



**HAL**  
open science

## Comparaison dialectométriques de parlers du Croissant avec d'autres parlers d'oc et d'oïl

Philippe Boula de Mareüil, Gilles Adda, Lori Lamel

### ► To cite this version:

Philippe Boula de Mareüil, Gilles Adda, Lori Lamel. Comparaison dialectométriques de parlers du Croissant avec d'autres parlers d'oc et d'oïl. *Le Croissant linguistique entre oc, oïl et francoprovençal : des mots à la grammaire, des parlers aux aires*, 2021. hal-03318765

**HAL Id: hal-03318765**

**<https://hal.science/hal-03318765v1>**

Submitted on 10 Aug 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# **Comparaison dialectométriques de parlers du Croissant avec d'autres parlers d'oc et d'oïl**

Philippe Boula de Mareüil, Gilles Adda, Lori Lamel

Université Paris-Saclay, CNRS, LIMSI, Orsay, France

## **Résumé**

Le passage à l'écrit des dialectes du Croissant, comme pour nombre de langues régionales ou minoritaires, soulève d'importants problèmes théoriques et pratiques. La transcription phonétique n'est pas plus simple ni plus rapide. Nous proposons dans ce travail de comparer les transcriptions orthographiques d'une part, phonétiques d'autre part, d'enregistrements audio actuels, réalisés à l'intérieur et à proximité du Croissant. Plus précisément, nous avons demandé à une trentaine de locuteurs des domaines d'oc et d'oïl de traduire un même texte (la fable d'Ésope « La bise et le soleil ») dans leur dialecte ou langue régionale. Sur cette base, nous avons calculé un taux symétrisé de différences entre variétés, que nous avons représenté graphiquement par des techniques d'analyse de données. Les résultats que nous présentons permettent de visualiser une sorte de distance entre les parlers du Croissant, les (autres) parlers d'oc et les parlers d'oïl.

**Mots-clés :** dialectométrie, transcription phonétique, alignement automatique, géolinguistique, langues en contact et en danger

## **Introduction**

La transcription phonétique de toute langue, de tout dialecte, est une tâche difficile et coûteuse en temps (Pike, 1982). La transcription orthographique, pour les langues régionales ou minoritaires, de France ou d'ailleurs, soulève également d'importants problèmes théoriques et pratiques. Il en va ainsi des parlers du Croissant, terme utilisé par Ronjat (1913) pour dénommer cette région du centre de la France, en raison de sa forme, qui constitue une aire de transition entre domaines d'oc et d'oïl. L'hybridation des formes, les fluctuations, les interférences que l'on observe d'un village à l'autre, à quelques kilomètres de distance, font de cette zone de turbulence linguistique un véritable laboratoire pour étudier la variation (Brun-Trigaud, 1992 ; Quint, 1998). La délimitation du domaine d'oc occupe la presque totalité de l'introduction de l'essai de Ronjat (1913) et l'objet même du rapport de Tourtoulon et Bringuier (1876), lequel aboutit à une cartographie de l'aire qui nous intéresse ici en sous-aires plus ou moins d'oc ou d'oïl suivant l'échelle : Oc – A1 –

A2 – A3 – B1 – B2 – Oïl. La polémique faisait rage à l'époque (Rousselot, 1891), et le texte où G. Paris (1888) exposait sa doctrine continuiste est assez connu.

G. Paris (1888 : 172) rejetait la notion même de frontière dialectale, voire la séparation entre langues d'oïl et d'oc : « la science nous apprend qu'il n'y a pas deux Frances, qu'aucune limite réelle ne sépare les Français du Nord de ceux du Midi et que d'un bout à l'autre du sol national nos parlers populaires étendent une vaste tapisserie dont les couleurs variées se fondent sur tous les points en nuances insensiblement dégradées. » Son élève J. Gilliéron, un des auteurs de l'*Atlas linguistique de la France* (Gilliéron & Edmont, 1902–1910) partageait du reste cette vision continuiste. Face à un continuum linguistique qui partage une importante base, où tout n'est que fluctuation autour d'un socle commun, le concept de frontière est certes à manipuler avec précaution (Saussure, 1916 ; Kloss, 1967). Tourtoulon et Bringuier (1876), au demeurant, ont mis en évidence pour la partition oc/oïl que certaines erreurs de classification provenaient de problème de conventions orthographiques. Il n'existe pas de standard orthographique unanimement accepté pour les parlers du Croissant et les dialectes voisins, à prédominance orale (Caubet *et al.*, 2002). Des solutions à ce problème seront mises à l'épreuve dans cet article, à travers la transcription d'un même texte d'une centaine de mots, que nous commencerons par présenter.

Nous proposons dans ce travail de comparer les transcriptions orthographiques d'une part, phonétiques d'autre part, d'enregistrements audio actuels, réalisés à l'intérieur et à proximité du Croissant. Plus précisément, nous avons demandé à une trentaine de locuteurs des domaines d'oc et d'oïl de traduire dans leur dialecte ou langue régionale le texte utilisé depuis un siècle par l'Association phonétique internationale (API) afin d'illustrer nombre de langues du monde : la fable d'Ésope « La bise et le soleil » (environ une minute de parole). Il s'agissait de 10 locuteurs du Bourbonnais, de la Marche et de la Charente, dont les productions transcrites orthographiquement ont été comparées à celles de 10 locuteurs du domaine nord-occitan (transcrites en graphie alibertine dite « classique »), à celles de 10 locuteurs du domaine poitevin-saintongeais (transcrites en graphie normalisée) et à celles de 6 locuteurs du domaine des parlers centraux d'oïl (transcrites en graphie plus ou moins personnelle, notamment en berrichon). Nous reviendrons sur ces graphies avant de montrer comment ces données peuvent être utilisées à des fins de dialectométrie, c'est-à-dire de mesures objectives de différences entre dialectes (Séguy, 1973 ; Goebel, 2002 ; Heeringa, 2004 ; Nerbonne *et al.*, 2007).

Nous décrirons d'abord nos points d'enquête et le protocole adopté. Nous présenterons ensuite une première analyse des données enregistrées,

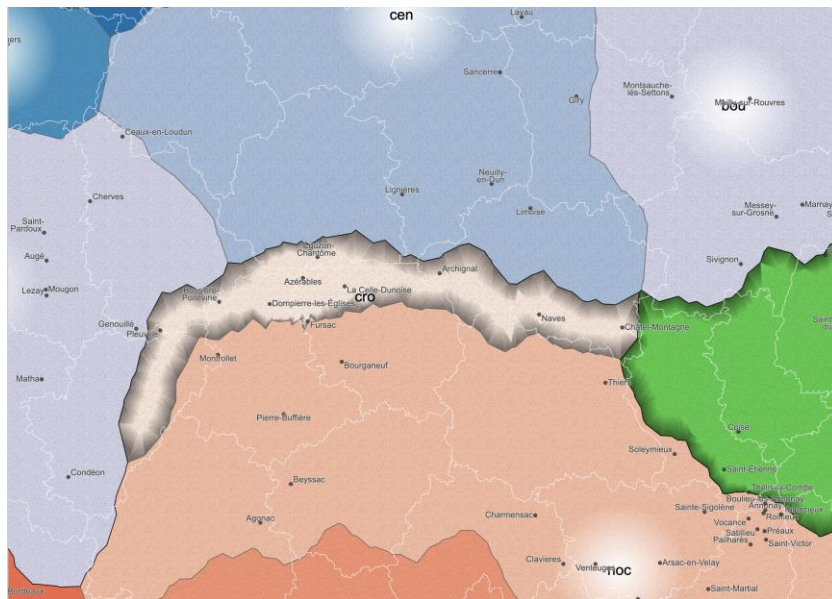
transcrites phonétiquement par des systèmes de conversion graphème-phonème et alignée automatiquement avec le signal audio : une méthodologie qui peut être appliquée à d'autres langues et dialectes.

## 1. Points d'enquête, locuteurs et protocole

La fable « La bise et le soleil » (disponible dans une centaine de langues ou dialectes sur le site de l'API) est le matériau dont nous sommes partis pour constituer un *Atlas sonore des langues régionales de France*, lequel prend la forme d'un site web dont l'objectif est de mettre en valeur notre patrimoine linguistique (Boula de Mareüil *et al.*, 2017). Nous voulions également faire traduire la parabole de l'enfant prodigue, utilisée depuis le XIX<sup>e</sup> siècle dans nombre de travaux de dialectologie, mais avons dû y renoncer, certains locuteurs s'y refusant. En revanche, nous avons fait traduire le texte de l'API à plus de 300 locuteurs, cartographiés sur le site <https://atlas.limsi.fr>. Une option permet de zoomer sur le Croissant, faisant apparaître des points d'enquête qui, à l'échelle de la France, seraient trop rapprochés les uns des autres. Ceux que nous avons retenus pour cette étude sont consignés dans le tableau 1, avec les abréviations qui seront reprises dans des figures ultérieures. La figure 1 permet de les visualiser sur la carte telle qu'elle apparaît à l'adresse <https://atlas.limsi.fr/?tab=cro>, avec les frontières de départements qui permettent de délimiter les régions administratives. Les limites du Croissant, elles, apparaissent floutées de façon à suggérer le caractère transitoire de cette aire linguistique.

**Tableau 1** : points d'enquête retenus en nord-occitan, poitevin-saintongeais, dans le Croissant et les parler centraux d'oïl. Les trois premières lettres des communes, en gras, seront reprises dans certaines figures.

<b>Nord-occitan</b>	<b>Poitevin</b>	<b>Croissant</b>	<b>Parlers centraux</b>
<b>M</b> ontrollet	Sablonceaux	<b>P</b> leville	<b>L</b> ignières
<b>A</b> gonac	<b>M</b> atha	<b>B</b> ussière-Poitevine	<b>N</b> eUILly-en-Dun
<b>P</b> ierre-Buffière	<b>S</b> aint-Pardoux	<b>D</b> ompierre-les-Églises	<b>S</b> ancerre
<b>B</b> eyssac	<b>A</b> ugé	<b>A</b> zérables	<b>L</b> avau
<b>B</b> ourganeuf	<b>M</b> ougon	<b>F</b> ursac	<b>L</b> imoise
<b>C</b> harmensac	<b>L</b> ezay	<b>É</b> guzon-Chantôme	<b>G</b> iry
<b>T</b> hiers	<b>C</b> ondéon	<b>L</b> a Celle-Dunoise	
<b>S</b> oleymieux	<b>C</b> herves	<b>A</b> rchignat	
<b>S</b> ainte-Sigolène	<b>C</b> eaux-en-Loudun	<b>N</b> aves	
<b>P</b> ailharès	<b>G</b> enouillé	<b>C</b> hâtel-Montagne	



**Figure 1** : carte des points d'enquête à l'intérieur et autour du Croissant.

Dans le tableau 2, d'Ouest en Est pour les communes du Croissant, on peut voir que cinq départements sont représentés, de même que les cinq sous-aires distinguées par Tourtoulon et Bringuier (1876). Certaines communes sont hors-zone, en raison du décès de l'un des collaborateurs (Bringuier, en 1875) ; mais figurent Fursac et Éguzon, considérés comme à la limite des domaines d'oc et d'oïl, respectivement. Dans le domaine des parlers centraux d'oïl, essentiellement bourbonnais et berrichons, nous n'avons été en mesure de cartographier que 6 locuteurs, malgré nos efforts, tant la proximité de Paris a pu faire plus qu'ailleurs disparaître les dialectes. En revanche, nous avons pu collecter des enregistrements en poitevin-saintongeais dans les quatre départements de l'ancienne région Poitou-Charente ainsi qu'en nord-occitan dans 10 départements : tous sont représentés dans l'échantillon retenu ici.

**Tableau 2** : communes analysées dans le Croissant, avec le département et la sous-aire dans la classification de Tourtoulon et Bringuier (1876) auxquelles elles correspondent, et le début de la fable d'Ésope (« La bise et le soleil se disputaient » en français). En italique sont rapportées les transcriptions francisantes qui ont eu notre préférence dans l'affichage de notre atlas sonore — même si des transcriptions en graphie occitane classique ont également été fournies.

Commune	Département	Ss-aire	Début de la fable
Pleuville	Charente	B2	<i>L'ébisail et le souleuil se pigoun'niant</i>
Bussière-Poitevine	Haute-Vienne	B1	<i>La bise et le souleil se disputiant</i>
Dompierre-les-Églises	Haute-Vienne	A1	La bise e le solelh se disputevan
Azérables	Creuse	A3	La bise e le solelh se disputièn
Fursac	Creuse	Oc	La bise e le solelh se disputavan
Éguzon-Chantôme	Indre	Oïl-B2	<i>Le vent et le souleuil se fâchèsyint</i>
La Celle-Dunoise	Creuse	A2	La bise e le solelh se disputavan
Archignat	Allier	—	La bise e le solelh se disputevon
Naves	Allier	—	La bise e le solelh ios disputián
Châtel-Montagne	Allier	—	La bise e le solelh se disputián

Les locuteurs enregistrés, entre 2015 et 2018, étaient quasiment tous retraités — ou tous, dans le Croissant, où aucun informateur ne pouvait être qualifié de « néo-locuteur ». Ils avaient exercé diverses professions, mais étaient tous issus du monde rural. Il leur a été demandé de traduire la fable soit directement avec le texte français sous les yeux, soit à partir de notes que les locuteurs avaient préféré prendre, soit encore après que l'enquêteur eut lu l'histoire, phrase par phrase en français. Les enregistrements ont pour la plupart été effectués au domicile des locuteurs ou dans une pièce calme, avec un enregistreur Zoom — leur intensité a ensuite été normalisée. Ils étaient associés à un consentement signé pour une libre diffusion.

## 2. Transcription orthographique

Pour les enregistrements en nord-occitan et dans les parlers du Croissant, des transcriptions orthographiques ont été fournies, selon la graphie occitane dite « classique », par deux experts, qui ont pu également s'inspirer de ce que les locuteurs avaient écrit. Fondée sur les travaux d'Alibert (1935) sur le languedocien (intermédiaire entre les autres dialectes), cette norme sans proprement d'assise territoriale favorise l'origine et l'histoire, plutôt que le lien avec la prononciation : il en résulte une orthographe avec une forte composante étymologique complétée par des règles à valeur phonologique. Le travail d'Alibert a ensuite été prolongé par le *Conselh de la Lengua Occitana* (CLO), qui a

apporté des précisions et spécifié des règles pour la transcription de différentes variétés occitanes (Sumien, 2007) : recommandations pour transcrire les sons /s/, /z/, les mots en *-atge-*, par exemple, les emprunts, etc. La norme préconisée admet des usages fréquents comme l'omission de signes diacritiques pour distinguer /e/~ε/ en limousin (opposition neutralisée sous forme de l'archiphonème /E/), ou la chute de consonnes intervocaliques en nord-occitan<sup>1</sup>.

Pour les enregistrements en poitevin-saintongeais, les locuteurs ont utilisé eux-mêmes la graphie « normalisée » UPCP-Métive, de type diasystémique comme la graphie occitane (Chevrier & Gautier, 2002). Cette orthographe se veut englobante, notant par exemple en *-àe* les infinitifs du 1<sup>er</sup> groupe, qui peuvent être prononcés en [a] ou en [e]. Les transcriptions ont par la suite été revues et éventuellement corrigées par une spécialiste. Pour les enregistrements dans les parlers centraux d'oïl, en revanche, nous ne disposons que des graphies individuelles, francisantes, des locuteurs, avec lesquels nous nous sommes contentés d'interagir pour harmoniser certaines incohérences, comme *pu* à côté de *pus*.

### 3. Première analyse des données du Croissant

Nous nous contenterons ci-dessous de quelques remarques, immédiatement appelées par la lecture du début de la fable, dans le tableau 2. D'abord, si tel locuteur utilise le mot *vent* pour désigner la bise, c'est son choix de traduction, que nous avons respecté<sup>2</sup>. Ensuite, si le mot *solelh* peut donner l'image d'une grande homogénéité, c'est que la graphie a délibérément été normalisée. Il n'appartient pas à ce niveau de transcription orthographique, pour nous, de noter les moindres différences de prononciation plus ou moins ténues et difficiles à retranscrire de façon cohérente (ex. [s(u)lœj]<sup>3</sup> ou [sulaj], [sulɛ(j)] ou

---

<sup>1</sup> Pour une problématisation de l'utilisation des règles et conventions orthographiques sur le présent corpus, nous renvoyons à Boula de Mareüil *et al.* (2017). Par exemple, une lettre muette comme le 't' sera transcrite dans un participe passé en occitan (ex. *finit*) mais pas en français — et inversement pour la 3<sup>e</sup> personne du pluriel des verbes, en français (ex. *sont*) mais pas en occitan.

<sup>2</sup> De rares locuteurs (parmi les quelque 300 que nous avons enregistrés) ont également traduit le français « bise » par des équivalents de « vent » dans d'autres dialectes, nord-occitans mais aussi liguriens, picards ainsi que dans des créoles à base lexicale française. Voir de même le tableau 3 pour le mot *manteau*.

<sup>3</sup>Un relecteur ou une relectrice anonyme nous fait remarquer que les formes du type [s(u)lœj] avec [œ], issues du latin SOL avec le suffixe diminutif -ICULUS/\*-UCULUS/-ELLU, ont une origine précise : dans ces formes [œ] peut se

[sulej]). Un dictionnaire de prononciations pourra inclure toutes ces variantes si on le souhaite, comme nous le verrons plus tard. Dans les paragraphes qui suivent, nous ne nous attarderons pas sur des détails de prononciation (nasalisation ou non des voyelles, affrication/palatalisation ou non des consonnes, etc.) ; en revanche, on ne peut pas faire l'économie de décrire un certain nombre de phénomènes.

Comme on peut s'en apercevoir dans le tableau 2, les imparfaits montrent la plus grande richesse. Si la plupart des parlers dégagent un yod à la 3<sup>e</sup> personne du pluriel des imparfaits des verbes du 1<sup>er</sup> groupe, on a un /v/ que l'on retrouve au singulier, avec des terminaisons en /av/ à Fursac, en /ɔv/ à La Celle-Dunoise et en /ɛv/ à Dompierre-les-Églises ainsi qu'à Archignat. Ces formes résultent d'évolutions irrégulières de l'imparfait latin (Zink, 1989), qui ne montrent pas une distribution aréale claire du type /ɛv/ à l'Est, /ɔv/ à l'Ouest, /av/ plus au sud du domaine considéré ici (cf. tableau 3). Chose unique à Éguzon-Chantôme, on trouve un /z/ au singulier comme au pluriel, avec des imparfaits en /ez/ (ex. /bufez/ « soufflait »). Là où on a un /j/ au pluriel, on a un /i:/ au singulier à Pleuville et à Bussière-Poitevine, un /ø/ à Azérables, un /ɔ/ à Châtel-Montagne et /jɔ/ (ex. /bufjɔ/ « soufflait ») à Naves. Ces trois dernières terminaisons se retrouvent dans les localités correspondantes au conditionnel, tandis qu'ailleurs c'est un /i/ qui est généralisé (ex. /arivri/ « arriverait »). Nos locuteurs du Croissant, par ailleurs, ont utilisé aussi bien le passé simple que le passé composé, parfois les deux dans le même texte — sans rechigner, dans certains points d'enquête, à recourir à l'auxiliaire *avoir* pour traduire « sont tombé » ou « s'est mise », caractéristique qui se retrouve dans d'autres endroits du Croissant (Quint, 1995).

Au niveau morphologique, également, la plupart des parlers du Croissant ont des infinitifs et des participes passés des verbes du 1<sup>er</sup> groupe en /a/. Seul Pleuville (aux confins du Poitou méridional) a un /ɛ/ dans les deux cas ; Azérables (Marche) et Châtel-Montagne (Bourbonnais) ont un /a/ à l'infinitif et un /ɛ/ au participe passé, tandis qu'à Naves (Bourbonnais), c'est l'inverse. Au niveau morpho-syntaxique, enfin, on peut constater dans le tableau 2 que Naves est le seul point d'enquête à avoir un pronom réfléchi pluriel différent du singulier : *ios disputián* (« se disputaient »). Tous les parlers du Croissant collectés sont à pronom sujet obligatoire, comme le français mais à la différence des

---

développer diachroniquement (dans le cadre de la déclinaison bicasuelle) de la désinence du nominatif *-eus* de *-ICULUS*, mais aussi de la forme oblique *-oil* de *\*-UCULUS* : on trouve la forme avec l'oblique de *\*-UCULU* dans la région contiguë de l'Est. Cependant, on ne relève la forme avec [œ] que dans l'ouest du domaine — à Éguzon-Chantôme, Azérables et Pleuville, dans les données traitées ici.



parlers occitans méridionaux, qui sont pro-drop. Les pronoms personnels sujets peuvent cependant prendre diverses formes, comme du reste les pronoms personnels compléments : il est remarquable, par exemple, que pour traduire « autour de lui », on a tantôt des formes de type *lu*, tantôt des formes de type *se*, sans que la répartition géographique n'apparaisse très clairement.

Le tableau 3 résume, pour quelques mots de la fable en français, les prononciations observées dans différents enregistrements du Croissant. Cette table illustre des choix particuliers de traduction évoqués plus haut, du type *capòt* plutôt que *mantèl* ou *bofar* plutôt que *soflar*. Ces différences peuvent avoir une incidence sur les mesures dialectométriques. Cependant, on trouve aussi le type *bouffer*, par exemple, dans des parlers d'oïl. Il ne semble donc pas a priori que ces différentes stratégies de traduction soient un problème trop important.

**Tableau 3 :** traductions (sous forme phonétique) des mots *soleil*, *manteau*, *souffler* et *soufflait* dans 10 communes du Croissant.

Commune	<i>soleil</i>	<i>manteau</i>	<i>souffler, soufflait</i>
Pleuville	sulœj	mâte	byfe, byfi:
Bussière-Poitevine	sulej	mâto	sufja, sufji:
Dompierre-les-Églises	sulaj	mâtaj	sufla, suflev
Azérables	slœj	mâte	sufja, sufjø
Fursac	sulaj	mâ(n)te	byfa, byfav
Éguzon-Chantôme	sulœj	kapo	bufa, bufez
La Celle-Dunoise	sulaj	mâtjo	sufja, sufjev
Archignat	sulaj	mâtjo	bufa, bufèv
Naves	sulaj	mêtso	bufe, bufjø
Châtel-Montagne	sulɛ	mâtjo	sufla, suflo

## 4. Exploitation dialectométrique

### 4.1. Méthodologie

Sur la base de nos données, nous avons calculé un taux symétrisé de différences entre variétés, que nous avons représenté graphiquement par des techniques d'analyse de données, comme dans Boula de Mareuil *et al.* (2017). Comme deux versions *A* et *B* n'ont pas nécessairement le même nombre de mots ou de caractères, les taux de différences de *A* par rapport à *B* et de *B* par rapport à *A* peuvent être inégaux : c'est pourquoi nous avons simplement calculé la moyenne des deux, paire par paire. Les résultats ainsi obtenus permettent de visualiser une sorte de distance entre les parlers du Croissant, les (autres) parlers d'oc et les parlers d'oïl. Les différences inter-variétés portant sur les caractères et sur les mots ont été visualisées par échelonnement multi-dimensionnel (MDS), une technique

qui permet de réduire les dimensions de la variation (ici,  $n = 36$ ) à un petit nombre de variables (2 ou 3) non-corrélées entre elles. Les figures ci-dessous ont été obtenues à l'aide du logiciel R (R Development Core Team, 2018), utilisant l'algorithme classique de la librairie *mds*, à partir de la matrice de dissimilitudes générée automatiquement. L'échelonnement renvoie un ensemble de points représentant les données de telle sorte que les distances (euclidiennes) entre ces points soient les plus proches possibles des dissimilitudes entre les données.

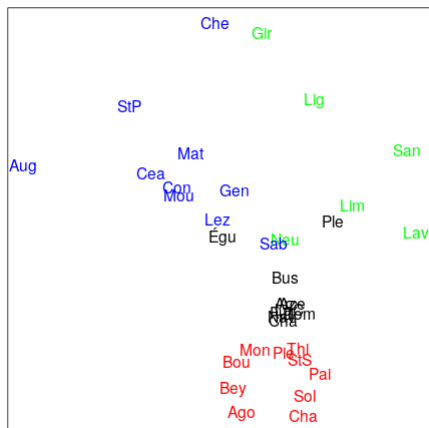
La même approche a ensuite été appliquée aux transcriptions phonétiques obtenues automatiquement par alignement forcé, avec des modèles acoustiques du français et un dictionnaire de prononciations incluant des variantes de prononciation, elles-mêmes générées par des systèmes de conversion graphème-phonème développés pour le français et l'occitan. Le dictionnaire de prononciations est une simple liste de mots avec, pour chaque entrée, une ou plusieurs prononciations : par exemple *mantèl* [mantɛl, mâtɛl]<sup>4</sup>. Avec quelques ajustements, des modèles acoustiques indépendants du contexte ont été utilisés, représentant des ensembles de 35 phones, avec 3 unités supplémentaires pour les silences, les respirations et les hésitations. Les sorties du système d'alignement doivent encore être corrigées, mais comme les frontières entre les unités sont assez bien placées, le processus s'en trouvera accéléré. Des expériences, en attendant, peuvent être menées, dont les résultats peuvent de même être projetés sur une carte de géographie, suivant une approche qui d'habitude en dialectométrie est réservée à des mots isolés issus d'atlas linguistiques plus ou moins anciens (Tillinger, 2016 ; Brun-Trigaud, à paraître).

## 4.2. Résultats

On obtient des patrons similaires avec des MDS calculés à partir des différences entre mot ou entre caractères. Nous présenterons dans la figure 2 les résultats obtenus sur la base des caractères, plus proches en nombre des phonèmes. Dans cette figure et les suivantes, les points du Croissant sont représentés en noir, ceux du domaine nord-occitan en rouge, ceux du domaine poitevin-saintongeais en bleu et les parlars centraux d'oïl en vert. Les étiquettes sont centrées sur les coordonnées des points issus du MDS.

---

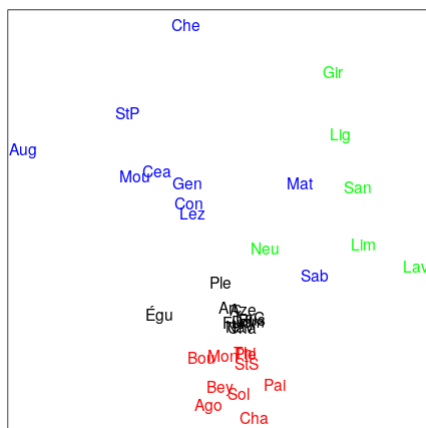
<sup>4</sup> La présence d'un appendice consonantique après la voyelle nasale [mâtɛl] n'a pas été prise en considération (cf. tableau 3).



**Figure 2 :** plan issu de l'échelonnement multi-dimensionnel sur la base des caractères (avec une graphie francisante pour Ple, Bus et Égu).

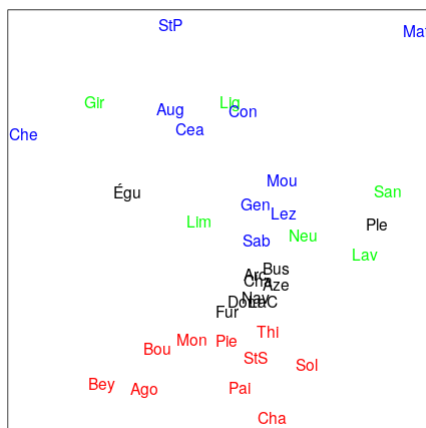
On constate que les points nord-occitans sont proches, de même que les points du Croissant marchois et bourbonnais, qui forment un groupe compact distinct. L'exception que constituent éventuellement Pleuville, Bussière-Poitevine et Éguzon-Chantôme sera examinée ci-après. De façon intéressante, les points les plus septentrionaux du domaine nord-occitan se retrouvent les plus proches du Croissant, dans le plan issu du MDS (formant un groupe également assez compact), tandis que les plus méridionaux se trouvent plus bas dans ce plan (Agonac, Beyssac, Charmensac, Soleymieux et Pailharès). Le poitevin-saintongeais et le berrichon-bourbonnais se trouvent quant à eux plus éclatés, formant des espaces distincts : poitevin-saintongeais à gauche (incluant le point de Pleuville), berrichon-bourbonnais à droite. En tout état de cause, la division entre oïl (en haut) et oc (en bas) est bien respectée, avec le Croissant au centre.

L'effet de la graphie est illustré dans la figure 3, pour les trois points d'enquête pour lesquels des transcriptions suivant les conventions d'oïl et d'oc ont été fournies : Pleuville, Bussière-Poitevine et Éguzon-Chantôme, avec respectivement pour le début de la fable en oc « l'ebisalh e le solelh se pignonhèn », « la bise e le solelh se disputian » et « le vent e le solelh se faschesien ». On observe que ces trois points se rapprochent du paquet correspondant au Croissant, dans le plan issu du MDS (jusqu'à se fondre avec lui pour Bussière-Poitevine), mais que Pleuville reste le plus proche du domaine d'oïl. De l'aveu même du locuteur, son dialecte était plus occitan dans son enfance et a évolué vers le poitevin méridional au cours du temps.



**Figure 3** : plan issu de l'échelonnement multi-dimensionnel sur la base des caractères (avec une graphie occitanisante pour Ple, Bus et Égu).

Enfin, l'échelonnement multi-dimensionnel appliqué au résultat de la conversion graphème-phonème est représenté dans la figure 4 — peu importe alors la graphie adoptée pour les trois points à la lisière du Croissant. Une certaine instabilité peut être constatée, par rapport aux deux figures précédentes ; mais dans l'ensemble, les regroupements sont respectés et cohérents. Un cluster très compact rassemble toujours les points du Croissants, hormis Éguzon-Chantôme et Pleuville, qui se trouvent en orbite parmi des points du domaine d'oïl. Dans la partie inférieure du plan, ne se trouvent que les points du domaine nord-occitan, avec au-dessous de l'ensemble croissantin des points correspondant à des localités géographiquement proches du Croissant (comme Thiers, Pierre-Buffière, Montrollet ou Bourganeuf), tandis qu'en bas apparaissent des points plus méridionaux comme Charmensac ou Pailhères.



**Figure 4** : plan issu de l'échelonnement multi-dimensionnel sur la base des phonèmes.

## Conclusion

L'atlas sonore que nous avons développé, qui visait à mettre en valeur la richesse de notre patrimoine linguistique, permet d'illustrer certains phénomènes sur une base comparable — bien connue des phonéticiens. En matière de transcription orthographique, on se heurte avec une langue comme l'occitan (a fortiori dans une zone de variation comme le Croissant) à une double difficulté : la transparence graphie-phonie d'une part, la correspondance entre les différentes variétés d'autre part. L'étude de cas présentée ici montre l'importance des conventions orthographiques pour transcrire ne serait-ce qu'un bref enregistrement audio : elle a mis en évidence la difficulté mais également la faisabilité de la tâche, ainsi que son exploitabilité à des fins de dialectométrie. La cohérence des résultats obtenus, par des techniques d'analyse de données, de conversion graphème-phonème et d'alignement automatique, nous incite à accréditer la validité de l'approche et à vouloir l'étendre à davantage de points d'enquêtes.

Nous reconnaissons que la normalisation orthographique/phonologique élimine une partie de l'information, peut faire perdre des indices diatopiquement cernés et applatit les différences. C'est une des fonctions de l'orthographe que d'effacer une certaine variation à caractère géolinguistique, ce pour quoi nous avons tenu à comparer les résultats à base de distances objectives (qui ne rendent pas explicitement compte de traits linguistiques) sur les transcriptions phonétiques (obtenues automatiquement). Même si, d'un point de vue typologique, quelques facteurs de microvariation phonétique continuent à être cachés (comme

des traces d'un appendice consonantique après une voyelle nasale), les calculs dialectométriques que nous avons représentés graphiquement valident globalement l'approche exploitée. La méthode permet de revisiter un vieux problème : celui de la limite entre oïl et oc.

La question de la proximité des langues est depuis longtemps l'occasion de spéculations politiques et sociales. Dans ce contexte, les résultats fournis par des algorithmes de classification peuvent être instrumentalisés à diverses fins, et nous préférons ne pas pousser trop loin des conclusions qui pourraient être surinterprétées. Il serait particulièrement imprudent de dessiner des limites à l'intérieur ou autour du Croissant, zone d'interférence où s'interpénètrent des influences d'oc et d'oïl, où s'entrecroisent de nombreux traits linguistiques : on l'a vu avec Pleuville (Charente), qui a pu évoluer du nord-occitan vers le poitevin méridional. Découper ainsi le territoire, la quête est vaine dans des espaces de transition qui peuvent être ouvertes, perméables et mouvantes (Saussure, 1916 : 278–279). Nous nous risquons et nous bornerons simplement à dire que, sans nier l'existence de lignes de fracture qui partagent la Marche et le Bourbonnais, cette étude suggère que le Croissant constitue une entité autonome, plus proche des parlers nord-occitans que des parlers d'oïl.

### **Remerciements**

Cette recherche s'insère dans les projets suivants, gérés par l'Agence Nationale de la Recherche : ANR-17-CE27-0001-01 (Projet « Les parlers du Croissant : une approche multidisciplinaire du contact oc-oïl ») et ANR-10-LABX-0083 (programme « Investissements d'Avenir », Labex EFL, Axe 3, Opération LC4 – « Les parlers du Croissant : une aire de contact entre oc et oïl »). Il contribue à l'IdEx Université de Paris – ANR-18-IDEX-0001.

### **Références**

- Alibert, L. 1935. *Gramatica occitana segon los parlars lengadocians*. Toulouse : Societat d'Estudis Occitans.
- Boula de Mareüil, P., Sichel-Bazin, R., Quint, N., Adda, G. 2017. Norme et variation à l'âge des corpus informatisés pour les langues régionales de France. In C. Feuillard (éd.), *Usage, norme et codification : de la diversité des situations à l'utilisation du numérique*. Bruxelles : EME Éditions (pp. 217–222).
- Brun-Trigaud, G. 1992. « Les enquêtes dialectologiques sur les parlers du Croissant : corpus et témoins », *Langue française* 93 : 23–52.

- Brun-Trigaud, G. à paraître. « Études dialectométriques sur le lexique des parlers du Croissant ». 2<sup>es</sup> *Rencontres sur les parlers du Croissant*, Montluçon.
- Chevrier, J.-J. & Gautier, M. 2002 *Le poitevin-saintongeais : langue d'oïl méridionale*. La Crèche : Gestes Éditions.
- Gilliéron, J. & Edmont, E. 1902–1910. *Atlas linguistique de la France*, Paris : Champion.
- Goebel, H. 2002. « Analyse dialectométrique des structures de profondeur de l'ALF », *Revue de linguistique romane* 66(261–262) : 5–63.
- Heeringa, W. 2004. *Measuring dialect pronunciation differences using Levenstein distance*. Thèse de doctorat, Rijksuniversiteit, Groningen.
- Kloss, H. 1967. « “Abstand” languages and “Ausbau” languages », *Anthropological Linguistics* 9(7) : 29–41.
- Nerbonne, J., Kleiweg, P., Heeringa, W., Manni, F. 2007. Projecting dialect differences to geography: bootstrap clustering vs. noisy clustering. In C. Preisach, L. Schmidt-Thieme, H. Burkhardt, R. Decker (éds.), *Data analysis, machine learning, and applications*. Berlin : Springer (pp. 647–654).
- Pike, K. 1982. *Phonetics*. Ann Arbor : Michigan Press.
- Quint, N. 1996. *Grammaire du parler occitan nord-limousin marchois de Gartempe et de Saint-Sylvain-Montaigut (Creuse) : étude phonétique, morphologique et lexicale*. Limoges : Cercle Limousin d'Études Occitanes et d'Action Populaire.
- Quint, N. 1998. « Aperçu d'un parler occitan de frontière : le marchois », *Bulletin de l'association internationale d'études occitanes* 14 : 126–135.
- R Development Core Team. 2018. R: A language and environment for statistical computing. Vienne : R Foundation for Statistical Computing.
- Rousselot, abbé J.-P. 1891. *Les modifications phonétiques du langage étudiées dans le patois d'une famille de Cellefrouin, Charente*. Paris : Thèse présentée à la Faculté des lettres de Paris.
- Ronjat, J. 1913. *Essai de syntaxe des parlers provençaux modernes*. Paris : Imprimerie nationale.
- Saussure, F. de. 1916. *Cours de linguistique générale*, Paris : Payot.
- Séguy, J. 1973. « La dialectométrie dans l'Atlas linguistique de la Gascogne », *Revue de linguistique romane* 37 : 1–24.
- Sumien, D. 2007. « Preconizacions del Conselh de la Lengua Occitana », *Lingüística Occitana* 6 : 1–157.
- Tillinger, G. 2016. « Étude des frontières linguistiques à l'intérieur de la zone d'interférence appelée “Croissant” ». *Atti del XXVIII Congresso internazionale di linguistica e filologia romana*, Rome (pp. 1037–1052).

- Tourtoulon, Ch. de et Bringuier, O. 1876. *Étude sur la limite géographique de la langue d'oc et de la langue d'oïl. Rapport destiné au ministre de l'instruction publique des cultes et des beaux-arts.* Paris : Imprimerie nationale.
- Zink, G. 1989. *Morphologie du français médiéval.* Paris : Presses Universitaires de France.