. N. Delone, 1890-1980) dont le nom russe a été francisé après coup en *Delaunay*.

(→ tesselage [ou : polygonisation] de Voronoi)²⁴ quitte à les prolonger jusqu'à leur fusion réciproque.

Dans le premier des deux cas, il en résulte un *treillis triangulaire*, alors que le résultat graphique du second est une *mosaïque polygonale* (polygonisation de Voronoi).

Les interpoints de mémoire lalannienne ont donc deux apparences²⁵ – *géométrique* et *géolinguistique* – auxquelles correspondent deux fonctions linguistiques opposées : celle de la *distanciation* (ou *démarcation*) par les *côtés de polygone* (voir la planche 7), et celle de la *communication* (inter-individuelle) à travers les *côtés de triangle* (voir la planche 8).

À l'aide de ces deux principes géométriques nous avons tiré, des 532 points de notre réseau de recherche, 1 453 interpoints qui, dans la perspective de l'EDMS, correspondent à autant de valeurs de *distance* (voir les deux cartes de la planche 7) ou de *similarité* (voir les deux cartes de la planche 8). Leur mise en carte aboutit à deux messages géolinguistiques complémentaires : alors que la planche 7 signale le *cloisonnement* interponctuel général, la planche 8 renvoie à la *connectivité* interponctuelle. Bien que, dans les deux cas, les implantations spatiales soient très claires, l'impact visuel du *cloisonnement* tel qu'il ressort des deux cartes de la planche 7, l'emporte de loin sur celui de la *connectivité*, visible sur les deux cartes de la planche 8.

Sur la carte de droite (en mode β) de la planche 7, la vertu égalisatrice du mode-DM bêta ressort avec clarté par des bourrelets de frontière épais (en bleu foncé et clair) qui, tant à l'ouest qu'à l'est, sillonnent verticalement l'espace de l'Ibéro-Romania de façon iconiquement « souveraine », c.-à-d. sans être perturbés par des cloisonnements concurrents tels qu'ils foisonnent sur la carte de gauche (en mode α).

²⁴ Mathématicien russe (G. F. Voronoi, 1868-1908). La paternité de ce principe géométrique très utile, est multiple : voici les noms de deux autres pionniers concurrents : P. G. Dirichlet (1805-1859) et A. H. Thiessen (1872-1956).

²⁵ Cf. à ce sujet notre contribution de 1983.

Sur la planche 8 en revanche, le pendant géométrique des bourrelets isoglottiques omniprésents sur la planche 7, est à rechercher dans plusieurs failles triangulaires effilées, toutes en bleu foncé, qui s'inscrivent entre les trois domaines linguistiques majeurs et aussi la Galice. Une fois de plus, la carte de droite (en mode β) montre des résultats plus serrés.

4.3 Une carte à paramètre : la synopsis des « coefficients d'asymétrie de Fisher » (CAF) : voir la planche 9

Un des exploits majeurs de l'EDMS est la découverte que les N distributions de similarité stockées dans une MSim carrée ($N \times N$) sont porteuses non seulement d'informations linguistiques de type *synchronique*, mais aussi de type *diachronique*. Ces dernières peuvent être repérées surtout à travers la mise en carte synoptique des « paramètres typiques » des distributions de similarité tels le *minimum*, le *maximum*, la *moyenne arithmétique* (MA), l'*écart-type* ou le *coefficient d'asymétrie de Fisher* (CAF). Surtout la synopsis du CAF permet de mieux juger d'une des propriétés les plus saillantes de tout réseau d'atlas, à savoir de l'« enchevêtrement particulier » (EP) des AT.

Or, depuis les recherches de J. Gilliéron (et de beaucoup d'autres), il est bien connu que, dans l'espace, le « changement linguistique » se manifeste par de multiples « bousculades » parmi les nombreuses AT « en lice », toutes engagées dans des processus tant d'expansion que de retrait. Or, c'est justement le paramètre du CAF qui permet de déterminer les régions où, à l'intérieur du réseau de l'ALPI, ces « bousculades expansionnistes » ont été les plus fortes, et celles où ces « télescopages » étaient de nature plus discrète.

Sur la planche 9, les scores négatifs du CAF, tous coloriés en *bleu foncé* (etc.), renvoient à des régions marquées par d'intenses « télescopages expansifs » entre les AT rivales, alors que les scores positifs du CAF, tous coloriés en *rouge*, *orange* et *jaune*, signalent des zones où les AT régionales en lice menaient une « vie beaucoup plus calme ».

À la lumière de l'histoire des trois grands domaines linguistiques ibéro-romans, tous profondément marquée par les tourbillons pluri-séculaires de la Reconquête, il n'est pas étonnant de voir que les zones les plus « mouvementées » (toutes marquées en bleu foncé) se situent – sur toutes les deux cartes de la planche 9 – entre les deux Castille, au nord de l'Aragon et entre la Galice et les Asturies, alors que les parages plus calmes (marqués en rouge, orange et jaune) se confondent avec les domaines du portugais, du galicien et du catalan.

4.4 L'analyse corrélatrice (la similarité linguistique corrélée avec la proximité géographique) : voir la planche 10

La méthode-DM de l'analyse corrélatrice dépend d'une assomption théorique capitale : à savoir qu'il existe, dans les profondeurs d'une MD tirée d'un ALx, des structures globales – d'entrée de jeu invisibles à l'œil nu de l'observateur –, qui peuvent être considérées comme émanant d'un aménagement particulier de l'espace effectué par les humains par voie *linguistique* (mieux : *dialectale*).

Depuis un certain temps (2005) nous appelons cette activité anthropique « gestion basilectale de l'espace par l'*homo loquens* ». Il paraît donc indiqué d'examiner, pour le même réseau de N points, les liens corrélatifs qui existent entre la gestion *linguistique* de cet espace pratiquée par l'homme, et celle qui résulte – par voie *naturelle* – des distances (mieux : proximités) kilométriques réciproques entre les N localités du même réseau.

Une telle comparaison met donc en œuvre les dimensions de l'« anthropique » (= géolinguistique) et du « physique » (= géométrie).

En termes de taxométrie, ceci se fait moyennant N (ici: 532) calculs de corrélation à exécuter à l'aide de l'indice de corrélation de Bravais-Pearson $r(BP)$ entre deux séries de valeurs, toutes répertoriées au sein de deux matrices carrées aux dimensions $N \times N$: d'un côté il s'agit des N colonnes de la matrice de *similarité* (calculées moyennant l'indice IRI_{jk}), et de l'autre des N colonnes d'une matrice de

proximité analogue, dans les cellules de laquelle ont été stockées les valeurs non pas des *distances* kilométriques (*dist*) – calculées d’ailleurs selon le fameux théorème de Pythagore : $a^2 + b^2 = c^2$ –, mais de leurs équivalents de *proximité* (*prox*) établies selon une formule très simple: $dist + prox = 100$.

De cette « comparaison corrélative » résultent 532 scores-r(BP) susceptibles d’être mis en carte selon les standards cartographiques courants de l’EDMS : voir les deux cartes de la planche 10.

Pour bien comprendre le message linguistique de ces deux cartes choroplèthes aux structures très claires (et aussi très convergentes), il faut connaître la nature statistique des valeurs sur lesquelles elles reposent. Il s’agit de 532 scores de corrélation, tous oscillant à l’intérieur d’une fourchette numérique située entre -1 (en bleu foncé) et +1 (en rouge), qui signalent le taux de l’interdépendance entre ces deux modes de gérer l’espace.

Les couleurs *chaudes* (du rouge au jaune) signalent la présence d’un phénomène que l’on pourrait appeler le « pas cadencé » entre les deux sortes de gestion de l’espace, alors que les couleurs *froides* (allant du bleu foncé jusqu’au vert clair) signalent le contraire : l’existence de fortes dissonances (ou : antagonismes, voire frictions) entre ces deux sortes de gestion de l’espace.

Les zones en bleu foncé sont donc celles où se sont déroulées, dans le passé, les plus intenses confrontations et frictions linguistiques (et partant aussi sociales et politiques). Il s’agit des lignes de force de la Reconquête chrétienne entre les 9^e et 15^e siècles, alors que les plages en rouge et en orange renvoient à des contrées moins touchées par les turbulences relatives aux affrontements militaires à proprement parler pour ne pas oublier les vicissitudes de la ré-colonisation chrétienne.

Évidemment, les applications de telles comparaisons corrélatives sont nombreuses : c’est que les comparaisons entre différentes catégories linguistiques (p. ex. *phonétique vs. lexicque, vocalisme vs. consonantisme*, etc.) fournissent également d’excellents résultats dont l’intérêt linguistique est considérable (voir Goebel 2005).

4.5 *L’analyse dendrographique : la « classification ascendante hiérarchique » (CAH) (méthode de Ward) : voir les planches 11 et 12*

Comme, en ligne générale, l'approche « arborescente » fait partie intégrante de la linguistique historique depuis longtemps, on ne s'étonnera point qu'elle réapparait aussi en matière de DM. Du point de vue statistique, la CAH opère avec des algorithmes²⁶ dont le but est de pratiquer, à l'intérieur de $N/2 (N - 1)$ ²⁷ scores de similarités stockés dans une MSim, N fusions exclusivement binaires, d'abord entre des éléments (= dia- ou lococolectes) isolés et ensuite entre des groupements majeurs créés successivement par voie agglomérative.

Le travail fusionniste débute au niveau des scores les plus petits (c.-à-d. près des « feuilles » de l'arbre) pour aboutir – après une « ascension » le long de l'échelle des scores de similarité triés par leur taille – à la génération de la « racine » (ou du « tronc ») de l'arbre par le biais de la fusion des deux dernières valeurs de cette échelle.

Les ramifications de l'arbre sont toujours binaires ; le « branchage » de l'arbre consiste en une hiérarchie d'embranchements que nous appelons *dendrèmes*. Pour les propos de la DM, il est de toute première importance de ne jamais considérer isolément l'arbre issu d'une classification de type CAH, mais de l'utiliser toujours pour une « spatialisation », c.-à-d. pour la projection de ses dendrèmes sur la carte de l'espace en question, tout en y créant des configurations spatiales typiques, appelées par nous *chorèmes*. Voir les quatre cartes des planches 11 et 12 où les spatialisations occupent la moitié supérieure, alors que les arbres respectifs se trouvent en bas.

L'identité du coloriage des branches des arbres (« dendrèmes ») et de leurs équivalents spatiaux (« chorèmes ») assure la détection des relations qui existent entre ces deux niveaux heuristiques.

²⁶ Les arbres présentés sur les planches 11 et 12 ont été construits moyennant l'algorithme-CAH proposé en 1963 par le statisticien américain Joe Ward, Jr : cf. Chandon-Pinson (1981 : 123). Leur utilité linguistique est très grande. Ceci n'empêche que le logiciel VDM offre encore d'autres algorithmes de type CAH dont le rendement linguistique doit être défini au besoin.

²⁷ Toute MSim carrée ($N \times N$) consiste en deux moitiés symétriques égales ($N \times N/2$) avec une diagonale (N) occupée uniquement par la valeur 100 (ou 0). La somme des scores statistiquement utilisables est donc de $(N \times N)/2 - N = N/2 (N - 1)$.

À noter aussi la différence qui existe, du point de vue statistique, entre la variabilité numérique à l'intérieur des dendrèmes (ou chorèmes) [= variabilité « *intra-group* »], et celle qui s'établit *entre* les différents dendrèmes/chorèmes [= variabilité « *inter-group* »].

La logique des algorithmes utilisés dans le cadre de la CAH veut que les deux variabilités soient minimales au niveau des « feuilles », et maximales près de la « racine » (ou du « tronc ») de l'arbre.

Il ne faut non plus passer sous silence le fait qu'il existe également une relation directe entre cette variabilité changeante et la « fiabilité » (linguistique) des classifications ainsi établies. C'est que cette dernière diminue avec l'augmentation des deux espèces de variabilité mentionnées ci-dessus. Ceci signifie que les groupements situés près des feuilles, souvent minuscules et linguistiquement peu intéressants, sont – au moins en théorie – plus « sûrs » que ceux qui à cause de leur position près de la racine, intéressent le linguiste en premier lieu.

Évidemment, il en résulte un certain dilemme, gênant, certes, du point de vue statistique, mais acceptable du point de vue heuristique.

En « lisant » un arbre-CAH dans une perspective diachronique il faut donc le « dépouiller » : soit *en gros* (voir la planche 11, à l'aide de seulement quatre dendrèmes-chorèmes), soit *au détail* (voir la planche 12, avec 10 dendrèmes-chorèmes), toujours en partant de la racine (ou du tronc), c'est-à-dire allant de droite à gauche.

Ce faisant, l'on constate qu'à la première bifurcation après la racine, le bloc castillan (en bleu foncé) se sépare du catalan (en vert) et de l'ensemble luso-galicien (en rouge et jaune) qui, lui, se désagrège à la bifurcation suivante. Grosso modo ce constat est le même pour toutes les quatre cartes de planches 11 et 12, bien qu'avec certains écarts quant à la géométrie des respectives bifurcations qui, eux, ressortent bien d'une considération comparée des deux arborescences.

Plus on entre dans les profondeurs de l'arbre, plus la logique et la géométrie des bifurcations changent. C'est ainsi que les ramifications générées dans les deux modes

(α et β) divergent de plus en plus, et que les chorèmes respectifs se rapetissent continuellement.

Ceci n'empêche que même des dendrèmes plutôt éloignés de la racine de l'arbre peuvent créer des chorèmes linguistiquement très intéressants dont un bel exemple se trouve au sein des deux arbres de la planche 12. Il s'agit du plus petit des deux dendrèmes asturiens, coloriés en bleu foncé et violet qui, au moment de la spatialisation, crée un chorème linéiforme en violet qui, tout en se coinçant entre la Galice et l'Asturies, s'étend jusqu'à la zone Eo-Navia de l'ETLEN.

Évidemment, le meilleur moyen pour dépouiller une arborescence-CAH est d'utiliser le logiciel VDM et d'exécuter cette tâche directement sur l'écran.

5. Conclusion

Le nouveau volet du logiciel VDM s'intègre parfaitement dans la logique méthodique de l'EDMS qui, par définition, est multiple et, de ce fait, bien loin d'être un circuit clos. Il offre de nouvelles voies pour la reconnaissance des fameuses « structures de profondeur » cachées dans le total des planches d'un atlas linguistique.

C'est le bon moment pour souligner, une fois de plus, que le logiciel VDM est à la libre disposition de tous ceux qui en ont besoin et en font la demande.²⁸

Remerciements

- Confection des Figures 1-3 : Werner Goebel, Vienne.
- Collaboration technique au projet-ALPI : Pavel Smečka.

Abréviations couramment utilisées

²⁸ Voici notre adresse courriel: hans.goebel@plus.ac.at.

ALx = Atlas linguistique (quelconque)
AT = aire taxatoire (aire de diffusion d'un « taxat » ou type linguistique)
CAH = classification ascendante hiérarchique
CM = carte muette (pour le dépouillement « taxatoire » des CO d'un atlas linguistique)
CO = carte originale (d'un atlas linguistique)
CSim = carte de similarité (un des six volets taxométriques de l'EDMS)
CT = carte de travail (tirée, par le biais d'une analyse typisante ou « taxatoire », d'une CO d'un ALx)
Dist = distance
DM = dialectométrie, dialectométrique
EDMS = École Dialectométrique de Salzburg
EP = enchevêtrement particulier (cf. 3.)
IRI_{jk} = Indice Relatif d'Identité (entre les points d'atlas j et k)
MA = Moyenne arithmétique
MD = matrice de données
MDist = matrice de distance
MEDMW = algorithme d'intervallisation (cf. 4.)
MINMWMAX = algorithme d'intervallisation (cf. 4.)
MSim = matrice de similarité (avec les dimensions N × N)
PR = point de référence (ou : de repère)
Prox = proximité
r (BP) = indice de corrélation de Bravais-Pearson
VDM = Visual DialectoMetry (logiciel dialectométrique créé dans les années 1997-2000 par Edgar Haimerl pour les besoins de l'EDMS).

References

AIS = JABERG, Karl & Jakob JUD (1928-1940) *Sprach- und Sachatlas Italiens und der Südschweiz*, 8 vol. Zofingen : Ringier (réimpression : Nendeln : Krauss, 1971) (deux présences dans le web: a) <<https://www3.pd.istc.cnr.it/navigais-web/?map=1>>; b) <<https://www.ais-reloaded.uzh.ch/>>).

ALF = GILLIÉRON, Jules & Edmond EDMONT (1902-1910) *Atlas linguistique de la France*. 10 vol., Paris: Champion (réimpression: Bologne: Forni, 1968) [consultable trois fois dans le web: Innsbruck: <<https://diglib.uibk.ac.at/urn:nbn:at:at-ubi:2-4568>>; Toulouse: <<http://symila.univ-tlse2.fr/alf>>; Grenoble: <<https://lig-tdcge.imag.fr/cartodialect5/#/>>].

CHANDON, Jean-Louis & Suzanne PINSON (1981) *Analyse typologique. Théories et applications*, Paris / New York / Barcelone / Milan: Masson.

CHRISTMANN, Hans Helmut (1971) « Lautgesetze und Wortgeschichte. Zu dem Satz „Jedes Wort hat seine eigene Geschichte” », in Wolf-Dieter Stempel & Eugenio Coseriu (ed.), *Sprache und Geschichte. Festschrift für Harri Meier zum 65. Geburtstag*, München: Fink, 119-124.

DEES, Anthonij (1980) *Atlas des formes et des constructions des chartes françaises du 13^e siècle* (*Beihefte der Zeitschrift für romanische Philologie*, vol. 178), Tübingen: Niemeyer.

ETLEN = D'ANDRÉS DÍAZ, Ramón (dir.), Fernando ÁLVAREZ-BALBUENA GARCÍA, Xosé Miguel SUÁREZ FERNÁNDEZ & Miguel Rodríguez Monteavaro (coll.) (2017) *Estudiu de la transición llingüística na zona Eo-Navia, Asturias (ETLEN). Atlas llingüísticu dialectográficu, horiométricu, dialectométricu*, Uvieu / Oviedo: Universidá d'Uviéu.

GILLIERON, Jules (avec Jean MONGIN) (1905) *Scier dans la Gaule romane du sud et de l'est : étude de géographie linguistique*, Paris: Champion (réimpression de l'original avec une traduction italienne en vue par Lorenzo Massobrio: « *Segare* » nella *Gallia romanza meridionale e orientale*, Novi Ligure: Grafica editoriale universitaria, 1990).

GOEBL, Hans (1981) « Éléments d'analyse dialectométrique (avec application à l'AIS) », *Revue de linguistique romane*, 45, 349-432.

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

GOEBL, Hans (1984) *Dialektometrische Studien. Anhand italaromanischer, rätoromanischer und galloromanischer Sprachmaterialien aus AIS und ALF* (*Beihefte der Zeitschrift für romanische Philologie*, vol. 191-193), Tübingen: Niemeyer.

GOEBL, Hans (1986) „Typophilie und Typophobie. Zu zwei problembeladenen Argumentationstraditionen innerhalb der Questione ladina“, in Günter Holtus & Kurt Ringger (ed.), *Raetia antiqua et moderna. W. Th. ELWERT zum 80. Geburtstag*, Tübingen: Niemeyer, 513-536.

GOEBL, Hans (2000) « La dialectométrie de l'ALF: présentation des premiers résultats », *Linguistica*, 40, 209-236.

GOEBL, Hans (2002) « Analyse dialectométrique des structures de profondeur de l'ALF », *Revue de linguistique romane*, 66, 5-63.

GOEBL, Hans (2003) « Regards dialectométriques sur les données de l'Atlas linguistique de la France (ALF): relations quantitatives et structures de profondeur », *Estudis Romànics*, XXV, 59-121.

GOEBL, Hans (2004) „Das „Croissant“ - eine Nachschau im Abstand von 90 Jahren (mit zwei dialektométrisch erstellten Farbkarten)“, in Volker Noll & Sylvia Thiele (ed.), *Sprachkontakte in der Romania. Zum 75. Geburtstag von Gustav Ineichen*, Tübingen: Niemeyer, 159-172.

GOEBL, Hans (2005) « La dialectométrie corrélatrice. Un nouvel outil pour l'étude de l'aménagement dialectal de l'espace par l'homme », *Revue de linguistique romane*, 69, 321-367 (cartes en couleurs: 356-367).

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

<<https://raco.cat/index.php/Dialectologia/article/view/198805/266004>>

GOEBL, Hans (2008b) « La dialettomettrizzazione integrale dell'ALS. Presentazione dei primi risultati », *Revue de linguistique romane*, 72, 25-113 (avec 48 cartes en couleurs).

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

GOEBL, Hans (2009b) « Dialectometry and quantitative mapping », in Alfred Lameli, Roland Kehrein & Stefan Rabanus (ed.), *Language and Space. An International Handbook of Linguistic Variation, vol. 2: Language Mapping* (Handbücher der Sprach- und Kommunikationswissenschaft [HSK] 30.2.), Berlin: de Gruyter 2010; 1^{ère} partie: 433-457 (texte); seconde partie: 2201-2212 (cartes).

GOEBL, Hans (2013) « La dialectométrisation del ALPI: rápida presentación de los resultados », in Emili Casanova Herrero & Cesáreo Calvo Rigual (ed.), *Actas del XXVI Congreso Internacional de Lingüística y de Filología Románicas* (Valencia 2010), Berlin, Boston: De Gruyter 2013, vol. VI, 143-154 (avec sept cartes en couleurs).

GOEBL, Hans (2014) « L'impact de la polynymie des cartes d'atlas sur le résultat de calculs dialectométriques », in Polska Akademia Umiejetnosci. Instytut Filologii Romanskiej Uniwersytetu Jagiellonskiego (ed.), *Linguistique romane et Linguistique indo-européenne. Mélanges offerts à Witold Mańczak à l'occasion de son 90^e anniversaire*, Kraków / Cracovie: Polska Akademia Umiejetnosci. Instytut Filologii Romanskiej Uniwersytetu Jagiellonskiego, 243-260 (avec 10 cartes en couleurs).

GOEBL, Hans, (2018a) « La face cachée de la géographie linguistique. Bref aperçu sur les « cartes muettes » produites pour l'ALF, l' AIS et le FEW », *Revue de linguistique romane*, 82, 5-63 (avec 16 cartes en couleurs).

GOEBL, Hans (2018b) "Dialectometry", in Charles Boberg, John Nerbonne & Dominic Watt (ed.), *The Handbook of Dialectology*, Hoboken, New Jersey: John Wiley & Sons, 123-142 (avec quatre planches en couleurs).

GOEBL, Hans (2022) « Un nouveau rejeton de l' « École Dialectométrique de Salzburg »: brève présentation du mode « bêta » de la dialectométrie de Salzburg », in: Vanderci de Andrade Aguilera, Fabiane Cristina Altino & Conceição de Maria de Araújo Ramos (org.), *Estudos dialetais brasileiros e europeus: uma homenagem a João Saramago*, Campo Grande: Editora UFMS, vol. 1, 101-140.

GOEBL, Hans & Pavel SMEČKA (2016a) « L'interprétation dialectométrique des atlas scripturaux d'Anthonij Dees », *Revue de linguistique romane*, 80, 321-368 (avec 50 cartes en couleurs).

GOEBL, Hans & Pavel SMEČKA (2016b) « The Quantitative Nature of Working Maps (WM) and Taxatorial Areas (TA): A Brief Look at two Basic Units of Salzburg Dialectometry (S-DM) », in Emmerich Kelih, Róisín Knight, Ján Mačutek & Andrew Wilson (ed.), *Issues in Quantitative Linguistics 4. Dedicated to Reinhard Köhler on the occasion of his 65th birthday*, Lüdenscheid: RAM-Verlag (Studies in Quantitative Linguistics, vol. 23), 113-127 (avec quatre cartes en couleurs).

GOEBL, Hans & Pavel SMEČKA (2017) « Trois regards dialectométriques sur l'aménagement géolinguistique du domaine d'oïl, basés sur une synthèse des données médiévales réunies par Anthonij Dees en 1980 et 1983, et celles de l'ALF », in Andres M. Kristol (ed.), *La mise à l'écrit et ses conséquences. Actes du troisième colloque „Repenser l'histoire du français“*, Université de Neuchâtel, 5-6 juin 2014. Tübingen: Francke, 15-49 (avec 12 planches en couleurs).

HAAG, Carl (ou: Karl) (1898) *Die Mundarten des oberen Neckar- und Donautales (Schwäbisch-alemannisches Grenzgebiet: Baarmundarten)*, Reutlingen: Hutzler.

JABERG, Karl (1906) „Zum Atlas linguistique de la France“, *Zeitschrift für romanische Philologie*, 30, 512.

JABERG, Karl (1908) *Sprachgeographie. Ein Beitrag zum Verständnis des Atlas linguistique de la France*, Aarau: Sauerländer (version espagnole: *Geografía lingüística, Ensayo de interpretación del „Atlas lingüístico de Francia“*, Granada: Universidad de Granada, 1959).

LALANNE, abbé Théobald (1949-1952) *L'indépendance des aires linguistiques en Gascogne maritime*, Berceau de Saint Vincent de Paul: chez l'auteur (réimpression: Labatut: Atelier Histoire Trait d'Union, 2018).

PARIS, Gaston (1888) « Les parlers de France », *Revue des patois gallo-romans* 2, 161-175 (aussi in Gaston Paris, *Mélanges linguistiques. Latin vulgaire et langues romanes, langue française, notes étymologiques*, publiés par Mario Roques, Paris: Champion, 1909, 432-448).

VIDESOTT, Paul (2009) *Padania scrittologica: Analisi scrittologica e scrittometriche di testi in italiano settentrionale antico dalle origini al 1525* (Beihefte zur Zeitschrift für romanische Philologie, vol. 343), Tübingen: Niemeyer.

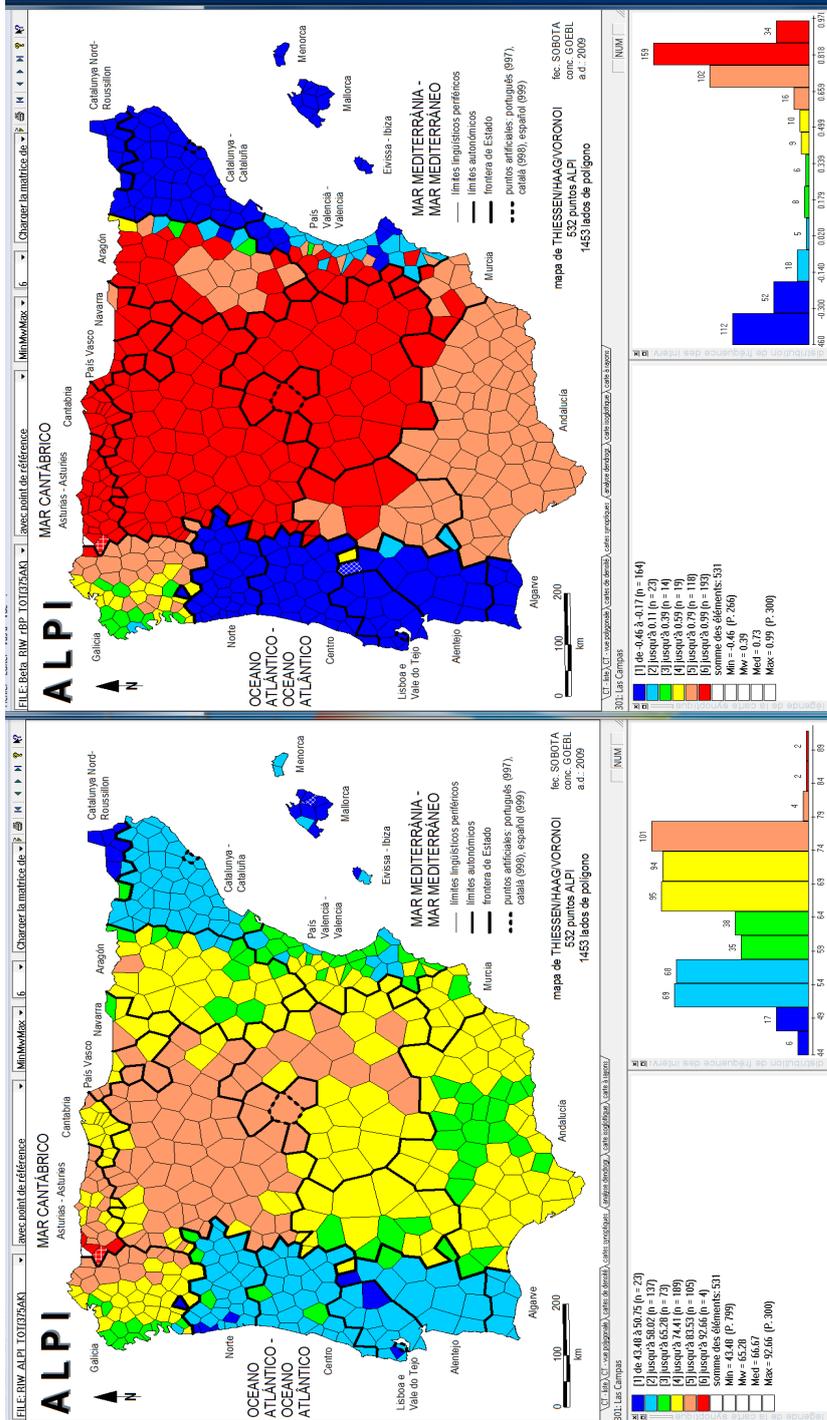


Planche 1. Deux cartes de similarité relatives au P.-ALPI 302 (Campas, Asturies)
 Corpus : 375 CT; *algorithme d' intervallisation*: MINMWMAX 6-tuple
 Calcul de similarité : à gauche (« alpha ») : $|RI|_{jk}$; à droite (« beta ») : $|RI|_{jk} + r$ (BP)

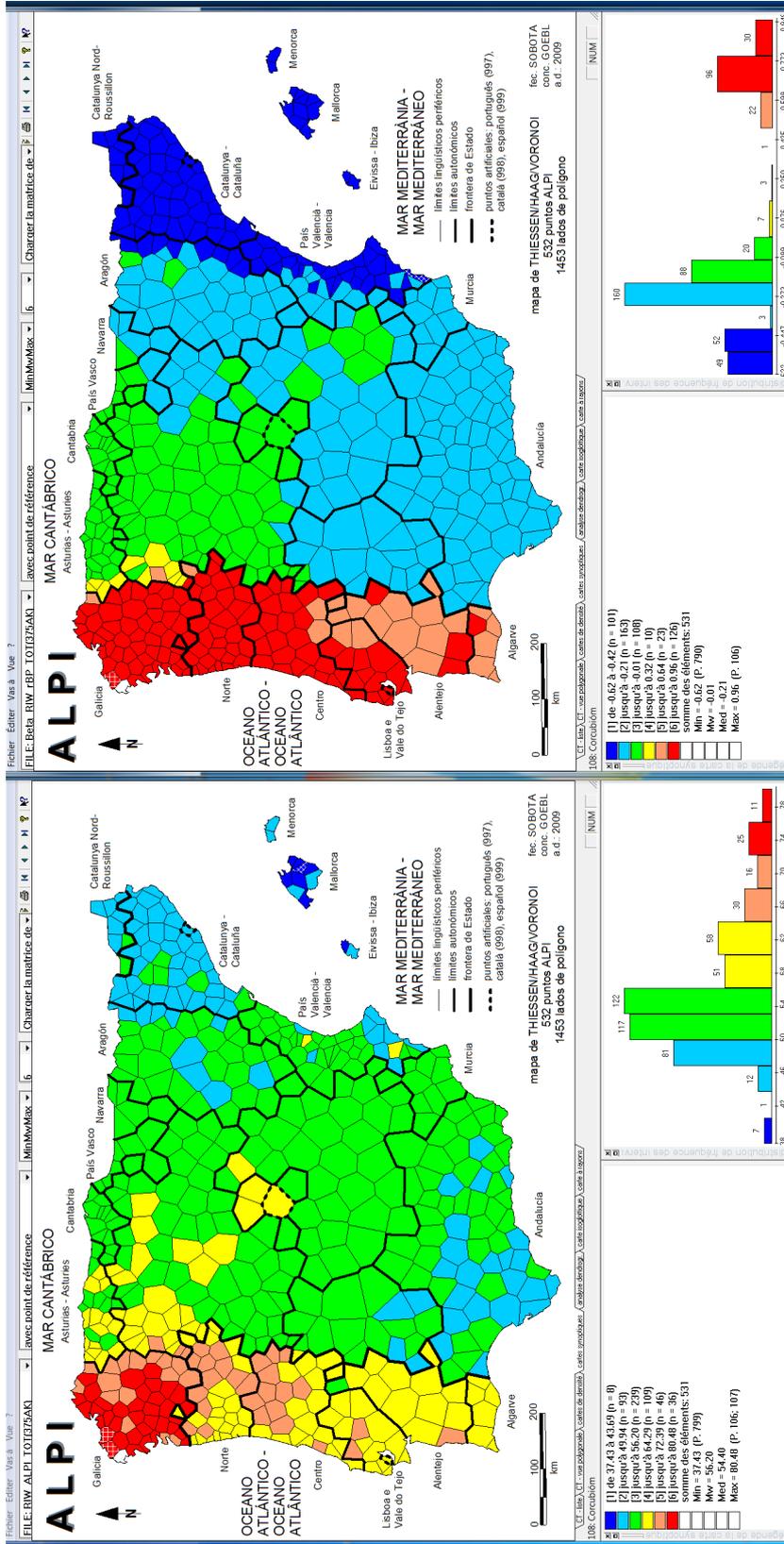


Planche 2. Deux cartes de similarité relatives au P.-ALPI 108 (Corcubióm, Galice)
Corpus : 375 CT; algorithme d' intervalisation: MINMWMAX 6-tuple

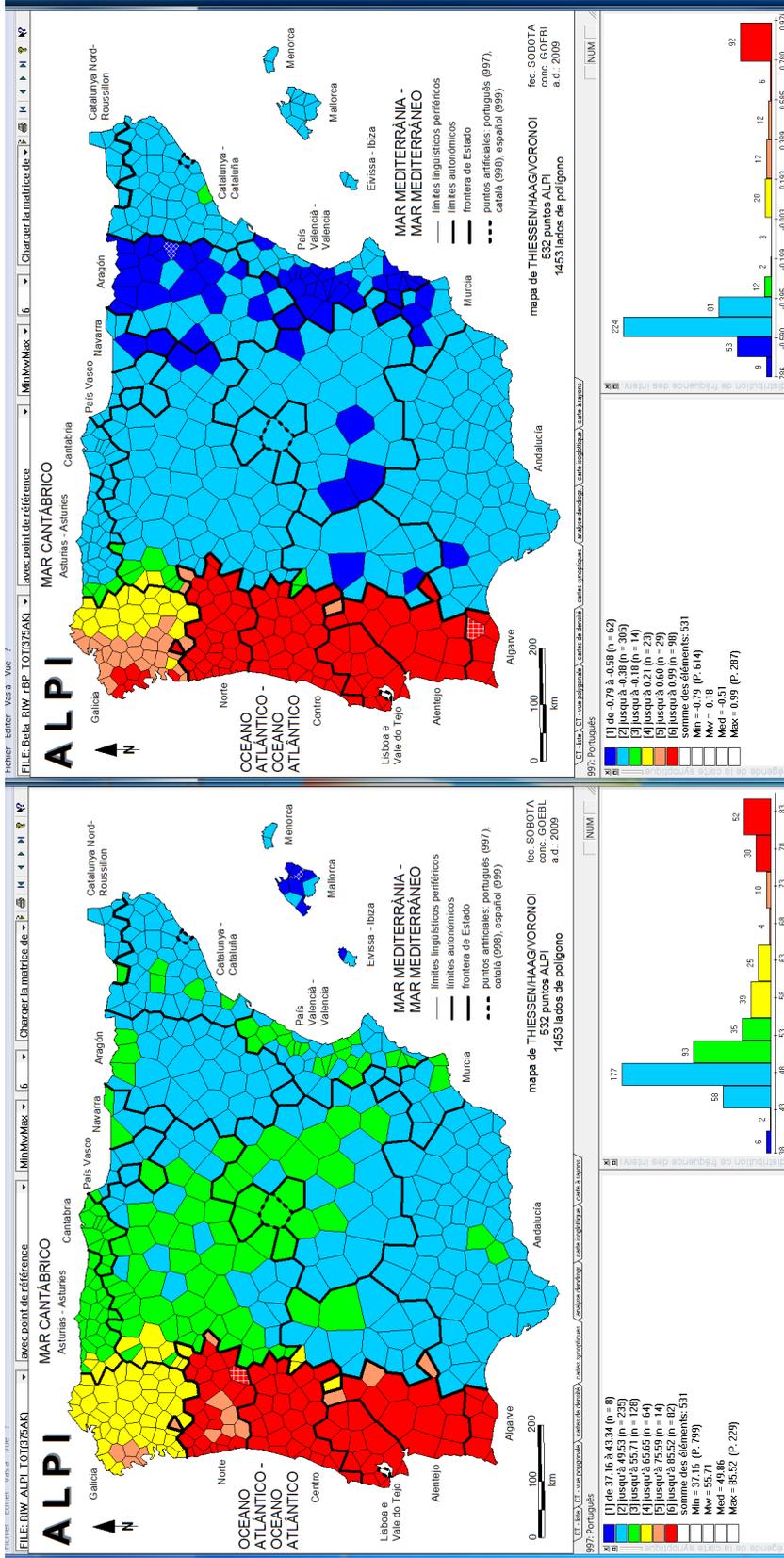


Planche 3. Deux cartes de similarité relatives aux P.-ALPI 274, à gauche (Almargem, au nord de Lisbonne) et au P.-ALPI artificiel 997, à droite (portugais standard)
 Corpus : 375 CT ; *algorithme d' intervalisation*: MINMWMAX 6-tuple
 Calcul de similarité : à gauche (« alpha »): IRI_{jk} ; à droite (« béta »): $IRI_{jk} + r$ (BP)

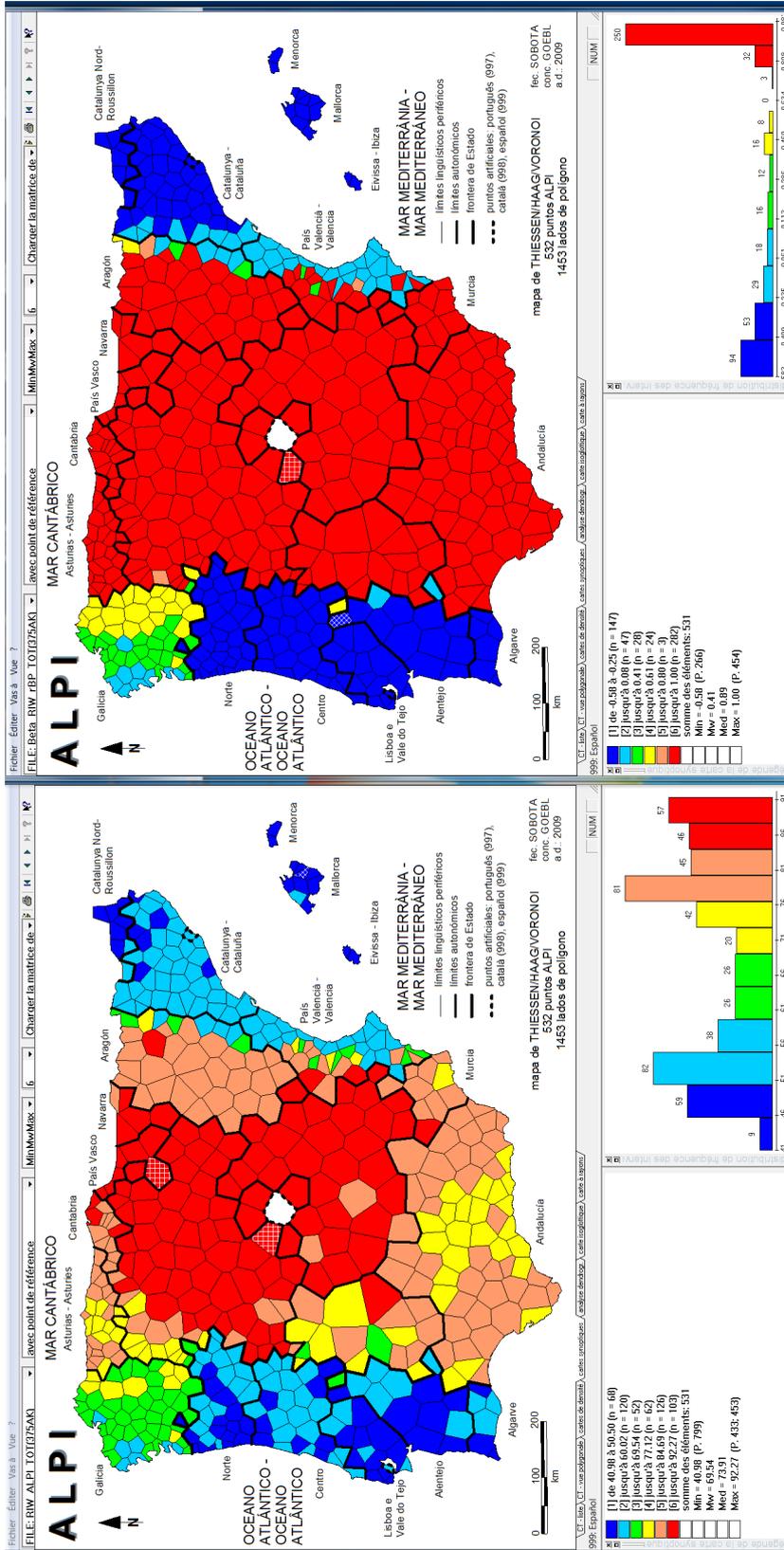


Planche 4. Deux cartes de similarité relatives au P.-ALPI artificiel 999 (espagnol standard)
Corpus : 375 CT; *algorithme d' intervalisation*: MINMWMAX 6-tuple
Calcul de similarité : à gauche (« alpha ») : IRI_{jk} ; à droite (« beta ») : $IRI_{jk} + r$ (BP)

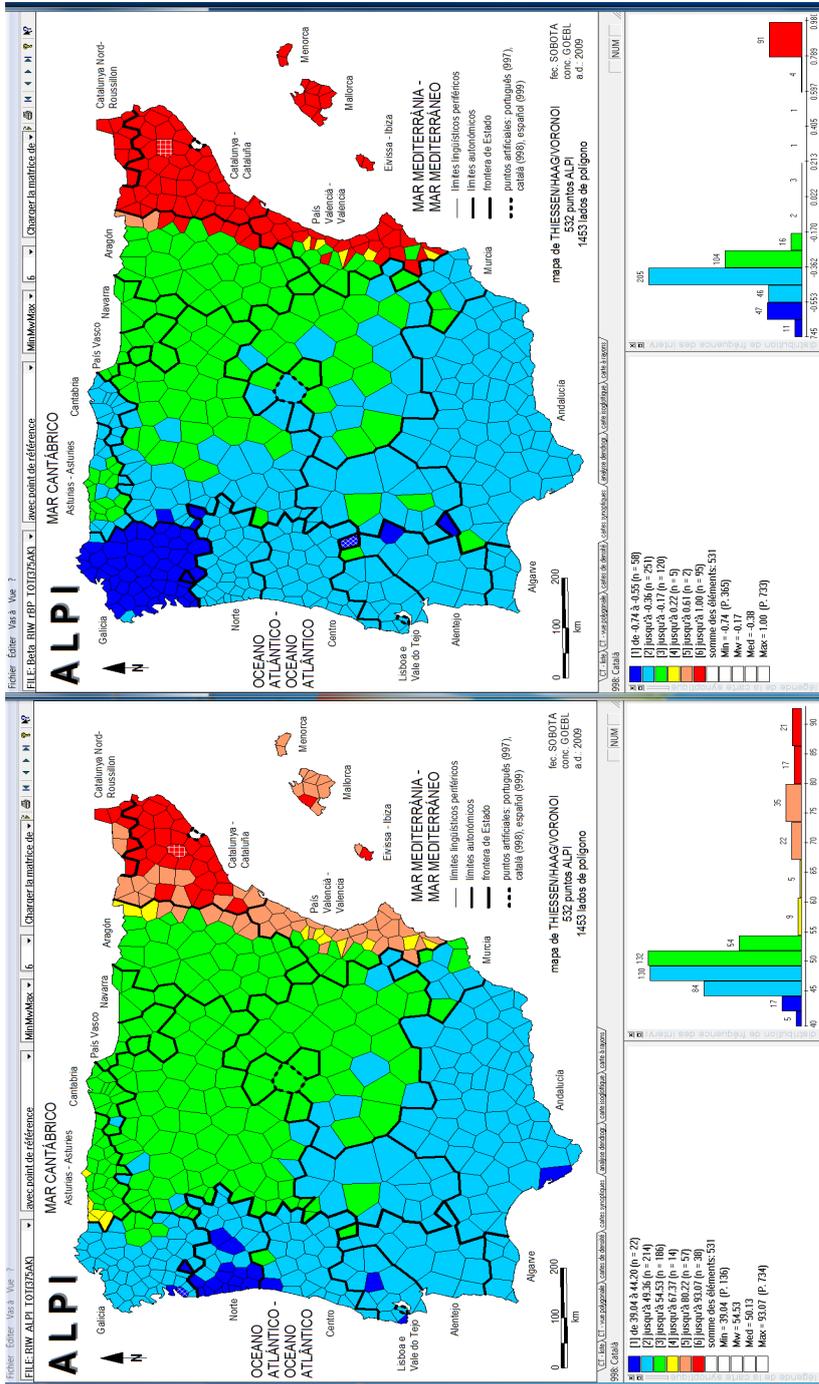


Planche 5. Deux cartes de similarité relatives au P.-ALPI 998 artificiel (catalan standard)
 Corpus : 375 CT; *algorithme d' intervalisation*: MINMWMAX 6-tuple
 Calcul de similarité : à gauche (« alpha ») : IR_{ijk} ; à droite (« beta ») : $IR_{ijk} + r (BP)$

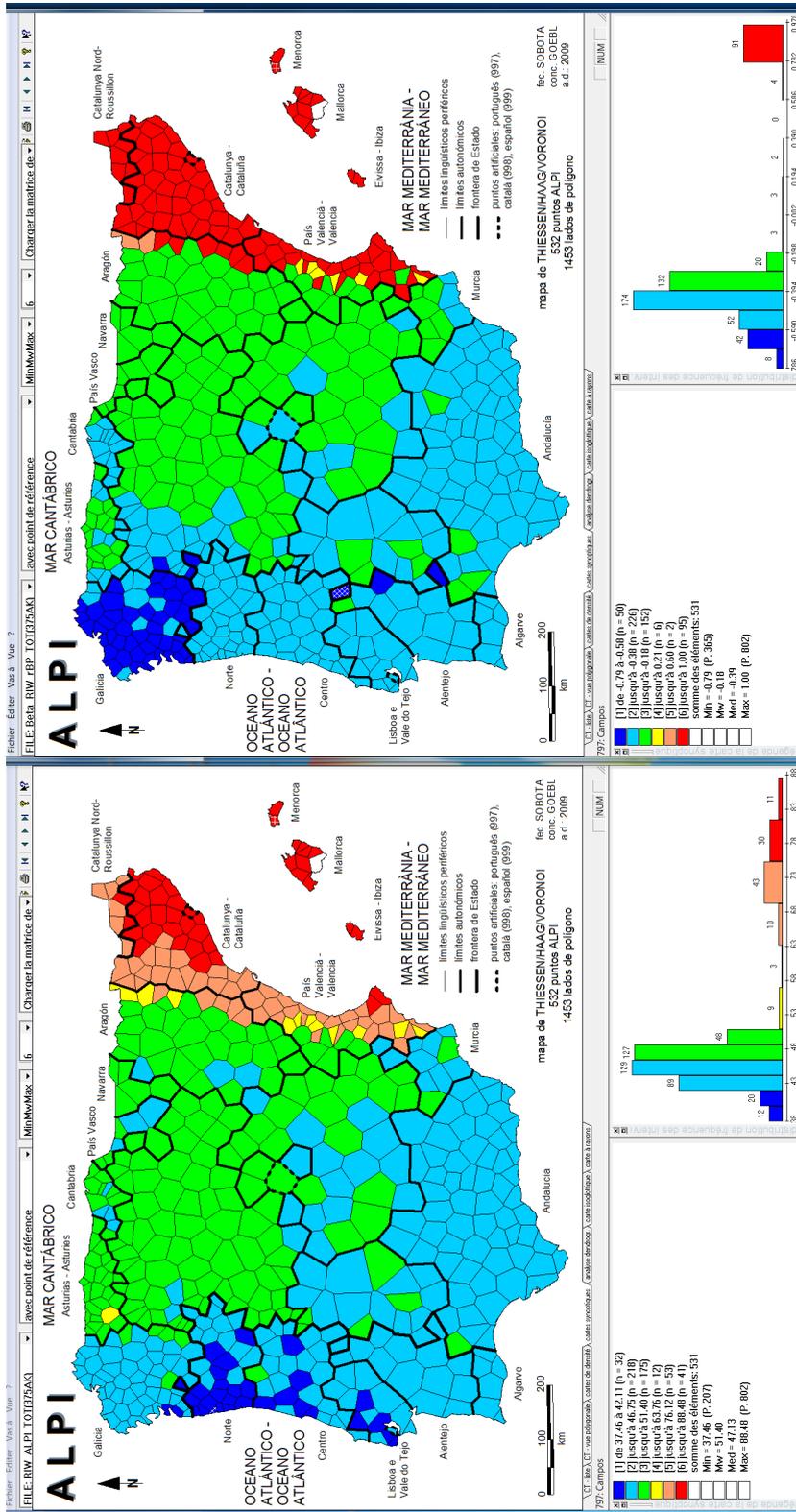


Planche 6. Deux cartes de similarité relatives au P.-ALPI 797 (Campos, Majorque)
 Corpus : 375 CT; algorithm d' intervallisation: MINMWMAX 6-tuple
 Calcul de similarité : à gauche (« alpha »): $|RI|_{jk}$; à droite (« béta »): $|RI|_{jk} + r$ (BP)

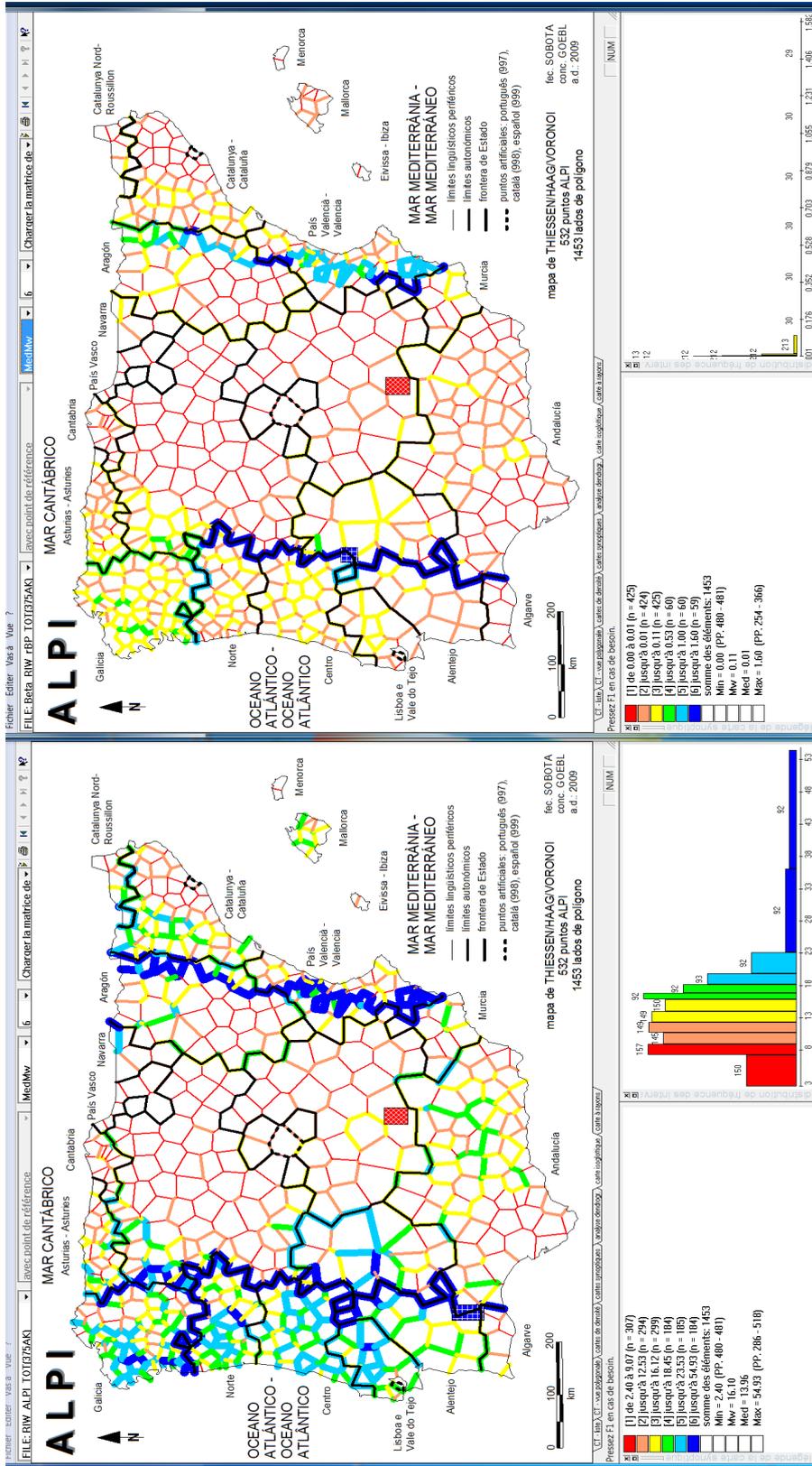


Planche 7. Deux analyses interponctuelles en fonction *discriminatoire* (→ cartes isoglotiques)
 Corpus : 375 CT ; *algorithm* d' *intervalisation*: MEDMW 6-tuple

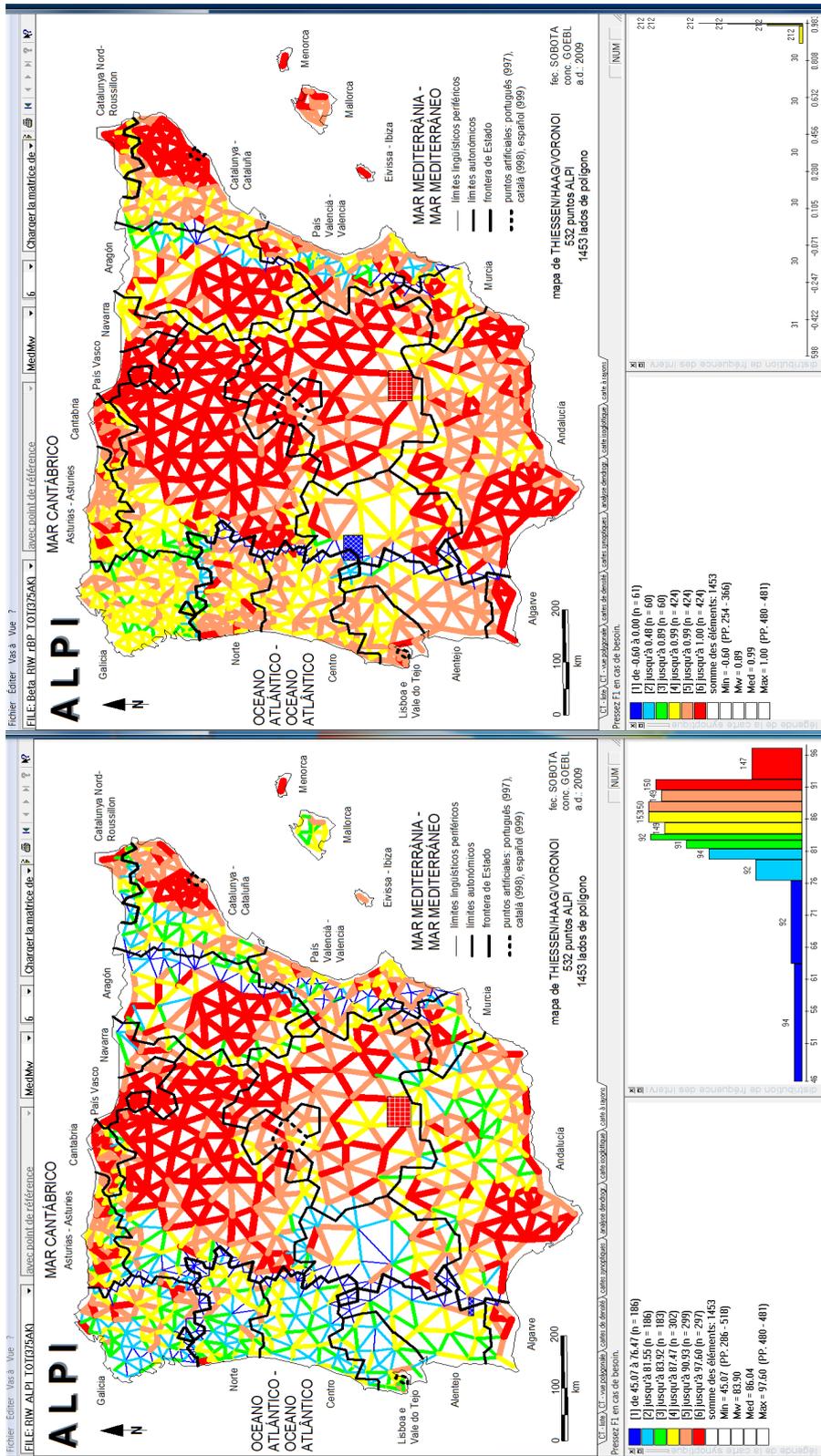


Planche 8. Deux analyses interponctuelles en fonction *communicative* (→ cartes à rayons)
Corpus : 375 CT ; *algorithme d' intervalisation*: MEDMW 6-tuple

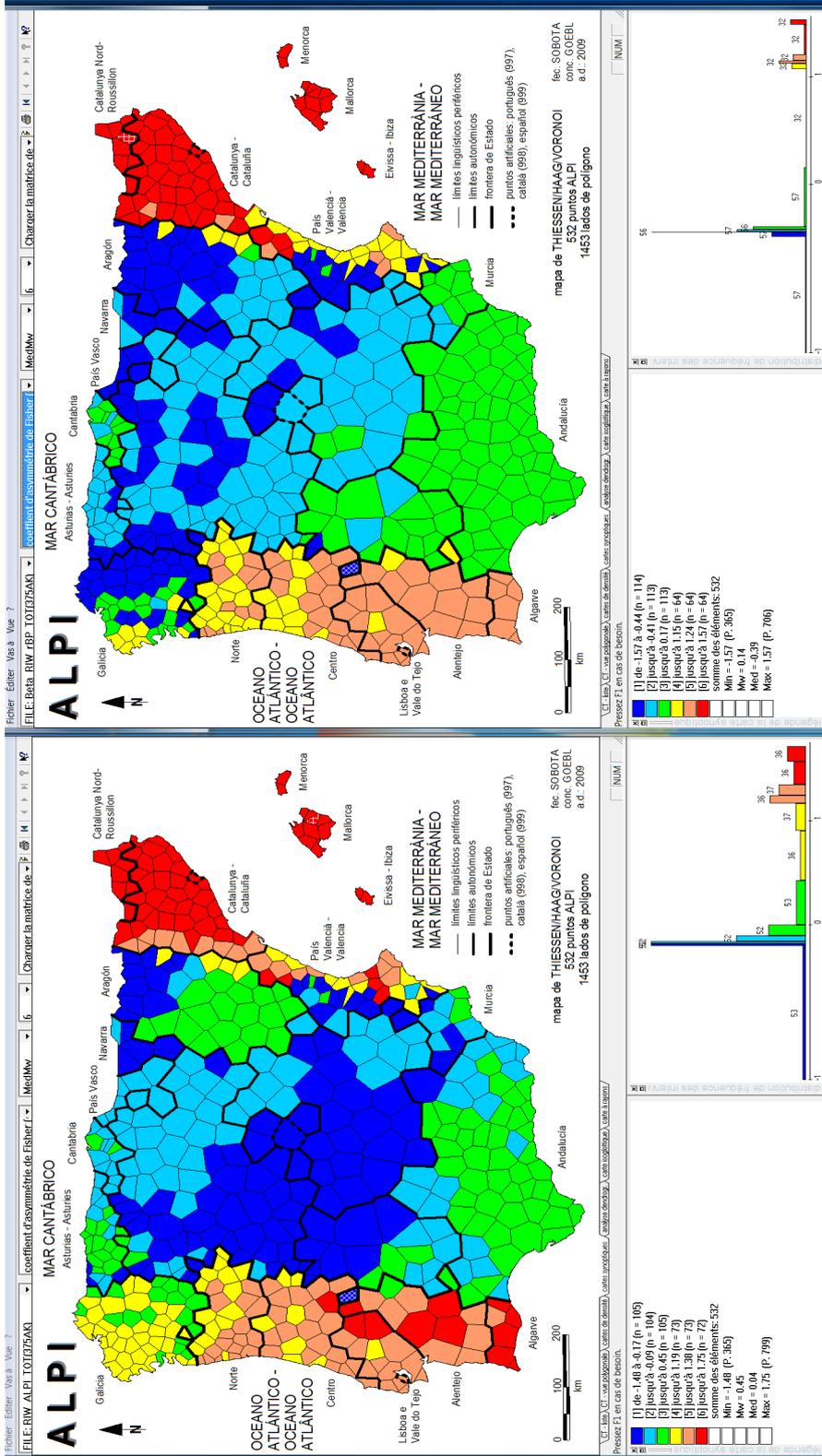


Planche 9. Deux synopses des scores du « Coefficient d' Asymétrie de Fisher » (CAF)
 Corpus : 375 CT; *algorithme d' intervallisation*: MEDMW 6-tuple
Calcul de similarité : à gauche (« alpha »): IRI_{jk} ; à droite (« beta »): $IRI_{jk} + r$ (BP)

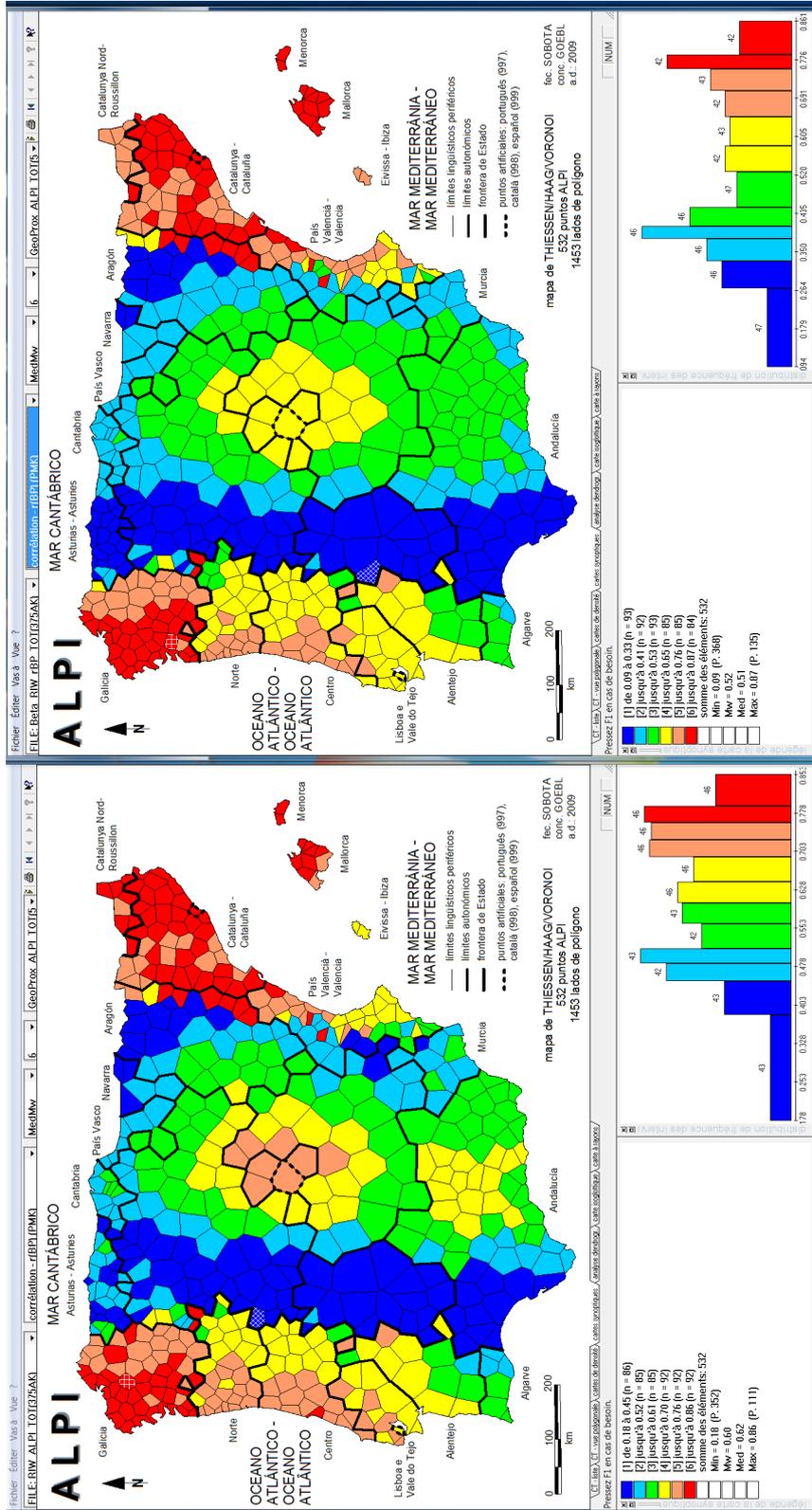


Planche 10. Deux cartes à corrélations (entre la similarité linguistique [A] et la proximité géographique $prox_{pyrtho} [= 100 - dist_{pyrtho}]$ [B])

Corpus A (à gauche et à droite): 375 CT; algorithm d' intervalisation: MEDMW 6-tuple

Calcul de similarité A : à gauche (« alpha »): IRI_{jk} ; à droite (« bêta »): $IRI_{jk} + r$ (BP)

Calculs de la proximité géographique B : à gauche et à droite: $100 - dist_{pyrtho}$

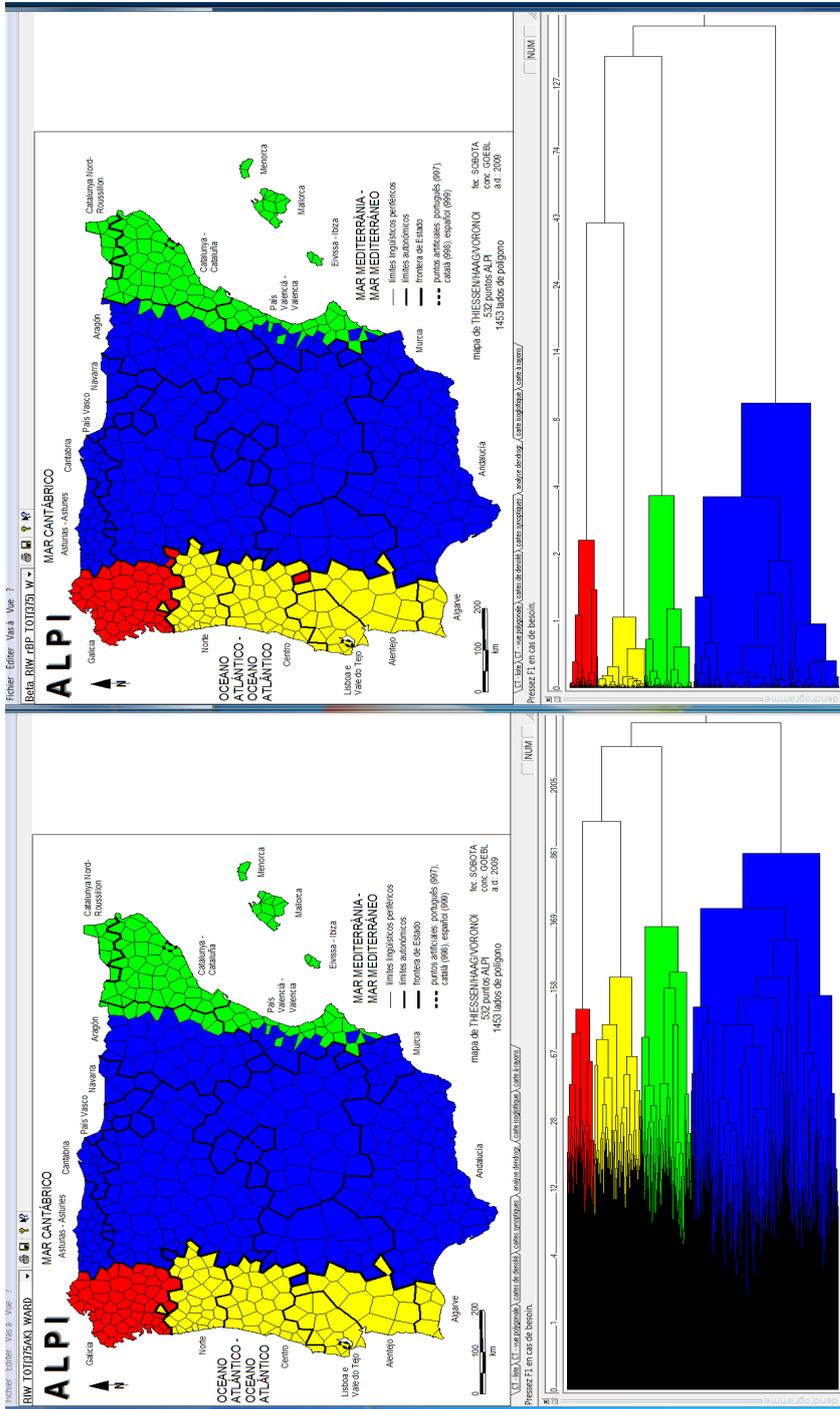


Planche 11. Deux « Classifications Ascendantes Hiérarchiques » (CAH) selon la méthode de Joe Ward Jr.

Corpus : 375 CT; Nombre des dendrèmes-chorèmes coloriés : 4

Calcul de similarité : à gauche (« alpha »): IRI_{jk} ; à droite (« bêta »): $IRI_{jk} + r$ (BP)

