

Jean-Léo LEONARD, Guylaine BRUN-TRIGAUD, Flore PICARD

17 Atlas linguistiques et perspectives dialectométriques.

RESUME : Les atlas du grand projet NALF relatifs au domaine occitan sont présentés dans les grandes lignes, ainsi que la base de données THESOC, qui les réunit sous forme numérique, permettant un éventail de traitements et d'objectivation des systèmes dialectaux et des grandes dynamiques aréologiques du diasystème occitan. Une application des outils les plus récents de géolinguistique et taxinomie dialectale assistées par ordinateur (Gabmap, Gephi, cladistique) est proposée, à l'échelle de l'ensemble du domaine, en mettant en regard les aires supradialectales de P. Bec (1963, 1973) et les dialectes communément reconnus de l'occitan avec les résultats de ces traitements dialectométriques, qui permettent des formes diversifiées de visualisation des regroupements à diverses échelles. Quatre algorithmes sont retenus dans Gabmap¹ (*Classification Ascendante Hiérarchique, ou méthode de Ward, Complete linkage, ou « voisin le plus éloigné », moyenne de groupes pondérée et non pondérée*), faisant alterner des regroupements réductionnistes, ou supradialectaux, aussi bien que l'individuation de dialectes et de sous-dialectes. Les lignes de force épistémologiques de la dialectométrie – née en domaine occitan, à partir des travaux pionniers de Théobald Lalanne et de Jean Séguy – servent de prisme à la description et à l'explication des structures émergentes et de profondeur du réseau dialectal occitan, qui représentent la diversité interne de l'occitan, de multiples points de vue, selon les modes de calcul de la distance linguistique, afin de caractériser le *fait dialectal* dans le domaine occitan, dans ses structures de profondeur autant que de surface.

1. Introduction

1.1. Le défi des « grandes fresques » : enjeux épistémologiques et sociétaux

1.2. Les atlas linguistiques régionaux et leur valorisation dans le Thesaurus Occitan

2. Dialectométrie : un survol historique

3. Géométrie variable des divisions dialectales du domaine occitan

4. Étude de cas de modélisations vicariantes

4.1. Dialectométrie *versus* Cladistique

4.2. Extension méthodologique du domaine de GabMap

5. Conclusion et perspectives.

6. Références

7. Annexe bibliographique

¹ Le nom est forgé sur l'anglais *gab* « bagout, éloquence » et *map* « carte ».

1. Introduction

1.1 Le défi des « grandes fresques » : enjeux épistémologiques et sociétaux

Le présent chapitre aura moins pour objectif de proposer un inventaire historique des théories et méthodes de la dialectométrie ou de la dialectologie quantitative, que de proposer un éventail de techniques et méthodes pour non pas seulement établir, mais aussi pour revisiter les classifications dialectales. Quiconque est coutumier des débats sur les blogs et les réseaux sociaux dans la sphère culturelle occitane ou des langues d'oc – quel que soit le glottonyme plus ou moins fédérateur qu'on décide d'employer – sait que les questions de classification (Sumien 2009), mais surtout, d'individuation des différents dialectes ou régiolectes du réseau dialectal gallo-roman méridional, sont l'objet d'âpres discussions. C'est d'ailleurs là une opportunité inespérée pour les chercheurs et les érudits de tirer parti de cette remise en cause des objets de connaissance connus et reconnus (ou canoniques) : « arbres généalogiques » des grands dialectes et de leurs sous-dialectes, cartes géolinguistiques, tracés d'aires dialectales ou ensembles supradialectaux, comme dans les cartes de P. Bec, ci-après. Force est de constater qu'il n'existe pas, dans ce domaine, de vérité intangible. Le point de vue crée l'objet², et un espace dialectal, qu'il soit dessiné sur une carte ou hiérarchisé sous forme de *Stammbaum* avec intention généalogique explicite ou comme topologie taxinomique (nos résultats, présentés dans ce chapitre), ne sera toujours qu'un instantané, un construit discutable, dont la valeur est double : à la fois objet de connaissance et objet de discussion, sur la base de faits concurrents, ou d'autres formes de modélisation, de visualisation. L'important est de pouvoir chaque fois remonter à la source des observations, à l'angle de vue qui donne la perspective aboutissant au résultat rendu visible sous forme d'objets de connaissance à géométrie variable, comme nous allons le voir. D'autant plus que la somme des données dont nous disposons désormais sur la variation du gallo-roman méridional est, quoiqu'on puisse en dire par désir de mieux faire ou de faire bien plus, d'ores et déjà colossale. Le THESOC³, ou *Thesaurus Occitan* (cf. Brun-Trigaud 2014 et présentation *infra*) constitue l'une de ces sources abondantes de données analysables à l'échelle de l'ensemble du

² À ce titre, Nerbonne et Kretzschmar (2003) opposent les notions d'*attributive dialect* vs *blind dialect*, littéralement « dialecte projeté » vs « dialecte invisible » que le linguiste Edward Sapir aurait pu traduire en son temps par « phénodialecte » vs « cryptodialecte », comme il opposait les « phénotypes » aux « cryptotypes » dans les formes sensibles des langues du monde : le premier terme correspond au dialecte visible, clairement individué par la tradition savante ou la tradition orale et le blason populaire, tandis que le second correspond aux structures profondes d'un domaine dialectal, auxquelles on ne peut accéder que par des outils d'analyse sophistiqués. La dialectométrie confirme souvent les premiers, mais elle excelle particulièrement à faire émerger les seconds – c'est là son pouvoir heuristique.

³ Cf. <https://bcl.cnrs.fr/rubrique40> et <https://bcl.cnrs.fr/article49>.

domaine, outre les sommes de données considérables réunies par le passé par d'autres projets ambitieux, comme les quatre volumes de la *Grammaire historique [sic] des parlers provençaux modernes*, de Jules Ronjat (1930-37), ou même les données de l'*Atlas Linguistique de la France* (1902-1910), sans compter les innombrables monographies ou atlas de régions et microrégions du domaine d'oc. Malgré le sentiment de frustration qu'on pourrait ressentir, face à l'ampleur de la tâche de documentation de la variation dialectale de ce domaine, si riche, parce que chargé d'histoire et d'une expérience collective unique, on notera cependant :

a) la somme colossale de données linguistiques que les linguistes peuvent rassembler pour des langues et dialectes minoritaires, ou vulnérables, voire en danger, lorsqu'ils bénéficient d'investissements conséquents de la part de l'État, notamment à travers le réseau des universités (Toulouse, Montpellier, Perpignan, Nice, Salzburg, etc.), dont certaines se distinguent non seulement par l'ampleur des matériaux recueillis, mais aussi par la contribution intellectuelle, à la science, par la fondation de véritables écoles de pensée en dialectologie – comme c'est le cas à Toulouse, autour de Jean Séguy, mais aussi à Montpellier à travers la Société des Langues romanes et la revue éponyme, à Nice par l'école de dialectologie structurale de Jean-Philippe Dalbera et désormais, à travers le THESOC.

b) Même s'il s'agit d'une *tâche sans fin*, comme toute entreprise descriptive ambitieuse en linguistique, constituer les données, en termes de collecte, mais aussi d'édition, puis de valorisation informatique, sous forme de base de données, implique de les rendre exploitables à grande échelle, notamment en vue de grandes synthèses, comme celles que permet la dialectométrie.

c) L'éventail des réalisations possibles, grâce aux bases de données ainsi élaborées, à travers la dialectologie quantitative en tant qu'approche *taxinomique* (classificatoire), intéresse autant la linguistique générale que la linguistique appliquée et, partant, la dialectologie générale et appliquée, au service des langues (*codification, standardisation*) et, partant, de leurs usagers (associations, locuteurs, artistes et créateurs). Libre à chacun de trouver réponse à ses questions et d'opter pour d'autres représentations ; il n'en reste pas moins que les résultats obtenus à l'aide des algorithmes dont nous disposons désormais doivent être considérés comme autant d'opportunités pour faire avancer aussi bien l'épistémologie des Sciences du langage que les connaissances factuelles sur des domaines linguistiques complexes, comme le domaine occitan, au sein des langues romanes – observatoire privilégié de la linguistique empirique, tant diachronique que diasystémique.

Nous utiliserons à cette fin l'éventail de *perspectives* et d'*objets de connaissance*, plutôt que de *construits canoniques*, deux méthodes

complémentaires : la cladistique et la dialectométrie par « distance d'édition » (algorithme de Levenshtein), à l'aide de GabMap⁴ et de Gephi⁵ (v. *infra*).

Après une rapide présentation des atlas linguistiques régionaux occitans à partir de l'expérience de leur valorisation par le THESOC, nous montrerons, par l'usage possible de ces ressources numérisées, plusieurs explorations des résultats offerts par la dialectométrie⁶ pour la classification des parlers occitans. Nous mettrons en regard deux perspectives complémentaires : d'une part, la dialectométrie de « nouvelle génération », avec l'algorithme de Levenshtein (ou « distance d'édition »)⁷ et deux méthodes de visualisation (cartes et graphes, ci-dessous) à partir de données lexicales et phonologiques et d'autre part, la méthode cladistique (voir Brun-Trigaud et *alii* 2014).

1.2. Les atlas linguistiques régionaux et leur valorisation dans le Thesaurus Occitan

Les atlas linguistiques qui couvrent le domaine occitan, font partie, comme ceux du reste du domaine gallo-roman, de deux campagnes d'enquêtes réalisées sur place auprès de locuteurs, généralement ruraux. La première campagne s'est déroulée de 1897 à 1901 pour aboutir à l'élaboration de l'*Atlas Linguistique de la France* (ALF) publié de 1902 à 1910 par J. Gilliéron et E. Edmont. Dans cet atlas, 209 points concernent le domaine occitan et 10 celui du *Croissant*. Près de 2100 notions sont traitées dans 1921 cartes. La seconde campagne, dite des Atlas régionaux ou Nouveaux Atlas Linguistiques de la France (NALF), a été lancée par A. Dauzat en 1939 (Simoni-Aurembou 1998). Véritablement commencées après la seconde guerre mondiale dans le cadre du CNRS, les enquêtes furent menées, peu à peu, par aires linguistiques jusqu'à récemment et conduisirent à la publication de six atlas, dont les domaines sont entièrement compris dans l'aire occitane :

- **ALG** : J. Séguy, *Atlas linguistique et ethnographique de la Gascogne* (1954-1973, 180 pts, 6 vol., 2531 cartes),
- **ALMC** : P. Nauton, *Atlas linguistique et ethnographique du Massif Central* (1957-1961), 55 pts, 3 vol., 1899 cartes),
- **ALAL** : J.-Cl. Potte, *Atlas linguistique et ethnographique de l'Auvergne et du Limousin* (1975-1992, 76 pts, 3 vol., 1736 cartes),
- **ALP** : J.-Cl. Bouvier et Cl. Martel, *Atlas linguistique et ethnographique de la Provence* (1975-2016, 170 pts, 4 vol., 1358 cartes),

⁴ Voir <https://dev.clarin.nl/node/1917>, et <https://www.sciencedirect.com/science/article/pii/S0024384115000315>.

⁵ Gephi est un logiciel de visualisation par graphe disponible sur <https://gephi.org/>.

⁶ Voir Goebel 1982 pour une présentation des fondements de la méthode dialectométrique.

⁷ Cf. Levenshtein (1966) – l'article fondateur – et Leinonen & al (2016),

- **ALLOc** : X. Ravier, *Atlas linguistique et ethnographique du Languedoc Occidental* (1978-1993, 131 pts, 4 vol., 1198 cartes),
- **ALLOr** : J. Boisgontier, *Atlas linguistique et ethnographique du Languedoc Oriental* (1981-1986, 86 pts, 3 vol., 980 cartes).

On peut adjoindre à cet ensemble les données périphériques de cinq autres atlas incorporant des données occitanes dans leurs marges : ALLy, P. Gardette, *Atlas du Lyonnais* (1950-1956) ; ALPO, H. Guiter, *Atlas des Pyrénées Orientales* (1966) ; ALCe, P. Dubuisson, *Atlas du Centre* (1971-1982) ; ALO, G. Massignon et B. Horiot, *Atlas de l'Ouest* (1971-1983) ; ALJA, J.-B. Martin et G. Tuailon, *Atlas du Jura et des Alpes du nord* (1971-1978).

Dans les années 1990, après l'échec d'une tentative de mise en place de l'informatisation, sans doute prématurée étant donnée la complexité de la saisie des données en phonétique, le CNRS mit fin au groupe des Atlas Régionaux, laissant inédites un grand nombre de données. C'est dans ce contexte qu'est né le projet du *Thesaurus Occitan* (THESOC), créé par J.-Ph. Dalbera, son but étant de rassembler au sein d'une même base accessible, toutes les données des atlas régionaux du domaine occitan (Dalbera et alii 2012). Actuellement, le THESOC comprend près de 1,5 millions de data et il continue d'intégrer les données inédites, les rendant ainsi accessibles, notamment via son site Web (<http://thesaurus.unice.fr>).

L'ensemble des atlas régionaux constitue un total d'environ 950 points et plus de 8000 notions, cependant ces chiffres ne reflètent guère la réalité des faits comparables. Cette exigence de comparabilité avait pourtant été prônée par Dauzat dès son appel « Le questionnaire, modelé en principe sur celui de Gilliéron-Edmont, comprendra une partie commune à toute la France (la majorité des mots et toutes les phrases) et une partie régionale » (Dauzat 1937, 101). Mais très vite, le questionnaire mis au point par Dauzat fut rapidement délaissé (cf. Gardette 1968, 29-31 et Séguy 1973a, 77-78). D'après eux, cet abandon s'explique par plusieurs raisons : d'une part, le questionnaire était en définitive trop court pour rendre compte de la richesse lexicologique des « patois », d'autre part, le manque de suite logique dans les chapitres et les notions, était difficile à mettre en place pour des conversations suivies avec les locuteurs. Enfin, le manque de précision dans les questions nouvelles ("types de haches", ...), risquait d'engendrer une confusion dans les données récoltées.

Finalement, les enquêtes en domaine occitan ont été effectuées à l'aide trois questionnaires différents qui ont eu chacun leurs émules : chronologiquement, Gardette, qui avait commencé ses enquêtes avant la mise au point du questionnaire de Dauzat, avait son propre questionnaire qui a été utilisé également par l'ALO, puis l'ALG est le seul à avoir suivi celui de Dauzat (vol. 1, 2, 3) avant de reprendre une nouvelle campagne d'enquêtes complémentaires, notamment plus complètes

pour la morphologie (vol. 4, 5, 6). Enfin Nauton qui avait également mis au point son propre questionnaire pour l'ALMC, (Nauton 1956, 57-62) a été suivi par l'ALAL, l'ALP, l'ALLOc et l'ALCe, et dans une certaine mesure par l'ALLOr qui avait démarré lui-même avec un autre questionnaire mis au point par Louis Michel avant de se rallier à celui de Nauton. Cependant, dans l'état actuel des publications des atlas régionaux, seulement 400 cartes sont communes aux atlas strictement compris dans le domaine occitan.

En dehors des NALF, d'autres atlas consacrés à des espaces plus petits que ceux des Atlas régionaux ont été mis en chantier dans le domaine occitan : Millardet (1910) pour les Landes, Camproux (1978) pour le Gévaudan, Camps (1985) pour le Biterrois, enfin ceux de Hallig pour la Lozère et de Balmayer (région de Montpellier) sont resté inédits.

La discipline dialectométrique est née, initialement, en terres occitanes, par le biais de l'équipe de l'ALG, dirigé par Jean Séguy⁸. De ce point de vue, la contribution du domaine occitan à la dialectologie générale et appliquée a été immense, et continue de perdurer et de prospérer, comme le montre la bibliographie jointe en annexe.

2. Dialectométrie : survol historique appliqué à la présente orientation

Comme nous l'avons annoncé plus haut, il n'est pas question ici de présenter une synthèse historique, ni même un état des lieux, de la dialectométrie et des méthodes quantitatives en géolinguistique, ou en dialectologie computationnelle. Tout au plus pouvons-nous, dans cette section, expliciter en quoi l'approche retenue ici, avec ses données et objets de connaissance réunis ad hoc, en tant qu'avancées récentes – et non en tant que compilation –, s'inscrit dans un contexte épistémologique, méthodologique et empirique plus large. Pour un survol récent des méthodes dialectométriques, nous renvoyons à un article de synthèse récent de Martijn Wieling et de John Nerbonne (2015), dans une perspective généraliste ; on lira également avec intérêt le récit de l'application des méthodes dialectométriques dans une zone de contacts interdialectaux bien délimitée au sein d'un domaine comme celui du catalan, par Valls *et al.* (2013), et Wieling *et al.* (2018), à l'aide d'un corpus *ad hoc*.

En correspondance avec les remarques *supra* sur la conception et l'élaboration en amont des questionnaires du NALF et les conséquences en aval, on lira avec

⁸ J. K. Chambers (2016) rappelle cependant qu'on peut remonter au linguiste polonais Jan Czekanowski (1882-1965) pour une approche quantitative en dialectologie analogue à celle de Jean Séguy, 45 ans plus tôt, appliquée aux langues slaves ; n'oublions pas non plus que Séguy se disait redevable à son collaborateur de l'A.L.G pour les Landes gasconnes, l'abbé Théobald Lalanne (v. *infra*).

intérêt l'article décisif de Hans Goebel (2014) sur l'analyse dialectométrique de deux grandes séries de cartes de l'ancêtre atlantographique du NALF : l'*Atlas linguistique de France* de Gilliéron et Edmont (1902-1910), avec application de la dialectométrie salzbourgeoise, à travers une comparaison terme à terme des résultats du traitement des deux séries, donnant lieu à une vingtaine de cartes, tant dans une perspective interponctuelle (un point comparé dans sa gradation chromatique à tous les autres points du réseau), que du point de vue de la condensation des aires, selon les paramètres (phonologie, lexique, etc.).

Dans un article de 1986, Dennis Philps présentait une synthèse historique des différentes méthodes de calcul de la similarité, et de son pendant dialectique, la distance dialectale, qui ont émergé dans la décennie qui a suivi les articles fondateurs de Jean Séguy (1971, 1973a, b). Cette contribution est d'un grand intérêt pédagogique, en ce qui concerne l'élaboration des algorithmes (Philps donne même à voir un script, ou routine de programmation, en annexe). Comme Jean-Louis Fossat, et conformément à la terminologie de l'École de dialectologie de Toulouse, il désigne les composantes de la langue telles que la phonologie (principalement la phonétique historique, dans les atlas linguistiques), la morphologie (verbale ou pronomiale), le lexique, sous le terme de *paramètre*, que nous reprendrons ici. L'auteur donne des clés de lecture des classifications hiérarchiques : il considère les branches externes de dendrogrammes comme dotées de « poids structurant » –, selon la logique du *tertium comparationis* : plus la branche externe (ou le dendrème périphérique à un clade) est intégrée dans une macro-structure interne, plus on peut la considérer comme dotée de caractères endémiques, tandis que plus elle est externe à l'ensemble de la taxinomie, plus son « poids structurant » est génétiquement et typologiquement représentatif d'une différence de dialecte ou de langue – l'auteur illustre cette caractérisation à l'aide d'exemples gascons. Alors que la dialectométrie de Jean Séguy (1971, 1973a-b) s'intéressait aux similarités (fonction de *communication* du dialecte) et aux différences (fonction de *démarcation*), dans une dynamique de relations interponctuelles dans un réseau dialectal (triangulation), à partir de la recherche sur le *gradient de gasconité* (qu'on pourrait aujourd'hui appeler « condensation typologique » des aires)⁹, se développent dans l'École de Toulouse des recherches sur la notion de *centralité* dans un réseau dialectal – cette quête traverse l'ensemble des thèses de Denis Philps (1976, 1985) – cf. Fossat (1978) qui, en outre, exprime des doutes sur le caractère sinon heuristique, du moins probant de la phonétique historique, en tant que

⁹ Nous verrons qu'un tel construit de dialectologue (le « gradient de gasconité ») s'avère peu heuristique aujourd'hui à la lumière des résultats de Gabmap : il y avait là une quête du noyau typologique d'un dialecte, d'ailleurs induit en amont par le choix de variables spécifiques, relevant d'une sélection d'isoglosses connues par la démarche qualitative ou philologique. Tout au plus trouvera-t-on, en revanche, dans nos résultats, des *chorèmes*, ou aires individuelles homogènes émergentes, sans préjuger d'un quelconque gradient typologique programmé en amont.

paramètre dialectométrique. Notons que, de ce point de vue, même si le choix des entrées de la base de données THESOC que nous allons utiliser a été motivé, en amont, par des considérations diachroniques, il n'en reste pas moins que le traitement gabaritique auquel la distance d'édition soumet les chaînes de caractères dans des alignements matriciels (ou gabarits, cf. tableaux 1 et 2 *infra*) revient à un traitement comparatif terme à terme, paradigmatique, et donc synchronique, des segments. C'est ce que nous appellerons plus loin l'approche *réalisationnelle*, en opérant sur la calculabilité des faits relevant de l'*absence*, de l'*identité* ou de la *substitution* dans une séquence calibrée de caractères, à l'échelle du mot, ou lexème.

Du point de vue épistémologique, il est important de retenir la célèbre division entre deux grands modes opératoires, en matière de saisie de toute réalité en sciences : comme le rappelle Hans Goebel dans nombre de ses contributions, l'approche de type *lumping* s'oppose à celle de type *splitting* (en anglais : *to lump* 'rassembler tout en un seul morceau' ; *to split* 'éclater, disperser'). On pourrait transposer cette polarité par les termes *réductionnisme* versus *atomisme* : entre *rassembler* contre (sub)*diviser* le réel dans les représentations qu'on en donne à voir. Mais le terme d'*atomisme* est connoté (souvent à tort, car aucun détail en soi ne nuit à l'observation) péjorativement, si bien que mieux vaudrait parler de « finesse de grain », à défaut de « détaillisme » – terme qui pourrait également prêter à connotation péjorative. Or, il s'agit de *Gestalt*, ou perception du réel, à partir de représentations (ou schèmes mentaux), de perspectives.

Une question s'avère centrale, au-delà de la recherche sur l'interponctualité et les *structures de profondeur des réseaux dialectaux*, dans une perspective réticulaire (tout réseau de localités, ou variétés) et taxinomique (tout classement et critère de classement de ces variétés) – caractéristique de la dialectométrie toulousaine et salzbourgeoise – : le dilemme de la catégorisation (et, partant, de la mesure) d'objets distincts, en termes de *langue, dialecte, sous-dialecte, parler*¹⁰. Les dialectes existent-ils ? Ou bien doit-on se contenter de l'idée que seul le *fait dialectal* existe, en relation avec un continuum, aux frontières ou délimitations si variables, si brouillées par les entrelacs d'affinités et de divergence de proximité ou à distance, que toute dichotomie serait sinon vaine, du moins peu heuristique ? L'école de Toulouse, autour de Séguy, utilisait habilement la lexie complexe du *fait dialectal*, pour élever le niveau d'observation, sans préjuger de ce qui pouvait être considéré comme dialecte ou sous-dialecte. L'inspirateur de la dialectologie quantitative, Théobald Lalanne, avait montré la voie en ce sens, lui qui intitulait son

¹⁰ V. note *supra*, concernant la hiérarchisation interne des domaines ou réseaux dialectaux. Le fondateur du paradigme de la recherche sur les *diasystèmes* (métagrammaires intégrant l'ensemble de la variation structurale d'un domaine ou réseau dialectal), Uriel Weinreich, avait écarté cette question en donnant le terme « variety » comme équivalent de « dialect » – dans son acception anglo-saxonne, qui correspond à « parler » (Weinreich 1954 : 389). Pour nous, cela revient paradoxalement à se priver de hiérarchie, autrement dit, de *structure interne* du diasystème.

rapport de recherches en 1949, publié à compte d'auteur à Saint-Vincent de Paul, « l'indépendance des aires linguistiques en Gascogne maritime » : toutes les aires s'avéraient selon lui « indépendantes », donc non réductibles à des ensembles supérieurs cohérents désignables sous le terme de « dialecte », ou *a fortiori* sous-dialecte, etc.

Pas même le « gascon noir » des Landes gasconnes maritimes, pourtant si nettement opposable au « gascon clair » par son vocalisme¹¹, sur le plan ethnolinguistique, à en croire le jugement des locuteurs eux-mêmes, ne voyait son individualité corroborée par les calculs systématiques, méthodiques, organisés selon des critères structuraux explicites (paramètre lexical dans le premier fascicule, paramètre de la phonétique historique dans le deuxième fascicule : « palatalisation de 'j' vers 'x', évolution $y > j$, conjointement $z > j$, chuintement $ks \text{ lat.} > x$ gascon (ch), etc.). À ce titre, Lalanne serait quasi-structuraliste, pré-labovien tant il se montre attentif aux séries structurales, qu'il traite quantitativement comme des variables phonologiques à part entière – et non plus seulement comme des séries évolutives. Mais son diagnostic rejoint celui de Paul Meyer et de Gaston Paris avant lui : les dialectes n'existent pas. Pas même le « gascon noir », en dépit du sentiment linguistique des locuteurs ou des érudits ou des linguistes.

On pouvait donc attendre de la dialectométrie qu'elle aide à résoudre ce dilemme. Pour Hans Goebel comme pour Henri Guiter avant lui, et parallèlement à Jean Séguy (dont toute l'œuvre dialectométrique peut être considérée, selon l'expression employée par Xavier Ravier, comme « une traversée du langage gascon », certes, mais, au-delà, comme un inlassable questionnement sur le statut dialectal du gascon au sein de la langue d'oc), les dialectes et sous-dialectes existent

¹¹ On consultera utilement Wikipedia (https://fr.wikipedia.org/wiki/Gascon_Noir), au sujet de la localisation de ces deux aires dialectales, constituées en deux sous-dialectes du gascon maritime : « **Le Gascon Noir**, parfois connu sous le nom de **parlar negre** ou plus localement **parlar negue** ([par'la 'nøɣrə, par'la 'nøɣə], le « parler noir ») est une variété de gascon parlée dans la moitié occidentale du département des Landes, dans l'agglomération bayonnaise (où se parle une variante appelée *gascon de Bayonne*), ainsi que dans le sud-ouest de la Gironde (hors bassin d'Arcachon). C'est le parler du Pays de Born, du Marensin, du Maremne, du Seignanx, du Pays d'Orthe, de la Grande Lande, du val de l'Eyre et d'une partie de la Chalosse. On l'appelle également **gascon maritime** ou **landais**. Il s'oppose au *parlar clar* (parler clair), usité plus à l'est et au nord. » Ces caractérisations sont résolument qualitatives, en tant que variables isophones fondées sur des différences de timbre vocalique. Malgré le grand rendement fonctionnel de ces variables cependant, elles se trouvent lissées, dans une approche quantitative. On ne trouvera trace de ces deux aires isophones qu'à un niveau de granularité très fin d'analyse dialectométrique, que nous n'aborderons pas ici.

bel et bien. La dialectométrie salzbourgeoise permet, outre la visualisation des relations interponctuelles entre chaque point d'atlas, de rendre compte de la condensation synoptique des coefficients de similarité, en tant que composante des *structures atlantographiques profondes* – dans une perspective, du reste, universaliste, qui est une préoccupation constante, hautement louable, de ce paradigme de recherche. La méthode dite « globale » de Guiter le permet également.

Depuis, les instruments aussi bien que les échelles de mesure en dialectométrie se sont considérablement diversifiés, dissuadant d'appliquer cette gamme de catégories épilinguistiques (langue, dialecte, sous-dialecte, parler, etc.) à tout résultat quantitatif, mais en son temps, ce premier questionnement quantitatif du « fait dialectal » a joué un rôle salubre, en rupture avec l'approche *continuiste* de Meyer et Paris, de nature purement impressionniste, et qui finissait par devenir une doctrine – un dogme. C'est donc là un des acquis de la dialectométrie : la découverte et la recherche ininterrompue, infatigable, de critères objectifs, quantitatifs, pour la détection de systèmes et sous-systèmes discrets dans les *diasystèmes* (Weinreich 1954) – tels que dialectes, sous-dialectes et variétés ou parlers, opposables à langues. Enfin, une constante que l'on retrouve dans toutes les études dialectométriques (et dont Jagueneau (1987) est un exemple-clé, dans son souci de distinguer entre *types* ou *caractères* poitevin-saintongeais vs d'oïl vs occitan), à travers l'approche taxinomique, n'est autre que la perspective typologique : au fond, les résultats des calculs dialectométriques rendent davantage compte de convergences typologiques entre aires et dialectes et leurs subdivisions, que de déterminismes généalogiques – à la différence de la classification de type *Stammbaum* des comparatistes du XIX^e siècle. Au sortir du processus, ce sont davantage des clusters ou regroupements d'attributs structuraux – de traits dialectaux cumulés et hiérarchisés, rétenteurs ou innovants, autrement dit, des *classes diasystémiques* – que des *unités génétiques*, que l'on obtient, et ce même lorsqu'on fait usage d'outils utilisés par les généticiens, comme la cladistique. Car les langues sont des *systèmes sémiotiques dynamiques* – non des organismes vivants.

3. Géométrie variable des divisions dialectales du domaine occitan

Parmi les méthodes récentes, ou méthodes de saisie émergente de la complexité géolinguistique, la *distance de Levenshtein*, aussi appelée *distance d'édition*, est utilisée principalement pour estimer la distance phonologique entre deux variétés dialectales (en l'occurrence, deux points d'atlas linguistique, comme ici, de la banque de données THESOC). Elle calcule la différence entre deux chaînes de caractères, qui est définie comme le nombre minimum d'opérations de *substitution*, d'*insertion* et de *suppression* de caractères nécessaires pour passer de l'une à l'autre. En calculant cette distance entre des cognats provenant de deux parlers distincts, on

obtient une mesure de la différence entre ces parlers. Comme pour toutes les méthodes dites quantitatives, qui fonctionnent par approximation, plus la quantité de données est grande, plus les résultats seront affinés et, en principe, fiables.

ALP 2	a	b	ε	λ	ɔ
ALP 8	a	b	e	λ	ɔ
Distance	0	0	1	0	0

Tableau 1 : Grille d'alignement automatisé des chaînes segmentales par Gabmap : entrée *abelha* 'abeille' : variétés ALP 2 vs ALP 8

Pour le lexème *abeille*, les cognats des points ALP 2 et ALP 8 ont une distance de Levenshtein de 1 ou un « score » de 1, au terme du calcul), c'est à dire qu'il y a un seul caractère de différence, comme reflété dans le tableau 1. Entre les cognats d'ALP 8 et ALP 3, la distance est de 2, correspondant à une *suppression* et une *substitution* (tableau 2) :

ALP 8	a	b	e	λ	ɔ
ALP 3		b	ε	λ	ɔ
Distance	1	0	1	0	0

Tableau 2 : Grille d'alignement automatisé des chaînes segmentales par Gabmap : entrée *abelha* 'abeille' : variétés ALP 8 vs ALP 3

On calcule la moyenne des distances obtenues pour chaque paire de dialectes, en normalisant les résultats selon la longueur des mots, pour que les mots les plus longs, qui contiennent donc souvent le plus de changements, ne prennent pas un poids démesuré. La distance finale entre deux dialectes peut s'exprimer en décimales ou en pourcentage, comme dans l'extrait *infra*.

	ALP 1	ALP 2	ALP 3	ALP 6	ALP 7	ALP 8	ALP 10	ALP 11	ALP 12
ALP 1		28	28	31	31	33	34	34	35
ALP 2	28		17	22	20	25	25	25	27
ALP 3	28	17		23	16	22	22	21	23
ALP 6	31	22	23		21	30	24	24	26
ALP 7	31	20	16	21		36	25	21	24
ALP 8	33	25	22	30	36		41	38	39
ALP 10	34	25	22	24	25	41		21	30
ALP 11	34	25	21	24	21	38	21		34
ALP 12	35	27	23	26	24	39	30	34	

Tableau 3 : Extrait des résultats de l'algorithme de distance, sous forme de matrice

Dans le cas de réseaux dialectaux de petite taille, les chiffres bruts obtenus par cette méthode peuvent être interprétés en l'état, mais cette méthode s'avère moins intuitive pour de plus grands réseaux. Il existe diverses manières de visualiser les résultats (cf. cartes *infra*), soit en utilisant des cartes géographiques, soit en utilisant des graphes de réseau (cf. graphes Gephi *infra*). Les deux méthodes aboutissent à des apports distincts, donc nous les expérimenterons toutes deux afin de les mettre en regard. Les cartes géographiques seront réalisées avec l'application Gabmap. L'École de dialectométrie de Groningen s'est constituée autour de John Nerbonne (2001) et Wilbert Heeringa (2004). Ces derniers utilisent les principes de l'algorithme de Levenshtein, ou distance d'édition (*edit distance*, en anglais). Au sein de cette école, Peter Kleiweg a développé en 2004 un logiciel téléchargeable et une plateforme en ligne RuG/L04 (<http://www.let.rug.nl/kleiweg/L04/>) basée sur le langage R (logiciel libre de traitement des données et d'analyses statistiques) dont une partie des fonctions est devenue, depuis 2010, une application en ligne Gabmap (<http://www.gabmap.nl>) (Nerbonne *et alii* 2011), assez largement utilisée dans la communauté scientifique.

Gabmap offre à la fois des outils pour l'analyse des données et plusieurs modes de représentation des résultats. L'application comprend des méthodes de mesure des distances linguistiques basées sur des transcriptions phonétiques, des données dialectales numériques (par exemple, les fréquences de formants) et des données catégorielles (par exemple, des variables syntaxiques ou morphologiques). L'analyse est réalisée à l'aide de quatre algorithmes (cf application *infra*). Les distances linguistiques peuvent être cartographiées au moyen d'une mise à l'échelle multidimensionnelle (*Multidimensional Scaling* : MDS) afin de visualiser les continuums dialectaux, ou au moyen d'une analyse de clusters (regroupements) afin de visualiser les groupes dialectaux. Il est également possible d'intégrer des réponses multiples (Aurrekoetxea *et al.* 2020) qui permettent de prendre en compte la variabilité dans les calculs de traitement des distances linguistiques.

Nous utiliserons dans ce qui suit la gamme des quatre algorithmes que Gabmap met à la disposition des usagers. Il existe de nombreux algorithmes qui peuvent donner des résultats très différents en fonction de la nature des données (Prokić/Nerbonne 2008). Gabmap en comporte quatre : *Complete Link*, *Group Average*, *Weighted Average* et *Ward's Method*. *Complete Link*, « lien complet », aussi appelé *Furthest Neighbour*, autrement dit la méthode du « voisin le plus éloigné », part des paires d'objets à la fois les plus éloignés et les plus semblables en les fusionnant deux à deux à travers un processus itératif, sur l'ensemble du spectre des distances obtenues pour les variétés comparées une à une.

Group Average, la « Moyenne de Groupe », (ou la « Méthode de Groupe par Paires non pondérées » utilisant des Moyennes Arithmétiques, UPGMA) appartient

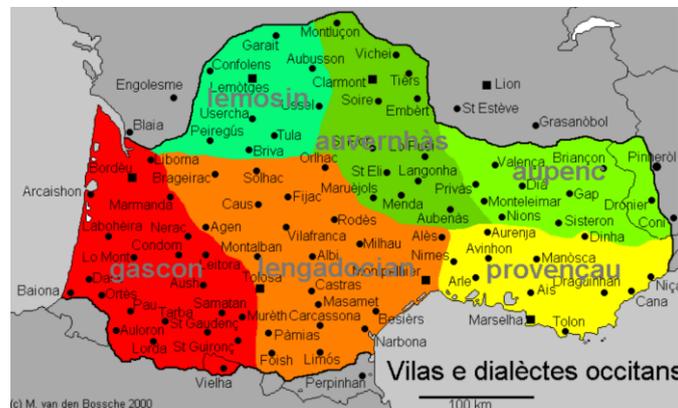
à un groupe de méthodes de regroupements moyens. Dans la « Moyenne de Groupe », la distance entre deux groupes, quels qu'ils soient, correspond à la moyenne des distances entre les membres des deux groupes en question. La moyenne est pondérée naturellement, en fonction de la taille des groupes. En conséquence, les groupes les plus petits seront moins pondérés et les plus grands le seront davantage. *Weighted Average*, « Moyenne Pondérée » (ou la Méthode de Groupe de Paire Pondérée utilisant des Moyennes Arithmétiques, WPGMA), tout comme la « Moyenne de Groupe », calcule la distance entre les deux groupes comme la moyenne des distances entre tous les membres de deux groupes. Mais dans la « Moyenne Pondérée », les groupes fusionnés maintiennent un poids égal indépendamment du nombre des membres de chaque groupe. Quant à la *Ward's Method* ou Méthode de John Ward, ou *Classification Ascendante Hiérarchique* (CAH), cette méthode est également connue comme la « méthode du désaccord minimal ». À chaque étape de l'analyse, les groupes qui s'amalgament sont ceux qui aboutissent à la plus petite augmentation du carré de la somme des distances de chaque élément à partir de la moyenne de son groupe. Cette méthode utilise une approche d'analyse des variations pour calculer les distances entre les groupes. Elle tend à créer des groupes de taille analogue.

Les graphes que nous utiliserons à des fins de visualisation alternative des dynamiques de *réticularité* (émergence de réseaux d'affinités/similarités) sont réalisés à l'aide du logiciel Gephi. Dans un graphe de réseau, les dialectes sont représentés par des points et les distances sont interprétées comme des liens entre ces points, dont le poids est inversement proportionnel à la distance calculée par l'algorithme de Levenshtein (poids minimal = distance maximale). La taille des points correspond à une approximation de leur poids dans le réseau, c'est-à-dire du nombre de liens forts avec d'autres dialectes : plus un point est épais dans le graphe, plus il est connecté à d'autres points.

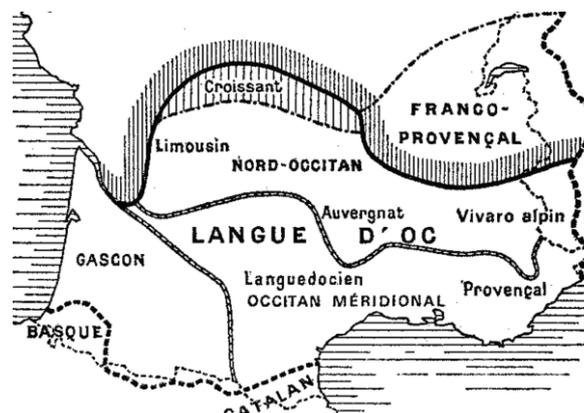
Le graphe est ensuite « seuillé », ou étagé (déployé en fonction de *seuils* par rapport à la moyenne normalisée des similarités ou convergences), afin de faciliter la visualisation des relations entre variétés, sous-dialectes, dialectes, voire langues (franges gallo-italiques alpines à l'extrême est, parlers « gavaches » au nord du gascon) : seuls les liens dont la valeur est située au dessus du seuil déterminé sont visibles et pris en compte dans la visualisation. La localisation des points à chaque seuil est calculée selon le poids et le nombre de ses liens, rapprochant ainsi les dialectes les plus similaires. Cela permet de faire apparaître des clusters, des groupements de variétés par leur similarité, constituant autant de dialectes et sous-dialectes. Pour cela, nous utilisons deux algorithmes : l'algorithme Force Atlas est utilisé pour les *sets* ou séries de données avec un petit nombre de points (quelques dizaines), et l'algorithme de Fruchterman-Reingold est utilisé pour les plus grand ensembles de données (plusieurs centaines de points), Les deux se servent du

principe de forces d'attraction-répulsion (*force-directed*), mais l'algorithme de Fruchterman-Reingold dispose les points au sein d'un cercle par un système d'approximation permettant à la fois une visualisation plus compréhensible d'un graphe de grande taille et une meilleure vitesse de traitement.

Nous partirons d'une classification « canonique » du domaine occitan dans l'ensemble des parlers gallo-romans : la carte de la figure 1 est un classique du genre. Elle est de type *splitting* (ou *détailliste*), tandis que les cartes 2 et 3 sont du type *lumping* (ou *réductionniste*), dans la mesure où elles représentent un niveau de divisions du rang des aires supradialectales. La carte 2, de Pierre Bec, relève d'un compromis entre les deux perspectives – de loin et de près –, puisque sept dialectes sont indiqués, à titre de repérage, mais sans lignes de démarcation interne.



Carte 1. Les langues de France et le continuum dialectal gallo-roman, selon M. van den Bossche (Source : Wikipedia)



Carte 2. Délimitation de la diversité dialectale de l'occitan selon des critères phonologiques, par P. Bec (31973 : 9)

Cette carte, de nature isoglossique (par faisceaux d'isoglosses, ici stylisés en lignes de partage, sur la base d'une nomenclature de variables phonologiques) postule une tripartition sur la base de coordonnées géographiques et historiques, selon une logique de bandes transversales (nord-occitan vs occitan méridional) d'une part, et d'une tierce zone au sud-ouest : le gascon. Des entités (dialectes ?

sous-dialectes ?) de rang inférieur sont spécifiées dans ces grands ensembles : languedocien au centre, provençal au sud-est, vivaro-alpin au nord-est, auvergnat (mal placé ici, sans doute pour des raisons de commodité typographique) et limousin, au centre-nord et au nord-ouest, respectivement. Pierre Bec retient donc sept entités dialectales au total, réparties sur trois macro-aires. La méthodologie utilisée par l'auteur étant qualitative, fondée sur une hiérarchie non explicitée sous forme arborescente, la structure interne de cette taxinomie (ou choix de classification d'un ensemble d'objets observables) n'est pas donnée avec cette représentation cartographique. La convergence relative, voire parfois terme à terme, avec les topologies et taxinomies générées par Gabmap sur la base de nos données THESOC va nous permettre de discerner des ensembles hiérarchisés (une *structure interne* du *diasystème*, comme *expression* du réseau dialectal du NALF, pour le même paramètre phonologique, notamment).

3.1. Aires phonologiques

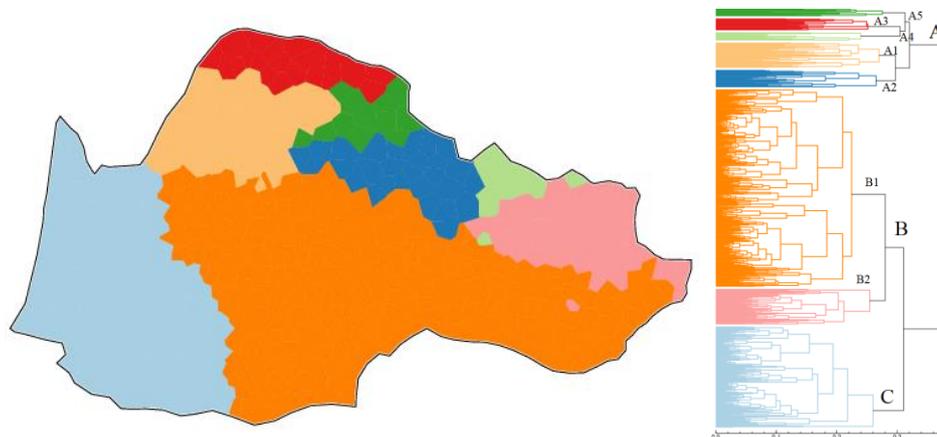
Nos résultats avec Gabmap, appliqué à notre base de données phonologique THESOC¹², à l'aide de l'algorithme *Complete Link* (ou « Voisin le plus éloigné »), génèrent un *objet de connaissance* très congruent avec la tripartition de P. Bec, en retenant 8 intervalles (les 7 entités retenues par Bec + 1, à titre de marge potentielle, ou « joker », en quelque sorte, permettant de saisir enclaves et micro-aires alloglottes, d'oïl ou de gallo-italique), comme le montre la carte 3a *infra*.

Voici, à titre indicatif, un échantillon des données utilisées (données de l'ALAL) :

¹² Les entrées de la base de données phonologique sont les suivantes : *abeille, agneau, agneler, aiguille, âne, arbre, blé, bœuf, caille, chapeau, châtaignier, cheminée, chemise, cheval, chèvre, ciel, ciseaux, clef, cou couteau, cuisse, cul, dé, drap, eau, échelle, escalier, étoile, feu, feuille, fiel, fil, foie, foin, font, fourche, froid, genou, gerbe, gland, guêpe, laine, lait, lessis* (lessive avec cendre et eau), *lièvre, lune, miel, mouche, mule, mûr, mûre, nid, noyer, œil, œuf, oie, oiseau, pie, pierre, pleuvoir, plier, pré, puits, soleil, souper, suer, suie, vache, veau, vent, vipère*. Ces lexies ont été retenues en fonction de leur fort pouvoir diagnostique du point de vue de la phonétique historique. Le calcul réalisé par Gabmap est de type « réalisationnel » : c'est à l'échelle du mot, tel qu'il a été réalisé lors de l'enquête de terrain, sous sa forme transcrite en alphabet phonétique, qu'opère le calcul. En cela, l'approche de Gabmap (et de tout traitement par distance d'édition) diffère radicalement de l'approche par « cartes-types » de la dialectométrie salzbourgeoise, qui indexe de manière binaire l'absence vs la présence d'un trait/type. Le bilan des données traitées est le suivant : 662 points d'enquêtes du NALF, 71 entrées ou lexies, 44 748 comparaisons ou tests (*instances*, en anglais), 245 029 caractères, dont 99 uniques, pour 240 562 occurrences ou *tokens*, dont 203 uniques.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1		abeille	agneau	agneler	aiguille	âne	arbre	blé	boeuf	caille	chapeau	châtaignier	cheminée	chemise	cheval
2	ALAL 1	ab œAo	ap̄ e		g yAo	'æz'	'abr*	bl̄ o	bjoz*	k'ɛj	tsap e	tsatanj œj	tsamin œj	tsamj o	tsav o
3	ALAL 10	b'œAo	ap̄ i	'niil a	g yAo	'az'	'æbr*	bl̄ o	b'œ*	k'æAo	tsap i	tsastan' aj	tsamin' ad	tsam' i'z	tsav o
4	ALAL 11	abij'i	ap̄ e	nel' o	lgij' j	'an	'æbe	bl̄ o	bj' y	k'aj	tsop e	tsatan i	tsamin' œj	tsamij' i'z	tsav œ
5	ALAL 12	ab'ejjo	ap̄ i	njel' o	agij' j	'an	'æbr	bl̄ o	b' y / b' œ	k'ɛj	tsop e	tsatan' i	tsamin' œb	tsam' i'z	tsav œ
6	ALAL 13		ap̄ o	'niil a	g yj	'on	'æbr	bl̄ o	b œ	k'œj	tsop i	tsatop' i*	famin' œd	fam' i'z	fav' o
7	ALAL 14		ap̄ o		g yj	'on	'œbe*	bl̄ o	b' œ* / b' œ	k'œj	tsap ej / tsœj	tsatan' i*	tsamin' ad	tsam' i'z	tsav o
8	ALAL 15	b'it*	ap̄ e	niil' o	g yj	'an	'æbe*	bl̄ o	b' œ* / b' œ	k'œj	tsap ej	tsatan' i*	tsamin' ad	tsam' i'z	tsav œ
9	ALAL 16	ab'ïks	ap̄ e	apil' o	g yj	'az	'æbe*	bl̄ o	b' œ* / b' œ	k'œj	tsap e	tsatan' e	tsamin' ad	tsam' i'z	tsav a*
10	ALAL 17	abœk' o	ap̄ e	opil' o	og yAo	'æze	'æbei	bl̄ o	b' œ*	k'æAo	tsop e	tsastan' ej	tsamin' ad	tsom' i'z	tsav aj
11	ALAL 18	ab'ej	ap̄ e	'niil a	ag yj	'az	'æbe*	bl̄ o	b' œ* / b' œ	k'œj	tsap e	tsastan' e	tsamin' ad / tsam' i'z	tsav œ	tsav œ
12	ALAL 19	ab'ïks	ap̄ e	niil' o	g yj	'an	'æbe*	bl̄ o	b' œ* / b' œ	k'œj	tsap e	tsatan' e	tsamin' ad	tsam' i'z	tsav œ
13	ALAL 2	b'œAo	ap̄ e	apcl' o	g yj	'æn	'æbe*	bl̄ o	bj' œ	k'œj	tsap e	tsatan' e	tsamin' œj / t'famin' o	tsav œ	tsav œ
14	ALAL 20		ap̄ e	ap'el' a	g yj	'an	'æbe*	bl̄ a	bj' o	k'æAo	tsap e	tsatan' ie	famin' ad	fam' i'z	fav' œ
15	ALAL 21		ap̄ o		ag yj	'on	'abr*	b'æ a	b'œ o / b'œ' o	k'œj	tsap i*	tsatan' i*	famin' ad	fam' i'z	fav' o
16	ALAL 22	ab'j' œ	ap̄ o		ig yj	'on	'obr*	bj' a	bj' o	k'œj	tsap i*	tsatan' i*	famin' ad	fam' i'z	fav' o
17	ALAL 23	ab'œj	ap̄ œ		ag yj	'on	'æbe' / 'æbe	bl̄ a	b' œ	k'aj	tsap œ	tsatan' e	fmin' ad / fam' œm' i'z	tsav o / fv' o	tsav o
18	ALAL 24				y yj	'on	'obr*	bj' a	b' œ	k'œj	tsap i*	tsatan' i*	famin' ad	fam' i'z	fav' o
19	ALAL 25	b'œAo	ap̄ œ	nel' o	ag yj	'on	'obr*	bl̄ o	b'œ' o / bj' œ	k'œj	tsap i*	tsatan' i*	famin' ad	fam' i'z	fav' œ
20	ALAL 26	ab'ijj	ap̄ e	apil' o	ag yj	'æz'	'æbe*	bl̄ o	b' œ*	k'œj	tsap e	tsastan' e	famin' ad / s:am' i'z	tsav o	tsav o
21	ALAL 27	ab'ijj	ap̄ e	apil' o	ag yj	'œz'	'æbe*	bl̄ o	bj' o	k'œj	tsap e	tsastan' e	samin' ad	sam' i'z	tsav o
22	ALAL 28	ab'œAo	ap̄ e	niil' o	g yj	'æz'	'æbeœ	bl̄ o	bj' œ	k'æAo	tsop e	tsastan' e	tsamin' ad	tsam' i'z	tsav œ
23	ALAL 29	b'ïœ	ap̄ el	apil' a	ig yj	'æzj	'æbe	bl̄ a	b' œ	k'æAo	tsap e	tsastan' e	tsamin' adœ	tsam' i'z	tsav al
24	ALAL 3	b'ejj	ap̄ e		g yj	'an	'æbe	bl̄ o	bjo*	k'œj	tsap e	tsastan' e	famin' œj / fam' i'z	tsav o	tsav o
25	ALAL 30	ab'ïœ	ap̄ el	apil' a	eg yj	'æzj	'æbre	bl̄ o	b' u	k'æAo	tsap el	tsastan' i	tsœmin' adœ	tsam' i'z	tsav al

Figure 1. Un fragment de la base de données THESOC d'un échantillon d'entrées (ALAL) retenu pour l'analyse des distances et regroupements selon le critère de la phonologie (consonantisme, vocalisme, structure syllabique, etc).

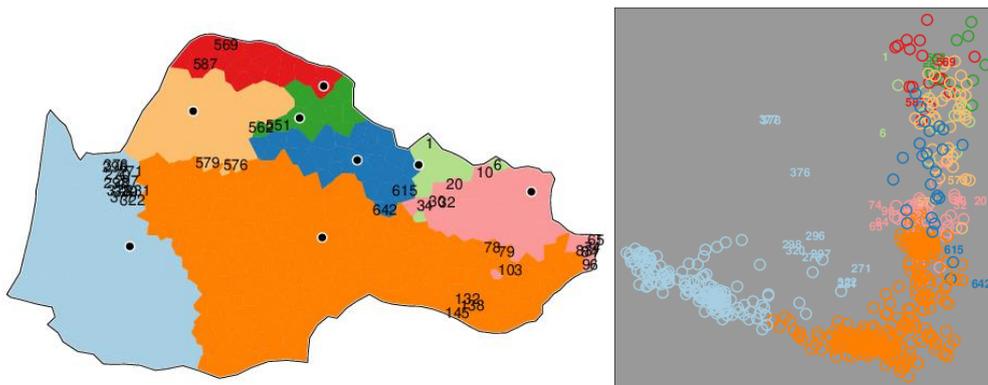


Carte 3a et dendrogramme : fonction *Complete Link* de Gabmap, données phonologiques, THESOC

Outre la convergence frappante entre les deux cartes – celle de P. Bec (carte 2 *supra*) et la nôtre (carte 3a), à l'aide de Gabmap – les deux topologies s'avèrent congruentes, sur un plan plus abstrait, d'ordre épistémologique, en vertu du principe élémentaire que *la carte n'est pas le territoire*. Nous avons là une topologie (la carte 3a), couplée à une taxinomie (le dendrogramme indexé de A à C). Le dendrogramme obtenu à l'aide de la fonction *Complete Link* de Gabmap apporte en outre de précieuses informations : le dendrème ou clade¹³ (ou *branche*, *cluster*) C – le gascon – est emboîté avec le grand clade B correspondant à « l'occitan méridional » de P. Bec qui, on l'a vu, regroupe provençal et languedocien, ou

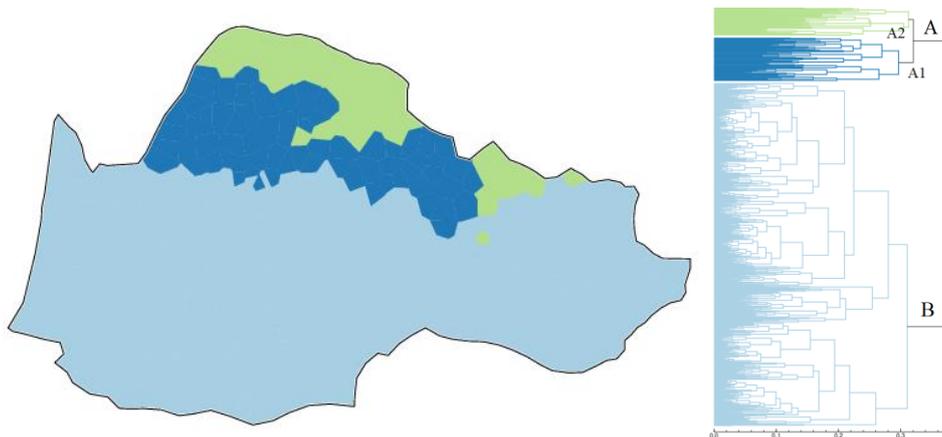
¹³ *Dendrème* est le terme utilisé par la dialectométrie salzbourgeoise, mais le terme de clade est plus commun dans la tradition taxinomique francophone, indépendamment de la cladistique à proprement parler et de son mode de calcul par « distances patristiques », dont il sera question ci-dessous. De même, dans la terminologie salzbourgeoise, les *chorème* correspondent aux aires dégagées par les cartes, correspondant aux *dendrèmes*, comme en 3a.

occitan central. Cet ensemble se subdivise en deux : B1 (languedocien, ou le noyau de l'occitan central) et B2 (le vivaro-alpin de P. Bec, que ce dernier classait cependant dans son ensemble « nord-occitan »). Le reste s'accorde avec la topologie et la taxinomie de P. Bec : la vaste, mais relativement étroite bande nord-occitane (tous les clades de l'ensemble A, sauf A3, valant pour le *Croissant*). Les entités limousine et auvergnate apparaissent clairement (respectivement, A1 d'une part, A2 et A5 d'autre part), même si l'intrication des structures internes locales est plus complexe qu'une simple congruence terme à terme, puisque A3 ou le *Croissant* est enchâssé dans un sous-ensemble de clades où c'est le nord-auvergnat (A5) qui tient lieu de membre externe, ou de clade *structurant*, selon la terminologie de Philps (1986). La fonction de validation de la cohésion des regroupements dans Gabmap (*Cluster Validation* sous *discrete clustering*, soit « regroupements pertinents », dans le menu) permet de vérifier que les données des points 569 (ALAL 52) et 587 (ALAL 71) à hauteur du clade A3 ont pu poser des problèmes de calculabilité, de même que les points auvergnats 562 (ALAL 45) et 551 (ALAL 34) (cf. carte 3b et sa topologie de positionnement scalaire ou MDS, en mode *Complete link*, sur 8 intervalles). La carte 3b permet de visualiser les autres zones litigieuses, sur le plan de la hiérarchisation par regroupements. On retiendra la leçon de ce « correctif », en gardant à l'esprit qu'une machine à clusters telle que Gabmap, comme toute mécanique taxinomique automatisée, donnera toujours des agglomérats, dont la robustesse doit être testée, en cas d'indécidabilité ou d'aberrations – comme c'est le cas pour ces exemples en 3b. Ce qui compte donc, au-delà de ces détails de robustesse de sous-zones spécifiques au sein de la structure générale, c'est la robustesse des grands regroupement, tels que A (nord-occitan) contre B (occitan méridional + vivaro-alpin) et C (gascon), en congruence avec la carte 2 de P. Bec. On rendra grâce à un logiciel en libre accès tel que Gabmap d'avoir prévu ses propres conditions de falsification, en donnant les moyens à l'utilisateur de détecter les zones de superposition et donc d'indécidabilité relative des regroupements qu'il propose.



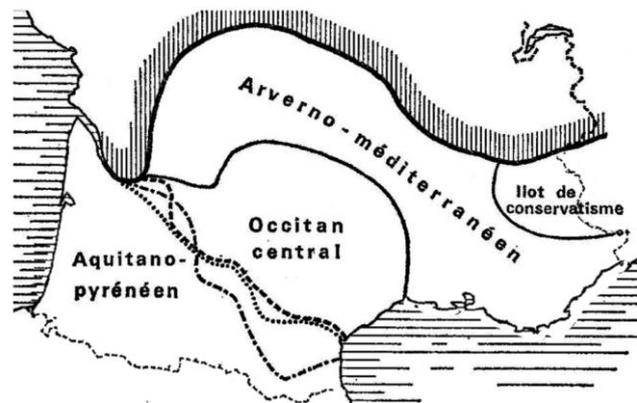
Carte 3b et positionnement multiscalaire pour falsification des résultats de la fonction *Complete Link* de Gabmap, données phonologiques, THESOC, 8 intervalles.

L'algorithme *Complete link* compte parmi les plus réducteurs (*lumping*) de la série de quatre fonctions de calcul proposées par Gabmap. Son pouvoir de rassemblement des aires dialectales permet de discerner d'emblée les grandes aires correspondant à un paramètre tel que la phonologie. La carte 3c et son dendrogramme rendent ainsi compte de la topologie par larges bandes de cet état des choses dialectométrique : l'occitan méridional unitaire dans la grande bande transversale au sud du domaine, qui rassemble la zone aquitaine (gasconne) et la zone orientale (provençale et vivaro-alpine), d'une part (clade B), et un nord-occitan d'autre part, qui se divise en deux bandes longitudinales : l'une pleine, au sud (clade A1), l'autre discontinue, au nord (clade A2).



Carte 3c et dendrogramme, 3 intervalles, fonction *Complete Link* de Gabmap, données phonologiques, THESOC

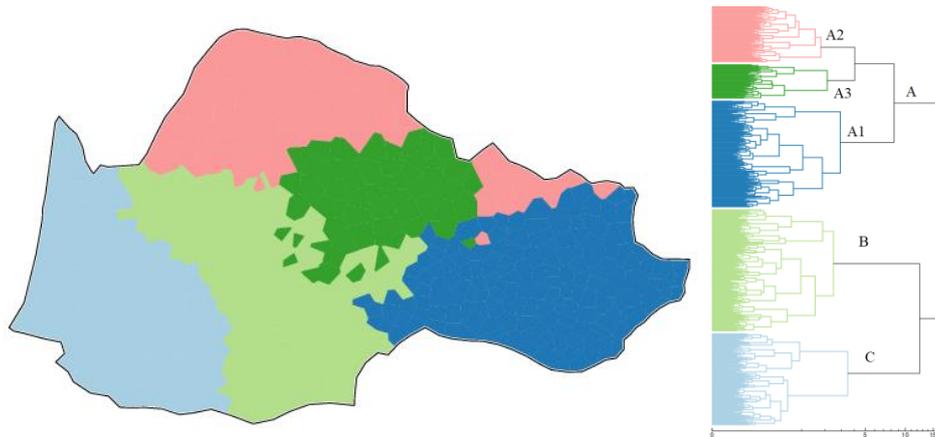
Il existe une deuxième représentation canonique – ou standard – de la structuration supradialectale du domaine occitan par P. Bec (1973, 38), qui opère une quadripartition, entre une bande orientée nord-ouest/sud-est, que l'auteur dénomme « arverno-méditerranéen », et un bloc composite comprenant « l'occitan central » et « l'aquitano-pyrénéen » au sud-ouest. Cet ensemble massif est flanqué, sur sa périphérie nord-orientale d'un « ilot de conservatisme » (*sic*) : cf. carte 4.



Carte 4. Quadripartition supradialectale du domaine occitan selon P. Bec (1973, 38)¹⁴

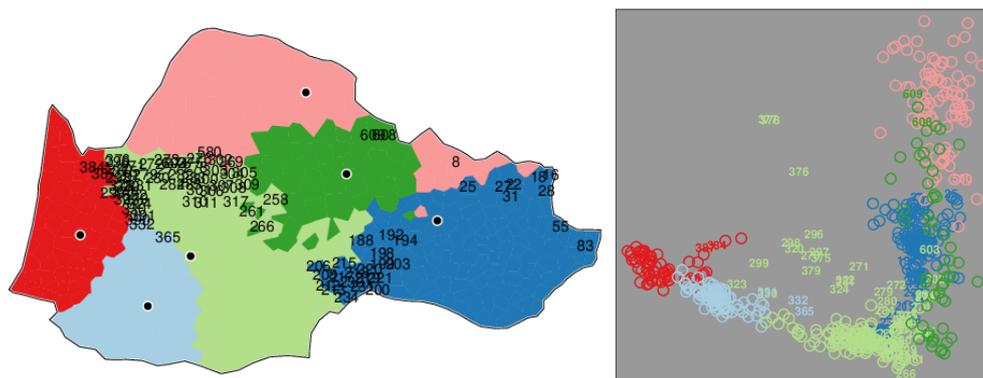
Cet état des choses qualitatif trouve sa correspondance quantitative dans l'application de la *Méthode de Ward*, ou *Classification ascendante hiérarchique* (CAH) sur Gabmap, comme dans la carte 5a ci-dessous, rendant compte, en retenant le même paramétrage que P. Bec (la phonologie), d'un relief dialectal plus accidenté (*splitting*) :

¹⁴ Il importe de donner ici le détail des critères sur lesquels P. Bec fonde ces grandes divisions du domaine occitan : la ligne épaisse continue qui sépare l'arverno-méditerranée de l'occitan central est celle du *bétacisme* (traitement de la labiale comme approximante ou comme obstruante) : *vi(n)* vs *bi(n)* 'vin'. Le bloc centre-occidental de l'aquitano-pyrénéen se subdivise, quant à lui – toujours selon cet auteur – en aquitano-pyrénéen à proprement parler, au sud-ouest (le gascon) vs l'occitan central, sur la base des critères suivants : palatalisation primaire *zur* vs *dzur* 'jour' (ligne médiane, en pointillés), palatalisation conditionnée ou secondaire du premier degré *fajta* vs *fajfa* 'faite' (ligne discontinue épaisse), et du second degré *fejft* / *fajft* (*fajf*) *idem* (ligne discontinue en points et tirets, plus au sud). Les résultats du traitement quantitatif confirment en partie ces macro-divisions (cf. carte 5c, par la méthode des probabilités), quoiqu'avec des tracés en partie différents, car les zones floutées ne descendent pas jusqu'à la côte méditerranéenne dans nos relevés, et la zone du tracé en ligne claire de Bec par le simple *bétacisme* correspond à des amas complexes et composites de poches interférentielles, dans cette même carte 5b *infra*. La dialectométrie confirme donc les macro-aires de la dialectologie traditionnelle, ou qualitative, mais en fonction de la loi des grands nombres, au lieu de se fonder sur des « lois phonétiques » en lignes claires. En cela, elles sont d'autant plus heuristiques, car elles incitent à rechercher une multiplicité de variables plus ou moins interdépendantes, au lieu de segments territoriaux *ad hoc*, aisément corrélables à des facteurs externes connus.



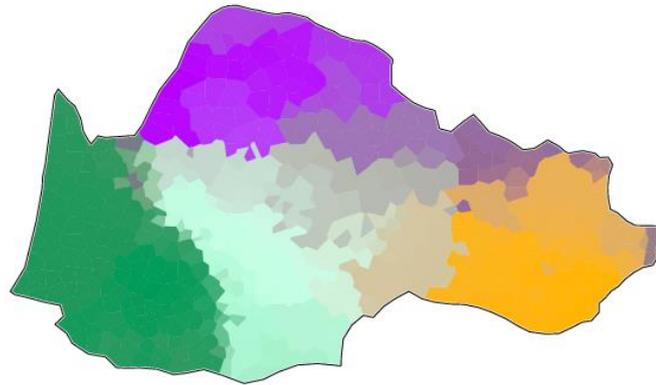
Carte 5a. Méthode de Ward (CAH) appliquée aux données phonologiques du THESOC : cinq intervalles.

Là encore, la congruence avec l'état des choses qualitatif postulé par P. Bec est satisfaisant – conditions *sine qua non* pour tester la validité de méthodes innovantes ou alternatives, en l'occurrence, quantitatives. L'ensemble A (ou bloc A) avec ses agrégats A1 (correspondant au provençal) d'une part et A2 (une frange discontinue nord-occitane, regroupant *Croissant* + limousin, ainsi qu'un fragment du vivaro-alpin, valant ici pour un fragment de « l'îlot de conservatisme » de P. Bec), associé au clade A3 (l'auvergnat), d'autre part, correspond au grand ensemble « arverno-méditerranéen » de P. Bec et à son « îlot de conservatisme » au nord-est, tandis que les clades B et C correspondent respectivement à « l'occitan central » et à « l'aquitano-pyrénéen » à proprement parler de P. Bec. Mais si l'on devait se contenter uniquement d'une simple congruence, nous resterions sur un simple constat, sans apport de nouvelles connaissances, et sans questionnement de la structuration du « fait dialectal » occitan à grande échelle. Il y a plus, et mieux : non seulement le traitement dialectométrique nous fournit une véritable taxinomie, avec une structure interne explicite, une hiérarchisation des aires et des regroupements ou classes de dialectes et sous-dialectes, telle que nous venons de la résumer, mais il nous signale également les zones grises d'indécidabilité, en fonction du corpus traité (cf. carte 5b de validation par positionnement multiscalair). On détecte ainsi cinq zones litigieuses, dont une correspond terme à terme avec l'embouchure de la Garonne, entre « aquitano-pyrénéen » (gascon) et « occitan central » (languedocien occidental).



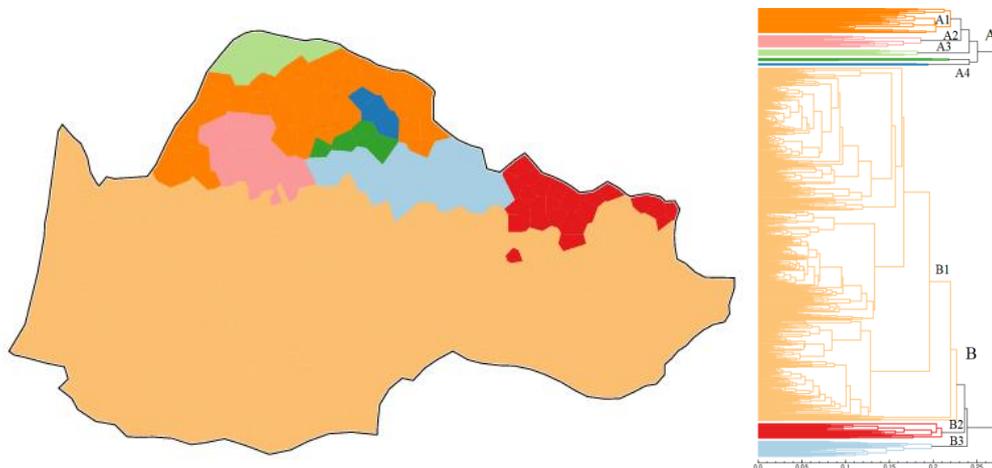
Carte 5b et positionnement multiscalaire pour falsification des résultats de la fonction *Méthode de Ward* de Gabmap, données phonologiques, THESOC, 5 intervalles.

Une autre zone grise se situe à la transition entre le bloc arverno-limousin, au nord-ouest, et le nord de « l'occitan central » selon P. Bec ; une autre encore apparaît au sud-est central, à la jonction entre le languedocien oriental et le provençal. Enfin, la zone, au demeurant définie de manière floue par l'étiquette « d'îlot de conservatisme » dans la carte 4 de P. Bec, est également parsemée de points superposables. Ce que la démarche qualitative finit par cacher, faute de pouvoir rendre compte de chaque détail, la méthode dialectométrique peut en rendre compte, en pointant vers un biais statistique, dans son monde quantitatif. La carte 5d rend explicitement compte des dynamiques de chevauchement et de saturation des aires, de manière particulièrement détaillée, à l'aide de la méthode des probabilités, faisant ressortir la « logique floue » des regroupements à la bordure (aux transitions) des aires dialectales. Elle permet d'accéder à une finesse de grain géolinguistique à laquelle les cartes à isoglosses, qualitatives, ne peuvent prétendre, puisque plus celles-ci multiplient les entrelacs, plus les cartes deviennent illisibles – aboutissant à un effet contreproductif, du point de vue de la vocation primordiale des artefacts cartographiques, qui est de faciliter la visualisation de phénomènes spatiaux complexes, aussi bien que de proposer une synthèse géographique interprétable, notamment à l'aide des divisions dialectales canoniques et de la géographie humaine (démographie, géohistoire, versants d'activités et d'échanges, etc.).



Carte 5c. Méthode des probabilités, ou « logique floue » des regroupements, application de Gabmap sur les données phonologiques du THESOC, 5 intervalles.

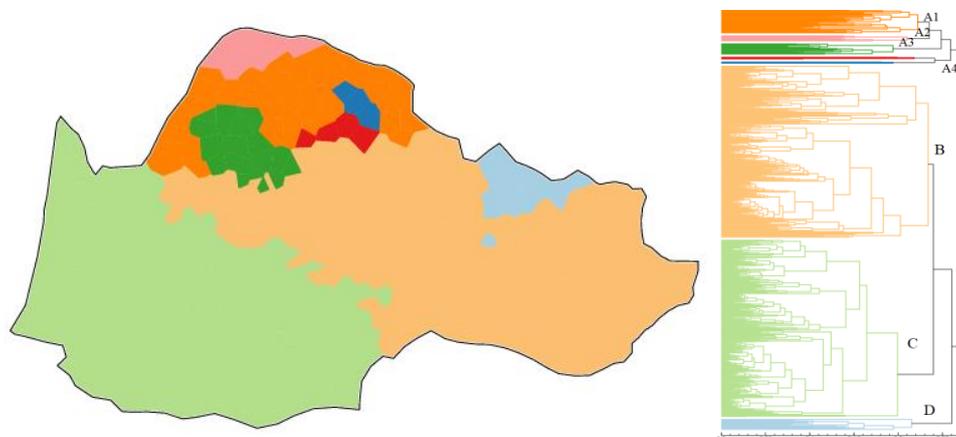
D'autres regroupements sont possibles, à l'aide des deux autres algorithmes offerts par Gabmap, que sont la moyenne de groupes (*group average*) d'une part, et la moyenne pondérée (*weighted average*), comme dans les cartes et dendrogrammes suivants. Les cartes 6a-b viennent apporter une autre focale (ou point de vue *vicariant*, autrement dit, interactif et complémentaire) aux précédents états des choses dialectométriques (Gabmap) et isoglottiques (P. Bec), selon une logique réductionniste (*lumping*).



Cartes 6a. Moyenne de groupes (non pondérée), phonologie, THESOC, 8 intervalles

La structure interne du dendrogramme de la carte 6a, avant pondération (qu'on trouvera en 6b), est davantage entrelacée : le *Croissant* (A3) revêt le rôle de clade structurant (position de cluster externe) en relation avec le limousin (clades A1-2), qui intègre deux variétés d'auvergnat (branches en A4). Dans cette configuration, c'est l'auvergnat méridional (clade B3) et l'ouest du vivaro-alpin (clade B2) qui s'assemblent à la périphérie du complexe « occitan méridional » (bloc B1), d'est en ouest – ce qui explique pourquoi le nord occitan s'assemble, selon ce point de vue

vicariant, avec le sud, sa périphérie sud-auvergnate confinant avec le vivaro-alpin occidental.



Carte 6b. Moyenne pondérée, **phonologie**, THESOC, 8 intervalles

La moyenne pondérée (carte 6b) lisse ces emboîtements, ou entrelacs : elle confirme la dichotomie entre un nord-occitan retranché en une masse compacte tout au nord (clade A et ses embranchements) contre le reste du domaine au centre et au sud. Là où cette masse centrale et transversale connaissait une tripartition (carte 6a), celle-ci se résorbe en une bipartition interne (trois sous-groupes de ce qui était B1 dans la carte en question) et dans son ensemble (B2-3 en 6a) ; dans la carte 6b, deux sous-ensembles se dégagent à l'est et à l'ouest (blocs B et C), valant pour une aire transversale sud-orientale et une aire aquitano-languedocienne, de manière congruente avec la carte 4, de P. Bec. Ce grand ensemble méridional est flanqué, au nord, d'un petit sous-ensemble (clade D, carte 6b), congruent avec un fragment du vivaro-alpin de la première carte de P. Bec, la carte 2 – mais qui se trouve trop à l'ouest de « l'îlot de conservatisme » de la deuxième carte du même auteur (la carte 4) pour prétendre se réclamer de cette entité périphérique de P. Bec. L'apport de la hiérarchisation des ensembles et des sous-ensembles par les aires de la carte 6b est indéniable, sans pour autant contredire les grandes tendances observées à l'aide de la méthode isoglottique. En outre, ces divers objets de connaissance militent tous en faveur d'une interprétation, sur le plan systémique : autant le sud du domaine tend à être relativement unitaire – à converger, par unité dans la diversité, d'est en ouest –, autant le nord semble bien plus hétéroclite, composite. Mais, loin d'être irréductible à un modèle explicatif ou à une caractérisation externe (histoire, géographie, etc.), deux tropismes se dégagent de cette diversité interne : d'une part, au nord-ouest, une chaîne de sous-dialectes de type limousin, dont les parlers du *Croissant* constituent une zone de transition au nord, au contact avec les parlers d'oïl, d'autre part, une chaîne de sous-dialectes de type auvergnat et vivaro-alpin occidental, orientés vers le sud-est (provençal, couloir rhodanien).

Voilà donc ce que l'on peut dire d'un *système fermé*, comme la phonologie et, de manière générale, étant donné la dimension réalisationnelle des données traitées par Gabmap (des variables phonologiques, calculées par alignements ou distance d'édition, à l'échelle de lexèmes, et non selon la logique des « cartes-types », par caractères abstraits). Les deux cartes canoniques de P. Bec (cartes 2 et 4 *supra*) voient certes confirmées leurs entités dialectales à plusieurs niveaux (macro- ou supradialectal et dialectal), mais considérablement amendées et précisées dans le détail de la structure interne des relations systémiques entre ces entités.

Dans la prochaine section, nous allons appliquer Gabmap au lexique, qui se définit comme un *système ouvert*, à la différence du paramètre phonologique : alors que l'inventaire des phonèmes d'une langue ou d'une variété dialectale est par définition limité, le nombre de mots (ou *lexèmes*, *lexies*) composant le vocabulaire n'est pas sujet aux mêmes restrictions (par exemple, le nombre de mots dans une œuvre comme *Le tiers livre*, de François Rabelais, s'élève à pas moins de 11 362 items¹⁵, alors que l'inventaire phonémique du moyen français aussi bien que du français moderne ne dépasse guère 36 items au total, dont moins d'une vingtaine de consonnes). Ce paramètre de la diversité lexicale nous permettra dans ce qui va suivre de redimensionner les aires, les configurations ethno-/géolinguistiques, et de poursuivre l'exploitation des données du THESOC à l'aide de Gabmap. Nous accorderons une particulière attention, une fois de plus, à la congruence avec les aires canoniques de la dialectologie, et apporterons, là encore, des résultats taxinomiques heuristiques – une hiérarchisation de la structure interne du diasystème, au lieu de simples configurations aréologiques issues de l'agencement d'isoglosses.

3.2. Aires lexicales

Les données lexicales (lexies différenciées) sont traitées d'une manière différente par Gabmap, en tant que mots-types, à l'aide d'une fonction de traitement « catégoriel », au lieu d'un calcul sur des chaînes de caractères, comme précédemment, pour la phonologie. Nous donnons à voir ci-dessous un extrait de quelques items retenus, pour l'ALAL¹⁶.

¹⁵ Obtenu par <http://ancilla.unice.fr/pages/bases/#>.

¹⁶ Les entrées de la base de données sont les suivantes : *abeille, aiguille, aiguiser (une faux), allumer, âne, ânesse, anse, araignée, araire, arbre, armoire, aubépine, avoine, balai, balle de céréales, battoir, beaucoup, bêler, belette, bélier, betterave, beugler, blaireau (animal), blé, bœuf, bois de chauffage, borne, bouc, bouche, boue, brebis, buis, caillé, caille (fem.), canard, cane, carotte, cerise, chaintre, chaise, chapeau, chardon, châtaignier, chatière, chaudron, chauffe-lit, chauve-souris, cheminée, chemise, chêne, chenille, cheval, cheveu, chèvre, chien, chiendent, chiffon, chiffonnier, chou, ciel, ciseaux, clef, coin à fendre le bois, coq, corbeau, corne, cou, coudre, couteau, couver, crapaud, crémaillère, cuillère, cul, cuvier, dard, dé (à coudre), descendre, dindon, drap, eau, eau de lessive, échelle, éclair, écorce, écureuil, épingle, escalier, étable, éteule, étoile, évier, fagot, faucher, faucille, faux (outil), fendre (du bois), fenêtre, fermer, feu, feuille, fiel, foie de porc, fontaine, fourche, fourmi, fumier, genévrier, gerbe, gland du chêne, gond, gousse d'ail, grenouille,*

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		abeille	aiguiser (une)	âne	anesse	anse	araignée	arbre	avoine	bêler	belette	bélier	betterave
2	ALAL 1	abelha	esmolar	ase		coana	aranhada	arbre	cívada	belar	mostèla	moton	betaraba
3	ALAL 10	belha	filar	ase	ânesse	ansa	aranha	arbre	cívada	bialar	beleta	bedan	carròta
4	ALAL 11	abelha	filar	âne	mitra	ansa	aranha	arbre	avena	belar	beleta	aret	carròta
5	ALAL 12	abelha	gusar	âne	mitra	ansa	aranha	arbre	avena	bialar	beleta	aret	carròta
6	ALAL 13	mosca	agusar	âne	ânesse	ansa	aranhada	arbre	avena	bialar	beleta	moton	betaraba
7	ALAL 14	mosca	afilar	âne	mitra	ansa	aranhada	arbre	avena	belar	beleta	bélier	betaraba
8	ALAL 15	belha	filar	âne	mitra	ansa	aranhada	arbre	avena	belar	beleta	moton	carròta
9	ALAL 16	abelha	repassar	ase	sauma	ansa	aranhada	arbre	cívada	bialar	beleta	aret	carròta
10	ALAL 17	abelha	afilar	ase	sauma	ansa	aranhada	arbre	cívada	bialar	mostèla	moton	betterave
11	ALAL 18	abelha	filar	ase	sauma	ansa	aranhada	arbre	cívada	belar	beleta	moton	betaraba
12	ALAL 19	abelha	filar	âne	baude	ansa	aranhada	arbre	cívada	bialar	beleta	aret	betaraba
13	ALAL 2	belha	esmolar	âne	mitra	ansa	aranhada	arbre	cívada	bialar	mostèla		betaraba
14	ALAL 20	mosca	filar	âne	baude	ansa	aranhada	arbre	avena	bialar	beleta	aret	betaraba
15	ALAL 21	mosca	filar	âne	baude	ansa	aranhada	arbre	avena		beleta	moton	betterave
16	ALAL 22	abelha	afilar	âne	baude	ansa	aranhada	arbre	avena	belar	beleta	bélier	betterave
17	ALAL 23	abelha	repassar	âne	borrica	ansa	verin	arbre	avena	belar	beleta	bélier	bleta
18	ALAL 25	belha	filar	âne	baude	ansa	aranhada	arbre	avena	bialar	beleta	bélier	betaraba
19	ALAL 26	abelha	filar	ase	sauma	ansa	aranhada	arbre	cívada	bialar	beleta	aret	betaraba
20	ALAL 27	abelha	filar	ase	sauma	ansa	aranhada	arbre	cívada	bialar	beleta	aret	betaraba
21	ALAL 28	abelha	afilar	ase	sauma	carba	aranhada	arbre	cívada	bialar	beleta	moton	betaraba
22	ALAL 29	belha	afilar	ase		carba	aranhada	arbre	cívada	bialar	bèlola	moton	bletaraba
23	ALAL 3	belha	filar	âne	mitra	ansa	aranhada	arbre	cívada	mialar	beleta	moton	carròta
24	ALAL 30	abelha	repassar	ase	sauma	carba	aranhada	arbre	cívada	bialar	bèlola	aret	bledaraba
25	ALAL 31	belha	filar	ase	sauma	gansa	aranhada	arbre	cívada	belar	beleta	aret	betaraba
26	ALAL 32	belha	afustar	âne	baude	ansa	aranha	arbre	avena	belar	beleta	bélier	betaraba
27	ALAL 33	belha	afilar	âne	borrica	ansa	aranha	arbre	avena	belar	beleta	moton	betterave

Figure 2. Extraits (ALAL) de la base de données THESOC concernant le lexique commun dans le NALF

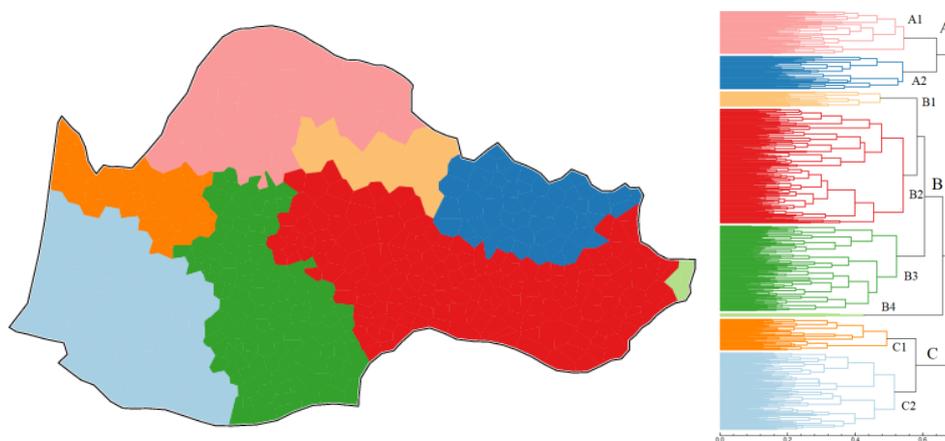
Au terme de leur étude critique des aires supradialectales occitanes, à l'aide d'une base de données similaire, Brun-Trigaud/Malfatto (2013) concluaient :

« Enfin, à la lumière des phénomènes que nous avons vus ici, on peut se poser la question de la légitimité des limites taxinomiques telles qu'elles sont établies habituellement. On le sait, la composante lexicale n'y est jamais prise en compte, alors qu'elle joue pourtant, nous semble-t-il, un rôle assez structurant dans l'intercompréhension. De notre point de vue, le lexique crée des affinités et des aires de cohésion entre les différents dialectes et peut modifier sensiblement l'architecture traditionnellement admise » (Brun-Trigaud/Malfatto 2013, 309).

« Création d'affinités » et « aires de cohésion » sont en effet les deux vertus heuristiques principales que nous allons retrouver constamment au fil de nos analyses de dialectologie quantitative. Ces effets de l'analyse quantitative du paramètre lexical permettent d'atténuer et de recadrer l'inévitable tentation de recourir à des définitions en termes de dialectes ou de sous-dialectes précis, dans une quête des corrélats externes. Qu'il soit donc clair d'emblée que, lorsque nous

griffer, grillon, guêpe, haie, hameau, heure, jabot, jambon, jardin, lait, lard, lessive, lèvres, lien de la gerbe, lièvre, lit, loquet, louche, lune, maison, manche de la faux, manger, marmite, mettre bas (brebis), miel, mite, moissonner, monter, mordre, mouche (insecte), mouchoir, mule, mulet (mammifère), mur, muselière, nager, neige, neiger, noisetier, noisette, nombril, noyer (arbre), nuage, œil, œuf de poule, oie, oignon, oiseau, omelette, orage, oreiller, ortie, oseille, ouvrir, peler, petit-lait, pétrin, pie (oiseau), pierre, pinson, pis (vache, brebis), pleuvoir, plier, poêle, poivre, pomme de terre, porc, poule, poumons (animal), poussière, pré, puce, puits, queue, radis, rat, rayon de miel, regarder, renard, repas de l'après-midi, repas de midi, repas du matin, repas du soir, ronce, rosée, ruche, ruisseau, s'asseoir, sabot, sanglier, scie, scier, se noyer, seau, seigle, semer, soleil, soufflet, suer, suie, taupe, taureau, tête, tonner, tonnerre, tourner (lait), traire, trayon, truie, tuile, vache, veau, vent, vent du nord, ver de terre, vermoulu, verrat, vessie de porc, vipère. Le bilan du traitement se décrit comme suit : 645 points d'enquêtes ou localités du NALF, 232 items, 144 025 comparaisons terme à terme.

aurons recours ici à de telles comparaisons, en termes d'entités dialectales connues, nous le ferons surtout afin de faciliter la lisibilité des résultats obtenus à l'aide de la méthode quantitative et de son éventail de solutions techniques – algorithmiques. Il faut se garder d'essentialiser les objets de connaissance produits à l'aide de ces outils – tout autant qu'à l'aide de la démarche isoglottique, puisqu'elle est indissociablement liée par des choix de variables, en amont, qui induisent ses résultats, conformément au paradoxe de l'observateur, particulièrement impliqué, en tant que démiurge d'aires, dans une telle pratique. Ici, le démiurge créateur d'aires est l'ordinateur, en fonction des intrants (*inputs*) dont on le nourrit pour réaliser ses opérations, aboutissant à des sorties (*outputs*) taxinomiques. Cette neutralité de la machine est une aubaine, qu'il serait dommage de gâcher en revenant aux réflexes classificatoires essentialistes, en quête perpétuelle d'explications par la géohistoire. Cependant, force est de reconnaître que ces *affinités émergentes* et ces aires de cohésion que font apparaître les algorithmes évoquent de manière tentante aussi bien des aires géopolitiques internes au tissu territorial du réseau dialectal occitan (géohistoire) que des versants liés aux flux d'activités, d'échanges et de populations (interactions des sociétés), dans l'étendue aussi vaste que diversifiée, du point de vue des milieux naturels et des condensations sociohistoriques potentielles, du domaine occitan.

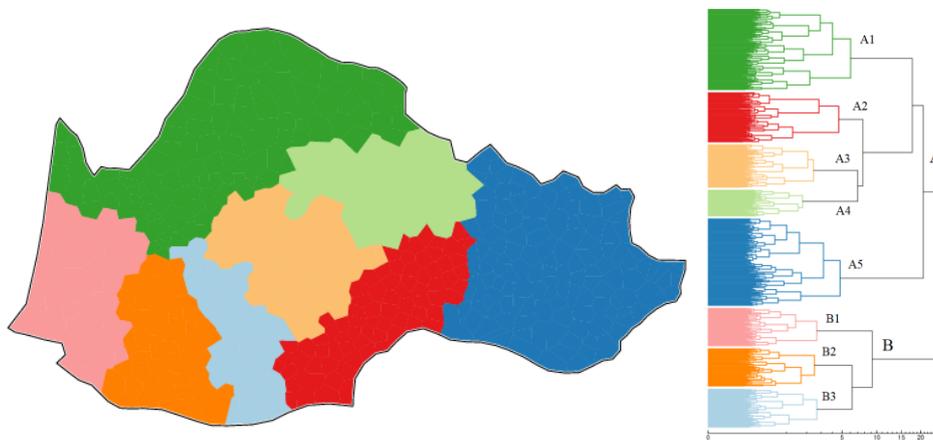


Carte 7. *Complete link* : données lexicales THESOC, 8 intervalles

Avec la carte 7, qui applique la méthode du « voisin de plus distant » (*complete link*), nous retrouvons la dichotomie nord-occitan (clade A) contre le reste du domaine (clades B-C), avec cependant des relations d'intrication plus accidentées que dans l'espace taxinomique précédent, qui relevait des systèmes fermés. Avec le paramètre du lexique, les blocs sont plus compacts, plus catégoriels, et s'avèrent davantage saillants, potentiellement liés à des aires culturelles (facteurs externes). Sur le plan des aires supradialectales, on retrouve l'aire « arverno-méditerranéenne » de P. Bec (cf. carte 4), dans l'agencement des clades B1-3, mais s'arrêtant au pied du bastion limousin que nous venons d'individuer au terme de

notre précédent parcours : le nord-auvergnat (B1) sert de marche, ou aire-tampon, au bloc massif du clade B2, et brise l'unité géographique du clade A, qui assemble limousin et vivaro-alpin. Cette aire transversale orientale B2 est adossée à une deuxième aire massive à l'ouest : celle du languedocien centre-occidental. Ce grand ensemble est associé, en position externe dans la configuration du dendrogramme, à une aire aquitaine (clade C), subdivisée en deux aires : l'une au sud (gascon central et pyrénéen : clade C2), l'autre au nord, entre Garonne et Dordogne, formant un crochet orienté vers le haut, qui rappelle l'ouest de l'ancienne province de la Guyenne (cf. Dartigue 1950), en continuité avec le Périgord (clade C1). Le clade B4 vaut pour la variété niçarde, à la périphérie du grand bloc centre-oriental. Nous venons d'éprouver le pouvoir fédérateur de l'algorithme *Complete link*, particulièrement heuristique pour suggérer des relations à grande distance, comme nous l'avons signalé déjà à deux reprises. Ce sont moins désormais les réseaux ethno-sociolinguistiques, corrélés aux aires phonologiques, que saisissait particulièrement bien l'approche précédente, à partir du paramètre phonologique, que nous saisissons ici avec le lexique, que des aires culturelles potentielles, bien qu'il faille se garder de toute conclusion hâtive.

Voyons ce que donne, à son tour, un algorithme moins fédérateur, relevant du *splitting*, en termes de saisie, mais considéré comme standard pour donner des subdivisions nettes des composantes taxinomiques d'un espace dialectal : la méthode hiérarchique ascendante de John Ward (CAH).



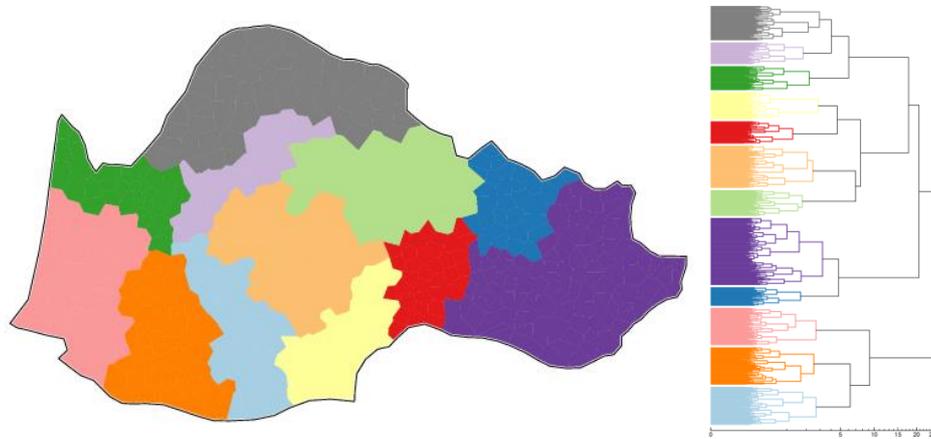
Carte 8a. *Classification ascendante hiérarchique (méthode de Ward) : données lexicales THESOC, 8 intervalles*

Dans cette carte 8a stratégique, car générée par la CAH – qui est l'algorithme jugé optimal pour révéler les divisions en dialectes, voire en sous-dialectes –, le monde se divise en deux (*dichotomie* entre le bloc A et le bloc B dans le dendrogramme), comme dans la carte 4, de P. Bec, qui faisait courir une ligne de division faite d'isoglosses enchevêtrées au sud-ouest, à hauteur de la Garonne. Cette ligne de démarcation sépare le bloc massif « arverno-méditerranéen » + « occitan

central » de « l'aquitano-pyrénéen », dont le cœur du gascon, à l'extrême sud-est (clade B1) joue le rôle d'entité structurante (rappelons que P. Bec avait pourtant fondé ces marco-divisions sur quelques isoglosses que nous avons détaillées en note, *supra*, alors qu'ici, c'est l'analyse du lexique qui en confirme les tracés *a posteriori*). De même, dans le macro-bloc centre-oriental, le provençal (clade A5) revêt ce rôle structurant, tout comme, au nord, ce que nous pouvons désormais appeler le guyenno-limousin (clade A1), gravite autour de trois aires centrales (A2-4). On voit d'emblée que ces aires forment des sortes de régions, que l'on pourra corrélérer, le cas échéant, à des périodes historiques, autrement dit, à des régions géopolitiques¹⁷ internes à l'espace dialectal occitan – les spécialistes des facteurs externes sauront mieux que nous y déceler des pistes de recherche. Les dialectes et sous-dialectes canoniques sont mieux représentés ici par le système ouvert qu'est le lexique que dans les précédents résultats fondés sur ce système fermé qu'est la phonologie : A5 correspond bien au provençal, en intégrant le vivaro-alpin (le clade en question est d'ailleurs bifide : la bipartition interne oppose le dauphinois, au nord-ouest, au reste de la Provence), A2 correspond au languedocien méditerranéen plutôt qu'au languedocien central ou occidental, A3 au rouergat, A4 à l'auvergnat, B3 au pays de Toulouse, etc. La teneur heuristique, en termes de facteurs externes (*déterminisme*), de cette taxinomie à huit niveaux semble donc élevée, même si l'on ne peut qu'enjoindre le lecteur à la prudence : l'aire de B2 par exemple, au sud-ouest, doit davantage s'interpréter comme déterminée par une dynamique de versant de l'Adour, ou comme aire de transition avec le languedocien occidental, que comme un effet de l'incidence du béarnais, qu'une subdivision en douze intervalles ne fait pas apparaître, alors qu'à ce niveau l'opposition entre les binômes Montpellier-Nîmes et Béziers-Narbonne émergent nettement (carte 8b *infra*). Ce niveau de grain fin, à douze intervalles, atteint l'extrême limite de ce qu'il convient d'expérimenter en taxinomie dialectométrique, mais nous en retenons ici les résultats à des fins heuristiques, étant donné le potentiel que recèle la CAH pour la délimitation des dialectes, mais aussi des sous-dialectes et des regroupements plus ou moins discrets de parlers. Les travaux de Louis Balmayer (1971), sur les isoglosses montpelliéraines, ainsi que les relevés isoglottiques très précis, complétés par des analyses dialectométriques à l'aide de la méthode globale de

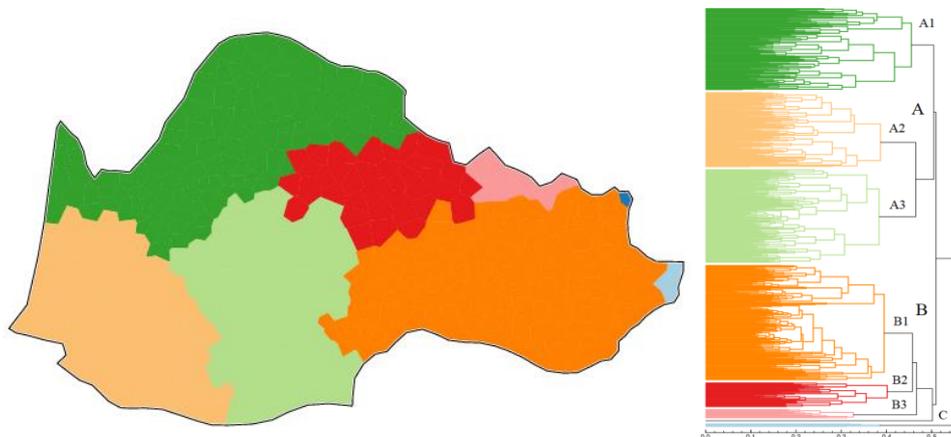
¹⁷ Nous entendons par « géopolitique » ici les anciennes provinces, comtés, etc. qui ont constitué, au Moyen-âge et, dans une moindre mesure, à la Renaissance, l'espace occitan, durant sa période de pleine vitalité. Voir par exemple l'étude de cas de Balmayer (1971), sur la corrélation avérée entre les limites des seigneuries de Montpellier et d'Aumelas, possessions de la dynastie des Guilhem, et les traits caractéristiques de l'occitan montpelliérain et de son arrière-pays, dans un cercle Sète-Pézenas-Gignac-Prades-Montpellier-Pérols, en lien avec l'hégémonie aragonaise, de 1204 à 1349 (Balmayer 1971, 92). Les corrélations externes en fonction de sous-limites géohistoriques de cette dimension ne sont certes pas systématiques en géolinguistique, mais elles ne sont pas exceptionnelles pour autant, selon la granularité des instruments de mesure qu'on applique sur un corpus donné.

Guiter proposés par Christian Camps (1991) confirment la pertinence de ce niveau d'analyse, à des échelles encore plus réduites – nous reviendrons prochainement sur le caractère fractal des micro-aires en domaine d'oc.



Carte 8b. Classification ascendante hiérarchique (méthode de Ward) : données lexicales THESOC, 12 intervalles

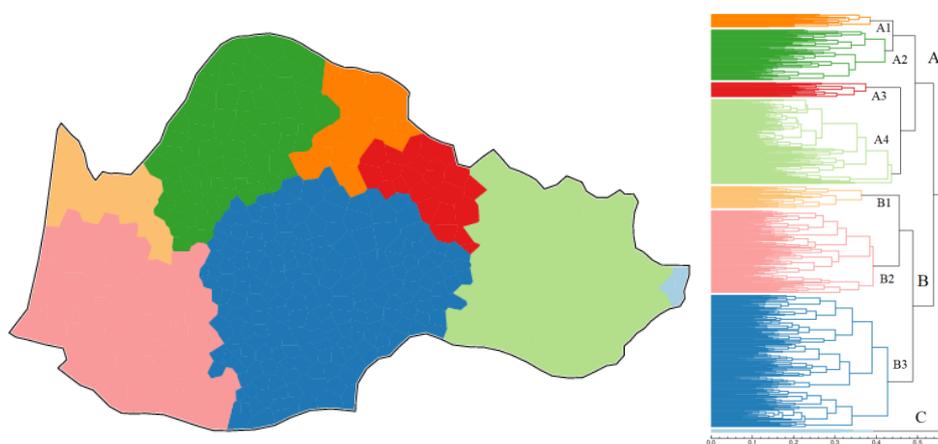
Que nous disent, pour le paramètre du lexique, les deux algorithmes fédérateurs que sont la *moyenne de groupes* et la *moyenne pondérée* ?



Carte 9a. Moyenne de groupes : données lexicales THESOC, 8 intervalles

L'algorithme de la *moyenne de groupes* non pondérée redimensionne les configurations d'une manière radicale, en renforçant les dynamiques que nous avons observées initialement, avec la méthode du *voisin le plus éloigné* (*complete link*). Le monde se divise bien en deux, entre un bloc occidental (bloc A) et un bloc oriental (bloc B) une fois de plus, mais ce sont ici deux masses de poids analogue, avec des structures internes de proportions différentes : le bloc A rassemble le gascon (clade A2) et le languedocien central (clade A3) ; les deux aires sont surmontées, au nord, d'une grande aire guyenno-limousine (clade A1) – le crochet prend ici la forme d'une longue spatule évasée, sur le plan graphique. À cet

ensemble de blocs de taille relativement analogues s’oppose, à l’est, un ensemble plus disproportionné : le clade B1 (provenço-cévennois ou provençal et le secteur le plus oriental du languedocien) d’une part et le clade B2 (auvergnat) d’autre part, de dimensions modestes. Une frange de dauphinois atteste d’un noyau vivaro-alpin (clade B3, en position externe, structurante, selon les termes de Philps 1986). Enfin, un clade C regroupe, à l’extrémité orientale, deux parlers de contact, éminemment périphériques : le briançonnais (ALP 20) et l’arrière-pays niçois (ALP 76, 86, 95, 111) – ces derniers, de type gallo-italique¹⁸. Cette configuration n’est jamais qu’une variante de la carte 4, associant « l’aquitano-pyrénéen » de P. Bec et « l’occitan central », en incorporant la partie nord de « l’arverno-méditerranéen ». On voit à quel point les deux configurations supradialectales proposées par P. Bec restent heuristiques, quel que soit le paramètre étudié.



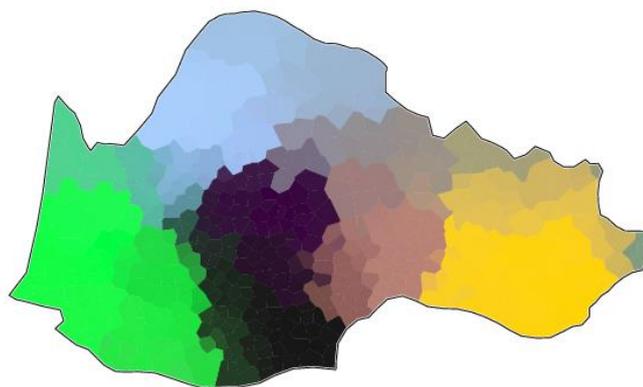
Carte 9b. *Moyenne pondérée* : données lexicales THESOC, 8 intervalles

Nous venons de remarquer que l’aire guyenno-limousine était rattachée au bloc occidental, dans l’état dialectométrique de la carte 9a, par *moyenne de groupe* non pondérée (*group average*). Mais on aurait tort de penser qu’il s’agit, dans l’abstraction bien entendu, d’une « annexion » de la cloche nord-occitane par le bloc occidental de la langue d’oc : tout au plus, s’agit-il d’une surface dialectométrique parmi d’autres. La *moyenne de groupes pondérée* propose une configuration asymétrique sur ce point. Un bloc A oriental (auvergnat occidental : clade A1 et limousin en tant que clade A2, associés au provençal : clade A4 et à l’auvergnat oriental en qualité de clade A3) reproduit la bande transversale orientée Nord-Ouest-Sud-est de l’arverno-méditerranéen de P. Bec (carte 4), de manière

¹⁸ Un relecteur nous fait remarquer à bon escient que « 76 (La Brigue) [et] (86) Saorge sont des parlers liguriens, mais le parler de Molinet 95 est un parler alpin appartenant à l’aire mentonasque ; le mentonasque 111 est généralement défini comme un parler de transition entre occitan et ligurien, à dominante occitane alpine, cf. Notamment : Revest (2009, 2011) et Dalbera (1995). »

parfaitement congruente, tandis qu'un bloc B occidental réunit trois clades semblables à ceux déjà observés : à l'ouest, la pointe médocaine de la Guyenne historique (clade B1) et à son sud, le gascon (clade B2), formant ainsi l'aire « aquitano-pyrénéenne » de P. Bec ; au centre du domaine, « l'occitan central » dans toute son extension (clade B3). Il y a convergence avec quasiment toutes les limites esquissées par P. Bec dans la carte 4 *supra*, hormis pour son « ilot de conservatisme » à l'est du vivaro-alpin – ce qui contribue, carte après carte et dendrogramme après dendrogramme, à invalider cette aire « conservatrice ». Le clade C dans l'arrière-pays niçois est cette fois positionné à l'extérieur du clade majeur qui regroupe les blocs A et B, confirmant le caractère allogène (ligure, donc gallo-italique) du clade C. Les vertus heuristiques de ce dernier algorithme, en tant que test de *lumping* des agrégats, sont donc confirmées, au point d'optimiser les questions de frontières de langues (ou *liminalité* génétique, sur le plan linguistique).

Nous avons évoqué et illustré avec l'étude du paramètre phonologique le pouvoir heuristique remarquable des cartes dialectométriques à continuum, de type « logique floue », que l'application Gabmap permet de générer. Voyons maintenant ce que cette approche donne pour le paramètre lexical (carte 10). Cette représentation affine considérablement les divisions, rendant compte des paliers entre les aires que nous venons d'examiner. Le floutage et la superposition « écaillée » des aires fait en particulier ressortir le rayonnement ou la *résonance transitionnelle*, en quelque sorte, de l'aire du « languedocien central », qui se divise d'ailleurs en deux, dans sa partie orientale, au contact avec le provençal – lui-même très chargé en transitions dans sa partie nord, celle du vivaro-alpin, qui se fond progressivement dans l'auvergnat sud-oriental.



Carte 10. Regroupement stochastique par positionnement multiscalaire : lexique THESOC

Il reste à rendre compte de quelques éléments techniques quant à la distribution des *taxats*, ou *types*, ou *traits dialectaux*, dans les deux bases de données THESOC

analysées à l'aide de Gabmap. La figure 3 montre les profils relativement asymétriques des deux séries de données : le paramètre phonologique, à gauche, présente, comme on peut l'attendre d'un système fermé, une plus grande densité de faible éventail de variantes, tandis que le paramètre du lexique (système ouvert) est plus diversifié. On remarque également dans la relation entre la distance géographique et la distance dialectale – vieille question qui préoccupait grandement Jean Séguy aux débuts de la dialectométrie (Séguy 1971), à laquelle Gabmap a prévu de donner une réponse algorithmique – une plus forte corrélation entre distance géométrique et distance linguistique, en ce qui concerne le paramètre lexical – ce qui conforte son potentiel diagnostic des entités *lectales* (*dialectales, sous-dialectales*)¹⁹.

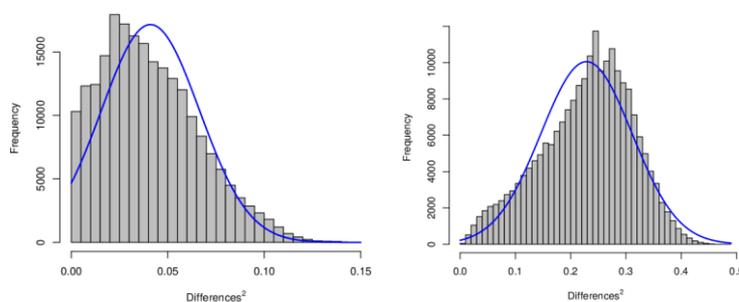


Figure 3. Structure des deux bases de données issues du THESOC : à gauche, paramètre phonologique ; à droite, paramètre lexical.

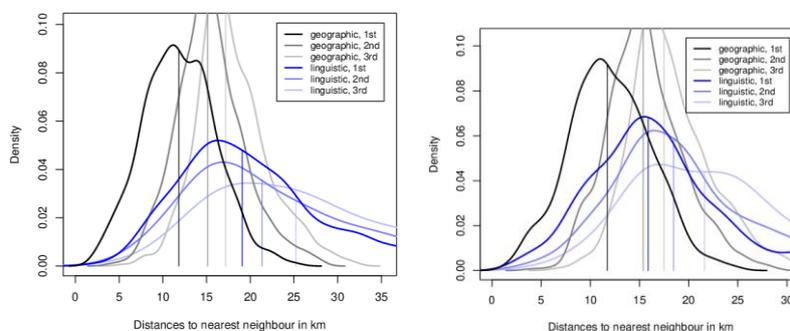


Figure 4. Relation entre la distance géographique et la distance dialectale (par voisin le plus proche) : à gauche, paramètre phonologique, à droite, paramètre lexical.

¹⁹ Cette question de la relation entre distance linguistique, de manière générale, et distance géographique, a été intégrée dans la version la plus récente de Gabmap, qui permet même de pondérer la diffusion de la variation dans un espace géolectal, en indexant le poids démographique des villes dans la géographie humaine d'un réseau dialectal, cf. Nerbonne *et al.* (2005) pour un rapport détaillé sur la dimension exploratoire et méthodologique de cette recherche. Ici, nous nous sommes contentés dans la figure 4 des résultats triviaux, non pondérés.

Au terme de ce survol des objets de connaissance générés par Gabmap à partir des données du THESOC, on constate la validité, dans les grandes lignes, des deux configurations supradialectales proposées en son temps par P. Bec. Mais la gamme algorithmique de Gabmap, avec ses modalités de calcul, dont trois relevant du *lumping* (approche réductionniste : *complete link*, *moyenne de groupes*, *pondérée* et *non pondérée*), et un autre permettant le *splitting* (approche détailliste, ou componentielle : *méthode de Ward*), agit à la façon d'un kaléidoscope, apportant de l'épaisseur et de la vicariance à ces structures robustes. On peut certes y lire en filigrane l'histoire des réseaux ethno/sociolinguistiques occitans, à travers le paramétrage phonologique, mais aussi les tendances à l'individuation ou émergence ainsi qu'au rayonnement d'aires culturelles, voire géopolitiques, dans la longue durée, à travers le paramètre du lexique. Cependant, il convient de rester prudent quant à l'interprétation des résultats. On ne gagne rien à chercher confirmation de l'existence intangible de son propre dialecte contre tous les autres – on a tout à gagner, au contraire, à comprendre l'intrication structurale et, à l'occasion, géohistorique, des réseaux et des aires qui ont façonné le domaine occitan au cours des âges, selon la logique de la *tabula rasa*, cher aussi bien à Théobald Lalanne qu'à Jean Séguy.

4. Étude de cas de modélisations et de visualisations vicariantes

4.1. Dialectométrie *versus* Cladistique

Alors que la dialectométrie salzbourgeoise utilise des cartes-types comme intrants et que GabMap calcule les alignements, absences, ajouts et substitutions de segments un à un, sans les hiérarchiser, ni tenir compte de ce qu'on sait des changements phonétiques²⁰, selon un traitement au niveau des gabarits lexicaux, par distance d'édition, et alors que la pondération des variables reste le parent pauvre des opérations de calcul, une autre méthode est possible. Elle a fait ses preuves dans des domaines aussi solides que la génétique des populations – nous voulons parler de la *cladistique*. Son mode opératoire se caractérise par les distances patristiques (états dérivés dans un graphe de génération de *caractères*, ou *traits évolutifs*, ici *traits dialectaux* : A dérivant B, qui dérive C et D, etc.). Puisque le changement phonétique (autrement dit, le paramètre phonologique) détermine puissamment la diversification dialectale, il va de soi qu'il serait souhaitable de pouvoir faire dériver les états diachroniques successifs les uns des autres, dans un corpus de données dialectales tel que le THESOC. La dialectométrie dans sa première

²⁰ Cette remarque est moins vraie depuis les nouvelles moutures du logiciel en ligne : les concepteurs ont prévu de tenir compte de l'allophonie segmentale (comptant par exemple le score de 0,5 au lieu de 1), dans les versions récentes. Si la cladistique permet une pondération plus sophistiquée, il n'en reste pas moins qu'elle se fait encore à la main, de manière non automatisée, sur des critères relevant de la démarche qualitative du linguiste, dans la construction des graphes de dérivation diachronique, comme dans la figure 5 *infra*.

mouture (Séguy 1971, 1973a-b) ne permettait pas de le faire, et la dialectométrie salburgeoise a depuis innové sur ce plan, mais par des fenêtres de calcul par rapport à la médiane d'une distribution²¹. Mais la cladistique (Darlu et Tassy 1993 en sciences du vivant, Gaillard-Corvaglia *et al.* 2008 en dialectologie), quant à elle, permet ce traitement de manière consubstantielle à sa démarche, ce qui implique que le linguiste investit en amont son savoir philologique et comparatiste dans l'édification des matrices numériques sur lesquelles porteront les calculs, à partir d'états phonologiques dérivés en diachronie, comme nous allons le voir pour une étude de cas restreinte, limitée à l'ALLOc.

Sérier les variables du point de vue structural, et non pas selon un calcul trivial objet par objet (consonne par consonne, voyelle par voyelle, etc., selon une logique d'alignements gabaritiques, comme dans Gabmap) devient possible grâce à l'approche cladistique (cf. Brun-Trigaud *et al.* 2014 pour une application à un segment du réseau dialectal occitan), qui explicite la démarche et le mode de calcul, pour des cartes de l'ALLOc. Voici un exemple de hiérarchisation des données en amont, afin de constituer une matrice d'analyse cladistique, pour une entrée de la base de données issue de l'ALLOc :

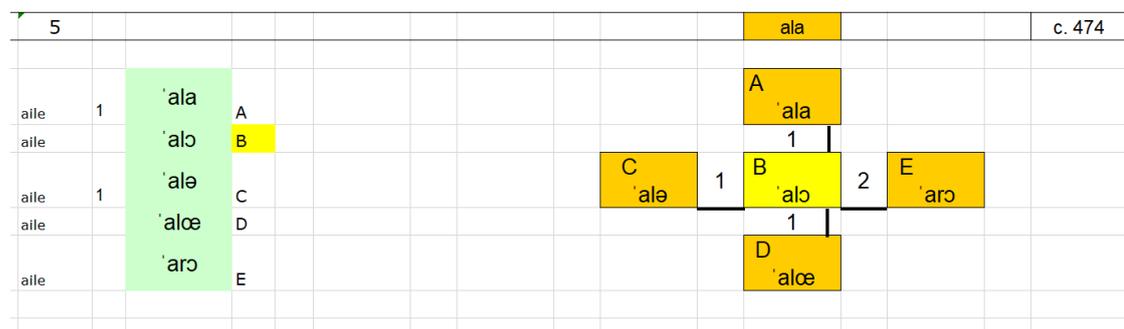


Figure 5. Stemma de dérivation diachronique avec pondération pour l'item ALA 'aile' de la base de données THESOC

Le stemma de la figure 5 se lit ainsi : de la forme ['ala], issue du latin ALA-, indexée au nœud A, découlent les formes B, C, D, E. Chacune des formes est en relation d'implication structurale avec les autres : par exemple, E implique B, pour implémenter le rhotacisme de la consonne médiane ($l > r$), alors que rien ne bouge par ailleurs, dans le vocalisme, d'autant plus que la voyelle basse atone doit être déjà rehaussée en une voyelle postérieure mi-basse, comme en B. On peut induire la forme D d'une antériorisation de cet état rehaussé de la voyelle post-tonique, atone, tout autant qu'on peut considérer que la réduction en une voyelle centrale,

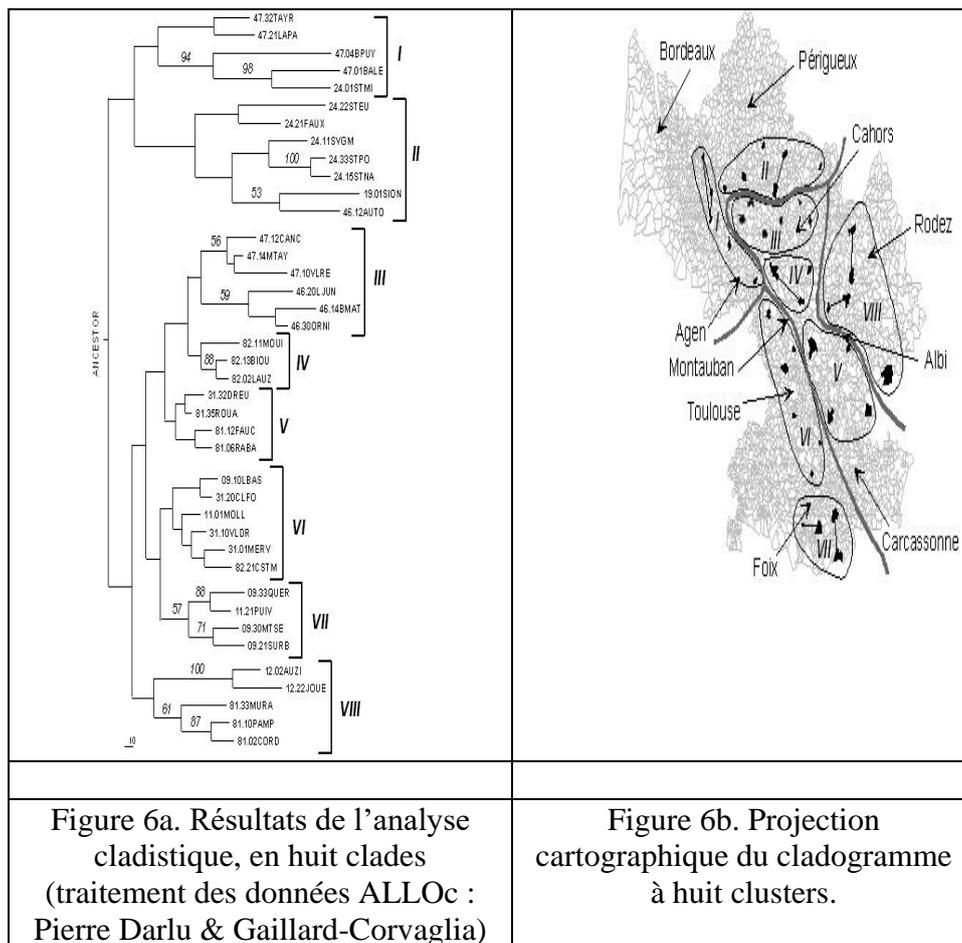
²¹ Cf.

dialektkarten.ch/dmviewer/popup.html#proj=alcd&lang=fr&id=info_interval_algorithms, consulté le 24/08/2022.

de type schwa, en découle également (c'est une hypothèse plus neutre, moins forcée, que de supposer que c'est l'état antériorisé de la voyelle en D qui donnerait le schwa du type ou caractère C). On pondère les liens : ici, c'est le rhotacisme (une latérale devient une rhotique, autrement dit, un « r ») qui est le fait marquant, on lui attribue donc un coefficient 2 au lieu d'un score 1 ailleurs, pour le détail des faits de vocalisme atone – étant donné que le A tonique reste stable. On notera que la forme B, qui correspond à celle de l'occitan standard, en termes de réalisation (la graphie est tout simplement < *ala* >), s'avère être la forme pivot.

[1]	abeille	8	[31]	coeurduporc	3	[61]	nid	6
[2]	agneau	6	[32]	côte	3	[62]	noyé	9
[3]	agnelé	10	[33]	cou	7	[63]	œil	7
[4]	aiguille	6	[34]	cuir	7	[64]	œuf	4
[5]	aile	4	[35]	cuisse	11	[65]	oiseau	10
[6]	âne	3	[36]	cul	7	[66]	pain	6
[7]	arbre	5	[37]	doigt	3	[67]	peau	4
[8]	bâtir	3	[38]	dos	8	[68]	pie	8
[9]	blé	4	[39]	eau	5	[69]	Pierre	1
[10]	boeuf	4	[40]	échelle	7	[70]	plier	6
[11]	bois	8	[41]	escalier	16	[71]	poirier	5
[12]	bouc	2	[42]	étoile	10	[72]	pondre	7
[13]	buis	6	[43]	farine	3	[73]	porte	1
[14]	chambre	7	[44]	feu	10	[74]	pré	1
[15]	chanvre	7	[45]	feuille	8	[75]	près	4
[16]	chapeau	7	[46]	fiel	7	[76]	puits	11
[17]	châtaigne	8	[47]	fil	6	[77]	roues	8
[18]	chatte	5	[48]	foiedeporc	8	[78]	sac	8
[19]	chaud	2	[49]	froid	4	[79]	sauce	7
[20]	chauffer	5	[50]	fromage	10	[80]	sel	4
[21]	chemin	2	[51]	jambe	3	[81]	soleil	9
[22]	chemise	6	[52]	laine	5	[82]	sueur	4
[23]	cheval	10	[53]	lait	12	[83]	taupe	5
[24]	cheveu	7	[54]	loup	4	[84]	vache	5
[25]	chèvre	4	[55]	lune	6	[85]	veau	9
[26]	chier	4	[56]	marteau	3	[86]	vent	5
[27]	ciel	5	[57]	miel	3	[87]	verre	7
[28]	cire	5	[58]	mule	3	[88]	vert	3
[29]	ciseaux	3	[59]	nettoyer	8	[89]	vitre	5
[30]	clef	1	[60]	nez	1	[90]	voler	3

Tableau 4. Les 90 mots-témoins de l'ALLOc et leur nombre de variantes phonologiques utilisés dans l'analyse cladistique



L'objet de connaissance produit par cette analyse (figure 6 *supra*) donne une image aisément interprétable de la structuration du diasystème du languedocien occidental dans son contexte occitan. On voit une aire centrale fragmentée en trois sous-aires – un languedocien central qui se subdivise en *quercinois* (Lot et Tarn-et-Garonne, en *albigois* (Tarn) et en *rouergat* (Aveyron) –, dont l'enveloppe, resserrée à la hauteur de Montauban, forme un couloir en carafe qui va de Carcassonne à Cahors. Au nord-ouest de cette aire centrale tripartite apparaît ce qu'on peut caractériser comme *occitan pré-septentrional girondin*. Ailleurs, à l'ouest et à l'est, l'aire centrale est flanquée d'une bande constituée de trois aires, distribuées entre Foix au sud, Toulouse au centre et Agen au nord-ouest d'une part, et de l'aire de Rodez au centre-est. Ces chorèmes²² suggèrent que les centres

²² Le *chorème* est une unité minimale constitutive de l'espace géographique : sur le plan conceptuel et géométrique ou sémiologique, ils se représentent comme des points, des lignes (isoglosses, etc.), des aires ou des réseaux (graphes) ; sur le plan des états observables comme des dynamiques de gravitation, de contact, de tropismes, de hiérarchie et d'échanges, ou distribution des ressources, etc., v. Brunet & Dolffus (1990 : 119), Léonard (2010, 2012 : 132-135). Signalons qu'en dialectométrie salburzgeoise cependant, ce concept revêt une acception plus large, et correspond à une aire sur une carte projetant les dendrèmes obtenus par l'analyse taxinomique.

urbains, les villes, ont eu une incidence, dans la moyenne ou longue durée historique, dans la condensation de variétés dialectales – car ici, nous sommes face à des sous-dialectes, voire à de simples regroupements de parlers, ou variétés locales ou micro-régionales de la composante occidentale du languedocien en tant que dialecte central du domaine occitan, comme nous l’avons observé précédemment. Ces chorèmes rappellent ce qu’on appelle en anglais les *Town dialects*, avec leurs arrière-pays – phénomène bien connu en Italie, où on parle de *dialetto urbano vs rustico*, ou dialecte urbain vs rural ou encore centre directeur vs arrière-pays, selon une approche centrifuge. Cependant, en domaine occitan, les villes moyennes (Rodez, Foix, Albi, etc.) doivent davantage être envisagées comme des épacentres que comme des centres rayonnants, dans cette configuration – à la différence de la situation géolinguistique italienne, où un ensemble de cités-Etats a imprimé sa marque au paysage dialectal, à l’époque médiévale²³. Cette dynamique de chorèmes de rang géographique inférieur (quasiment au niveau des « pays » ou des « terroirs ») n’était ni le but ni l’objet de nos analyses, dans le présent chapitre. C’est là l’un des paradoxes de la connaissance scientifique, qui veut que la science, quoique tendant à l’universel, revient par définition à une quête de savoir dans la complexité, présente à toutes les échelles – de l’immensément grand à l’infiniment petit. Ce survol des données de l’ALLOc à l’aide de la cladistique et de ses distances patristiques nous fait entrer dans la dimension fractale (structures minimales itératives, formant *a posteriori* des motifs cohérents et interdépendants) de la diversité dialectale. C’est cette nature fractale qui a fait si longtemps douter de l’existence des dialectes, alors qu’elle n’en est que la trame : ce sont les affinités réelles, objectives (paramètres et variables linguistiques) ainsi que les forces de cohésion structurale et systémique qui justifient, *a posteriori*, la possibilité d’une catégorisation et d’une hiérarchisation en langues, dialectes, sous-dialectes et parlers, comme nous y encourageait Henri Guiter dès les débuts de la dialectologie – et ce, en contradiction avec sa source d’inspiration : « l’indépendance des aires dialectales » de Théobald Lalanne, qui s’avouait désarmé face à ce caractère fractal du fait dialectal (v. Fossat 2016).

4.2. Extension méthodologique du domaine de GabMap

²³ On trouvera une étude très fine des agrégats et limites dialectales internes au domaine de l’ALLOc dans Camps (1986), qui associe les deux méthodes (qualitative et quantitative, selon la méthode globale de Guiter). Son étude montre une structuration de cet espace très homogène au centre et au sud (départements de l’Ariège, du Tarn et du Tarn-et-Garonne), qui contraste avec une intense fragmentation au nord (Dordogne, Lot, sud-est de l’Aveyron), à l’échelle de sous-dialectes. Nos résultats cladistiques sont congruents avec ceux de cette étude, quoique les nôtres permettent une hiérarchisation fine des aires, et une validation statistique des clades, selon un modèle de parcimonie. *Chorème* est souvent utilisé par Hans Goebel comme corrélat de *dendrème* : les premiers, en tant qu’aires sur des cartes (topologies), les seconds, en tant que clades dans des dendrogrammes (taxinomie).

L'avantage de la dialectométrie computationnelle, ou des méthodes de traitement algorithmique des données dialectales, tient à la multitude d'états des choses, de topologies, d'objets de connaissance, que cette solution technique est capable de traiter automatiquement, de manière quasiment instantanée, dès qu'on calibre pour elle les intrants, car non seulement différents algorithmes donnent des résultats différents²⁴, mais l'analyse des composantes de la langue (phonologie, morphologie, syntaxe, lexique) donnent également des résultats contrastés – c'est l'approche *paramétrique*, souvent mentionnée ici. Cette configuration en damier (de type *splitting*) diffère de celle obtenue à l'aide de l'algorithme *Complete Link supra*, en phonologie (de type *lumping*), qui convergeait davantage avec la carte canonique – la carte 2, de P. Bec. Reprenons ces divers éléments (figure 7, *infra*, qui confronte la réticulation par Gephi en 7a et la cartographie par Gabmap en 7b). Auparavant, notons que la méthode des connectogrammes par Gephi au seuil de similarité de 0,80 (figure 7a), qui fait apparaître des réseaux, s'avère très puissante pour montrer les relations de *fusion vs fission* (ou dépendance/appartenance/affinité vs autonomie/indépendance) des différents sous-domaines de l'occitan, tels que documentés par les atlas compilés dans la base de données THESOC, à travers les différents atlas régionaux, dont une synthèse dialectométrique est proposée ici.

On voit ainsi se dégager une macro-structure centrale, flanquée de trois structures périphériques, satellitaires : d'une part, le bloc central, ou noyau diasystémique constitué par l'ALLOr (en rouge), l'ALG (en vert), reliés par l'ALLOc (en violet), d'autre part l'ALP (en bleu), relié finement mais uniquement à l'ALLOr, ainsi que les sous-domaines du nord occitan, constitués de l'ALMC (en vert), solidement amarré à l'ALLOr et, de manière très diffuse, comme une nébuleuse, les localités de l'ALAL (en orange). En quelques mots, cela veut dire que le languedocien est bien le grand dialecte central, et que ses sous-dialectes sont des carrefours – le languedocien oriental, vers le provençal, à l'est, et vers l'auvergnat, au nord-est : le languedocien occidental, vers le gascon. Viennent ensuite deux ensembles : l'un très hétéroclite, nébuleux (le nord occitan : limousin

²⁴ Outre la démonstration supra de cette *puissance de vicariance* – ou diversité des points de vue –, on lira utilement Kassian (2015), qui applique une large gamme d'algorithmes sur les données des langues est-caucasiennes. On peut considérer cette étude comme un modèle du genre, déployant un large éventail de points de vue vicariants – une illustration de comment un chercheur « teste » ses résultats taxinomiques à l'aide de divers algorithmes dans un domaine linguistique.

et auvergnat), l'autre très compact et autonome : le provençal.

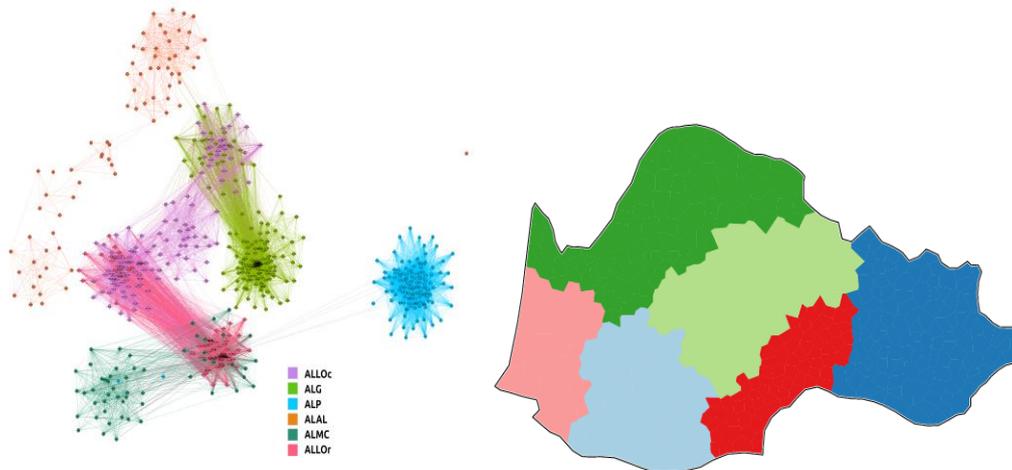


Figure 7a Gephi : indice 0,80 Figure 7b Méthode de Ward, 6 intervalles
Figure 7. Exemples de graphe (Gephi)²⁵ et de carte (GabMap : CAH) à partir des données lexicales de 235 cartes issues du THESOC.

Ces *objets de connaissance* sont autant de points de vue *vicariants* (alternatifs et complémentaires) sur les configurations dialectales du domaine occitan. On constate une fois de plus que la carte canonique, ou le modèle de référence (la carte 2), est certes utile, mais que ce n'est en rien une vérité absolue, gravée dans le marbre. En effet, la carte canonique des isoglosses phonologiques de P. Bec relève plutôt du réductionnisme, en définitive, car le provençal y est rassemblé avec une vaste aire centrale, partagée avec le languedocien. La carte GabMap en phonologie, par l'algorithme *Complete Link* (carte 3a, *supra*) allait dans le même sens, voire même plus loin encore. L'algorithme de la *Moyenne Groupée* (*Group Average*), carte 6b, s'aventurerait encore plus loin dans le *lumping*, ou réductionnisme, sur la base de l'échantillon de phonologie. En revanche, la carte en 7b, à droite de la figure 7 se situe sur le pôle opposé (*Méthode de Ward*, échantillon de phonologie), en termes de saisie : de manière atomiste, on retrouve les grands dialectes, mais on se retrouve aussi face à un éclatement de ceux-ci en sous-dialectes, en ce qui concerne notamment le languedocien. Le gascon est comme comprimé, et y perd sa partie nord, réunie avec le limousin. La phonologie, plus variable que le lexique en tant que système fermé dont la dérive évolutive est fortement déterminée par sa logique systémique propre, induit ces phénomènes de fragmentation.

²⁵ Les abréviations sont les mêmes que p.4, dans la présentation des atlas. NB : les couleurs de la carte à droite (GabMap) sont aléatoires et ne correspondent pas à celles du graphe –Gabmap ne permet pas de choisir les couleurs.

La figure 8 met en regard le résultat de l'analyse cladistique, sous forme d'aires correspondant aux clades du dendrogramme de la figure 6a (carte, à gauche), avec le connectogramme correspondant généré par Gephi (graphe, à droite). La réticulation par Gephi permet de mieux saisir les interactions entre les différents réseaux et sous-réseaux constituant la trame du continuum dialectal d'un sous-dialecte comme celui du languedocien occidental, à partir des données de l'ALLOc.

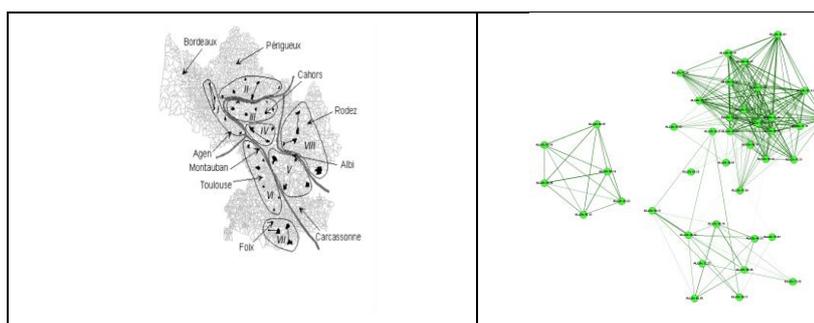


Figure 8. Résultats de l'analyse cladistique (carte, à gauche) et résultats par distance d'édition sur Gephi, seuil 0,86 (graphe, à droite). Mêmes données de l'ALLOc (par THESOC), dans les deux traitements taxinomiques

Une fois la confrontation d'objets de connaissances (cartes par isoglosses ou par traitement dialectométrique) effectuée, et que nous avons posé les bases de notre double regard (heuristique et contrastif) sur des objets de grande taille, concernant l'ensemble du domaine occitan, nous poussons l'exploration de la diversité interne de l'occitan encore plus loin, en termes de finesse de grain aréologique, à partir d'une étude de cas propice à la multiplicité des points de vue expérimentaux : l'ALLOc (Ravier 1978-1993), dont les données phonologiques sont envisagées à l'aide de deux méthodes – d'une part, la cladistique (Hennig 1950, 1966), fondée sur des distances dites patristiques (où les caractères génétiques ou, ici, phonologiques, sont dérivés les uns des autres dans des graphes évolutifs), d'autre part, la dialectométrie automatisée (distance de Levenshtein), visualisée par des graphes de similarité.

La diversité interne du réseau dialectal occitan tel que le représentent les données des atlas linguistiques (et, partant, du THESOC), se prête à une multitude de configurations géolinguistiques (divisions et subdivision en « dialectes » et sous-dialectes, ou encore « parlars », variétés)²⁶. Il n'y a pas une seule et unique

²⁶ Si la question des critères de hiérarchisation interne d'un domaine dialectal avait, aux débuts de la dialectométrie, trouvé une réponse simple par la « méthode globale » d'Henri Guiter (additionnant tous les paramètres : phonologique, morphologique, lexical, etc.), qui proposait cette échelle de classification : « Nous admettons d'ores et déjà qu'à partir de 20% de différences entre

classification (ou *taxinomie*) des dialectes occitans, définitivement tangible, mais plutôt *un éventail de possibles*, à différents degrés d'observation, en séparant ou au contraire en croisant les méthodes quantitatives, qui garantissent des *conditions de vérité*, sans pour autant imposer une solution unique. L'apport de cette approche à géométrie variable pour la recherche interdisciplinaire est considérable, en termes de *visualisation* – un enjeu capital face à tout objet complexe, comme nous n'avons cessé de le souligner.

4.3 Visualisation par graphes dynamiques et seuils normalisés : davantage sur Gephi

La visualisation par connectogrammes, ou flux de réseaux, par Gephi, est une solution technique puissante, que nous avons commencé à entrevoir ci-dessus (figures 7a et 8), elle offre des applications de traitement pour modéliser des systèmes complexes (cf. Patriarca/Heinsalu/Léonard 2020). En convertissant les scores de différences obtenus avec la distance d'édition (Gabmap) en similarités, ce logiciel rend compte des liens entre les objets analysés – ici comme ailleurs dans ce chapitre, entre les points des atlas occitans du NALF. L'application permet de choisir un seuil, en faisant défiler en continu les graphes de liens, de 0 à 1. Voici deux résultats obtenus à l'aide de ce procédé : pour la phonologie (seuils de 0,75 et 0,825), avec les mêmes données du THESOC que celles soumises au traitement dendrographique et cartographique à l'aide de Gabmap (distance d'édition) :

deux points il s'agit de parlars différents ; de 30%, de sous-dialectes différents ; de 50%, de dialectes différents ; de 80%, de langues différentes. » (Guiter 1973, 79). Cependant, aujourd'hui, la diversité des calculs sur 1, comme avec la distance d'édition, ou par distances patristiques, comme avec la cladistique, rend moins évidente une telle grille. Si l'on s'en tient aux classifications qualitatives, il est plus aisé de donner des exemples de sous-dialectes : le béarnais ou le gascon landais (notamment le « gascon noir ») comme *sous-dialectes* du *dialecte* gascon. Mais peut-on qualifier le vivaro-alpin de sous-dialecte du dialecte provençal ? Nos mesures *infra*, suggèrent que ce serait négliger la continuité de cette entité avec le nord-occitan. Une *variété* est plus facile à définir : c'est l'équivalent du « parler » dans son acception la plus commune : le parler d'une localité (un point d'atlas linguistique).

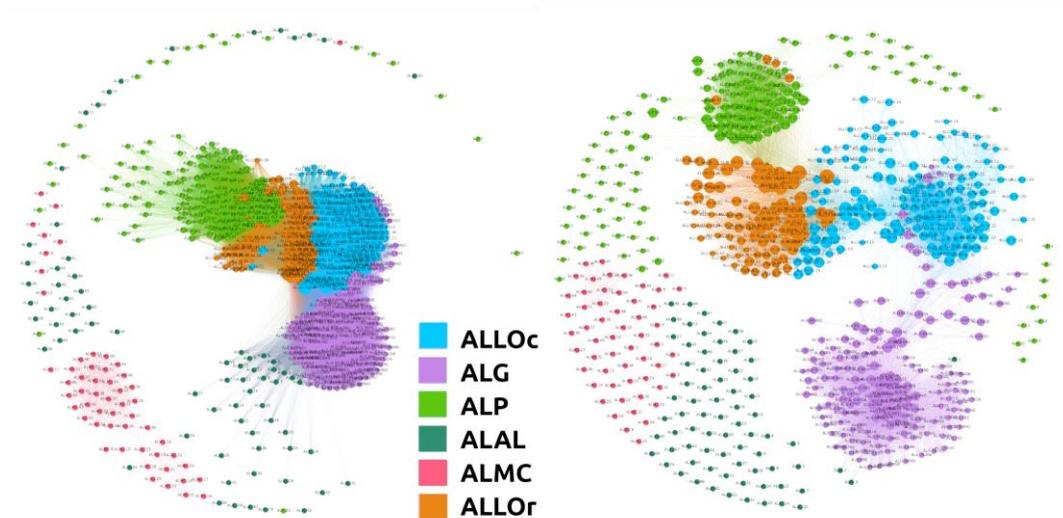


Figure 9. Seuils 0,75 et 0,825 : données phonologiques, THESOC, visualisation réticulaire à l'aide de Gephi

Les « mondes dialectométriques » que ces deux séries de réticulation font apparaître sous forme de « connectogrammes » sont autant d'états variables d'affinités et de forces cohésives. *A posteriori*, ils *simulent* des réseaux ethno-sociolinguistiques, si l'on veut qu'ils se prêtent à une telle lecture, certes, mais à condition (et à cette seule condition) de ne jamais oublier que ces *états agrégatifs* sont dynamiques, à géométrie variable, et que, même si de nettes tendances sont observables, aucun *moment* de la série ne doit être extrait arbitrairement de la chaîne des transformations qu'offre la modulation statistique, dépendant du seuil de similarité retenu. Ainsi, le seuil 0,75, relativement peu exigeant en *convergence structurale* (et implicitement, en *intercompréhension potentielle*), pour le paramètre phonologique réalisationnel, tel que le saisit la distance d'édition pour notre base de données THESOC, révèle un noyau compact au centre de la « galaxie dialectale occitane » – pour autant qu'on nous permette cette métaphore, à des fins de lisibilité des données –, constitué par le languedocien central (ALLOr, en orange et ALLOc en bleu), flanqué à ses deux extrémités du provençal (ALP, en vert) et du gascon (en violet). Le centre de gravité de ce noyau est bel et bien le languedocien, ou « occitan central » de P. Bec, qui intègre une frange gasconne à la droite du noyau ALLOc. Tels sont les phénomènes de concentration, ou de condensation des réseaux dialectaux, en structure interne du diasystème. Venons-en maintenant aux phénomènes de dispersion de ces blocs. Les deux dialectes – provençal et gascon – formant une polarité autour du centre de gravité, qui est incorporé, enchâssé littéralement dans leur secteur présentent une structure de dispersion notablement différente. Car le provençal se disperse en deux temps : d'abord comme autant de variétés ou parlers éjectés du noyau, sous forme d'un bouquet, puis sous forme d'une ceinture externe, dans ce troisième cercle qu'est la bordure constituant la

sphère externe – formé de nuages ou de noyaux issus de l’ALAL (pour le *Croissant* et le limousin), de l’ALMC (pour l’auvergnat) et de l’ALP (pour le vivaro-alpin et le dauphinois). Le gascon, quant à lui, semble suivre trois options : soit s’entrelacer sur sa frontière avec « l’occitan central », soit former un bouquet d’éjection de variétés limousines (ALAL), ou encore, dans la continuité de ce mouvement, s’affilier de loin à des parlers limousins de la ceinture externe du graphe – on retrouve la dimension guyennaise observée auparavant, avec les cartes et dendrogrammes de Gabmap.

Au seuil 0,825, pour le paramètre lexical, le caractère compact, cohésif, des deux composantes, orientale (ALLOr) et occidentale (ALLOc) de « l’occitan central » est confirmé, et une structure interne nette se dégage, pour ces deux blocs – particulièrement diversifiée pour le languedocien occidental –, de même qu’est notable la compacité et la propension à la transition avec le bloc central des deux blocs antithétiques que sont le provençal (ALP) et le gascon (ALG). De même, le caractère *diffus* – pour ne pas dire *volatile* – du sous-domaine nord-occitan (ALAL, ALMC, mais aussi la partie alpine de l’ALP), que faisaient bien apparaître les algorithmes réducteurs de Gabmap, est largement confirmé par ces connectogrammes : l’occitan septentrional forme désormais une épaisse ceinture – une sorte de corset – tout autour des blocs centraux, où le nuage de points de l’auvergnat (ALMC) se retrouve enchâssé dans le nuage limousin (ALAL). La frange externe du provençal, quant à elle, se répartit tous azimuts dans la bordure de la sphère externe.

Quel scénario, en revanche, nous propose le paramètre lexical, à partir des connectogrammes de Gephi ? De nature catégorielle stricte (*testa* contre *cap* “tête” sont en relation d’opposition terme à terme, à la différence de la complexité des contrastes phonétiques et phonologiques au sein d’un même lexème), le lexique induit un niveau-seuil de pertinence des agrégats réticulaires bien plus bas que pour le paramètre phonologique.

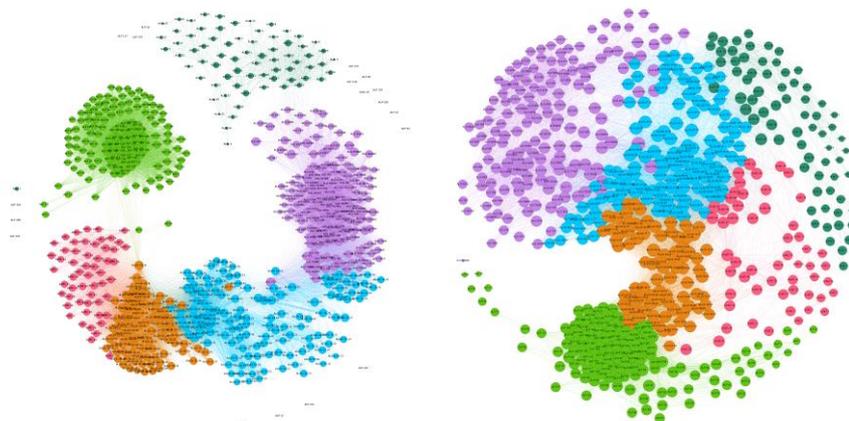


Figure 10. Seuils 0,65 et 0,875 : données lexicales, THESOC, visualisation réticulaire à l'aide de Gephi

Au seuil 0,65, une chaîne de blocs associe dans une grande « chenille » l'occitan central et ses deux composantes (ALLOr et ALLOc), flanquées, à chaque extrémité, de l'auvergnat (ALMC, à gauche) et du gascon (ALG, à droite), organiquement liés par un dense écheveau d'entrelacs. On dirait que c'est le languedocien oriental qui se distend en auvergnat, tout comme le languedocien occidental s'entrelace avec le gascon sur son flanc ouest. Sans surprise, on voit encore le provençal se détacher nettement de cette chaîne, quoiqu'en maintenant des liens avec le languedocien oriental (ALLOr). On distingue nettement un noyau très dense au centre de son agrégat, entouré d'une épaisse gangue de parlers, dont certains se disloquent, sans se rattacher à aucun autre – une tendance d'ores et déjà observée dans l'examen des résultats lexicaux à l'aide des algorithmes de Gabmap. Le *Croissant* et le limousin (ALAL) ne présentent pas de densité : leur structure est volatile, et les liens sont principalement orientés vers la zone de dispersion des liens émanant du gascon (ALG), rappelant l'aire guyenno-gasconne entrevue précédemment, avec Gabmap. Au seuil 0,875, une reconfiguration allant vers davantage de regroupements potentiels émerge, et surtout, accentue les effets de condensation *vs* dispersion des blocs : les aires nord-occitanes (ALAL et ALMC) se périphérisent et vaporisent leurs liens, tandis que les deux blocs périphériques que sont le gascon (ALG) et le provençal (ALP) se disloquent et forment ou contribuent à former la périsphère de l'ensemble.

5. Conclusion et perspectives

L'objectif de ce chapitre était à la fois de présenter les nouveaux atlas linguistiques occitans du NALF et de la valorisation qui en est faite par la base de données THESOC et de montrer les horizons de recherche que permettent de telles données. Nous avons donc déployé les multiples représentations taxinomiques et cartographiques présentées et commentées ci-dessus, qui sont à la fois congruentes avec les divisions dialectales traditionnelles du domaine occitan (par exemple, celles de P. Bec, en tant que divisions supradialectales canoniques), et novatrices par le « grain fin » des modélisations quantitatives obtenues grâce à divers traitements dialectométriques et cladistiques des données du THESOC, à différentes échelles de grandeur (l'ensemble des sous-domaines atlantographiques d'une part, un sous-domaine relativement central et très intriqué comme celui de l'ALLOc, d'autre part). Enfin, loin de surestimer la seule méthode quantitative, nous concluons ce chapitre en rappelant l'intercomplémentarité entre les deux approches empiriques fondamentales qu'entretiennent en géolinguistique, l'analyse qualitative, à l'aide d'isoglosses et d'aires, d'une part, et l'approche quantitative (dialectométrie et cladistique), d'autre part. Il est cependant clair pour nous que la dernière approche jouit d'un avantage incomparable sur la première : la capacité

raisonnée, justifiée sur le plan logique (notamment taxinomique) et mathématique, en tant que mode opératoire, de hiérarchiser, emboîter et même pondérer les aires ou entités dialectales (dialecte, sous-dialecte, variétés), selon une géométrie variable, plutôt que par essentialisation fondée sur des facteurs externes. De ce point de vue, la dialectométrie appartient à la grande famille des méthodes relevant du rationalisme appliqué, cher à l'épistémologue Gaston Bachelard (1949). La résolution du dilemme continuiste, qui voulait que les dialectes n'existent pas : seuls les parlers existent, dans une trame continue et insécable (ou « doctrine de Gaston Paris »), est envisageable par l'approche quantitative et les instruments de visualisation de la complexité que fournissent les applications par distance d'édition, notamment.

Car nous obtenons des taxinomies dialectales à divers niveaux, qui sondent ce que les dialectologues de l'ALG autour de Jean Séguy appelaient *le fait dialectal* : notre rationalisme, outillé de GabMap et d'algorithmes en tous genre, nous permet de faire apparaître sous des formes multiples, comme à travers un prisme, la diversité des classes de dialectes possibles au sein d'une langue. On pourrait se demander pourquoi le languedocien devrait-il être nécessairement la *lingua franca* occitane, en quelque sorte. Les résultats dialectométriques le montrent d'objet de connaissance en objet de connaissance, issus de divers traitements algorithmiques (*Complete link, Méthode de Ward, Moyenne pondérée, Moyenne des groupes* ; distances patristiques et graphes à seuils normalisés, par réticulation Gephi) : cette aire dialectale centrale, au-dessous du bouclier nord-occitan, se prête structurellement à jouer le rôle de « variété moyenne ». Sa réticularité, en termes actifs – de connectivité – avec le reste du réseau diasystémique, au sud du « bouclier arverne », est très dense, et elle s'élanche dans toutes les directions, tandis que la composante auvergnate est liée au provençal par le biais du vivaro-alpin. Ce sont là des faits, que le rationalisme métalinguistique et mathématique (isoglosses et dendrogrammes) fait apparaître, que la société peut adopter, accepter ou rejeter. Mais c'est alors un débat qui se tient dans une autre sphère.

Nous avons commencé à envisager la classification automatisée des langues par l'algorithme de Levenshtein, ou distance d'édition, en tant qu'outil de classification relevant de la linguistique générale, mais aussi de la dialectologie appliquée. La dialectométrie est avant tout outil réflexif, vicariant, à géométrie variable, d'autant plus nécessaire que la standardisation s'est faite, en occitan, sur la base du languedocien central. Nous avons donc ici « testé » ce choix par la dialectologie computationnelle, et on pourra d'autant plus juger de sa pertinence, en termes de centralité structurale : cette aire semble en effet fonctionner comme le pivot d'une certaine unité du sud du domaine. Nous avons vu que deux blocs s'opposent nettement : le sud, assez unitaire, contre le nord, très fragmenté et diversifié, et dont le centre le plus cohérent semble se situer dans l'aire limousine.

Nous avons vu également que, plutôt que d'essentialiser les dialectes (gascon, languedocien, provençal, limousin, auvergnat, etc.), il faut plutôt les prendre comme des « décideurs multiples », des compères du linguiste et de l'aménageur : ils permettent de « tester » la recevabilité ou au contraire le caractère aberrant des configurations obtenues. Nous avons procédé de manière méthodique, en faisant usage de diverses solutions algorithmiques et de diverses techniques de visualisation (Gephi), en les comparant, en les hiérarchisant, et en recherchant à toujours falsifier – dans le sens popperien, épistémologique du terme, les résultats, afin de tester leur pertinence et leur robustesse. Nous avons évité de répondre aux débats localistes (béarnais contre gascon, les deux contre le languedocien, provençal contre languedocien, etc.), qui ont certes leur raison d'être dans le cadre d'un débat d'idées, mais qui relèvent de l'opinion, sur la base de faits et d'arguments choisis en fonction de la finalité que se fixent les groupes de discussion. Or, nous avons vu qu'en dialectologie quantitative, on procède tout autrement : par ensembles raisonnés et argumentés de paramètres structuraux, en tenant le plus grand compte de la distribution complémentaire des phénomènes de variation. La linguistique appliquée doit tenir compte aussi de ces divisions et de ces individuations locales, car toute imposition arrogante d'une seule solution reste vouée à l'échec, en cas de polarisation. C'est là encore, un défi pour la dialectologie quantitative, qui est à la fois une méthode qui intéresse la recherche fondamentale en linguistique théorique et descriptive, et une science appliquée à finalité sociale, et ne peut faire l'économie d'une certaine réflexivité, et d'une écoute des débats sociaux, sans prendre parti pour autant. C'est d'ailleurs dans cette neutralité, cette indifférence ethnométhodologique nécessaire (Garfinkel 1967), que le choix du terme de *fait dialectal* s'est imposé autour de l'entreprise de l'ALG, et ce n'est pas non plus par hasard que c'est dans ce domaine périphérique si original qu'est le domaine du gascon, que l'effort de rationaliser davantage la définition des entités dialectales et des aires géolectales par la dialectométrie s'est imposée, autour des travaux de Jean Séguy et de Théobald Lalanne.

Références :

- ALAL = Potte, Jean-Claude (1975-1992), *Atlas linguistique et ethnographique de l'Auvergne et du Limousin*, Paris, Editions du CNRS (3 vol.).
- ALCe = Dubuisson, Pierrette (1971-1982), *Atlas linguistique et ethnographique du Centre*, Paris, Editions du CNRS (3 vol.).
- ALF = Gilliéron, Jules/Edmont, Edmond (1902-1910), *Atlas Linguistique de la France*, Paris, Champion.
- ALG = Séguy, Jean (1954-1973), *Atlas linguistique et ethnographique de la Gascogne*, Paris, Editions du CNRS (6 vol.).
- ALJA = Martin, Jean-Baptiste/Tuailon, Gaston (1971-1978), *Atlas linguistique et ethnographique du Jura et des Alpes du Nord*, Paris, Editions du CNRS (3 vol.).

- ALLOc = Ravier, Xavier (1978-1993), *Atlas linguistique et ethnographique du Languedoc Occidental*, Paris, Editions du CNRS (4 vol.).
- ALLOr = Boisgontier, Jacques (1981-1986), *Atlas linguistique et ethnographique du Languedoc Oriental*, Paris, Editions du CNRS (3 vol.).
- ALLy = Gardette, Pierre (1967-1976), *Atlas linguistique et ethnographique du Lyonnais*, Paris, Editions du CNRS (5 vol.).
- ALMC = Nauton, Pierre (1957-1961), *Atlas linguistique et ethnographique du Massif Central*, Paris, Editions du CNRS (4 vol.).
- ALO = Massignon, Geneviève/Horiot, Brigitte (1971-1983), *Atlas linguistique et ethnographique de l'Ouest*, Paris, Editions du CNRS (3 vol.).
- ALP = Bouvier, Jean-Claude/Martel, Claude (1975-1986), *Atlas linguistique et ethnographique de la Provence*, Paris, Editions du CNRS (3 vol.) et Bouvier, Jean-Claude/Martel, Claude/Brun-Trigaud, Guylaine (2016), *Forcalquier, Alpes de Lumière* (1 vol.).
- ALPO = Guter, Henri (1966), *Atlas linguistique des Pyrénées Orientales*, Paris, Editions du CNRS (1 vol.).
-
- Aurrekoetxea, Gotzon/Nerbonne, John/Rubio, Jesus (2020), *Unifying Analyses of Multiple Responses*, *Dialectologia* 25, 23-42.
- Bachelard, Gaston (1949), *Le rationalisme appliqué*, Paris, Presses Universitaires de France.
- Balmayer, Louis (1971), *Quelques caractéristiques de l'occitan montpelliérain*, *Revue de Linguistique Romane* 35, 90-96.
- Bec, Pierre (1963, ³1973), *La langue occitane*, Paris, Presses Universitaires de France, Que-Sais-Je ?
- Brun-Trigaud, Guylaine (2014), *Le Thesaurus Occitan, une base de données multimediale dédiée aux dialectes occitans*, *La Bretagne Linguistique* 18, 22-34. (hal-01360681)
- Brun-Trigaud, Guylaine/Malfatto, Albert (2013), *Limites dialectales vs limites lexicales du domaine occitan : un impossible accord ?*, in: Carrilho, Ernestina/Magro, Catarina /Alvarez, Xosé (edd.), *Current approaches to Limits and Areas in Dialectology*, Newcastle upon Tyne, Cambridge Scholar Publishing, 294-310.
- Brun-Trigaud, Guylaine/Darlu, Pierre/Gaillard-Corvaglia, Antonella/Sauzet, Patric/Léonard, Jean Léo (2014), *Exploration cladistique de l'ALLOc*, in: Alén Garabato, Carmen/Torreilles, Claire/Verny, Marie-Jeanne (edd.), *Actes du Xe Congrès de l'Association Internationale d'Etudes Occitanes*, Limoges, Lambert Lucas, vol. 2, 489-500.
- Brunet, Roger/Dolffus, Olivier (1990), *Mondes nouveaux*, Paris, Hachette.
- Camproux, Charles (1978), *Petit Atlas linguistique discursif du Gévaudan*, Montpellier, Centre d'Etudes Occitanes, Université Paul-Valéry.
- Camps, Christian (1985), *Atlas linguistique du Biterrois*, Toulouse, Institut d'études occitanes.
- Camps, Christian (1986), *Limites linguistiques d'après l'Atlas linguistique et ethnographique du Languedoc occidental*, in: *Actes du XVIIème Congrès International de Linguistique et de Philologie Romanes*, Aix-en-Provence, Université de Provence, vol. 6, 119-135.
- Camps, Christian (1991), *Limites linguistiques en Languedoc oriental*, in: Kremer, Dieter (ed.), *Actes du XVIIIème Congrès International de Linguistique et de Philologie Romanes*, Tübingen, Max-Niemeyer Verlag, 362-369.

- Chambers, J. K. *Liberating Dialectology*, In: Weiling, Martijn/ Kroon, Martin / van Noort, Gertjen/Bouma, Gosse (éds.) *From Semantics to Dialectometry: Festschrift in Honour of John Nerbonne*, College Publications, 85-92.
- Corvaglia-Gaillard, Antonella, Darlu, Pierre et Léonard, Jean Léo (2008), *Approche cladistique des vocalismes d'oïl à partir des données de l'A.L.F.*, *Bollettino dell'Atlante Linguistico Italiano* III Serie - Dispensa 32, 36-78.
- Dalbera, Jean-Philippe/Oliviéri, Michèle/Ranucci, Jean-Claude/Brun-Trigaud, Guylaine/Georges, Pierre-Aurélien (2012), *La base de données linguistique occitane Thesoc. Trésor patrimonial et instrument de recherche scientifique*, *Estudis Romànics* 34, 367-387.
- Darlu, Pierre, Tassy, Pascal (1993), *Reconstruction phylogénétique. Concepts et Méthodes*, Masson. Accessible sur <https://pascalneige.files.wordpress.com/2012/01/darlu-et-tassy-1993.pdf>; consulté le 24/08/2022.
- Dartigue, Charles (1950), *Histoire de la Guyenne*, Paris, Presses Universitaires de France, Que-Sais-Je ?
- Dauzat, Albert (1939), *Un nouvel atlas linguistique de la France*, *Le Français Moderne* 7, 97-101.
- Fossat, Jean-Louis (1978), *Etat des recherches dialectométriques sur le domaine gascon : "fonction maximale" et "fonction minimale" du dialecte*, In: Werlen, Iwar (ed.), *Probleme der schweizerischen Dialektologie*, Fribourg, Editions Universitaires, 109-139.
- Fossat, Jean-Louis (2016) *Le parcours de Th. Lalanne : du recueil de données (1947-1952) aux cartes synthétiques de l'Atlas linguistique de la Gascogne maritime*, *Géolinguistique* [Online], 16 | 2016, DOI: <https://doi.org/10.4000/geolinguistique.517>, consulté le 24/08/2024.
- Fruchterman, Thomas M. J./Reingold, Edward M. (1991), *Graph Drawing by Force-Directed Placement*, *Software – Practice & Experience*, Wiley 21 (11), 1129-1164.
- Gardette, Pierre (1968), *Atlas linguistique et ethnographique du Lyonnais. IV. Exposé méthodologique et tables*, Paris, Editions du CNRS.
- Garfinkel, Harold (1967), *Studies in Ethnomethodology*, Prentice-Hall, Englewood Cliffs.
- Gilliéron, Jules/Edmont, Edmond (1902-1910), *Atlas Linguistique de la France*, Paris, Champion.
- Goebel, Hans (1982), *Dialektometrie. Prinzipien und Methoden des Einsatzes der numerischen Taxonomie im Bereich der Dialektgeographie*, Vienne, Verlag der Öst. Akademie der Wissenschaften.
- Goebel, Hans/Smečka, Pavel (2014), *L'analyse dialectométrique des cartes de la série B de l'ALF*, *Revue de linguistique romane* 78, 439-497.
- Guitier, Henri (1973), *Atlas et frontières linguistiques*, in: *Les dialectes romans de France*, Paris, Editions du CNRS, 61-109.
- Heeringa, Wilbert (2004), *Measuring dialect pronunciation differences using Levenshtein distance*. Groningen, University of Groningen, Ph.D. dissertation.
- Hennig, Willi (1950), *Grundzüge einer theorie der Phylogenetischen systematik*, Berlin, Deutscher Zentralverlag.
- Hennig, Willi (1966), *Phylogenetic Systematics*, Urbana, University of Illinois Press.
- Jagueneau, Liliane (1987), *Structuration de l'espace linguistique entre Loire et Gironde : analyse dialectométrique des données phonétiques de l'Atlas linguistique et ethnographique de l'Ouest*, Thèse, Toulouse (sous la dir. de J.-L. Fossat).
- Kassian, Alexei (2015), *Towards a formal genealogical classification of the Lezgian languages (North Caucasus): testing various phylogenetic methods on lexical data*, *PLoS ONE* 10(2): e0116950. doi:10.1371/journal.pone.0116950.
- Leinonen, Therese/Çölterin, Çağrı/Nerbonne, John (2016), *Using Gabmap*, *Lingua* 178, 71-83.
- Lalanne, Théobald (1949), *L'indépendance des aires linguistiques en Gascogne maritime*, Saint-Vincent-de-Paul, chez l'auteur (2 vol).

- Léonard, Jean Léo (2010), *Chorèmes, aires et réseaux. Si l'Estonie m'était contée... A travers ses isoglosses*, in: Renaud, Patrick (éd.), *Actes du colloque Les situations de plurilinguisme en Europe comme objet de l'histoire*, Paris, L'Harmattan, 97-120.
- Léonard, Jean Léo (2012), *Éléments de dialectologie générale*, Paris, Michel Houdiard éditeur.
- Levenshtein, Vladimir I (1966), *Binary Codes Capable of Correcting Deletions, Insertions, and Reversals*, *Soviet Physics – Doklady* 10/8, 707-710.
- Millardet, Georges (1910), *Petit atlas linguistique d'une région des Landes. Contribution à la dialectologie gasconne*, Toulouse, E. Privat.
- Nerbonne, John/Colen, R./Gooskens, C./Kleiweg, P./Leinonen, T. (2011), *Gabmap - a web application for dialectology*, *Dialectologia: revista electrònica*, 65-89.
- Nerbonne, John/Heeringa, W. (2001), *Computational Comparison and Classification of Dialects*, *Dialectologia et Geolinguistica* 9, 69-83.
- Nerbonne, John/van Gemert, Ilse/Heeringa Wilbert (2005), *A Dialectometric View of Linguistic "Gravity"*, Alpha Informatica report, University of Groningen, accessible sur <https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.530.5957>, consulté le 24/08/2022.
- Nerbonne, John/Kretzschmar, W. (2003), *Introducing Computational Techniques in Dialectometry*, *Computers and the Humanities* 37, 245-255.
- Heinsalu Els/Patriarca Marco/Léonard Jean Léo (2020), *Languages in Space and Time. Models and Methods from Complex Systems Theory*, Cambridge University Press.
- Philps, Dennis (1976), *Balaguères-Bethmale-Biros : étude dialectométrique*. Toulouse, Institut d'études méridionales.
- Philps, Dennis (1985), *Atlas dialectométrique des Pyrénées Centrales*. Université de Toulouse 2 (Thèse d'Etat sous la dir. de J.-L. Fossat).
- Philps, Dennis (1986), *L'automatisation des atlas linguistiques*, in *Actes du XVllème congrès international de Linguistique Romane*, Aix-en-Provence : Université de Provence, vol. 6, 465-486.
- Prokić, Jelena/Nerbonne, John (2008), *Recognize Groups among Dialects*, In: Nerbonne, John/Gooskens, Charlotte/Kürschner, Sebastian/van Bezooijen, Renée (eds), *International Journal of Humanities and Arts Computing*, Special Issue on *Language Variation*, 153-172.
- Revest, Laurent (2009), *Le dialecte occitan alpin : aire d'extension et caractéristiques linguistiques*, thèse de doctorat, dir. Jean-Philippe Dalbera, Université Sophia Antipolis.
- Revest, Laurent (2011), *Caractéristiques linguistiques de l'alpin (ou gavot) maritime du pays mentonnais et des vallées des Paillons*. In : *Actes du 13^e congrès des Langues dialectales*, Monaco 10/01/2009, 99-160.
- Dalbera, Jean-Philippe (1995), *Polymorphisme et innovation dans l'aire occitane alpine : Le parler de Sainte-Agnès*, *Travaux du Cercle linguistique de Nice*, 17, 3-35.
- Ronjat, Jules (1930-1937), *Grammaire istorique [sic] des parlers provençaux modernes*, Macon, Protat (4 vol.).
- Séguy, Jean (1971), *La relation entre la distance spatiale et la distance lexicale*, *Revue de Linguistique Romane* 35, 335-357.
- Séguy, Jean (1973a), *La dialectométrie dans l'Atlas linguistique de la Gascogne*, *Revue de Linguistique Romane* 37, 1-24.
- Séguy, Jean (1973b), *La fonction minimale du dialecte*, in: *Les dialectes romans de France à la lumière des atlas régionaux*, Paris, Editions du CNRS, 27-42.
- Simoni-Aurembou, Marie-Rose (1998), *Les atlas linguistiques de la France par régions*, *Modèles linguistiques* 38 (XIX/2), 37-54.

- Sumien, Doumergue (2009), *Classificacion dei dialèctes occitans*, *Linguistica Occitana*, <http://linguistica-oc.com/wp-content/uploads/2013/07/Linguistica-occitana-7-Sumien.pdf>
- Esteve Valls, Martijn Wieling, John Nerbonne (2013), *Linguistic advergence and divergence in north-western Catalan: A dialectometric investigation of dialect leveling and border effects*, *Literary and Linguistic Computing*, (28)1, 119–146, <https://doi.org/10.1093/lc/fqs052>.
- Martijn Wieling, Esteve Valls, R.Harald Baayen, and John Nerbonne (2018), *Border effects among Catalan dialects*, In: Dirk Speelman, Kris Heylen and Dirk Geeraerts (eds.) *Mixed Effects Regression Models in Linguistics* Springer: Quantitative Methods in the Humanities and Social Sciences, 71-97.
- Wienreich Uriel (1954), *Is a structural dialectology possible?*, *Word*, 10(2-3), 388-400.
- Wieling, Martijn/Nerbonne, John (2015), *Advances in Dialectometry*, In: *Annual Review in Linguistics* 1(1), 243-264. doi:10.1146/annurev-linguist-030514-124930

Annexe : Essai de bibliographie de dialectométrie en domaine occitan

➔ **Lalanne, Théobald [1880-1962] :**

- (1949), *L'indépendance des aires linguistiques en Gascogne maritime*, Saint-Vincent-de-Paul, chez l'auteur (2 vol).
 (1951), *Une coupe dialectale : l'axe lexical Alpes-Océan (Plan du Var - Soustons)*, Annales du Midi 63 (n°14), 149-162.
 (1953), *Indices de polyonymie. Indices de polyphonie*, Le Français Moderne 21, 263-274.

➔ **Séguy, Jean [1914-1973] :**

- (1971), *La relation entre la distance spatiale et la distance lexicale*, Revue de Linguistique Romane 35, 335-357.
 (1973a), *La dialectométrie dans l'Atlas linguistique de la Gascogne*, Revue de Linguistique Romane 37, 1-24.
 (1973b), *La fonction minimale du dialecte*, in: *Les dialectes romans de France à la lumière des atlas régionaux*, Paris, Editions du CNRS, 27-42.

➔ **Fossat, Jean-Louis :**

- (1975), et Besche, Bruno/Garric, Marie-Noëlle (edd), *Dialectologie, dialectométrie, ethnosémantique, sociolinguistique, lexicométrie. Travaux de l'E.R.A. 352 C.N.R.S.*, Toulouse, Université de Toulouse II Le Mirail.
 (1976a), et Saudinos, Louis/Philps, Dennis (edd), *Toponymie et dialectométrie des Pyrénées centrales. Documents pour la recherche sociolinguistique et ethnolinguistique. Série microtoponymique. Documents de travail no 1 Larboust-Vallée d'Oueil*, Toulouse, Institut d'Etudes Méridionales.
 (1976b), et Philps, Dennis (edd), *Microdialectologie et dialectométrie des Pyrénées gasconnes. Documents pour la recherche méridionale. Série dialectologique, 1.* Toulouse, Université de Toulouse II-Le Mirail, Institut d'études méridionales.
 (1976c), *La microdialectométrie: Problèmes et méthodes*, in: Fossat, Jean-Louis/Philps, Dennis (edd), *Microdialectologie et dialectométrie des Pyrénées gasconnes. Documents pour la recherche méridionale. Série dialectologique, 1.* Toulouse, Université de Toulouse II-Le Mirail, Institut d'études méridionales.
 (1976d), *Petite introduction méthodologique aux études dialectométriques menées en Gascogne*, in: Fossat, Jean-Louis/Philps, Dennis (edd), *Microdialectologie et dialectométrie des Pyrénées gasconnes. Documents pour la recherche méridionale. Série dialectologique, 1.* Toulouse, Université de Toulouse II-Le Mirail, Institut d'études méridionales.
 (1977), *Vers un traitement automatique des données dialectologiques en dialectométrie*, in: Putschke, W. (ed), *Automatische Sprachkartographie*, Nos. 3-4 of *Germanistische Linguistik for 1977*, 311-334.
 (1978), *Etat des recherches dialectométriques sur le domaine gascon : "fonction maximale" et "fonction minimale" du dialecte*, In: Werlen, Iwar (ed.), *Probleme der schweizerischen Dialektologie*, Fribourg, Editions Universitaires, 109-139.
 (1979), *La théorie de la variation à la lumière des mesures dialectométriques de l'Atlas linguistique et ethnographique de la Gascogne*, *Cahiers de linguistique, d'orientalisme et de slavistique*, 51-64.
 (1980), *Variation et théorie des isogloses quantitatives*, *Cahiers du centre interdisciplinaire des sciences du langage. Sections de Linguistique générale et appliquée 2*, Toulouse, Université de Toulouse 2-Le Mirail, 155-164.

(1993), *Architecture d'une base de connaissance dialectale : contribution a la modélisation du fait dialectal occitan*, in: *Nazioarteko dialektologia biltzarra : agiriak. Actas del congreso internacional de dialectología*. Bilbao, Euskalzainda, 111-133.

Dans la mouvance de Jean-Louis Fossat :

Philps, Dennis :

- (1976a), *Balaguères-Bethmale-Biros : étude dialectométrique*. Université de Toulouse-Le Mirail. Institut d'études méridionales. Publié avec concours du Centre national de la recherche scientifique. Études romanes-E.R.A. 352 (Thèse de troisième cycle). Toulouse.
- (1976b), *Dialectometry*, in: Fossat, Jean-Louis/Philps, Dennis (edd), *Microdialectologie et dialectométrie des Pyrénées gasconnes. Documents pour la recherche méridionale. Série dialectologique, 1*. Toulouse, Université de Toulouse II-Le Mirail, Institut d'études méridionales.
- (1985), *Atlas dialectométrique des Pyrénées Centrales*. Thèse d'Etat. Toulouse (sous la dir. de J.-L. Fossat).
- (1986), *L'automatisation des atlas linguistiques*, in *Actes du XVIIème congrès international de Linguistique Romane*, Aix-en-Provence : Université de Provence, vol. 6, 465-486.

Jagueneau, Liliane :

- (1987), *Structuration de l'espace linguistique entre Loire et Gironde : analyse dialectométrique des données phonétiques de l'Atlas linguistique et ethnographique de l'Ouest*, Thèse, Toulouse (sous la dir. de J.-L. Fossat).

➔ **Guiter, Henri [1909-1994] :**

- (1973), *Atlas et frontières linguistiques*, in: *Les dialectes romans de France*, Paris, Editions du CNRS, 61-109.
- (1974), *Une vérification de loi linguistique par corrélation*, *Revue de Linguistique Romane* 38, 253-264.
- (1974), *Un metode estadistic en lingüística*, *Miscellanea Barcinonensia* 39, 39-65.
- (1975a), /Sardà, A., *L'atlas lingüístic de Catalunya i la fragmentació dialectal del català*, *Miscellanea Barcinonensia* 40, 93-112.
- (1975b), *Limites linguistiques du Rouergue septentrional*, in: *Etudes sur le Rouergue. Actes du 47e Congrès de la Fédération Historique du Languedoc méditerranéen-Roussillon (Rodez 1974)*, Rodez, Société des Lettres, sciences et arts de l'Aveyron, 15-20.
- (1979a), *La limite occidentale des parlers limousins d'après l'Atlas linguistique et ethnographique de l'Ouest*, in: *Actes du 102e Congrès national des Sociétés savantes (Limoges, 1977)*. *Bulletin Philologique et Historique du Comité des travaux historiques et scientifiques*, 11-20.
- (1979b), *Morphologie statistique et comparaison des langues romanes*, *Revue des langues romanes* 83, 1-26
- (1979c), *Critique et limites d'une méthode*, in: *Mélanges à la mémoire de Louis Michel*. Montpellier, Université Paul-Valéry, 261-267.
- (1980), *Limites linguistiques du Velay méridional*, in: *Actes du 52e Congrès de la Fédération Historique du Languedoc méditerranéen-Roussillon (Le Puy-en-Velay, 1979)*. *Bulletin de la Société Académique du Puy*, 109-116.
- (1981a), *Limites linguistiques dans la région bordelaise*, in: *Etudes sur la Gascogne au Moyen-Âge. Actes du 104e Congrès national des Sociétés savantes (Bordeaux)*, Paris, Bibliothèque Nationale, 59-67.
- (1981b), *Appréciation de l'importance des écarts en géolinguistique*, *Revue de Linguistique Romane* 45, 341-348.
- (1984), *Limites dialectales d'après l'Atlas de Provence*, in: *Actes du 105e Congrès national des Sociétés savantes (Grenoble, 1983)*. *Bulletin Philologique et Historique du Comité des travaux historiques et scientifiques. Année 1982-1984*, 23-31.
- (1985a), *Les méthodes quantitatives en géolinguistique sont-elles équivalentes ?*, in: *XVIe Congrès internacional de lingüística filologia romàniques ; Tom II, secció I Lingüística diacrònica i dialectologia*, Palma de Mallorca, Moll, 355-367.

- (1985b), *Corrélations de méthodes génético-linguistiques*, Cahiers de l'Institut de Linguistique de Louvain 11 (1-2), 125-145.
- (1985c) *Valeurs discriminantes linguistiques*, in: *Linguistique comparée et typologie des langues romanes. Actes du XVIIème Congrès International de Linguistique et Philologie Romanes (Aix-en-Provence, 1983)*, vol. 2, 223-249.
- (1989), *Structures dialectales à l'Est de Lyon*, in: *Actes du 112e Congrès national des Sociétés savantes (Lyon, 1987)*, Paris, Editions du CTHS, 265-277.
- (1991a), *Sur l'Atlas linguistique de l'Auvergne et du Limousin*, *Revue de Linguistique Romane* 55, 101-117.
- (1991b), *Applications d'une méthode géolinguistique en galloroman et ibéroroman*, in: Kremer, Dieter (ed), *Actes du XVIIIe Congrès International de Linguistique et de Philologie Romanes, université de Trèves 1986*, Tübingen, Niemeyer, t. III, 351-36.
- (1992), *Le sillon rhodanien*, in: *Contacts de langues, de civilisations et intertextualité. Actes du 3e Congrès international de l'AIEO, Montpellier, 1990*. Montpellier, Service des publications de la Recherche de l'Université de Montpellier, t.1, 219-230.
- (1995) *Sur la dialectologie poitevine*, in: Horiot, Brigitte (ed.), *Dialectologie et Littérature du domaine d'oïl occidental. Sur le thème des enquêtes dialectales. Actes du IVe Colloque de dialectologie du domaine d'oïl occidental (1990)*, Lyon, Université de Lyon III, 127-134.

Dans la mouvance d'Henri Guiter :

Camps, Christian :

- (1986), *Limites linguistiques dans le Biterrois*, in: *Actes du 110e Congrès national des Sociétés Savantes, Montpellier, 1985. Recherches sur l'histoire de Montpellier et du Languedoc*, Paris, Editions du CTHS, t. II, 229-239.
- (1986), *Limites linguistiques d'après l'Atlas linguistique et ethnographique du Languedoc occidental*, in: *Actes du XVIIe Congrès international de Linguistique et Philologie Romanes. Aix-en-Provence, 1983*. Université de Provence, t. 6, 120-135.
- (1991), *Limites linguistiques en Languedoc oriental*, in: *Actes du XVIIIe Congrès international de Linguistique et Philologie Romanes. Trèves, 1986*. Tübingen, Niemeyer, t. 3, 362-369.

Balmayer, Louis :

- (1971), *Quelques caractéristiques de l'occitan montpelliérain*, *Revue de Linguistique Romane* 35, 90-96.
- (1984), *Méthode de l'indice relatif moyen de cohérence appliquée à l'atlas linguistique du Montpelliérain*, in: Goebel, Hans (ed), *Dialectology*, Bochum, Brockmeyer, 70-81.
- (2000), *Lignes dialectométriques et isoglosses sur le domaine du Montpelliérain*, *Revue de linguistique romane* 64, 409-426.

➔ **Goebel, Hans** (travaux spécifiques ou au moins englobant et montrant la partition de l'occitan) :

- (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.
- (2004), *Das "Croissant" - eine Nachschau im Abstand von 90 Jahren (mit zwei dialektometrisch erstellten Farbkarten)*, in: Noll, V./Thiele, S. (edd), *Sprachkontakte in der Romania. Zum 75. Geburtstag von Gustav Ineichen*, Tübingen, Niemeyer, 159-172.
- (2009), *Quelques coups d'oeil dialectométriques sur l'Atlas linguistique de la France : structures de surface et structures de profondeur*, in: Dalbera-Stefanaggi, Marie-José/Simoni-Aurembou, Marie-Rose (edd), *Images de la langue : représentations spatiales, sémantiques et graphiques*, Paris, Editions du CTHS, 39-60.
- (2011a), *Introduction aux problèmes et méthodes de l'École dialectométrique de Salzbourg (avec des exemples gallo-, italo- et ibéroromans)*, in: Alvarez Pérez, Alfonso/Carrilho, Ernestina/Magro, Catarina (edd), *Proceedings of the International Symposium on Limits*

and Areas in Dialectology (LimiAr), Lisbon 2011, Lisboa, Centro de Linguística da Universidade de Lisboa, 117-166.

- (2011b), *Áreas, fronteras, similitudes y distancias: lección breve de geolingüística cuantitativa*, in: Andrés Díaz, Ramón de (coord.), *Lengua, ciencia y fronteras*, Uviéu/Oviedo: Ediciones Trabe / Universidá d'Uviéu [Anexos de Revista de Filoloxía Asturiana, II], 11-33.
- (2014), /Pavel Smečka, *L'analyse dialectométrique des cartes de la série B de l'ALF*, *Revue de linguistique romane* 78, 439-497.
- (2016), *Du chemin parcouru entre Coquebert de Montbret père et fils et la plus récente dialectométrie: une reconsidération critique*, in: Buchi, Éva/Chauveau, Jean-Paul/Pierrel, Jean-Marie (edd), *Actes du XXVIIe Congrès international de linguistique et philologie romanes (Nancy 2013)*, Strasbourg, ELiPhi, vol. I, 29-67.

Dans la mouvance de John Nerbonne :

Brun-Trigaud, Guylaine/Malfatto, Albert/Sauzet, Maguelone :

- (2020), *Essai de typologie des aires lexicales occitanes : regards dialectométriques*, in: Courouau, Jean-François (ed), *Fidélités et dissidences. Actes du XIIe congrès de l'Association Internationales d'Études Occitanes (Albi, 2017)*, Toulouse, Section française de l'Association internationale Etudes occitanes, vol. 1, 169-179.