

Quarante-quatre regards ‘ensemblistes’ sur les données de l’ALF. Une étude dialectométrique ‘à la salzbourgeoise’

RÉSUMÉ : Il s’agit de la présentation du nouveau module ‘aréologique’ du logiciel dialectométrique VDM, qui depuis 1997-2000 est utilisé par les adhérents de l’École Dialectométrique de Salzbourg’. Ce nouveau module exécute, en recourant à deux théorèmes centraux de la ‘Théorie des Ensembles’, la synthèse cartographique, finement circonstanciée, d’un grand nombre d’aires taxatoires’ de taille variable. Les cartes ‘à densité’ de type ‘ensembliste’ qui en résultent, jettent non seulement un nouveau jour sur l’‘enchevêtrement particulier’ des mots, phonèmes et morphèmes qui occupent les cartes originales de l’ALF, mais ouvrent aussi de nouvelles perspectives diachroniques sur la genèse des quatre grands domaines linguistiques de la Galloromania (i. e. des domaines d’Oïl, d’Oc, du Croissant et du francoprovençal).

MOTS CLÉS : dialectométrie, aréologie, enchevêtrement particulier, aires taxatoires, cartes à densité, cartes ensemblistes, ALF (« Atlas linguistique de la France »), Oïl, Oc, Croissant, francoprovençal.

1. Remarques préliminaires

L’adjectif *ensembliste* – d’usage courant chez les mathématiciens – est dérivé du terme « Théorie des *Ensembles* »¹ dont deux théorèmes élémentaires (*inclusion* et *réunion*) seront utilisés par la suite, alors que l’allusion à *Salzbourg* comme domicile des deux auteurs sert à orienter les idées des lecteurs vers les méthodes dialectométriques (DM) que nous avons développées depuis la fin des années 1970². Les deux dernières décennies ont en effet vu naître plusieurs autres méthodes-DM pour l’analyse quantitative des données d’un atlas linguistique qui – malgré la référence commune à la *dialectométrie* – n’ont rien à voir ou tant s’en faut, avec les méthodes-DM pratiquées par l’École dialectométrique de Salzbourg (EDMS)³.

2. La mouvance gilliéronienne de l’EDMS

Les relations entre l’EDMS et la géographie linguistique, telle que Jules Gilliéron (1854-1926) l’a présentée (et pratiquée) dans son étude classique sur

¹ Voir → *Théorie des Ensembles* sur « Wikipédia français » (dernière consultation : 11-5-2022).

² En dernière analyse, il faudrait associer les noms de *deux* villes universitaires à la genèse de la DM : celui de *Regensburg / Ratisbonne* (où H. Goebel a travaillé entre 1973 et 1981), et celui de *Salzburg / Salzbourg* (où il a été titulaire d’une chaire de linguistique romane entre 1982 et 2012).

³ Il s’agit des « Écoles dialectométriques » de Groningen (cf. Heeringa 2004) et d’Ulm-Augsburg (cf. Pickl 2013).

«Scier dans la Gaule romane du sud et de l'est» (publiée en 1905), résident surtout dans la préparation géo-classificatrice des données brutes tirées de l'ALF. Jamais J. Gilliéron n'aurait pu présenter ses idées novatrices sur le conflit entre les types lexicaux *SERRARE-scier*, *SERARE-fermer* et *RESECARE-couper* sans le recours aux cinq cartes interprétatives multicolores qui accompagnent son texte.

La confection de ces cartes reposait sur deux piliers dont le premier est de type *empirico-pratique* (par la typisation voire simplification des données originales fort disparates de l'ALF, et la définition successive d'un grand nombre d'aires linguistiques et leur visualisation moyennant le remplissage multicolore de «cartes muettes» [CM])⁴, et l'autre de nature *théorique* (par l'assomption de la dynamique spatio-temporelle de ces aires, toutes pourvues de tailles et de configurations très changeantes).

Quand, de 1997 à 2000, nous avons entrepris le codage-DM d'un nombre aussi grand que possible de cartes linguistiques de l'ALF, nous avons dû parcourir, pour la saisie informatique de leur immense variabilité, les mêmes étapes que J. Gilliéron, bien qu'en utilisant une autre terminologie: voir à ce sujet la Figure 1 ci-contre.

La première section qui y figure (A) et qui s'étend de la «Source originale» (= ALF) [1] à la «Matrice de données» [2] embrasse exactement les mêmes démarches typologiques et cartographiques que J. Gilliéron a dû effectuer en 1905 (ou peu avant), évidemment abstraction faite de la stricte orientation statistique et des nécessités computationnelles de notre propre travail.

C'est grâce à ces deux atouts 'modernes' que nous avons pu conférer une nouvelle facette à notre lecture-DM de l'ALF, à savoir celle de l'optique des 'données de masse' qui, elles, sont caractérisées par la polarité *ontologique* intrinsèque entre leurs *surfaces* et leurs *profondeurs* (sous-jacentes). À ce propos, il ne faut pas oublier que ces dernières sont régies par un grand nombre de régularités, voire de lois intrinsèques, qui ne sont perceptibles qu'au prix d'un traitement 'compresseur' des données brutes. Dans cet article, il ne sera question que du contenu de la 'matrice de données' telle qu'elle figure sur la Figure 1 (étape 2), et aucunement des autres maillons de la chaîne-DM (surtout avec les sections B et C).

⁴ Cf. Goebel 2018. L'outil heuristique de la CM est une particularité des études romanes. Il est entièrement inconnu chez les germanistes et anglicistes qui, de ce fait, ne peuvent pas profiter des nombreux avantages pratiques et théoriques qu'offre le recours systématique aux CM: cf. Goebel 2002b, passim.

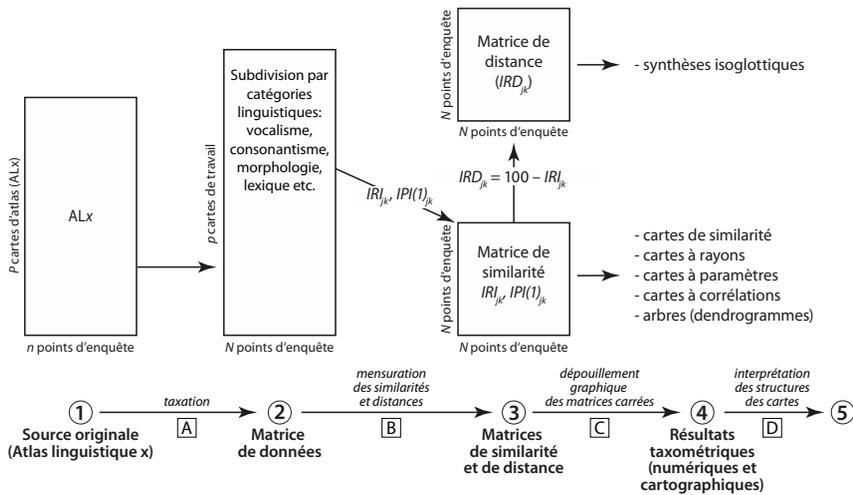


Figure 1: Alignement des démarches méthodiques de l'« École Dialectométrique de Salzbourg » (version alpha).

Encore une remarque d'inspiration cartographique : les cinq cartes illustratives, que Gilliéron avait ajoutées à son étude de 1905 sur l'onomasologie de « scier » dans le Sud et l'Est de la France, ont été réalisées à l'aide de CM fournies par l'imprimerie Protat & Frères à Mâcon auxquelles il a appliqué la méthode des 'cartes en aires'⁵ (cf. Brunet 1987, 71sq.). Durant toute sa vie, il est resté fidèle à cette technique tout en refusant, comme beaucoup de ses successeurs de langue française⁶, d'appliquer aussi la technique des 'cartes en courbes ou isolignes'⁷ (cf. *ibid.*, 76sq.), devenue d'usage courant surtout dans

⁵ De nos jours, ce type de carte est appelé communément 'carte choroplèthe' (all. *Choroplethenkarte*, angl. *choropleth map*).

⁶ Les rares exceptions sont deux membres de l'École lyonnaise (Pierre Gardette, 1906-1973 et Simone Escoffier, 1911-1990) de même qu'Henri Guiter (1909-1994) : cf. à titre d'exemple Gardette 1983 et Guiter 1973. Il est curieux de voir que même dans la grande exégèse synthétique de l'ALF, publiée en 2005 par G. Brun-Trigaud, Y. Le Berre et J. Le Dù, les 'cartes en courbes' manquent totalement.

⁷ La carte 'en aires' ou 'choroplèthe' détermine des zones – soit en couleurs soit en noir et blanc – qui se caractérisent par un ou plusieurs traits donnés et qui s'opposent par là aux autres aires de la même carte ; la carte 'en courbes' ou 'isoglottique' comporte des faisceaux d'isolignes (dont chacune correspond à une opposition entre deux traits linguistiques donnés) qui structurent le territoire. Leur valeur heuristique est très différente.

les pays de langue allemande (cf. Ettmayer 1924) et aussi, depuis Rosenqvist (1919), en Scandinavie.

Il suffit de jeter un rapide coup d'œil à l'excellente introduction à l'ALF de K. Jaberg (publiée en 1908), pourvue tant de cartes en *courbes* (synthèses isoglottiques) que de cartes en *aires* (cartes choroplèthes), pour non seulement constater cette différence, mais aussi pour se convaincre des avantages de l'application combinée de ces deux méthodes. Soulignons également l'impact quantitatif que Jaberg avait conféré aux quatre cartes en courbes de son livre, par la superposition systématique des tracés de plusieurs linéaments isoglottiques. Ce faisant, il était nettement en avance sur son ancien maître de l'École Pratique des Hautes Études de Paris⁸.

Dans cet article, nous utilisons la technique des 'cartes en plages' (Brunet 1987, 73sq.) qui – aux dires de cet auteur – «est une variété de carte en aires, mais qui offre une particularité supplémentaire et très intéressante: elle est *graduée*, et montre ainsi *les variations d'intensité d'un phénomène de surface*. Elle contient une hiérarchie, exprimée par l'échelle des *plages* de teintes adoptée».

3. Bref historique de la 'taxation' de 626 cartes originales, extraites des volumes 1-7 de l'ALF⁹

Ce travail, sans nul doute de type 'bénédictin', a été effectué entre 1997 et 2000 par un groupe de collaboratrices dévouées, tant estudiantines que diplômées¹⁰, et s'est déroulé parallèlement au développement informatique du logiciel VDM («Visual DialectoMetry»)¹¹ par les soins de notre ami Edgar Haimerl. C'est ainsi qu'il a été possible de tester la fonctionnalité de VDM en état de gestation à l'aide d'un stock de matériaux de plus en plus amples.

Notre but était alors, évidemment dans le cadre d'inévitables restrictions personnelles, logistiques et économiques, d'engranger un nombre aussi grand que possible de 'cartes de travail' (CT) embrassant toutes les catégories linguistiques. En fin de compte, nous avons pu tirer, des 626 cartes originales de

⁸ Karl Jaberg a suivi les cours de Jules Gilliéron à l'École Pratique des Hautes Études à Paris en 1900-1901 (cf. Pop/Pop 1959, 55 et 59).

⁹ Voir aussi les informations données dans nos contributions de 2000, 2002(a) et 2003.

¹⁰ En voici les 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.

¹¹ Voir Goebel 2004a.

l'ALF analysées (ou 'taxées') par l'équipe mentionnée ci-dessus,¹² 1 681 CT dont les structures internes, définies par le nombre des 'taxats' respectifs¹³, varient entre 2 et 90 taxats¹⁴ par CT. Voir la Figure 2 qui montre la courbe, fort éloquent, de la dépendance entre le taux de la polynymie taxatoire interne des CT et leur nombre. Il en résulte que le nombre des CT diminue au fur et à mesure que leur polynymie interne augmente. Cette régularité est telle qu'elle a induit le linguiste quantitatif allemand Gabriel Altmann (1931-2020) à en tirer une formule, baptisée par lui «loi de Goebel» (cf. Altmann 1983).

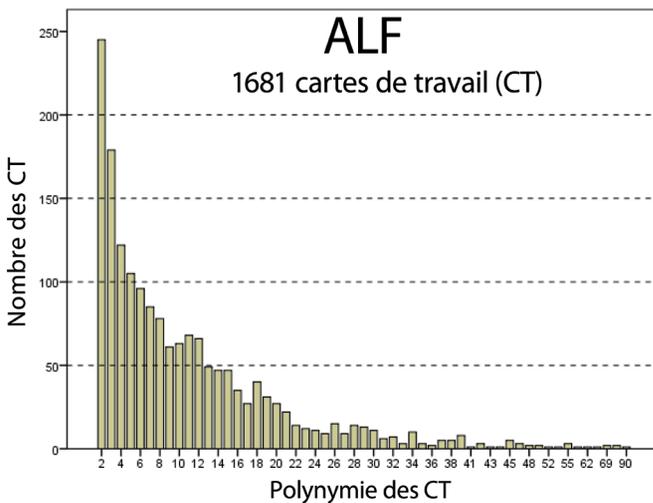


Figure 2: Relation entre la polynymie taxatoire des cartes de travail (CT) (→ axe x, en abscisse) et leur nombre absolu (→ axe y, en ordonnée), qui règne dans notre corpus-DM intégral tiré de l'ALF.

¹² Comme la charpente méthodologique de la DM a été construite sur le modèle d'une discipline biologique appelée (en anglais) «Numerical *Taxonomy*», nous l'avons pourvue de plusieurs termes basés sur le radical du verbe grec *táttein / tássein* («ordonner, arranger, agencer» etc.).

¹³ Dans son étude de 1905, Gilliéron utilise, pour la désignation de ces entités abstraites, le mot *type*: «Ces cartes doivent être coloriées comme celle des géologues: abstraction faite des différences secondaires, les vocables patois se groupent par *types*, et ces types distincts couvrent des aires distinctes qu'il convient de désigner aux yeux par une couleur.» (Gilliéron 1905, 3). K. Jaberg, par contre, appelle, dans son livre de 1908, ces «vocables patois» tout bonnement *Wörter* (mots); voir aussi la note 15.

¹⁴ Soulignons qu'à chaque «taxat» correspond toujours une «aire taxatoire» (AT) particulière.

Les 1 681 CT sont dépositaires de 19 328 taxats auxquels correspondent autant d'«aires taxatoires» (AT) dont chacune dispose d'une taille et d'une configuration spatiale très spécifique, voire unique.

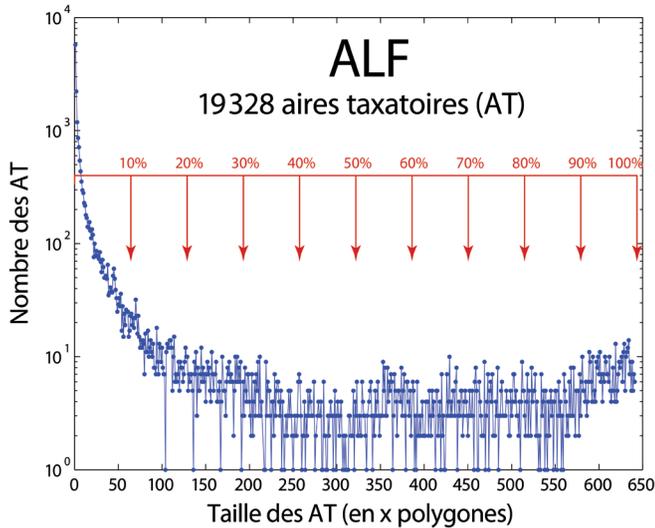


Figure 3 : Relation entre la taille des aires taxatoires (AT) (→ axe x, en abscisse) et leur nombre absolu (→ axe y, en ordonnée; avec échelle logarithmique), qui règne dans notre corpus-DM intégral tiré de l'ALF.

Sur l'ordonnée: score minimal: 1 (= 10^0), score maximal: 5 743 (= $10^{3,759}$).

La règle rouge superposée renvoie à la gradation (en % de $N = 641$) de la taille des AT.

Précisons que la taille de ces 19 328 AT varie entre 1 et 640 polygones ou points-ALF)¹⁵. Voir à ce sujet la Figure 3 dont le message global rejoint de très près celui de la Figure 2: alors qu'il y a beaucoup d'AT de dimensions réduites voire exigües, leur nombre diminue au fur et à mesure que leur taille s'approche du seuil maximal de 640.

Soulignons que c'est justement l'immense variabilité des AT – et de ce fait aussi des linéaments qui les entourent (= isoglosses) – qui n'a cessé, depuis l'avènement des atlas linguistiques, d'inquiéter les esprits des

¹⁵ Une AT couvrant l'étendue intégrale d'une CT et partant pourvue – dans notre analyse de l'ALF – de 641 polygones, ne comporterait aucune variation géolinguistique: elle est donc inutile, voire négligeable en matière de DM.

(géo)linguistes¹⁶. C'est que les AT – même celles de la même orientation linguistique – ne coïncident pratiquement jamais tout en infirmant l'espérance théorique des linguistes, imbus de la prétendue régularité des 'lois phonétiques', de voir se fondre les isoglosses en de solides faisceaux toujours compacts et bien ordonnés¹⁷. La vérité était, par contre, que les mêmes AT et leurs frontières formaient, à l'intérieur d'une matrice de données, un fouillis inextricable (→ « enchevêtrement particulier »)¹⁸ d'où sont nés, il y a plus d'un siècle, deux hypothèses de (très) mauvais augure :

- de la *non-existence des dialectes* (cf. p. ex. Paris 1888) et de
- l'*impossibilité de la classification* « scientifique » *des dialectes* (cf. p. ex. Schuchardt 1870/1900).

Avec l'avènement de la DM il s'est avéré que ces deux hypothèses étaient carrément fausses, et que malgré l'omniprésence de l'« enchevêtrement particulier » des AT (ou des taxats tout court), il est possible d'opérer, dans cette masse de données mouvantes, des classifications rationnelles (et scientifiquement très utiles) de toute sorte.

Soulignons, en guise de conclusion de ce paragraphe, que les calculs numériques de l'EDMS reposent sur deux unités de base qui, elles, constituent le contenu de toutes nos matrices de données (MD)¹⁹ :

- sur les CT qui forment les lignes de la MD, et
- sur les AT qui en sont les composantes : voir la Figure 1.

¹⁶ À l'exception de l'abbé Théobald Lalanne (1880-1952) à qui revient le grand mérite d'avoir découvert (et décrit dans les moindres détails) le principe de l'indépendance (et de l'enchevêtrement complexe) des aires linguistiques : cf. Lalanne 1949-1952 (réimpression : 2018).

Des innombrables passages relatifs à l'enchevêtrement prodigieux des aires linguistiques, disséminés dans sa riche production scientifique, nous ne citons, à titre d'exemple, que la suivante : « Ce chevauchement, ou mieux ce grouillement d'aires est loin d'être aussi simple qu'un cloisonnement de sous-dialectes bien parqués et bien stables, mais le réel est atrocement complexe et il est souvent imprudent de le qualifier. » (Lalanne 1949-1952 : publication originale : 1952, 100 ; réimpression : 2018, 139).

¹⁷ Voir à ce sujet le passage suivant (entièrement fondé) tiré de Jaberg 1908, 6 : « Das Lautgesetz, das da sagt, daß ein gewisser (lateinischer) Laut unter gewissen Bedingungen an einem bestimmten Orte sich gleichbleibt oder sich zu einem bestimmten neuen Laut entwickelt, ist eine Abstraktion. In Wirklichkeit hat jedes Wort seine besondere Geschichte. Gebiete gleicher lautlicher Entwicklung bei verschiedenen Wörtern decken sich auch in sprachlich widerstandsfähigen Gegenden nur annähernd. »

¹⁸ Nous utilisons ce concept depuis quelque vingt ans (cf. Goebl 2016, 117).

¹⁹ Voir à ce sujet aussi notre contribution anglaise de 2016.

4. L'application des méthodes de l'EDMS à différents corpus de CT et d'AT

Une étude globale de la variabilité des cartes de travail et des aires taxatoires implique, bien entendu, de les réunir en plusieurs (sous-)corpus et de les analyser selon différents critères :

- selon leur *nombre* (par l'extraction aléatoire d'un nombre préalablement défini de cartes de travail du corpus total) [ceci vaut seulement pour les CT],
- selon leur *appartenance catégorielle* (*vocalisme, consonantisme, lexique* etc.) [ceci vaut tant pour les CT que pour les AT]²⁰,
- selon leur *polynymie interne* [ceci vaut seulement pour les CT]²¹, et aussi
- selon leur *taille* [ceci vaut seulement pour les AT].

En ce qui concerne la variation aléatoire du nombre des CT analysées, nous avons trouvé, de concert avec les expériences de beaucoup d'autres sciences empiriques, que déjà avec le tiers ou même le quart des CT à disposition, l'on obtient les mêmes résultats classificatoires qu'avec le total des CT disponibles²². Les choses changent avec la considération comparée de corpus-CT différenciés selon leur appartenance *catégorielle*: là, il s'avère qu'il y a des différences notables entre les résultats d'une analyse-DM de corpus *phonétiques, lexicaux* ou *morpho-syntaxiques*²³. Il est d'ailleurs non moins vrai que ces différences donnent lieu à des explications et interprétations linguistiques très intéressantes et pertinentes.

Quant à la polynymie interne des CT, il s'est révélé qu'entre l'analyse-DM de sous-corpus munis uniquement de CT *oligo-nymiques*²⁴ d'un côté, et de CT *poly-nymiques* de l'autre, il n'y a pratiquement pas de différences²⁵. C'est ainsi qu'est née la supposition que les données des atlas linguistiques soient sillonnées de *structures de profondeur* très robustes, et que l'enchevêtrement

²⁰ Voir les exemples présentés dans nos contributions de 1984, 2000, 2002(a), 2003 et 2009.

²¹ Voir nos contributions de 2014 (pour l'ALF) et 2020 (pour l'AIS).

²² Cf. Goebel (1984 I, 197sqq.).

²³ L'appartenance catégorielle des 1 681 CT de notre corpus-DM relatif à l'ALF se présente comme suit: *phonétique* en général: 1 117 CT (dont 477 pour le *vocalisme* et 479 pour le *consonantisme*) et *lexique*: 471 CT.

²⁴ Ce sont des CT pourvues de très peu d'AT (< gr. *olígos* "petit, court, peu nombreux").

²⁵ Cf. Goebel 2014 (pour l'ALF) et Goebel / Smečka 2020 (pour l'AIS).

particulier' des AT, jadis considéré comme une tare regrettable de la géographie linguistique en tant que telle, semble être, malgré son apparence perturbatrice, un des facteurs constitutifs de cette admirable cohérence qui ressort autant des cartes (de travail) très *simples* que des cartes (de travail) *complexes* et comportant plusieurs dizaines de types. Cette supposition veut en plus que lesdites structures de profondeur soient l'œuvre d'une action anthropique particulière que nous appelons l'« aménagement basilectal de l'espace par l'Homo Loquens »²⁶, et que cette dernière soit régie, comme tant d'autres manifestations humaines, par des lois et régularités intrinsèques.

5. Le défi classificatoire de la superposition systématique d'AT de taille variable

Les expériences décrites dans le paragraphe précédent nous ont amenés à demander à notre ami Edgar Haimerl d'ajouter, à l'architecture originale de son logiciel VDM, une nouvelle composante, capable de traiter les AT (stockées dans n'importe quelle de nos matrices de données-DM), soit *isolément* soit de façon *globale*. C'est ainsi qu'est né, entre 2009 et 2011, le module « aréologique » de VDM dont sont issues les recherches présentées dans cet article. Ce module permet de pratiquer – toujours dans le sillage géo-classificatoire de J. Gilliéron, mais avec l'impact de l'informatique moderne (et aussi de la Théorie des Ensembles) – une stratigraphie aréologique extensive et d'utiliser ainsi la vertu spatio-temporelle des AT pour de nouvelles introspections diachroniques, souvent inattendues, voire même déroutantes.

Le principe 'ensembliste' de base consiste dans la sélection et la combinaison successive d'AT qui doivent répondre à deux conditions :

- qu'elles contiennent un ou plusieurs points d'atlas à définir au préalable (théorème de *l'inclusion*),
- que leurs tailles, définies en fractions (pourcentages) du réseau d'atlas en question (ici : $N = 641$), varient à l'intérieur de marges ('fourchettes') préalablement établies.

Le résultat de cette double opération ensembliste – sélection (par *inclusion*) et combinaison (par *réunion*) – est une carte choroplèthe multicolore qui, de par son profil 'paysagiste' tridimensionnel, fournit des informations très intéressantes à partir des points de vue de la géographie linguistique et de l'histoire de la Galloromania. Précisons en outre que ces cartes combinatoires

²⁶ Nous utilisons ce concept depuis 2005.

– appelées aussi «cartes à densité» – ont été réalisées, du point de vue cartographique, selon les principes visualisateurs bien établis de l'EDMS. Ceci vaut tant pour la définition (algorithmique) des paliers numériques que pour leur coloriage à l'aide de teintes empruntées au spectre solaire.

L'algorithme utilisé en la matière s'appelle MINMWMAX: il constitue la procédure visualisatrice standard de l'EMDS²⁷. Sa propriété quantitative consiste à former, des deux côtés de la moyenne arithmétique tirée de tous les scores numériques à visualiser²⁸, un nombre égal de classes (ou: d'intervalles) dont les deux séries de largeurs sont, respectivement, identiques.

6. Des CT à la réunion combinatoire de plusieurs AT

Voir les Planches 1 et 2.

La Planche 1 présente quatre CT qui servent à illustrer, à titre d'exemple, leur caractère *cartographique* et *classificatoire*. Elles ont été extraites au petit hasard de notre corpus-DM de l'ALF et utilisent le réseau polygonal de l'ALF (avec 641 polygones ou points d'enquête)²⁹ tel qu'il est incorporé dans le logiciel VDM. Les légendes multicolores renvoient aux différents *taxats* et leurs *aires* (AT) tout en indiquant, entre crochets, l'étendue spatiale de ces dernières. Les numéros (suivis de deux points) qui accompagnent les rectangles multicolores servent uniquement à l'identification des AT. La description linguistique des AT – repérable entre les numéros d'ordre et les crochets – suit de très près les transcriptions phonétiques répertoriées sur les cartes originales de l'ALF.

Toutes les quatre cartes de la Planche 1 sont polynomes: leur bigarrure taxatoire interne varie entre 9 (centre-gauche) et 17 (centre-droite). La carte de gauche se réfère aux dénominations galloromanes de la *brebis*: elle est donc d'ordre *lexical*. Les trois autres cartes sont toutes d'ordre *phonétique*; elles montrent la variation diatopique de trois nexus cruciaux de l'étymon latin MERCĀTU (à savoir les succédanés de -E- prétonique, de -Ā- accentué et de -C- prétonique) qui est à la base de toutes les formes répertoriées sur la carte 812 de l'ALF (*le marché*). Les couleurs *rose* et *bleu clair* renvoient, res-

²⁷ Cf. Goebel 1981, 365 s. (en français) et 1984, I, 93 ss. (en allemand).

²⁸ Ici, leur nombre peut varier – théoriquement – entre 1 et 641.

²⁹ Nous avons ajouté, aux 638 points d'atlas originaux de l'ALF, trois points d'atlas factices qui correspondent aux langues standard du français (P. 999), de l'italien (P. 998) et du catalan (P. 997). Rappelons que le réseau de l'ALF comprend 638 *points d'enquête* et 639 *enquêtes*. Ce petit décalage s'explique par le fait qu'Edmond Edmont a fait deux *enquêtes* au P.-ALF 284, Saint-Pol-sur-Ternoise, son village natal.

pectivement, à des AT 'nordistes' et 'sudistes'. Les autres teintes n'ont aucune connotation géographique.

De ces quatre CT il ressort avec une grande netteté :

- que la variabilité de la taille et de la configuration des AT est vraiment étonnante ;
- que les AT pourvues d'une superficie à la fois *grande* et *cohérente*, sont très clairement *minoritaires* ;
- que l'*émiettement* partiel ou total des AT est la règle.

Passons maintenant à la considération de la Planche 2 dont le but central est de montrer la logique, ensembliste et cartographique, et le résultat de la combinaison (par *réunion*) de trois AT de taille très inégale. Voici les critères de leur sélection opérée par le logiciel VDM :

- toutes les trois AT contiennent le polygone du P.-ALF 226 (critère de l'*inclusion*) ;
- la taille de l'AT de *gauche* est comprise dans une fourchette entre 1 et 64 points de l'ALF (= 1-10 % du réseau total de 641) ;
- la taille de l'AT du *centre-gauche* se situe dans une fourchette entre 256 et 320 points de l'ALF (= 40-50 %) ;
- la taille de l'AT du *centre-droite* se trouve à l'intérieur d'une fourchette allant de 577 à 641 points de l'ALF (= 90-100 %).

La carte de droite montre la réunion pure et simple de ces trois AT. C'est une carte «à densité» classique. Il s'agit là d'une méthode située à cheval entre l'*arithmétique* (par le cumul additif de la valeur 1) et la *cartographie* multicolore (à travers la visualisation échelonnée par le recours aux teintes graduées *vert*, *bleu clair* et *bleu foncé*). En dernière analyse, la carte de droite ressemble, de par ses trois dimensions cartésiennes (*longueur* [x], *largeur* [y] et *hauteur* [z]), à une petite montagne, dont le sommet principal – marqué en bleu foncé – se trouve en Île-de-France.

La nature linguistique de ces trois AT n'a en cela aucune importance. Leur sélection ne répond qu'à des fins démonstratives et s'est faite uniquement en fonction de leur taille géographique. Ceci n'empêche que ces trois AT font partie intégrante des AT cumulées sur les Planches 5 (fourchette : 0-10 %), 9 (fourchette : 40-50 %) et 14 (fourchette : 90-100 %), toujours en vue de la confection des cartes de gauche (toutes relatives au P.-ALF 226).

7. Construction et interprétation des cartes à AT combinées (cartes ‘à densité variable’)

La mise en carte combinée (par *réunion*) de plusieurs AT présuppose deux choix délibérés:

- d’un ou de plusieurs points d’atlas – tous inclus dans les AT en question – qui, de ce fait, agissent comme «pivot(s)» d’un corps géométrique imaginaire,
- de la taille des AT qui répondent à ce choix. C’est pourquoi le nouveau module ‘aréologique’ du logiciel VDM permet de régler avec précision la taille des AT superposables en fonction de la grandeur du réseau d’atlas de base (ici: 641 P.-ALF).

La combinaison de plusieurs AT groupées autour d’un noyau (ou ‘pivot’) géographique précis repose sur l’assomption qu’une éventuelle géo-typicité initiale de ce noyau puisse être renforcée moyennant la superposition d’AT supplémentaires de tailles et de silhouettes différentes. Une expérience géo-combinatoire de ce genre devrait donc:

- commencer par la réunion (ou: superposition) combinatoire de *toutes* les AT contenant le ‘pivot’ (ou ‘noyau’) en question (voir les Planches 3 et 4), et
- se poursuivre par la réunion *échelonnée* d’AT – contenant toujours le même ‘pivot d’épreuve’ – judicieusement sélectionnées en fonction de leur taille (voir les Planches 5-14).

7.1. Choix des quatre «pivots» d’inclusion

Nous avons retenu, pour une première démonstration sommaire de cette nouvelle méthode ensembliste, quatre points-ALF isolés, relatifs à autant de «zones-clés» de la Galloromania:

- au domaine d’Oïl: le P.-ALF 226 (Le Plessis-Piquet, Hauts-de-Seine);
- au domaine du Croissant: le P.-ALF 704 (Saint-Quentin, Creuse);
- au domaine d’Oc: le P.-ALF 744 (Valdériès, Tarn);
- au domaine francoprovençal: le P.-ALF 956 (Sixt, Haute-Savoie).

Soulignons que l’inclusion ensembliste de points-ALF *isolés* ne constitue qu’une première étape et qu’elle peut être approfondie par l’inclusion de *plusieurs* points-ALF contigus, voire même d’entières *pages* plus ou moins

cohérentes, dans les AT sujettes au tri ensembliste. Les expériences que nous avons déjà entreprises dans cette direction ont d'ailleurs fourni des résultats très prometteurs.

7.2. *Superposition ensembliste (par réunion) du total des AT disponibles*

Voir les Planches 3 (avec quatre profils de similarité) et 4 (avec quatre superpositions ensemblistes).

Précisons d'entrée de jeu que les deux 'tétraptyques' en question se ressemblent beaucoup à cause de la similarité intrinsèque des logiques mathématiques de leur construction: en effet, aux quatre 'points de référence' de la Planche 3 (avec autant de profils de similarité)³⁰ correspondent les 'points-pivots d'inclusion' de la Planche 4 (avec quatre combinaisons ensemblistes)³¹. Les profils choroplèthes des huit volets ('cartes') des deux tétraptyques sont très clairs. Précisons que les géo-types qui y apparaissent pourraient être générés aussi à partir d'autres 'points de repère' ou de 'pivots' situés dans leur voisinage.

On peut d'ailleurs prélever le nombre des AT combinées tant sur les cartes mêmes (voir la mention «Max» placée tout en bas de la légende numérique) que dans la légende verbale correspondante. Toujours est-il que de l'apparent chaos de quelque 1 670 AT³² superposées par voie ensembliste se dégagent des résultats globaux très ordonnés. La même chose vaut pour les quatre profils de similarité, tous basés sur les 1 681 CT de la matrice de données analysée, et créés par l'application de l'indice de similarité standard de l'EDMS, IRI_{jk} . Nous verrons par la suite que cet agencement bien ordonné dépend de

³⁰ Pour la mesure de la similarité et la confection des cartes y ayant trait, cf. Goebel (1981, 357sq.) [en français] et (1984, I, 74sq.) [en allemand].

³¹ Pour éviter des malentendus: les cartes 'ensemblistes' (comme celles de la Planche 4) sont le produit de quatre additions arithmétiques très banales (dont le principe a été démontré sur la Planche 2) qui produisent l'empilement pur et simple des superficies d'un grand nombre (entre 1665 et 1679) d'aires taxatoires (AT) de tailles et configurations très changeantes; les profils choroplèthes de ces cartes ensemblistes reposent donc sur $N = 641$ polygones coloriés. – Le calcul des cartes 'de similarité' (comme celles de la Planche 3) est nettement plus compliqué; il rend nécessaire une mensuration de similarité (portant sur $N = 641$ points d'atlas) qui, à l'aide d'un index de similarité préalablement déterminé (ici: IRI, mais il peut y en avoir d'autres) produit une matrice de similarité carrée (avec les dimensions extérieures N^2); des N vecteurs de cette matrice carrée l'on peut tirer autant (ici: $N = 641$) de cartes de similarité dont chacune repose sur $N-1$ scores visualisés et un polygone (ou point) laissé en blanc.

³² Pour les nombres exacts des AT superposées voir les quatre légendes numériques de la Planche 3, de gauche à droite: 1679, 1674, 1676 et 1665.

nombreux facteurs et qu'il est possible de déterminer avec précision l'influence de chacun d'entre eux.

7.3. *Présentation et interprétation de quarante superpositions ensemblistes (par réunion) d'AT de taille changeante*

Voir les Planches 5-14.

Notre dispositif expérimental prévoit la définition de dix paliers échelonnés quant à la taille des AT à superposer : le premier des dix paliers va de 0 à 10% (de la grandeur du réseau intégral de 641 PP.-ALF), le deuxième de 10 à 20% [etc.], et les deux derniers de 80 à 90% et de 90 à 100%. On verra qu'il en résulte, pour chacun des quatre points-pivots, des effets géo-typiques très éloquentes ; ceux-ci s'annoncent, dans les deux premières fourchettes (0-10% et 10-20%), par l'éclosion d'agglomérations macro-régionales déjà bien circonscrites, qui s'affermissent par la suite (fourchettes 20-30% et 30-40%) et finissent par s'estomper inévitablement à cause d'un 'brouillage blanc parasitaire'³³ global du signal géolinguistique. En règle générale, cela se produit à partir des fourchettes d'épreuve de 60-70% ou de 70-80%.

La succession de ces trois étapes (*éclosion – affermissement – brouillage*) a pu être observée dans toutes nos expériences ensemblistes. La leçon qui en dérive est que la formation de noyaux dialectaux bien typés ('géo-types') s'opère exclusivement à l'aide d'AT de taille petite ou moyenne : voir à ce propos, sur la Figure 3, la règle graduée en rouge.

7.3.1. *Superposition ensembliste (par réunion) d'AT situées entre 0 et 10% (de 641)*

Voir la Planche 5.

Pour les points-pivots 226 (Oïl ; à gauche), 744 (Oc ; au centre-droite) et 956 (francoprovençal ; à droite) surgissent des géo-types bien définis et aussi prévisibles. Mais ce qui frappe beaucoup, c'est l'orientation carrément *occitane* du géo-type *crescentain*³⁴ (au centre-gauche) gravitant autour du point-pivot 704. Cette orientation 'sudiste' vaut pour l'ensemble des points-pivots du Crois-

³³ Voir → *bruit numérique, bruit blanc* et *bruit additif gaussien* sur « Wikipédia français » (dernière consultation : 11-5-2022). (En allemand les mêmes phénomènes sont appelés *weißes Rauschen* : leur nom anglais est *white noise*.)

³⁴ Nous reprenons ici un néologisme – étymologiquement bien fondé – que nous avons utilisé déjà en 2002 (cf. Goebel 2002a, 22).

sant³⁵ jusqu'à la fourchette de 30-40% pour basculer, à partir de la fourchette 40-50%, rapidement dans une orientation très nettement *oïlique*.

Comme il est légitime d'associer une plus grande ancienneté linguistique aux AT de dimensions réduites, il semble bien que le Croissant puisse être défini comme un domaine linguistique *diachroniquement* hybride : avec d'anciennes *racines* d'Oc et un *feuillage* d'Oïl³⁶ de date plutôt récente.

Attention au nombre inégal des AT superposées : à gauche (→ P. 226) : 38 AT ; centre-gauche (→ P. 704) : 253 AT ; centre-droite (→ P. 744) : 205 AT ; à droite (→ P. 956) : 265 AT³⁷. Ce déséquilibre concerne la plupart des dix fourchettes d'épreuve.

7.3.2. *Superposition ensembliste (par réunion) d'AT situées entre 10 et 20% (de 641)*

Voir la Planche 6.

Les effets de compactage de la planche précédente se sont renforcés. Quant à la position 'taxatoire' du P. 956, l'étalement de provignements jaunes (dans la classe 4) le long de la marge nord-orientale de la carte est remarquable. Cet effet 'oïlisant' sera contrebalancé sur la planche suivante (n° 7) par une nette orientation 'sudiste'.

7.3.3. *Superposition ensembliste (par réunion) d'AT situées entre 20 et 30% (de 641)*

Voir la Planche 7.

Pas de changements pour les points-pivots 226, 704 et 744. Pour le point-pivot francoprovençal (956), une réorientation s'annonce vers le sud (domaine d'Oc).

7.3.4. *Superposition ensembliste (par réunion) d'AT situées entre 30 et 40% (de 641)*

Voir la Planche 8.

Trois des quatre points-pivots examinés (= 704, 744 et 956) disposent d'une orientation taxatoire carrément sudiste, dont deux (pour les PP. 704 et 744) sont très solides.

³⁵ Voir à ce sujet les deux cartes multicolores de notre contribution de 2004b.

³⁶ La bibliographie relative au Croissant est très vaste : nous ne citons à titre d'exemple que les contributions de G. Brun-Trigaud de 1990 et 2020, la nôtre de 2004(b) et le volume-recueil de date récente (2021), publié par L. Esher, M. Guérin, N. Quint et M. Russo.

³⁷ Ces chiffres apparaissent en bas des légendes numériques : voir la mention «Max».

7.3.5. *Superposition ensembliste (par réunion) d'AT situées entre 40 et 50% (de 641)*

Voir la Planche 9.

À partir de cette fourchette, l'orientation ensembliste précédente des points *crescentain* (704) et *francoprovençal* (956) vers le sud change complètement : tous les deux basculent vers le domaine d'Oïl. Il en résulte que les assises historiques du Croissant et du domaine francoprovençal sont bicéphales. Ce constat s'aligne grosso modo sur ce que nous savons déjà de la formation du francoprovençal³⁸, mais il apporte également des éléments importants aux discussions menées autour du Croissant³⁹.

7.3.6. *Superposition ensembliste (par réunion) d'AT situées entre 50 et 60% (de 641)*

Voir la Planche 10.

La nouvelle orientation oilique des profils taxatoires des points-pivots 704 (Croissant) et 956 (domaine francoprovençal) s'affermirait. À remarquer, pour le profil taxatoire du P. 744 (Oc; centre-droite), l'implantation de plusieurs polygones jaunes (dans l'intervalle 4) en plein domaine d'Oïl : c'est un signe avant-coureur de l'effondrement successif du géo-type occitan sous le coup du 'brouillage blanc parasite'.

7.3.7. *Superposition ensembliste (par réunion) d'AT situées entre 60 et 70% (de 641)*

Voir la Planche 11.

L'effondrement du géo-type occitan (relatif au P. 744) s'intensifie alors que les géo-types des trois autres points-pivots résistent encore.

7.3.8. *Superposition ensembliste (par réunion) d'AT situées entre 70 et 80% (de 641)*

Voir la Planche 12.

L'effet du 'brouillage blanc parasite' déstabilise aussi les géo-types taxatoires des points-pivots 226 (à gauche), 704 (au centre-gauche) et 956 (à droite) : à remarquer, dans les trois cas, que l'insertion de la Provence dans l'intervalle 3 revient à une 'neutralisation' (ou 'banalisation') géo-spatiale de cette région.

³⁸ Cf. Kristol 2016.

³⁹ Cf. Guérin 2020, § 2.

7.3.9. *Superposition ensembliste (par réunion) d'AT situées entre 80 et 90% (de 641)*

Voir la Planche 13.

Toutes les quatre cartes choroplèthes de cette planche ont été pleinement investies par le « brouillage blanc parasitaire » géo-spatial.

7.3.10. *Superposition ensembliste (par réunion) d'AT situées entre 90 et 100% (de 641)*

Voir la Planche 14.

Ce qui frappe au premier coup d'œil, c'est la quasi-identité des quatre profils choroplèthes. Il en résulte la disparition complète de toute typicité géo-spatiale. Signalons aussi que les effectifs totaux des AT superposées [456 (à gauche), 439 (au centre-gauche), 443 (au centre-droite) et 438 (à droite)] sont devenus très similaires.

8. Épilogue

La superposition ensembliste d'AT de tailles changeantes constitue un enrichissement très utile de la gamme méthodologique de l'EDMS et du logiciel VDM. Sa valeur réside non seulement dans sa vertu purement *géolinguistique* et partant *synchronique*, mais aussi dans sa grande utilité *diachronique*. Soulignons en outre que tant ces avantages spatio-temporels que les avatars de la triade géo-classificatoire (*éclosion – affermissement – brouillage*) mentionnée ci-dessus (cf. 7.3.) se sont manifestés non seulement dans nos analyses ensemblistes des données de l'ALF, mais aussi dans celles réalisées avec les données-DM tirées de l'AIS (cf. Goebel 2008), de l'ALDC (cf. Goebel 2013a) et de l'ALPI (cf. Goebel 2013b) : ce phénomène semble donc être un *universel linguistique*.

Il est hors de doute que le travail qu'on vient de lire est de type *global* et *quantitatif*. Pour pouvoir apprécier toute la portée scientifique de ses résultats, il faudrait compléter cette analyse *quantitative* par une étude philologique et linguistique détaillée de type *qualitatif*, qui devrait porter sur toutes les AT que nous avons réunies en synthèse dans les quarante-quatre profils choroplèthes dressés par voie ensembliste. Or, une telle étude *qualitative* – qui requerrait évidemment l'introspection et l'analyse circonstanciée d'un nombre inouï de silhouettes très variées d'AT⁴⁰ – est tout autre que chimérique. C'est

⁴⁰ Dans le moindre des cas, il s'agirait de plusieurs centaines d'AT. Avec VDM, il est possible d'examiner simultanément tant les aires taxatoires (AT) particulières que les CT d'où ces AT ont été extraites.

que le module aréologique du logiciel VDM sauvegarde, pour chaque requête ensembliste effectuée, les AT sélectionnées dans des fichiers particuliers dont la consultation visuelle sur l'écran est aisée pour qui dispose du logiciel VDM et aussi des données de l'ALF (ou d'autres atlas linguistiques) qui y sont répertoriées.

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.

Université de Salzbourg

Hans GOEBL
Pavel SMEČKA

9. Remerciements

Module aréologique du logiciel VDM: Edgar Haimerl, Ulm / Seattle.

Confection des Figures 1-3: Werner Goebel, Vienne, et Dumitru Kihai, Zurich.

10. Références bibliographiques et abréviations

10.1. Abréviations

AT = aire taxatoire

CM = carte muette

CT = carte de travail

DM = dialectométrie, dialectométrique

EDMS = École dialectométrique de Salzbourg

MD = matrice de données

IRI_{jk} = Indice Relatif d'Identité entre les points d'atlas j et k

P.-ALF = point(s) de l'ALF

VDM = Visual DialectoMetry (logiciel dialectométrique créé dans les années 1997-2000 par Edgar Haimerl pour les besoins de l'EDMS).

10.2. Répertoires

AIS = Jaberg, Karl / Jud, Jakob, 1928-1940. *Sprach- und Sachatlas Italiens und der Südschweiz*, 8 vols., Zofingen, Ringier (réimpression: Nendeln, Kraus, 1971) [consultable en ligne : <<https://navigais-web.pd.istc.cnr.it/?map=1401>>].

- ALDC = Veny, Joan / Pons i Griera, Lúdia, 2001-2018. *Atlas Lingüístic del Domini Català*, 9 vols., Barcelona, Institut d'Estudis Catalans [consultable en ligne : <<https://aldc.espais.iec.cat/>>].
- ALF = Gilliéron, Jules / Edmont, Edmond, 1902-1910. *Atlas linguistique de la France*, 10 vols., Paris, Champion (réimpression : Bologne, Forni, 1968) [consultable en ligne : 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/#/>>].
- ALPI = *Atlas lingüístico de la Península Ibérica*, 1962. Vol. I: *Fonética*, Madrid, CSIC [consultable en ligne : <<http://www.alpi.csic.es/>>].

10.3. Travaux

- Altmann, Gabriel, 1983. «Die Entstehung diatopischer Varianten. Ein stochastisches Modell», *Zeitschrift für Sprachwissenschaft* 4, 139-155.
- Brun-Trigaud, Guylaine, 1990. *Le croissant : le concept et le mot. Contribution à l'histoire de la dialectologie française au XIX^e siècle*, Lyon, Centre d'Études Linguistiques Jacques Goudet.
- Brun-Trigaud, Guylaine, 2020. *Les parlers de la Creuse. Frontière et carrefour*, Guéret, Société des sciences naturelles, archéologiques et historiques de la Creuse.
- Brun-Trigaud, Guylaine / Le Berre, Yves / Le Dù, Jean (2005). *Lectures de l'Atlas linguistique de la France de Gilliéron et Edmont. Du temps dans l'espace. Essai d'interprétation des cartes de l'Atlas linguistique de la France de Jules Gilliéron et Edmond Edmont augmenté de quelques cartes de l'Atlas linguistique de la Basse Bretagne de Pierre Le Roux*, Paris, CTHS.
- Brunet, Roger, 1987. *La carte, mode d'emploi*, Paris, Fayard/Reclus.
- Esher, Louise / Guérin, Maximilien / Quint, Nicolas / Russo, Michela (éd.), 2021. *Le Croissant linguistique : entre oc, oil et francoprovençal. Des mots à la grammaire, des parlers aux aires*, Paris, L'Harmattan.
- Ettmayer, Karl von, 1924. *Über das Wesen der Dialektbildung, erläutert an den Dialekten Frankreichs*, Vienne, Académie des Sciences de Vienne, Mémoires de la classe philologique et historique, vol. 66,3.
- Gardette, Pierre, 1983. *Études de géographie linguistique*, publiées par les soins de Brigitte Horiot / Marie-Rose Simoni-Aurembou et Georges Straka, Strasbourg/Paris, Société de Linguistique Romane/Klincksieck.
- Gilliéron, 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).
- Goebel, Hans, 1981. «Éléments d'analyse dialectométrique (avec application à l' AIS)», *RLiR* 45, 349-420.
- Goebel, Hans, 1984. *Dialektometrische Studien. Anhand italo-romanischer, rätoromanischer und galloromanischer Sprachmaterialien aus AIS und ALF*, 3 vol., Tübingen, Niemeyer.

- Goebel, Hans, 2000. «La dialectométrisation de l'ALF: présentation des premiers résultats», *Linguistica* 40, 209-236 (avec 12 cartes en couleurs).
- Goebel, Hans, 2002a. «Analyse dialectométrique des structures de profondeur de l'ALF», *RLiR* 66, 5-63. (avec 24 cartes en couleurs)
- Goebel, Hans, 2002b. «Sprachatlant: woher? womit? wozu? Einige buntgemischte Reflexionen am Gartenzaun zwischen Romanistik und Germanistik», in: Anreiter, Peter / Ernst, Peter / Hausner, Isolde / Kalb, Helmut (éd.), *Namen, Sprachen und Kulturen. Imena, Jeziki in Kulture. Festschrift für Heinz Dieter Pohl zum 60. Geburtstag*, Wien, Präsens, 257-274.
- Goebel, Hans, 2003. «Regards dialectométriques sur les données de l'Atlas linguistique de la France (ALF): relations quantitatives et structures de profondeur», *Estudis Romànics* 25, 59-121 (avec 24 cartes en couleurs).
- Goebel, Hans, 2004a. «VDM-Visual Dialectometry. Vorstellung eines dialektometrischen Software-Pakets auf CD-ROM (mit Beispielen zu ALF und Dees 1980)», in: Dahmen, Wolfgang et al. (éd.), *Romanistik und neue Medien* (Romanistisches Kolloquium XVI), Tübingen, Narr, 209-241.
- Goebel, Hans, 2004b. «Das "Croissant" – eine Nachschau im Abstand von 90 Jahren (mit zwei dialektometrisch erstellten Farbkarten)», in: Noll, Volker / Thiele, Sylvia (éd.), *Sprachkontakte in der Romania. Zum 75. Geburtstag von Gustav Ineichen*, Tübingen, Niemeyer, 159-172.
- Goebel, Hans, 2005. «La dialectométrie corrélative. Un nouvel outil pour l'étude de l'aménagement dialectal de l'espace par l'homme», *RLiR* 69, 321-367 (24 cartes en couleurs: 356-367).
- Goebel, Hans, 2008. «La dialettometrizzazione integrale dell' AIS. Presentazione dei primi risultati», *RLiR* 72, 25-113 (avec 48 cartes en couleurs).
- Goebel, Hans, 2009. «Quelques coups d'œil dialectométriques sur l'Atlas linguistique de la France: structures de surface et structures de profondeur», in: Dalbera-Stefanaggi, Marie-José / Simoni-Aurembou, Marie-Rose (éd.), *Images de la langue: représentations spatiales, sémantiques et graphiques*, Paris, Éditions du CTHS, 39-60.
- Goebel, Hans, 2013a. «La dialectometrització dels quatre primers volums de l'ALDC», *Estudis Romànics* 35, 87-116.
- Goebel, Hans, 2013b. «La dialectometrización del ALPI: rápida presentación de los resultados», in: Casanova Herrero, Emili / Calvo Rigual, Cesáreo (éd.), *Actas del XXVI Congreso Internacional de Lingüística y de Filología Románicas* (Valencia 2010), Berlin, Boston, de Gruyter, vol. 6, 143-154.
- Goebel, Hans, 2014. «L'impact de la polynymie des cartes d'atlas sur le résultat de calculs dialectométriques», in: AA.VV., *Linguistique romane et Linguistique indo-européenne. Mélanges offerts à Witold Mańczak à l'occasion de son 90^e anniversaire*, Kraków/Cracovie, Polska Akademia Umiejętności, Instytut Filologii Romanskiej Uniwersytetu Jagiellońskiego, 243-260 (avec dix cartes en couleurs).
- Goebel, Hans, 2018. «La face cachée de la géographie linguistique. Bref aperçu sur les «cartes muettes» produites pour l'ALF, l' AIS et le FEW», *RLiR* 82, 5-63 (avec 16 cartes en couleurs).

- Goebl, Hans / Smečka, Pavel, 2016. «The Quantitative Nature of Working Maps (WM) and Taxatorial Areas (TA): A Brief Look at two Basic Units of Salzburg Dialectometry (S-DM)», in: Kelih, Emmerich / Knight, Róisín / Mačutek, Ján / Wilson, Andrew (éd.), *Issues in Quantitative Linguistics 4. Dedicated to Reinhard Köhler on the occasion of his 65th birthday* (Studies in Quantitative Linguistics, vol. 23), Lüdenscheid, RAM-Verlag, 113-127 (avec quatre cartes en couleurs).
- Goebl, Hans / Smečka, Pavel, 2020. «Il ruolo della polinimia nelle ricerche dialettometriche di tipo salisburghese. Con esempi tratti dall' AIS », in: Iannàcaro, Gabriele / Videsott, Paul / Dell'Aquila, Vittorio / Poppi, Cesare (éd.), *Fassa, Ladinia e oltre. Studi in onore di Fabio Chiocchetti*, Vigo di Fassa / Vich, Istitut Cultural Ladin «Majon di Fascegn» (Mondo ladino 44, 2020), 243-281 (avec huit cartes en couleurs).
- Guérin, Maximilien, 2020. «Transmission et dynamique des parlers du Croissant», *Cahiers du Groupe d'études sur le plurilinguisme européen* (GEPE) 12, <<http://www.ouvroir.fr/cpe/index.php?id=1276>>.
- Güter, Henri, 1973. «Atlas et frontières linguistiques», in: Straka, Georges (éd.), *Les dialectes romans de France à la lumière des atlas régionaux*, Paris, CNRS, 61-109.
- Heeringa, Wilbert, Jan, 2004. *Measuring dialect pronunciation differences using Levenshtein Distance* (Groningen dissertations in linguistics 46), Groningen, Rijksuniversiteit.
- 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).
- Kristol, Andres, 2016. «Francoprovençal», in: Ledgeway, Adam / Maiden, Martin (éd.), *The Oxford Guide to the Romance Languages* (Oxford Guides to the World's Languages), Oxford, University Press, 350-362.
- 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: Paris, Gaston, *Mélanges linguistiques. Latin vulgaire et langues romanes, langue française, notes étymologiques*, publiés par Mario Roques, Paris, Champion, 1909, 432-448).
- Pickl, Simon, 2013. *Probabilistische Geolinguistik. Geostatistische Analysen lexikalischer Variation in Bayerisch-Schwaben* (Beihefte der Zeitschrift für Dialektologie und Linguistik, 154), Stuttgart, Franz Steiner.
- Pop, Sever / Pop, Rodica Doina, 1959. *Jules Gilliéron. Vie, enseignement, élèves, œuvres, souvenirs*, Louvain, Centre international de dialectologie générale.
- Schuchardt, Hugo, 1870 / 1900. *Über die Klassifikation der romanischen Mundarten* [Leçon inaugurale, tenue à Leipzig, le 30 avril 1870], Graz, Styria, 1900 [aussi in: Spitzer, Leo (éd.), *Hugo Schuchardt-Brevier. Ein Vademecum der allgemeinen Sprachwissenschaft*, Halle, Niemeyer, 1928² (réimpression: Darmstadt, Wissenschaftliche Buchgesellschaft, 1976), 166-88].

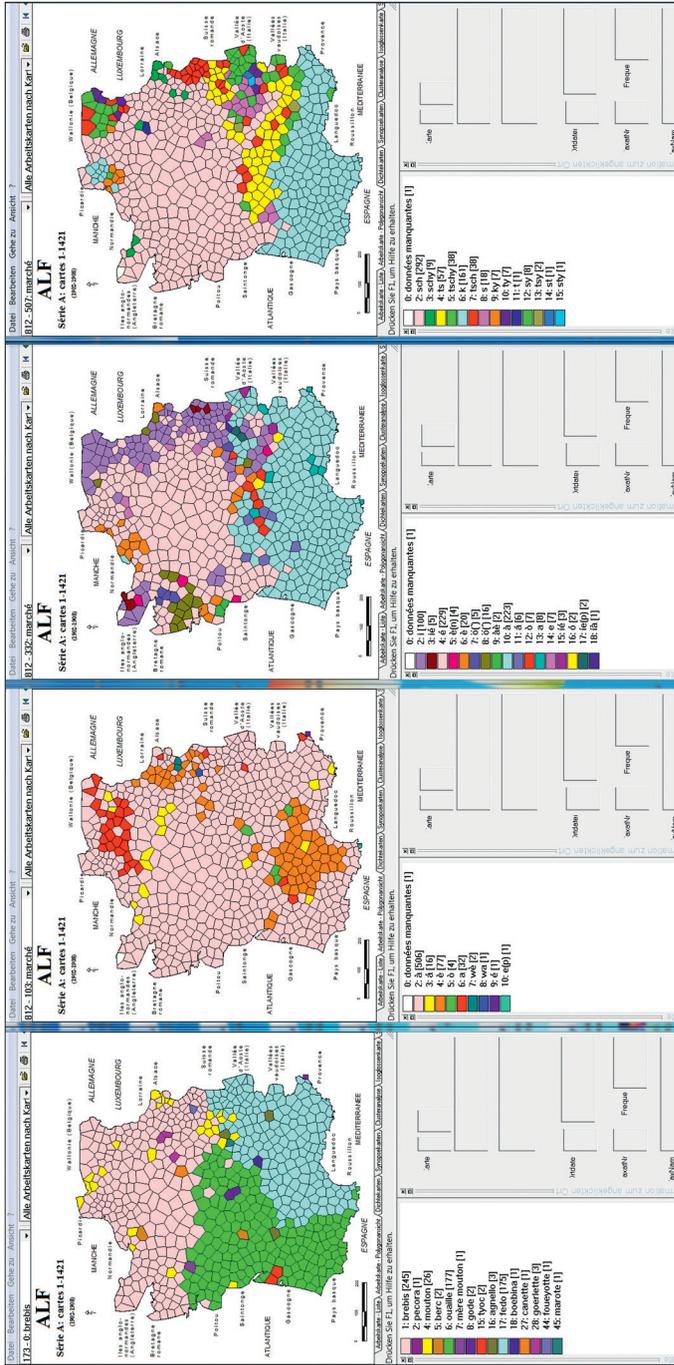


Planche 1 : Synopse de quatre cartes de travail (CT) tirées de l'ALF

- À gauche : ALF 173 *la brebis* (→ variation lexicale)
- Centre-gauche : ALF 812 *le marché* (MERCĀTU → variation phonétique: vocalisme prétonique)
- Centre-droite : ALF 812 *le marché* (MERCĀTU → variation phonétique, vocalisme tonique)
- À droite : ALF 812 *le marché* (MERCĀTU → variation phonétique, consonantisme prétonique)

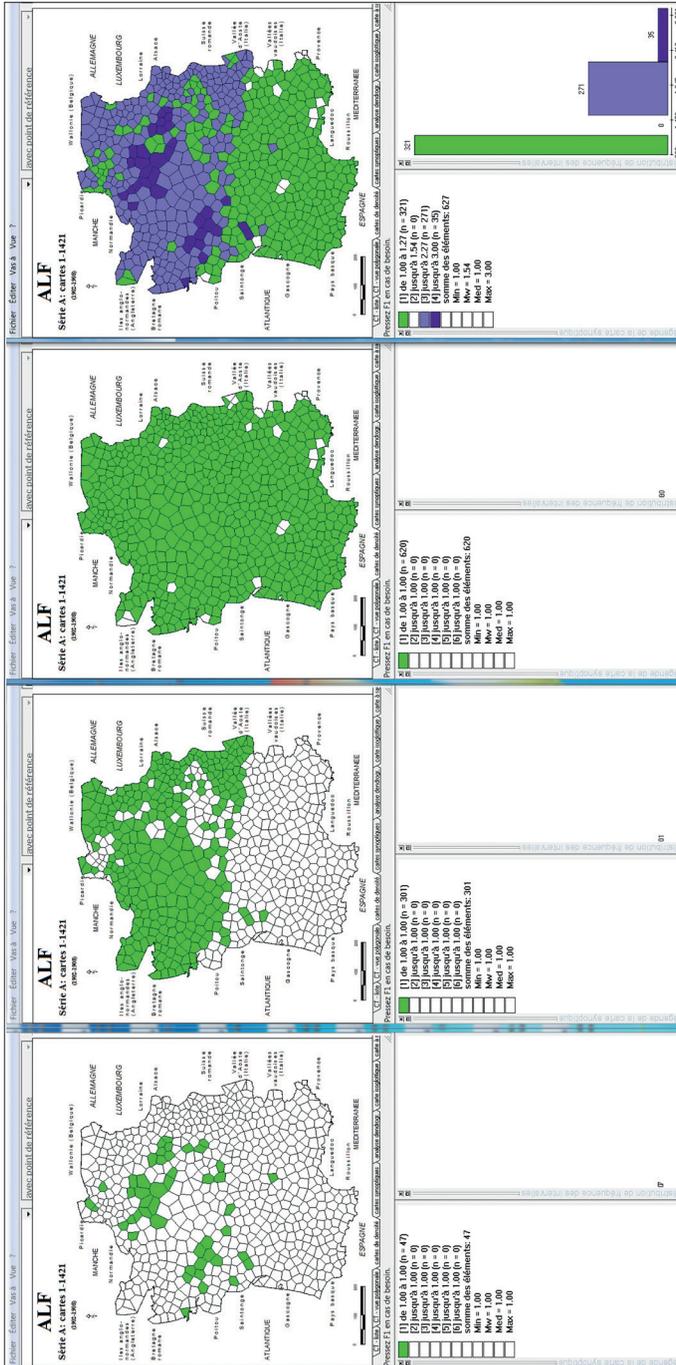


Planche 2: Illustration de la superposition (par *réunion*) iconique de trois aires taxonomiques (AT) de taille inégale

- À gauche : AT aux dimensions très *limitées* (avec 47 P.-ALF = 7,3% de 641); ALF 1132 *râteau*: -ÉLLU > ρ
- Centre-gauche : AT aux dimensions *moyennes* (avec 301 P.-ALF = 46,9% de 641); ALF 25 où *vas-tu*: où (< UBI)
- Centre-droite : AT aux dimensions *grandes* (avec 620 P.-ALF = 96,7% de 641); ALF 46 *Antoine*: -ONIU > n
- À droite : Superposition (par *réunion*) des superficies de ces trois AT

Algorithme de visualisation: MINMWMAX 4-tuple

La couleur *bleu foncé* symbolise l'enchevêtrement spatial le plus dense des trois AT en question.

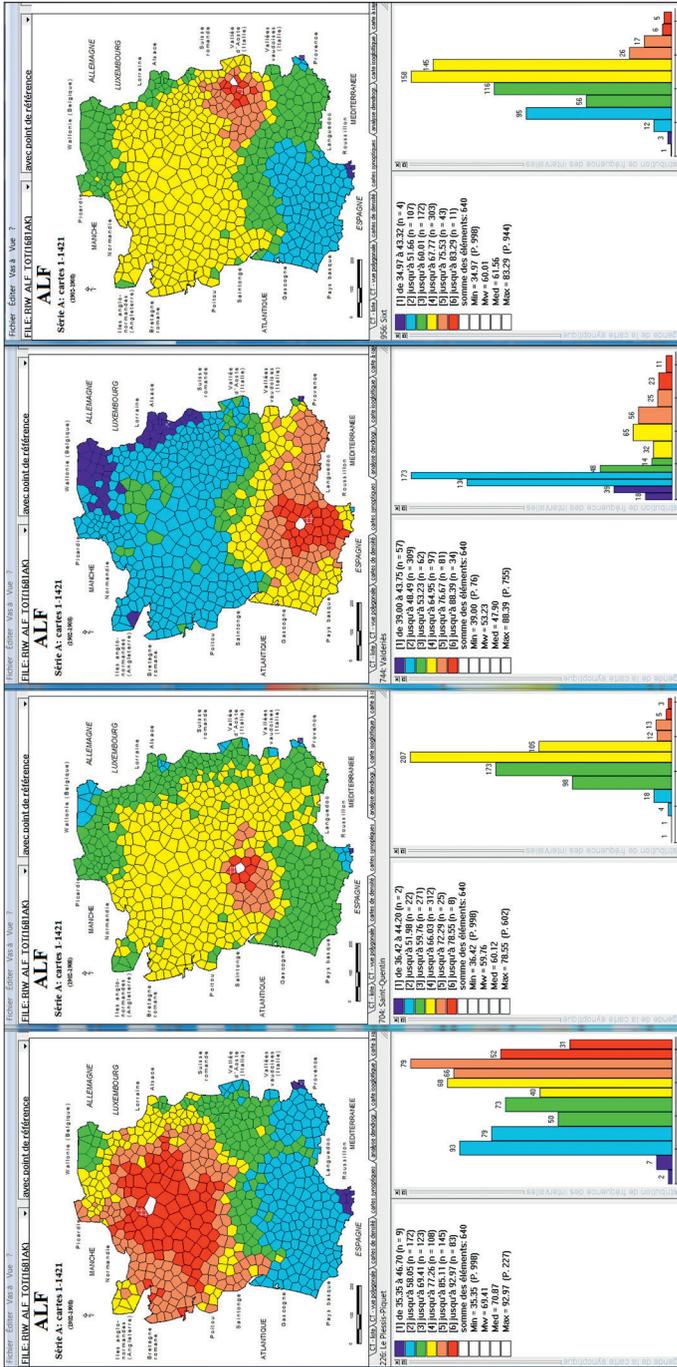


Planche 3 : Synopse de quatre cartes de similarité

Mesure de similarité: IRI_{jk} ; Corpus: 1681 CT (toutes les catégories linguistiques); algorithme de visualisation: MINMWMAX 6-tuple

- À gauche: point de référence: 226, Le Plessis-Piquet, Hauts-de-Seine
- Centre-gauche: point de référence: 704, Saint-Quentin, Creuse
- Centre-droite: point de référence: 744, Valdierès, Tarn
- À droite: point de référence: 956, Sixt, Haute-Savoie

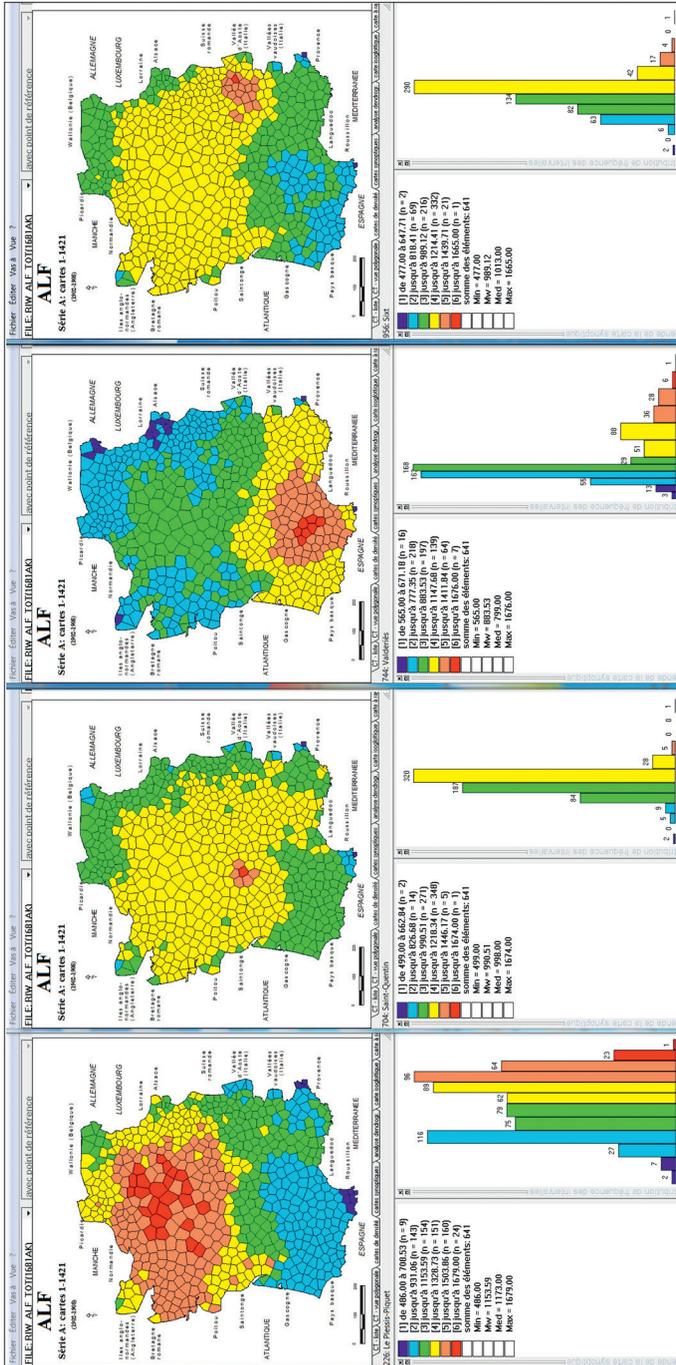


Planche 4: Synopse de quatre superpositions (par réunion) d'AT aux dimensions intégrales (0-100% de 641 P.-ALF)

Corpus: 1681 CT (toutes les catégories linguistiques); algorithme de visualisation: MINMWMAX 6-tuple

• À gauche: point d'inclusion: 226, Le Plessis-Piquet, Hauts-de-Seine; Σ AT empliées: 1679

• Centre-gauche: point d'inclusion: 704, Saint-Quentin, Creuse; Σ AT empliées: 1674

• Centre-droite: point d'inclusion: 744, Valdieriès, Tarn; Σ AT empliées: 1676

• À droite: point d'inclusion: 956, Sixt, Haute-Savoie; Σ AT empliées: 1665

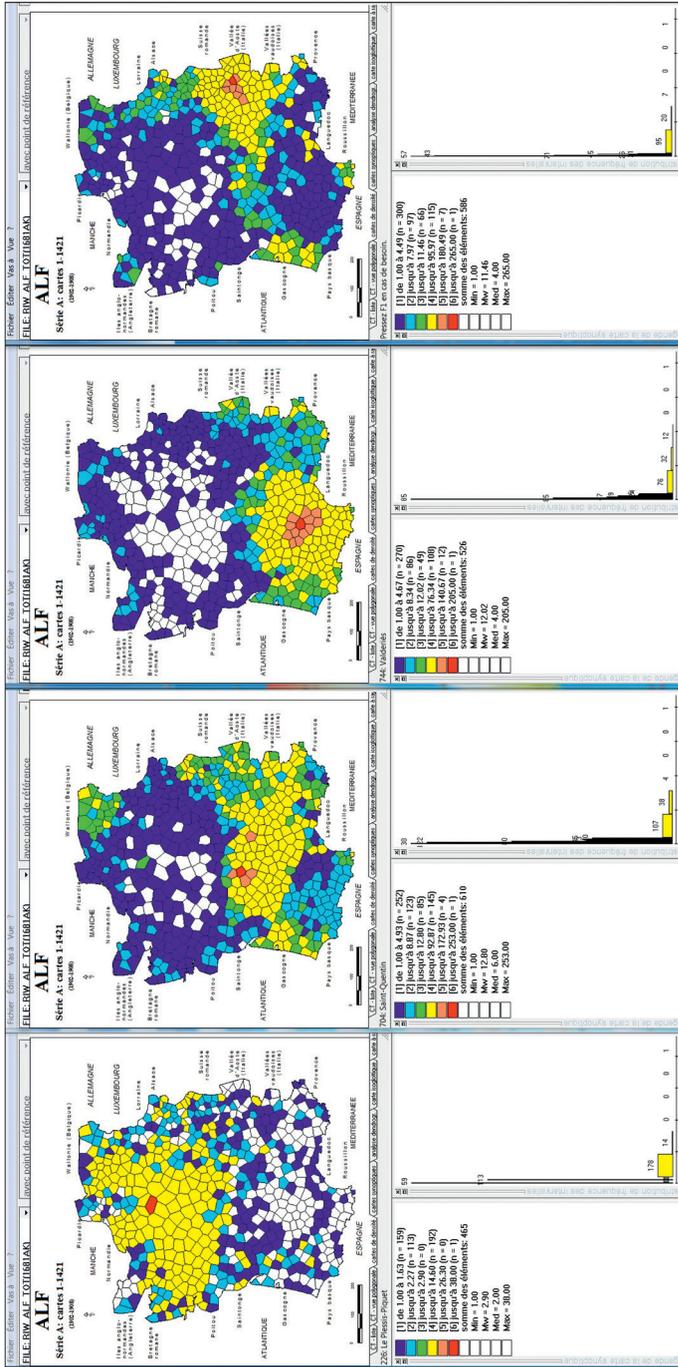


Planche 5 : Synopse de quatre superpositions (par réunion) d'AT aux dimensions limitées (0-10% de 641 P.-ALF)

Corpus : 1681 CT (toutes les catégories linguistiques) ; algorithme de visualisation : MINMWMAX 6-tuple

- À gauche : point d'inclusion : 226, Le Plessis-Piquet, Hauts-de-Seine ; Σ AT empliées : 38
- Centre-gauche : point d'inclusion : 704, Saint-Quentin, Creuse ; Σ AT empliées : 253
- Centre-droite : point d'inclusion : 744, Valderrès, Tarn ; Σ AT empliées : 205
- À droite : point d'inclusion : 956, Sixt, Haute-Savoie ; Σ AT empliées : 265

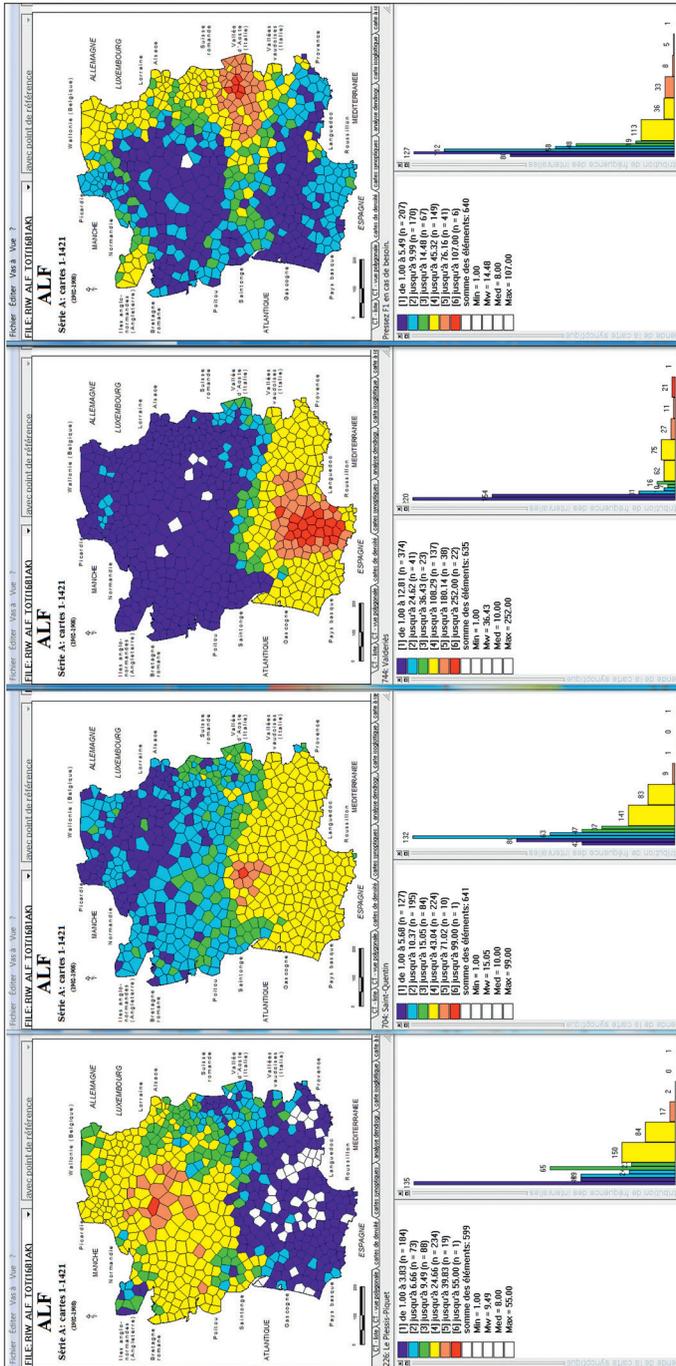


Planche 6 : Synopse de quatre superpositions (par réunion) d'AT aux dimensions limitées (10-20% de 641 P.-ALF)

Corpus : 1681 CT (toutes les catégories linguistiques) ; *algorithme de visualisation* : MINMWMAX 6-tuple

- À gauche : point d'inclusion : 226, Le Plessis-Piquet, Hauts-de-Seine; Σ AT empliées : 55
- Centre-gauche : point d'inclusion : 704, Saint-Quentin, Creuse; Σ AT empliées : 99
- Centre-droite : point d'inclusion : 744, Valdieriès, Tarn; Σ AT empliées : 252
- À droite : point d'inclusion : 956, Sixt, Haute-Savoie; Σ AT empliées : 107

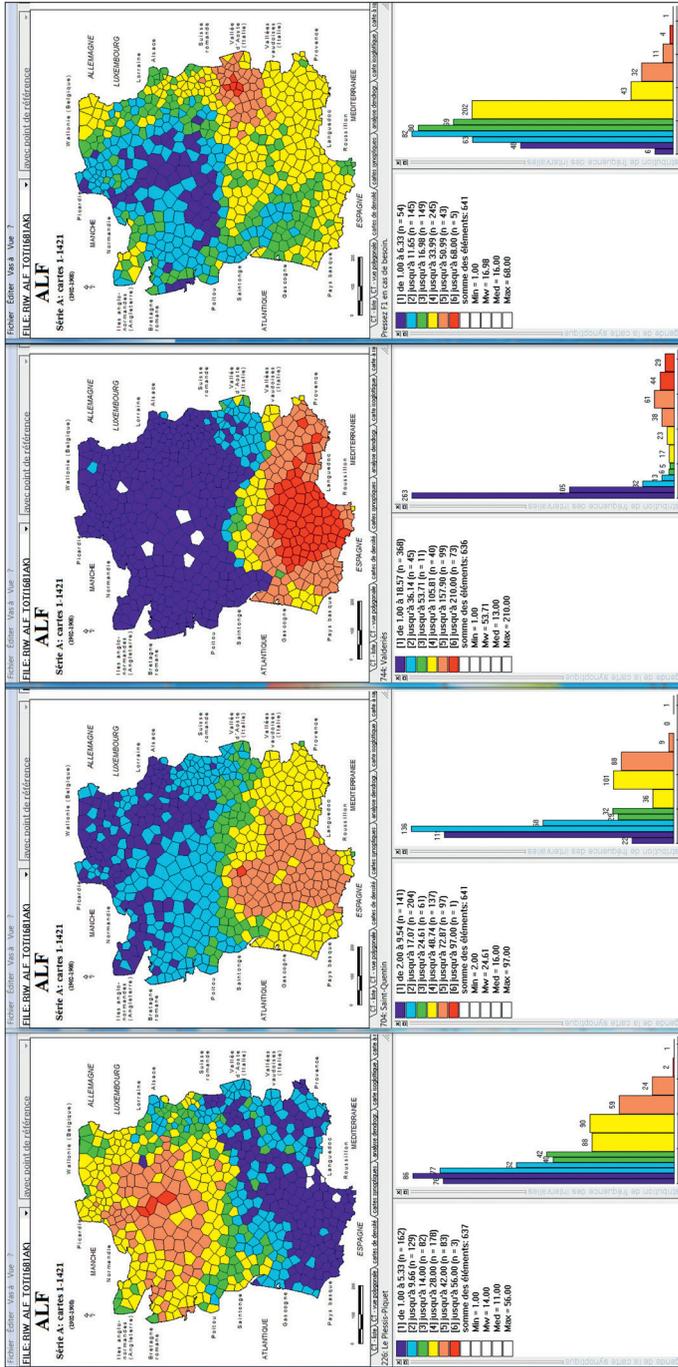


Planche 7: Synopse de quatre superpositions (par réunion) d'AT aux dimensions limitées (20-30% de 641 P.-ALF)

Corpus: 1681 CT (toutes les catégories linguistiques); algorithme de visualisation: MINMWMAX 6-tuple

- À gauche: point d'inclusion: 226, Le Plessis-Piquet, Hauts-de-Seine; Σ AT employées: 56
- Centre-gauche: point d'inclusion: 704, Saint-Quentin, Creuse; Σ AT employées: 97
- Centre-droite: point d'inclusion: 744, Valderêts, Tarn; Σ AT employées: 210
- À droite: point d'inclusion: 956, Sixt, Haute-Savoie; Σ AT employées: 68

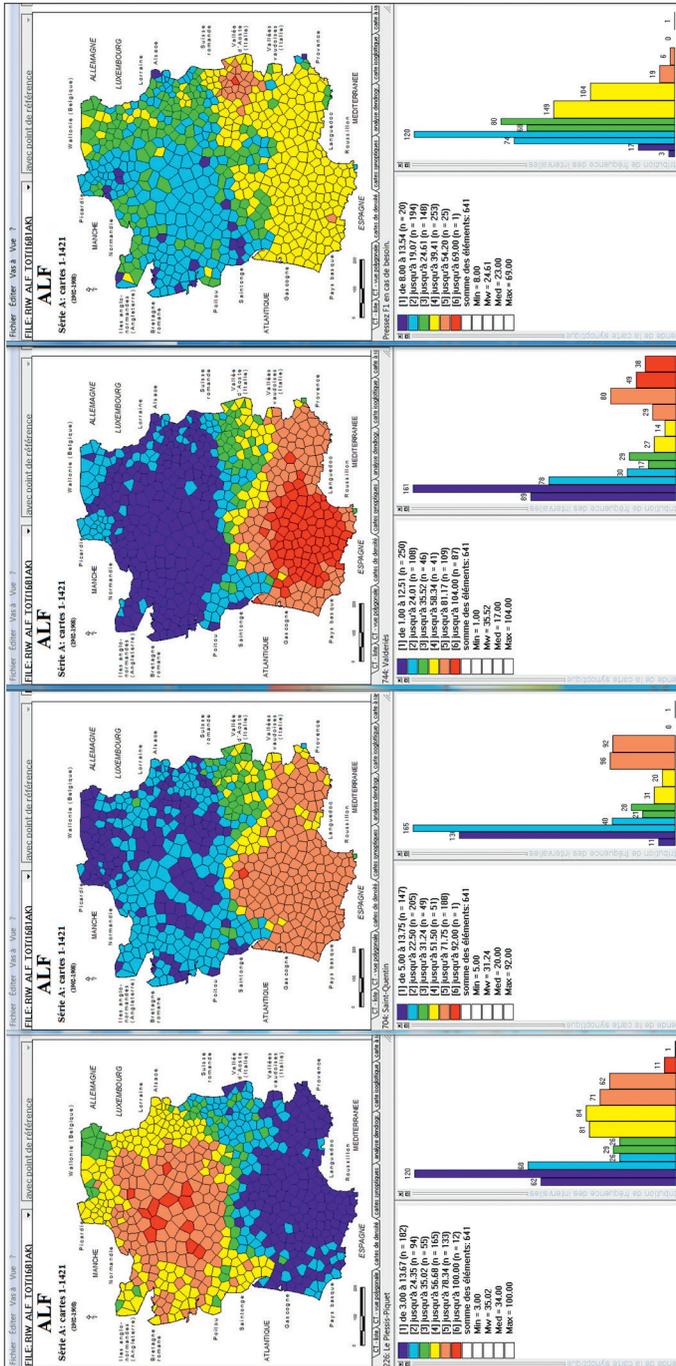


Planche 8: Synopse de quatre superpositions (par réunion) d'AT aux dimensions limitées (30-40% de 641 P.-ALF)

Corpus: 1681 CT (toutes les catégories linguistiques); *algorithme de visualisation*: MINMWMAX 6-tuple

• À gauche: point d'inclusion: 226, Le Plessis-Piquet, Hauts-de-Seine; Σ AT empliées: 100

• Centre-gauche: point d'inclusion: 704, Saint-Quentin, Creuse; Σ AT empliées: 92

• Centre-droite: point d'inclusion: 744, Valderiès, Tarn; Σ AT empliées: 104

• À droite: point d'inclusion: 956, Sixt, Haute-Savoie; Σ AT empliées: 69

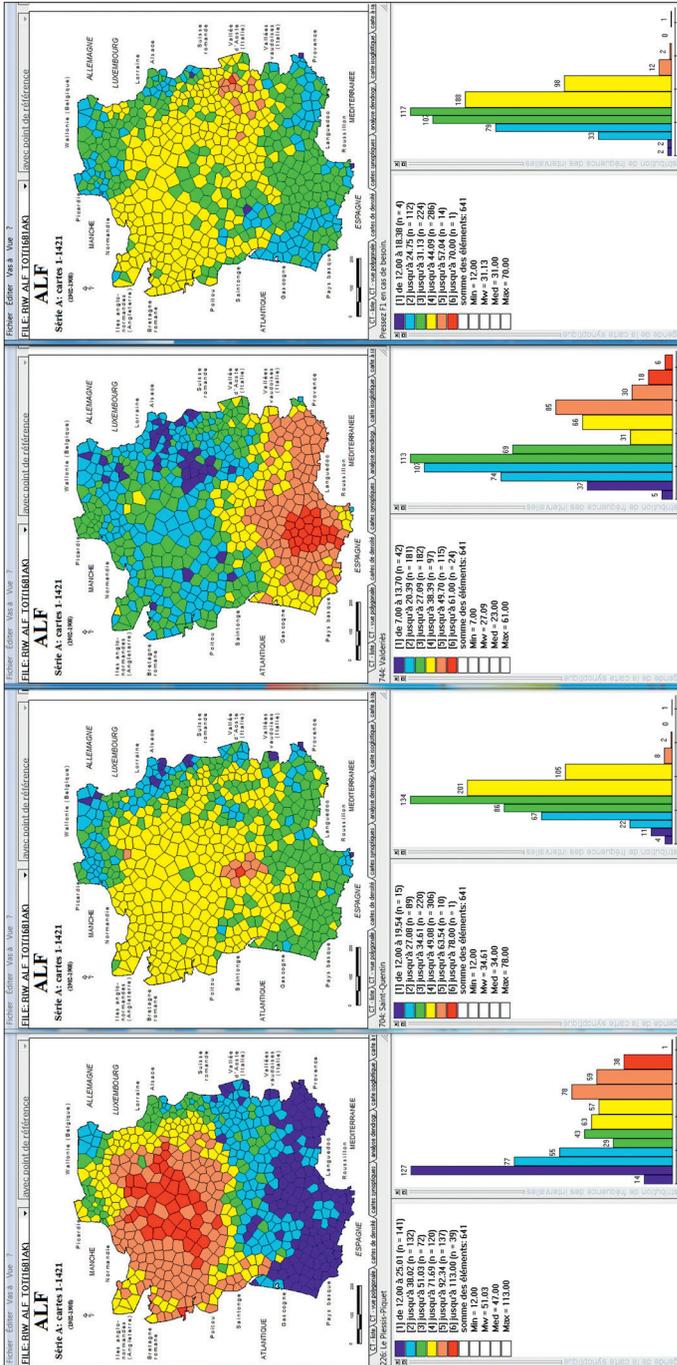


Planche 9 : Synopse de quatre superpositions (par réunion) d'AT aux dimensions limitées (40-50% de 641 P.-ALF)

Corpus : 1681 CT (toutes les catégories linguistiques) ; *algorithme de visualisation* : MINMWMAX 6-tuple

- À gauche : point d'inclusion : 226, Le Plessis-Piquet, Hauts-de-Seine ; Σ AT empliées : 113
- Centre-gauche : point d'inclusion : 704, Saint-Quentin, Creuse ; Σ AT empliées : 78
- Centre-droite : point d'inclusion : 744, Valenciennes, Tarn ; Σ AT empliées : 61
- À droite : point d'inclusion : 956, Sixt, Haute-Savoie ; Σ AT empliées : 70

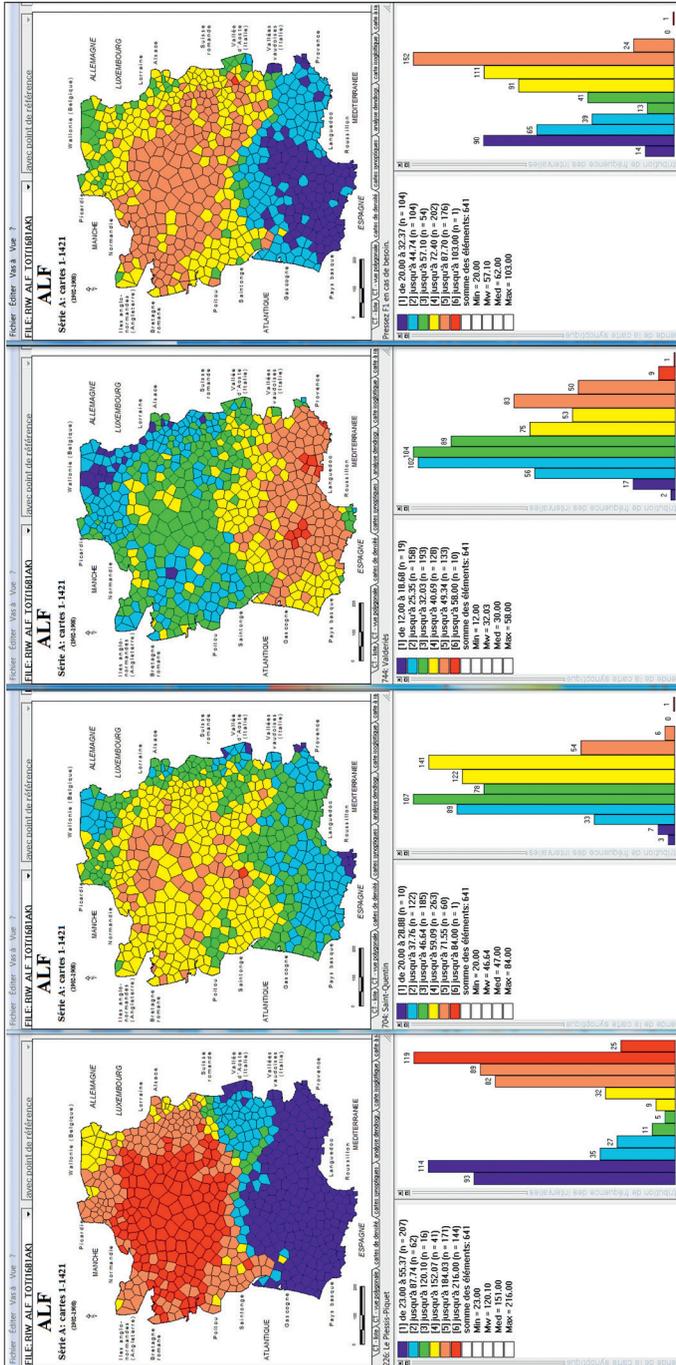


Planche 10: Synopse de quatre superpositions (par réunion) d'AT aux dimensions limitées (50-60% de 641 P-ALF)

Corpus: 1681 CT (toutes les catégories linguistiques); *algorithme de visualisation*: MINMWMAX 6-tuple

- À gauche: point d'inclusion: 226, Le Plessis-Piquet, Hauts-de-Seine; Σ AT empliées: 216
- Centre-gauche: point d'inclusion: 704, Saint-Quentin, Creuse; Σ AT empliées: 84
- Centre-droite: point d'inclusion: 744, Valdeyres, Tarn; Σ AT empliées: 58
- À droite: point d'inclusion: 956, Sixt, Haute-Savoie; Σ AT empliées: 103

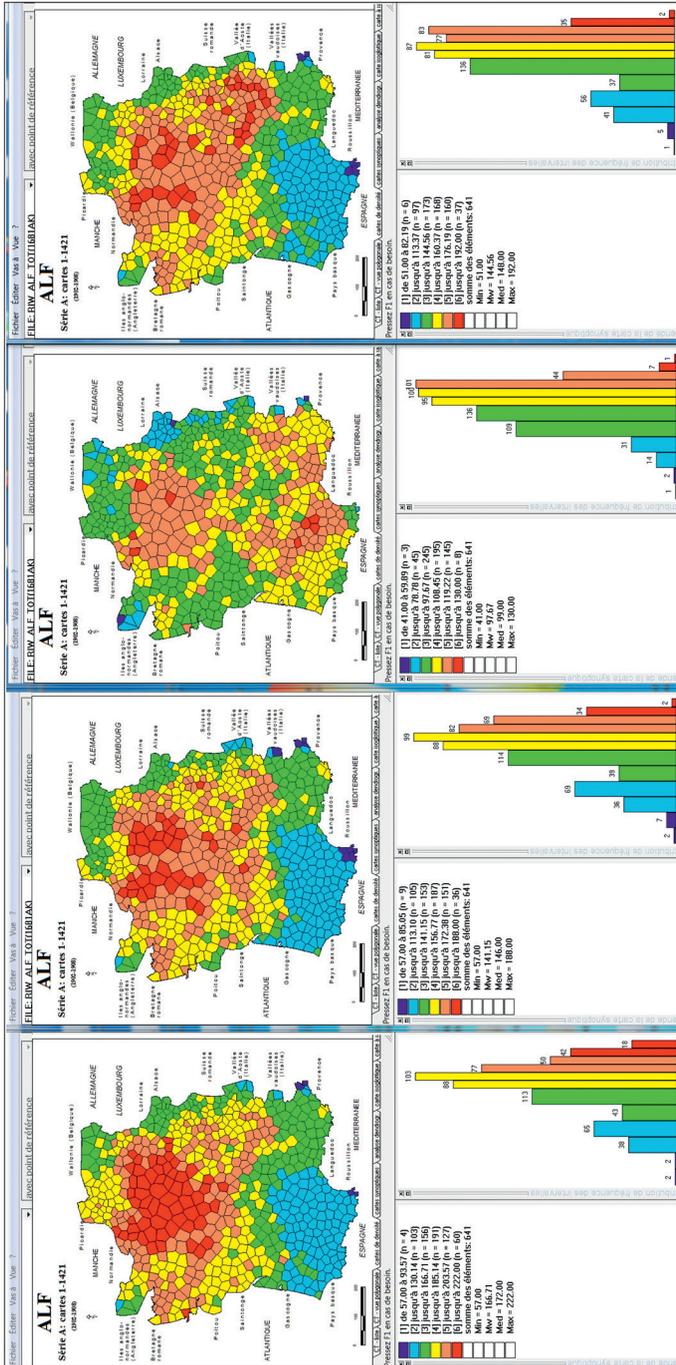


Planche 12: Synopse de quatre superpositions (par réunion) d'AT aux dimensions limitées (70-80% de 641 P.-ALF)

Corpus: 1681 CT (toutes les catégories linguistiques); algorithme de visualisation: MIN/MAX 6-tuple

- À gauche: point d'inclusion: 226, Le Plessis-Piquet, Hauts-de-Seine; Σ AT employées: 222
- Centre-gauche: point d'inclusion: 704, Saint-Quentin, Creuse; Σ AT employées: 188
- Centre-droite: point d'inclusion: 744, Valderiès, Tarn; Σ AT employées: 130
- À droite: point d'inclusion: 956, Sixt, Haute-Savoie; Σ AT employées: 192

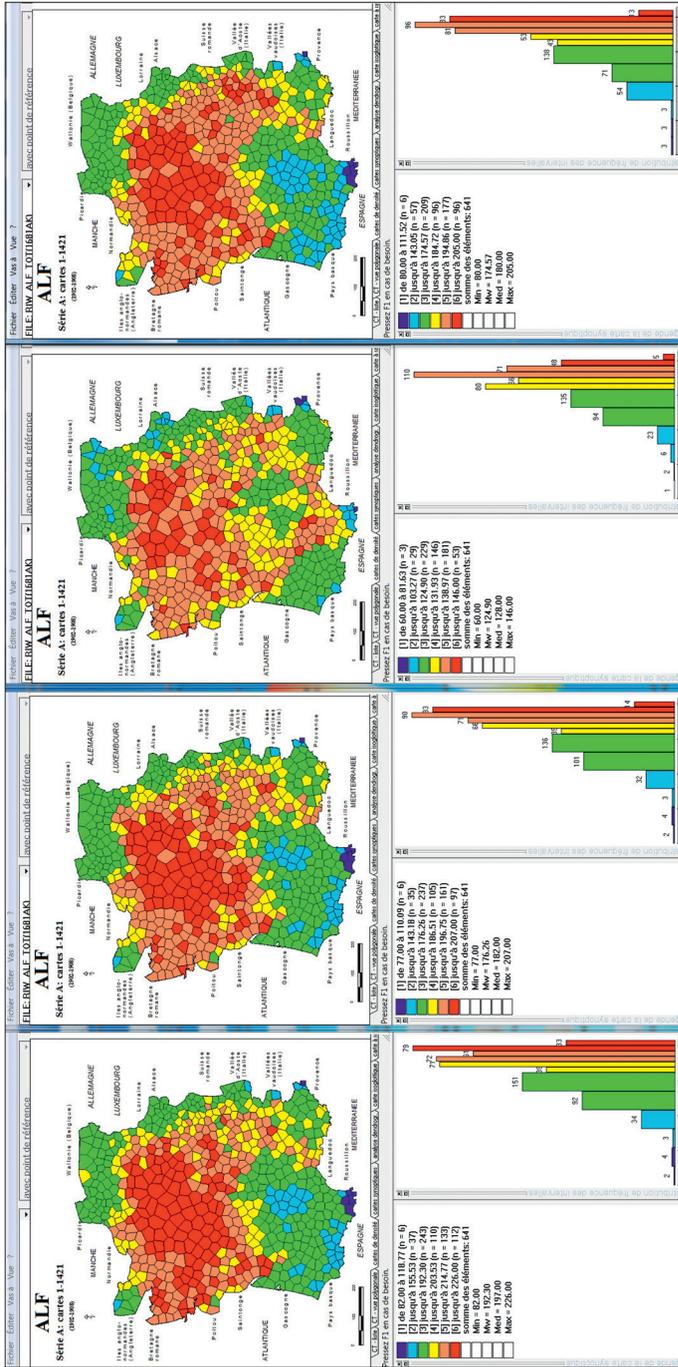


Planche 13: Synopse de quatre superpositions (par réunion) d'AT aux dimensions limitées (80-90% de 641 P.-ALF)

Corpus: 1681 CT (toutes les catégories linguistiques); algorithme de visualisation: MINMWMAX 6-tuple

- À gauche: point d'inclusion: 226, Le Plessis-Piquet, Hauts-de-Seine; Σ AT empliées: 226
- Centre-gauche: point d'inclusion: 704, Saint-Quentin, Creuse; Σ AT empliées: 207
- Centre-droite: point d'inclusion: 744, Valderiès, Tarn; Σ AT empliées: 146
- À droite: point d'inclusion: 956, Sixt, Haute-Savoie; Σ AT empliées: 205

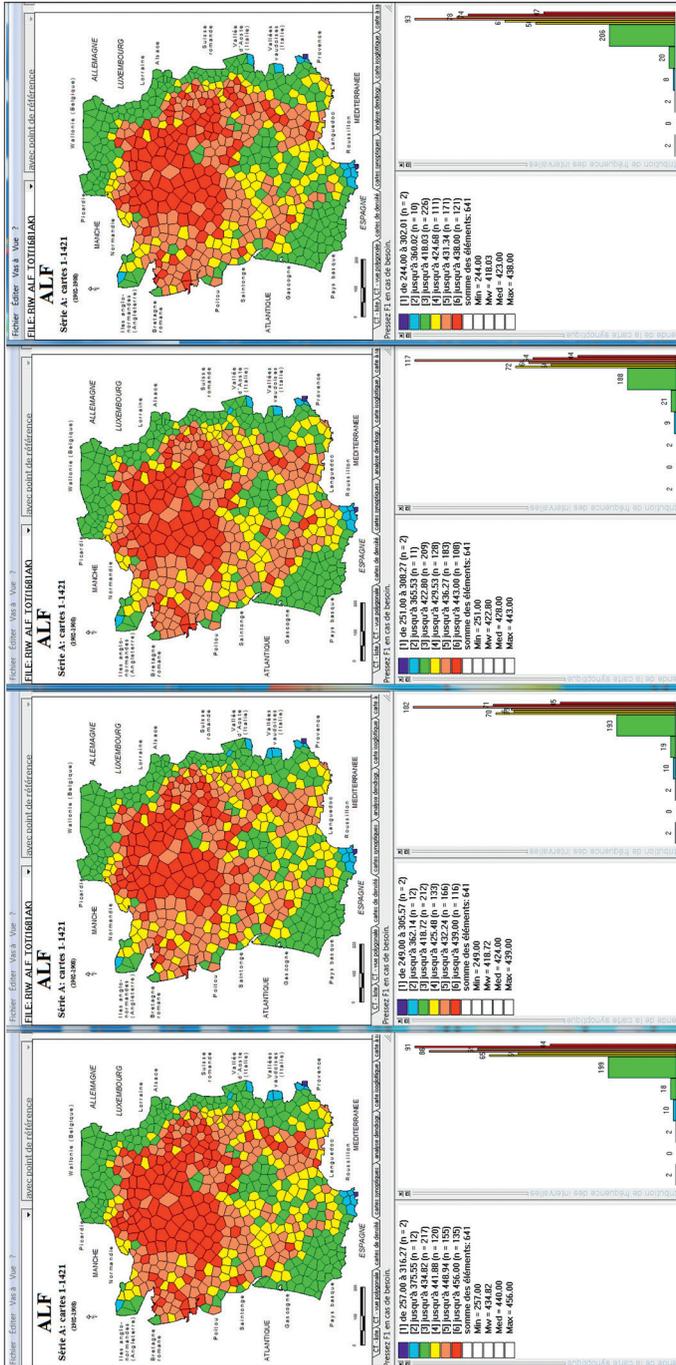


Planche 14: Synopse de quatre superpositions (par réunion) d'AT aux dimensions limitées (90-100% de 641 P.-ALF)

Corpus: 1681 CT (toutes les catégories linguistiques); algorithme de visualisation: MINMWMAX 6-tuple

- À gauche: point d'inclusion: 226, Le Plessis-Piquet, Hauts-de-Seine; Σ AT empliées: 456
- Centre-gauche: point d'inclusion: 704, Saint-Quentin, Creuse; Σ AT empliées: 439
- Centre-droite: point d'inclusion: 744, Valderiès, Tarn; Σ AT empliées: 443
- À droite: point d'inclusion: 956, Sixt, Haute-Savoie; Σ AT empliées: 438

