

ACTES DU **XII**^{e/n} **CONGRÈS**
de l'Associacion Internacionala d'Estudis Occitans
de l'Association Internationale d'Études Occitanes
ALBI, 10-15/07 2017

édités par Jean-François Courouau
en collaboration avec David Fabié
editats per Joan-Francés Courouau
en colaboracion amb Dàvid Fabié

Fid elitats
e t **é**lités
dissi déncias
dences

Section française de l'Association internationale d'études occitanes

SFAIEO

Vol. 1



Guylaine BRUN-TRIGAUD
Albert MALFATTO
Université Nice Sophia-Antipolis – CNRS-BCL

Maguelone SAUZET
Université de Neuchâtel
Centre de Dialectologie et d'Étude du Français Régional

Essai de typologie des aires lexicales occitanes : regards dialectométriques

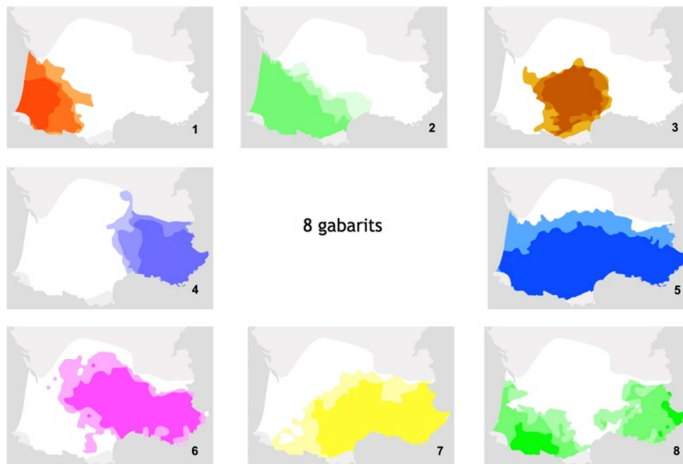
1. En amont

Dans Brun-Trigaud & Malfatto (2013), nous avons proposé un essai de typologie des aires lexicales occitanes. Nous y avons étudié les données dialectales issues des atlas linguistiques de la France par régions, qui, à l'époque comme aujourd'hui, sont disponibles dans la base de données en ligne du *Thesaurus Occitan*. Nous avons travaillé sur un corpus de 132 notions/cartes communes aux six atlas du domaine (voir NALF).

Six types principaux de distribution des aires avaient été déterminés, regroupant une grande partie du corpus, allant du plus simple « unilemme » (ex. « vache », avec un terme unique *vaca*) au plus complexe (ex. « coccinelle » et ses 42 types différents).

Puis nous avons déterminé huit gabarits principaux pour les aires lexicales (voir **Carte 1**) :

- 1) un espace occupant l'ouest de la Garonne, que nous avons appelé « gascon » (mais peut-être il serait plus convenable de dire « aquitain »),
- 2) un espace « aquitano-pyrénéen » débordant vers la Méditerranée, donc un peu plus large que le premier,
- 3) un espace « central » occupant une aire autour du Languedoc,
- 4) un espace « oriental » ou provençal au sens large, qui souvent déborde sur les parlers franco-provençaux,
- 5) un espace « transversal » qui exclut les parlers limousins et auvergnats
- 6) un espace en « diagonal » excluant les parlers gascons,
- 7) un espace « méditerranéen » épousant l'arc de cercle de la Méditerranée,
- 8) un espace « périphérique » regroupant des aires pyrénéennes et alpines élargies.



Carte 1 : huit gabarits

2. De nouvelles données et un nouvel outil

La publication du dernier volume de l'*Atlas de la Provence* (2016) nous permettant d'ajouter un grand nombre de cartes communes, nous avons décidé de reprendre ce corpus, avec la collaboration de Maguelone Sauzet. Et surtout, nous avons maintenant la possibilité d'utiliser un nouvel outil, une application en ligne du nom de *Gabmap*¹, ouverte à tous, dont le but essentiel est de faire de la dialectométrie, mais pas seulement... (Leinonen *et al* 2016).

L'utilisation de l'application est relativement simple : il suffit d'importer un fichier de GoogleMaps avec les coordonnées géolocalisées et un fichier de données au format texte correspondant aux points choisis.

Ici, nous avons retenus 645 points d'enquêtes sur les 684 initiaux, car nous avons éliminé, hélas, un certain nombre de points trop lacunaires pour être pris en compte, notamment du côté de la Provence.

Gabmap se charge automatiquement de réaliser la polygonisation du domaine, ce qui lui permet par la suite de présenter les résultats sous forme de cartes à aires.

Notre nouveau corpus comporte 232 cartes, soit 100 de plus que pour celui de 2013 : elles sont très variées, allant de la carte unilemme à la carte comportant plus de 70 lemmes pour le « cuvier ».

3. Acquisition des nouvelles données et lemmatisation

Afin de pouvoir observer les nouvelles données lexicales dont nous disposons, avec notre nouvel outil informatique *Gabmap*, nous avons construit un corpus enrichi en commençant par un travail de lemmatisation des formes phonétiques extraites de l'*ALP 4*, saisies et converties en API (Bouvier *et al* 2016).

Il était nécessaire de transformer ces données en matériau lexical qui convienne à la fois pour l'implémentation dans le THESOC et pour la mise à l'épreuve avec l'essai de typologie lexicale de 2013.

Nous présenterons ici plus particulièrement le travail de lemmatisation, consistant à choisir les formes de référence sous lesquelles on rassemble des variantes phonétiques rencontrées dans l'atlas.

Nous avons observé les conventions de transcription préconisées par le THESOC. Tout d'abord nous avons choisi comme lemmes lexicaux les formes retenues dans le *Dictionnaire Occitan-Français* (DOF) d'Alibert ou, en cas d'absence, une forme reposant sur les principes de la graphie classique. Pour rassembler les formes sous une même « étiquette lexicale », nous les avons « déshabill[ées] de [leur] variation dialectale » (Sauzet/Brun-Trigaud 2013, 21) : à l'intérieur d'une langue, on considère que des formes remontant au même étymon et partageant une même construction morphologique relèvent du même lemme. Ainsi, pour l'occitan, [k'abrɔ, f'æbr, f'j'ɛb, ts'abrɔe, f'j'urɔ] sont regroupées sous le lemme « cabra ».

Les règles semblent simples. Cependant, plusieurs difficultés conceptuelles mises en évidence demeurent (Sauzet/Brun-Trigaud 2013, 21) et l'on ne peut pas encore confier ces tâches à une machine, bien qu'il existe des tentatives pour les automatiser (cf. l'algorithme de J-P. Dalbera et D. Strazzabosco (Olivieri *et al* 2017) ou des projets exploratoires pour la reconnaissance automatique/OCR des formes phonétiques des atlas imprimés). C'est donc manuellement, forme après forme, que nous avons procédé à la lemmatisation du corpus.

Les difficultés rencontrées pour la lemmatisation des formes de l'*ALP 4* concernaient les problématiques habituelles. Notamment, les « zones grises » entre phonétique et morphologie, où il faut établir la forme lexicale la plus adaptée entre phénomènes d'aphèreses, d'interfixations sans valeur sémantique, d'agglutination ou de déglutination... Faut-il, par exemple, rassembler *gulha* et *aqulha* < lat. ACŪCŪLA ? Il a aussi fallu trancher parmi les formes hybrides rencontrées à la

¹ Son nom est forgé sur l'anglais *gab* « bagout, éloquence » et *map* « carte » : "les cartes qui parlent bien".

périphérie du domaine occitan ou les formes isolées dont on ne peut proposer qu'une prudente translittération.

D'un point de vue technique, il a également fallu adapter certaines spécificités inhérentes aux atlas linguistiques à l'outil informatique que nous voulions utiliser : supprimer les cartes et les points d'enquête trop lacunaires. Nous sommes également restés conscients que nous avons détourné *Gabmap* des fonctions pour lesquelles il a été conçu en lui faisant réaliser des cartes de distribution lexicale : l'application ne pouvait donc pas répondre à nos besoins les plus spécifiques.

Cependant, la lemmatisation lexicale permet de proposer un renouveau méthodologique intéressant. En effet, fixer une forme occitane comme lemme lexical « étiquette » est un instrument de neutralité interprétative, par rapport au choix de l'étymon, par exemple, qui peut rendre invisible des évolutions morphologiques internes à la langue ou poser des difficultés de reconstruction. Le choix d'un lemme occitan, même s'il nous confronte aux difficultés que nous avons évoquées plus haut et à la question du choix d'une norme graphique, nous permet de manipuler le lexique occitan, sans se perdre dans la dispersion des formes phonétiques. Au lieu de « tuer » la variété dialectale, un lemme occitan de référence permet finalement de conserver une certaine objectivité : « Les lemmes sont utiles pour manipuler la langue quand il s'agit [...] de discuter de la variation et de l'histoire des mots » (Sauzet/Brun-Trigaud 2013, 7), ils peuvent être considérés comme « un outil commode pour trier et référencer la masse de leurs variantes dialectales. » (Sauzet/Brun-Trigaud 2013, 32).

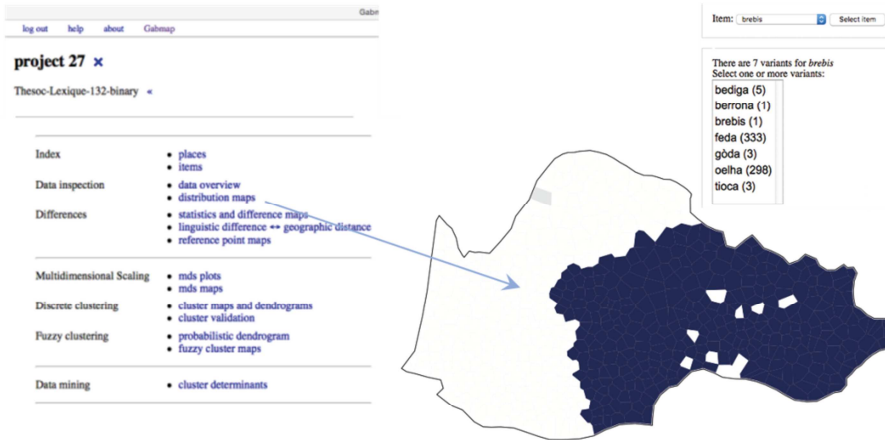
4. Création des cartes de distribution des aires lexicales : deux fonctions de Gabmap pour la cartographie

L'outil informatique à notre disposition s'est révélé être un véritable « couteau suisse », qui a accompagné les différentes étapes de notre projet. Nous allons voir comment l'application web *Gabmap* se prête à d'autres usages que sa fonction dialectométrique initiale.

Dans la deuxième partie de notre travail, nous avons utilisé *Gabmap* pour produire rapidement et facilement des cartes de distribution lexicale relatives à une notion (ou « carte », au sens dans lequel on l'entend traditionnellement dans les atlas linguistiques) et observer la diffusion d'un lexème dans l'espace occitan.

Grâce à deux fonctionnalités de *Gabmap*, il est possible de créer des cartes à aires, soit pour un seul lemme lexical, soit pour l'ensemble des lemmes relatifs à une notion. Une des différences entre ces deux fonctions réside dans le degré d'intervention, plus ou moins grand, dont on dispose sur les données importées dans l'application.

Tout d'abord, la fonction appelée « *distribution map* » nous permet d'isoler une notion lexicale et de projeter sur un fond de carte vierge l'aire de diffusion des lemmes relatifs à cette notion. Grâce à une liste déroulante, l'ensemble des « cartes » disponibles donne accès, pour chacune d'elles, à tous les lemmes lexicaux qui ont été saisis, en précisant leur fréquence respective. Ainsi, pour la carte « brebis », on rencontre *feda* (333 fois), *oelha* (298 fois), *bediga* (5 fois)... En sélectionnant d'un clic ou par une expression régulière dans cette liste, on projette la distribution d'un seul ou de plusieurs de ces lemmes, que l'on peut ensuite exporter dans différents formats de sortie. (**Carte 2**)

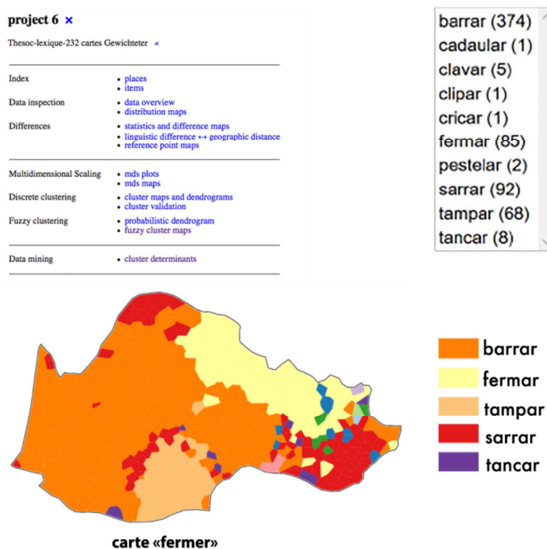


carte «brebis» : aire *fedas*

Carte 2 : Distribution de l'aire lexicale de *fedas* dans la carte « brebis » et interface de la fonction *distribution map*

Ainsi, l'application Gabmap permet de jouir d'une liberté importante dans la construction des cartes d'aires lexicales : choisir de faire certains regroupements, exclure ou inclure des formes lexicales. On peut, par exemple, rassembler toutes les formes provenant du même étymon ou bien toutes les formes qui présentent une construction morphologique identique. Par exemple, pour la carte « cuve à lessive », on peut rassembler *bugièr*, *bugadièr* et *bugadièra*, issus d'un étymon germanique *BÛKÒN (FEW 15/2, 9a).

À l'aide d'une autre fonction, « *cluster maps* », il est possible de laisser Gabmap créer tout seul des cartes de distribution complètes avec l'ensemble des formes relatives à une notion lexicale. Gabmap ne peut cependant créer une carte de distribution lexicale complète que pour un micro-corpus composé d'une seule notion et qui ne devra pas comporter plus de 12 lemmes. Il n'est alors pas possible d'intervenir sur la représentation cartographique, de modifier les formes regroupées (ni de choisir les couleurs attribuées à chaque aire lexicale) (**Carte 3**).



carte «fermer»

Carte 3 : Carte « fermer » construite automatiquement avec la fonction *cluster maps*

5. Les nouvelles cartes confrontées aux aires lexicales de 2013 : une coïncidence réelle mais pas systématique

L'application *Gabmap* est conçue pour une approche dialectométrique et quantitative : il s'agit de faire du *data mining* à l'intérieur de notre corpus. Mais grâce aux deux fonctions précédemment évoquées, il est possible de conserver une approche plus conventionnelle, « qualitative », dans le sens où l'on peut observer de près les nouvelles cartes ajoutées au corpus de 2013, en isolant chaque lexème et en cartographiant sa diffusion dans l'espace occitan.

La première observation que nous pouvons faire, après avoir passé en revue les nouvelles cartes, est encourageante et positive : dans notre corpus varié, il y a bien une coïncidence avec les gabarits de 2013 et nous retrouvons les aires lexicales identifiées alors.

Cependant, cette adéquation n'est pas systématique et certaines distributions sont surprenantes. La carte « allumer », par exemple, oppose les lemmes *alumar*, *alucar* et *atubar* selon une représentation que l'on s'explique difficilement avec la grille d'analyse traditionnelle des aires dialectales et des parlers occitans, héritée de la classification de Pierre Bec (1995).

Certaines aires lexicales sont bien plus récurrentes que d'autres. Cela n'est certainement pas imputable seulement à un effet de corpus : il est possible de voir là une cohésion plus forte de certains dialectes de l'espace occitan, du point de vue du lexique.

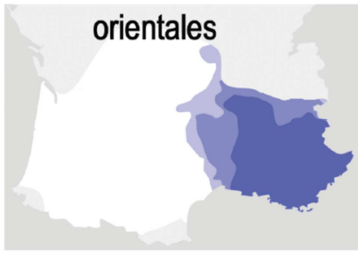
La spécificité lexicale du triangle aquitain, en particulier, est souvent mise en évidence, comme le montre pour la carte « mordre » la distribution du verbe *nhacar*. Nous avons également vu apparaître dans notre corpus des aires périphériques, qui constituent des espaces de conservation étymologique aux sud-est et sud-ouest du domaine occitan. La diffusion du lemme *escoba*, < lat. SCOPA, de la carte « balai » en est un exemple.

L'aire orientale est bien illustrée dans la carte « poêle » avec le lemme *sartan* (opposé à *padena* et *padèla*) ou bien dans la carte « cheveu », avec le lexème *cheveu* qui s'oppose en Provence au *pel* présent dans tout le reste de l'espace occitan. **(voir Carte 4)**

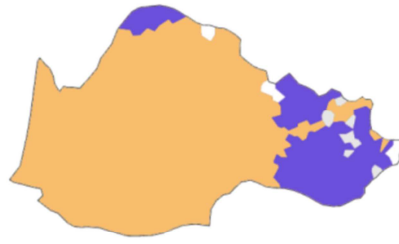
Au contraire, certains gabarits n'ont pas trouvé de correspondance dans notre nouveau corpus : nous n'avons pas retrouvé l'espace transversal, par exemple, qui exclut le domaine occitan septentrional.

Il faut également observer, pour de nombreux cas, des aires lexicales incomplètes, qui se rapprochent beaucoup des gabarits distingués de 2013 sans en épouser les contours, comme pour la diffusion de la *gaspà* de la carte « petit lait », un peu trop réduite pour correspondre au gabarit de l'aire centrale. Parfois, d'autres lexèmes semblent avoir mité une aire lexicale autrefois plus étendue, comme pour la diffusion de la *tina* de la carte « cuvier » qui semble avoir pu s'étendre autrefois sur une large diagonale excluant les parlers gascons. **(voir Carte 4)**

L'éclairage cartographique que *Gabmap* nous propose sur le lexique occitan, révèle des configurations intéressantes. On perçoit la complexité des mécanismes et des mouvements lexicaux qui se sont joués à l'intérieur même de la langue occitane : glissements sémantiques, innovations ou conservations collectives, influences des superstrats, substrats ou adstrats... Une prochaine étape pourrait consister à s'appuyer sur notre approche de typologie lexicale pour mieux comprendre la structure de l'espace occitan à travers le lexique et sa capacité autonome de fonctionnement.



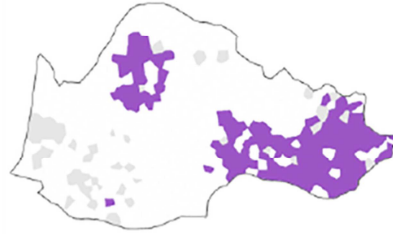
orientales



carte «cheveu» : aire *cheveu* (Vs. *pel*)



diagonales



carte «cuvier à lessive» : aire *tina*

Carte 4 : Deux comparaisons de distribution lexicale avec les gabarits de la typologie de 2013

6. Pistes de réflexions et extrapolations : de nouvelles lectures des aires lexicales grâce à Gabmap

En observant de près les aires de diffusion de chaque lexème, à l'intérieur d'une même notion, parmi une grande masse de données lexicales occitanes, nous avons conservé une approche géolinguistique conventionnelle. Notre outil, *Gabmap*, a cependant apporté un renouveau conceptuel, car il ouvre de nouvelles possibilités pour manipuler le lexique occitan.

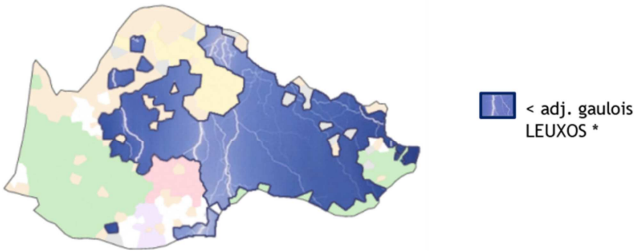
En effet, comme nous l'avons déjà indiqué « [...] aujourd'hui les outils informatiques et notamment graphiques permettent de faire sauter les verrous de la lisibilité des cartes de synthèse ou cumulatives et même de révéler des phénomènes restés jusqu'ici invisibles » (Brun-Trigaud 2012, 77).

À partir des fonctions de cartographie de *Gabmap*, nous avons pu proposer un éclairage original sur le lexique occitan. En rapprochant des lemmes de façon un peu inattendue, nous avons parfois observé une contiguïté, ou au contraire une concurrence, intéressantes dans les aires distributionnelles.

Ainsi, *Gabmap* permet d'essayer des extrapolations originales en rassemblant sur une carte des lemmes à priori disparates. Par exemple, pour la mosaïque de la carte « aubépine », on peut visualiser une aire de distribution contiguë constituée de lemmes *bartàs blanc*, *espina blanca*, *boisson blanc*... Il s'agit des appellations qui ont cherché à remotiver le nom à travers le trait sémantique de blancheur, originellement présent dans le mot étymologique (< latin ALBUS) perdu ou remotivé.

Nous avons également rendu visible la pertinence du rapprochement des formes *eslauç*, *iglauç*, *liuçada* en suivant les hypothèses de G. Tuailon (1981) ou E. Schüle (GPSR, VI, 255b, entrée *èludzi*). Malgré des traitements phonétiques complexes, il est possible de voir derrière ces formes un même étymon gaulois, l'adjectif *LEUXOS signifiant « clair, brillant » (FEW, 5, 263). En les projetant sur Gabmap, l'on a pu voir apparaître une aire diagonale. Cette configuration rappelle celles des cartes interprétées construites à partir de l'ALF et de l'*Atlas Linguistique de la Basse-Bretagne*, pour des notions comme « son de blé », « aulne » ou « dos », qui « semblent démontrer

l'existence d'une très ancienne continuité linguistique » entre les langues celtiques et romanes, du point de vue du lexique du moins (Le Dù / Brun-Trigaud 2013, 30-33). (**Carte 5**)



carte «éclair» : extrapolation étymologique

Carte 5 : Exemple d'extrapolation étymologique sur la carte « éclair »

La diffusion d'un mot ne peut pas être appréhendée comme un phénomène indépendant : les mots coexistent et sont corrélés aux pratiques humaines ainsi qu'aux réalités concrètes qu'ils désignent. Il faudra donc toujours rechercher les spécifications lexicales, interroger l'exclusivité lexicale d'une notion, la polysémie et la synonymie. Cela pourrait donner lieu à un renouveau de la lexicographie occitane, grâce à nos outils informatiques pour le traitement et la visualisation des données, en proposant une approche à la fois onomasiologique et sémasiologique.

Par exemple, *bofiga* de la carte « vessie (de porc) », qui correspond bien au modèle des aires périphériques de notre typologie, a également le sens de « cloque, grosseur sur la peau, phlyctène » attesté dans le *DOF*. De même, la répartition étonnante dans la carte « bouche » du lemme *gørja* qui coupe comme une faucille la zone occupée par le lemme *boca* peut être comparée avec la carte « lèvres », où *boca*, par métonymie, est aussi employé. Il faudra également interroger les contextes dans lesquels *gørja* peut désigner la « bouche » et *boca* les « lèvres », s'il y a d'autres raisons sémantiques qui expliquent cette diffusion. Les données des atlas, ici, ne suffisent pas et ce sont d'autres matériaux qui doivent nous aider à comprendre les interrogations que *Gabmap* nous a permis de mettre en lumière dans notre corpus.

7. Quelques mots sur la dialectométrie

Comme nous l'avons dit plus haut, le but principal de *Gabmap* est de fournir des résultats de dialectométrie qui est une discipline linguistique, initiée par Jean Séguy (1971, 1973) et Henri Guiter (1973) dans les années 1970 à Toulouse.

Elle permet de faire des mesures quantifiées à partir des oppositions entre plusieurs points de la carte dialectale.

Plus simplement, on compare les réponses de chaque point d'un réseau à celles de tous les autres points et on quantifie le nombre de différences trouvées.

Hélas, cette discipline a presque disparu du champ de la dialectologie française après la mort de Jean Séguy en 1976, même s'il ne faut pas oublier les travaux de Henri Guiter ou ceux de Jean-Louis Fossat (notamment Fossat, 1975).

C'est à Salzbourg, avec Hans Goebel, que la discipline a repris son essor. On lui doit l'idée d'ajouter la polygonisation pour la visualisation des données. Il a publié de nombreux travaux que l'on peut consulter en ligne, notamment sur l'*Atlas Linguistique de la France*.

Puis, une nouvelle école de dialectométrie est apparue dans les années 2000 sous l'égide de John Nerbonne, basé à Groningue au Pays Bas. Ce dernier a introduit une nouvelle façon de

traiter les données avec l'algorithme de Levenshtein et surtout a œuvré pour la mise à disposition pour tous d'abord du logiciel de traitement, puis de l'application en ligne que nous avons utilisée.

Néanmoins nous précisons que dans le cas présent du lexique, nous ne travaillons qu'en binaire et non pas avec l'algorithme de Levenshtein. *Gabmap* donne la possibilité de choisir entre plusieurs façons de traiter les données, au moment de leur implémentation.

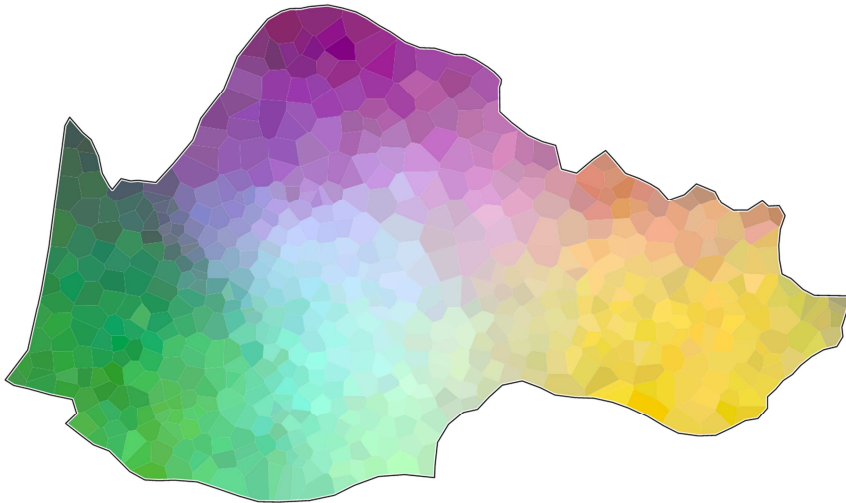
8. Résultats obtenus avec Gabmap

Dans l'espace imparti de cet exposé, nous ne présenterons que deux méthodes de visualisation du résultats des algorithmes de statistiques sur les données : d'une part, MDS, Multidimensional Scaling et d'autre part, Fuzzy Clustering pour les 150 000 data intégrés dans *Gabmap*.

En préambule, il faut d'abord donner quelques explications sur la classification ascendante hiérarchique (CAH) : le principe de l'analyse en groupes (ou clusters) est basé sur les algorithmes de classification ascendante hiérarchique (CAH). Les algorithmes opèrent, chacun à leur façon, car il en existe de plusieurs types, des agglomérations successives des éléments en fonctions de leurs affinités pour aboutir à un arbre de classification clairement structuré que l'on appelle dendrogramme.

Selon les types d'algorithmes utilisés, certains minimisent la distance entre les points et maximisent la distance entre les groupes pour faire émerger des frontières plus nettes, comme Fuzzy Clustering, alors que d'autres, comme Multidimensional Scaling (MDS), se focalisent davantage sur le continuum dialectal.

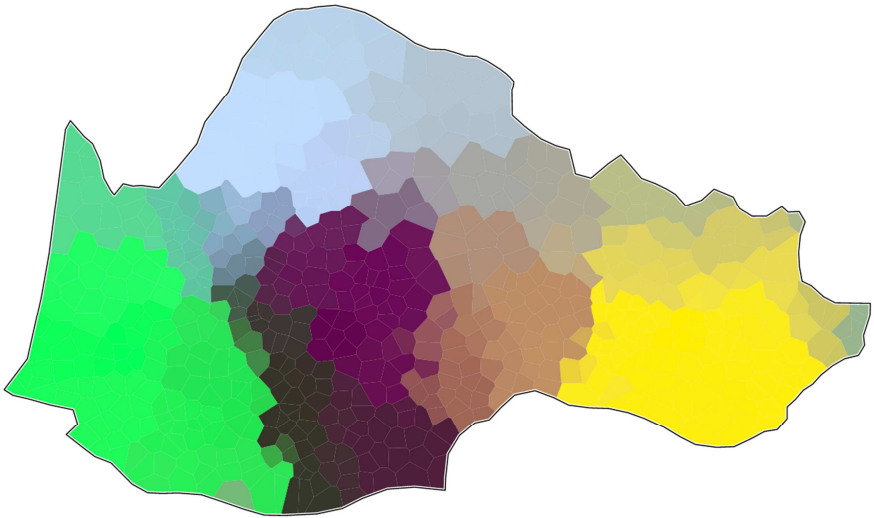
Ce dernier permet donc de montrer les relations linguistiques entre tous les points pour l'ensemble des données. Chaque point ayant une couleur spécifique, les points semblables du point de vue linguistique ont des couleurs proches. En principe, ce modèle de visualisation ne montre pas de rupture brusque entre les aires dialectales (cf. **Carte 6**).



Carte 6 : Multidimensional Scaling

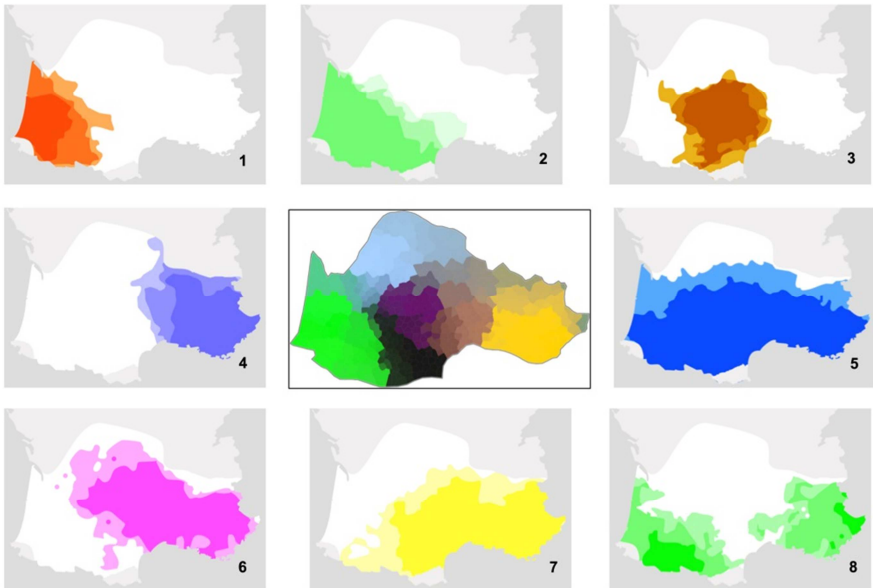
Cette carte montre un visage connu de la répartition des dialectes occitans : un groupe gascon (dégradé de verts), un groupe languedocien (dégradé de turquoise), un groupe arverno-limousin (dégradé de mauve) et un groupe provençal (dégradé de jaune).

Au contraire, Fuzzy Cluster est un outil de synthèse qui permet de mettre en avant les aires linguistiques tout en préservant l'aspect des zones servant de transition (**Carte 7**).



Carte 7 : Fuzzy Cluster

Dans cette carte aux aires plus marquées, on retrouve peu ou prou les gabarits que nous avons définis de manière plus empirique lors de notre expérience précédente (**Carte 8**).



Carte 8 : Gabarits et Fuzzy Clusters

Conclusions

En dépit du fait que nous nous inscrivions dans la dissidence vis-à-vis des pratiques actuelles de la dialectologie française, en pratiquant la dialectométrie qui, par ailleurs, connaît de belles perspectives, il nous semble intéressant de poursuivre cette démarche, d'une part, en continuant la lemmatisation des cartes communes disponibles, d'autre part, en étendant l'aire aux atlas périphériques qui comprennent quelques points occitans et enfin, établir des cartes dans la perspective d'un projet de synthèse sur le lexique en occitan.

Références bibliographiques

Abréviations

- ALF = GILLIERON, Jules / Edmond EDMONT (1902-1910). *Atlas linguistique de la France*, Paris, Champion.
- NALF = Nouveaux Atlas Linguistiques de la France
- ALAL : POTTE Jean-Claude (1975-1992). *Atlas Linguistique et Ethnographique de l'Auvergne et du Limousin*, Paris, Éditions du CNRS, 3 vol.
- ALG : SEGUY, Jean (1954-1974). *Atlas Linguistique et Ethnographique de la Gascogne*, Paris, Éditions du CNRS, 6 vol.
- ALLOc : RAVIER, Xavier (1978-1993). *Atlas Linguistique et Ethnographique du Languedoc Occidental*, Paris, Éditions du CNRS, 4 vol.
- ALLOr : BOISGONTIER, Jacques (1981-1986). *Atlas Linguistique et Ethnographique du Languedoc Oriental*, Paris, Éditions du CNRS, 3 vol.
- ALMC : NAUTON, Pierre (1957-1963). *Atlas Linguistique et Ethnographique du Massif Central*, Paris, Éditions du CNRS, 4 vol.
- ALP : BOUVIER, Jean-Claude / Claude MARTEL (1975-1986). *Atlas Linguistique et Ethnographique de la Provence*, Paris, Éditions du CNRS, 3 vol. et BOUVIER, Jean-Claude / Claude MARTEL / Guylaine BRUN-TRIGAUD (2016). *La langue d'oc telle qu'on la parle. Atlas Linguistique de la Provence*, Les Alpes de Lumière, 2016.
- DOF = ALIBERT, Louis (1966). *Dictionnaire occitan-français, sur la base des parlers languedociens*, Toulouse, Institut d'Estudis Occitans.
- FEW = von WARTBURG, Walter (1922-2002). *Französisches Etymologisches Wörterbuch. Eine darstellung des galloromanischen sprachschatzes*, Bonn/Leipzig/Bâle, Teubner/Klopp/Zbinden, 25 vols.
- GPSR = GAUCHAT, Louis et al. (1924-). *Glossaire des patois de la Suisse romande*, Neuchâtel, Attinger.

Études

- BEC Pierre (1963). *La langue occitane*, Paris, Presses Universitaires de France, coll. « Que sais-je ? » [⁶1995].
- BRUN-TRIGAUD, Guylaine (2012). « Essai de typologie des aires lexicales dans l'Atlas Linguistique du Centre », *Annales de Normandie. Association Les Annales de Normandie, Mélanges offerts à Catherine Bougy*, 62/2, 77-93.
- BRUN-TRIGAUD, Guylaine (2014). « Un usage particulier de l'algorithme de Damerau-Levenshtein dans le domaine occitan », in Federica Diémoz / Dorothée Aquino-Weber (eds.), *Toujours langue varie. Mélanges de linguistique historique du français et de dialectologie galloromane offerts à M. le Professeur Andres Kristol par ses collègues et anciens élèves*, Genève, Droz, 127-147.
- BRUN-TRIGAUD, Guylaine (2016). « Vers un renouveau des atlas linguistiques régionaux ? », *Géolinguistique* 16, 7-20.
- BRUN-TRIGAUD, Guylaine / Albert MALFATTO (2013). « Limites dialectales vs limites lexicales dans le domaine occitan : un impossible accord ? », in Ernestina Carrilho / Catarina Magro / Xosé Álvarez Perez (eds), *Current Approaches to Limits and Areas in Dialectology*, Cambridge Scholars Publishers, 293-310.
- BRUN-TRIGAUD, Guylaine / Tanguy SOLLIEC / Jean LE DÛ (2016). « A New Dialectometric Approach Applied to the Breton Language », in John Nerbonne / Marie-Hélène Côté / Remco Knooihuizen (eds.), *The Future of Dialects*, Berlin, Language Science Press, 135-153.
- FOSSAT, Jean-Louis (1975). *Dialectologie, dialectométrie, ethnosémantique, sociolinguistique*, Toulouse, Université de Toulouse II-Le Mirail.

- GUI TER, Henri (1973). « Atlas et frontières linguistiques », in Georges Straka / Pierre Gardette (éds.), *Les dialectes romans de la France à la lumière des atlas régionaux*, Paris, Éditions du CNRS, 61-109.
- LE DÙ, Jean / Guylaine BRUN-TRIGAUD (2013). « Les aires dialectales gallo-romanes à la lumière du celtique », in Ernestina Carrilho / Catarina Magro / Xosé Álvarez (éds.), *Current Approaches to Limits and Areas in Dialectology*, Cambridge Scholars Publishers, 27-48.
- LEINONEN, Therese / Çağrı ÇÖLTEKİN / John NERBONNE (2016). « Using Gabmap », *Lingua* 178, 71-83.
- NERBONNE, John / Rinke COLEN / Charlotte GOOSKENS / Peter KLEIWEG / Therese LEINONEN (2011). « Gabmap. A Web Application for Dialectology », *Dialectologia Special Issue II*, 65-89.
- NERBONNE, John / Wilbert HEERINGA (2001). « Computational Comparison and Classification of Dialects », *Dialectologia et Geolinguistica* 9, 69-83.
- NERBONNE, John / Wilbert HEERINGA (2010). « Measuring Dialect Differences », in Jürgen-Erich Schmidt / Peter Auer (éds.), *Language and Space. An International Handbook of Linguistic variation, I, Theories and Methods*, Berlin, Mouton/De Gruyter, 550-567.
- OLIVIERI, Michèle / Sylvain CASAGRANDE / Guylaine BRUN-TRIGAUD / Pierre-Aurélien GEORGES (2017). « Le Thesaurus occitan dans tous ses états », *Revue Française de Linguistique Appliquée* XXII/1, 89-102.
- ROHLFS, Gerhard (1964). « La langue d'oc, carrefour des langues romanes », *Revue de linguistique romane* 28, 95-102.
- SAUZET, Patrick / Guylaine BRUN-TRIGAUD (2013). « Le Thesaurus Occitan : entre atlas et dictionnaire », *Corpus* 12, 105-140.
- SÉGUY, Jean (1971). « La relation entre la distance spatiale et la distance lexicale », *Revue de linguistique romane* 35, 335-357.
- SÉGUY, Jean (1973). « La dialectométrie dans l'Atlas linguistique de la Gascogne », *Revue de linguistique romane* 37, 1-24.
- TUAILLON, Gaston (1981). « Atlas Linguarum Europae : "Eclair" », *Vox Romanica* 40, 1-12.

Liens

Gabmap : <http://www.gabmap.nl>

GOEBL, Hans (bibliographie) : https://www.sbg.ac.at/rom/people/prof/goebl/dm_publi.htm

Thesaurus Occitan : <http://thesaurus.unice.fr/index.html>