Dialectologia 2 (2009), 81-111.

LE PROJET BOURCIEZ: TRAITEMENT GÉOLINGUISTIQUE D'UN CORPUS DIALECTAL DE 1895[1]

Gotzon Aurrekoetxea[1] and Charles Videgain[2]

[1] Université du Pays Basque, Gasteiz-Vitoria

gotzon.aurrekoetxea@ehu.es

[2] Université de Pau et des Pays de l'Adour – CNRS UMR 5478

charles.videgain@univ-pau.fr

Abstract

This contribution shows the way to carry out the geolinguistic exploitation of texts gathered in the 19th century by E. Bourciez, using different automatic tools. After typing the manuscript in a word processor, the "Simple Concordance Program" was used to create the list of words of the 150 Basque texts of the collection. Afterwards, all of the words were introduced in a database. Then, the lemmatisation process was done, with a program made ad hoc. This step is necessary to use data in a geolinguistic way. And the final step was to exploit the data in the dialectometric way, using the VDM program conceived by Hans Goebl.

Keywords

Dialectology, dialectometry, database, linguistic technology.

1. Introduction

La dialectologie en domaine basque connaît depuis quelques années divers développements, aussi bien dans le cadre de la dialectologie traditionnelle que depuis une trentaine d'années en ayant recours à des méthodologies plus novatrices. Cela se manifeste en particulier dans l'attention prêtée à la classification des divers parlers ou dialectes.

Le travail qui suit présente l'exploitation d'un corpus dialectal datant de la fin du XIXème siècle et tente une représentation cartographique des objets linguistiques offert par un corpus aux items limités en nombre mais d'une grande densité du point de vue

[1] Ce programme de travail a bénéficié de l'aide de l'Université du Pays Basque (Vitoria-Gasteiz) sous le numéro de projet 00033.130H-14888/2002.

©Universitat de Barcelona

géographique. Nous y présentons une cartographie traditionnelle des données (sur support informatisé) puis un traitement inspiré des méthodes de la dialectométrie.

Nous avons publié trois ouvrages sur les textes basques recueillis par Edouard Bourciez vers 1895, le premier consacré au corpus lui-même, le second aux cartes lexicales, le troisième aux cartes grammaticales (Aurrekoetxea & Videgain 2004, Aurrekoetxea, Videgain & Iglesias, 2004 [2006], 2005 [2008]. L'aide qui nous a été fournie[2] nous a permis de réaliser cette trilogie.

2. L'édition des textes (saisie informatisée des textes manuscrits)

Le point de départ du projet est constitué par l'ensemble des textes recueillis dans un document intitulé *Recueil des Idiomes de la Région Gasconne*. Ce travail a commencé en 2002. Ce recueil est généralement désigné comme *Recueil Bourciez*.

Sur Bourciez lui-même, nous renvoyons à l'introduction de notre premier volume ou bien à un article bien informé d'Alain Viaut (Viaut 1995) sur ce professeur romaniste, contemporain à Bordeaux de Durkheim, qui eut assez d'autorité et constance pour obtenir de réunir cet ensemble remarquable de matériaux dans l'Académie de Bordeaux.

Edouard Bourciez réunit en 1894 un total de 4444 versions de la *Parabole de l'Enfant prodigue* (dont les versions basques). Il sut utiliser pour cela les services de l'Education Nationale et par le biais des Inspecteurs d'Académie il obtint une réponse de tous les instituteurs sollicités. Tous les instituteurs furent tenus de retourner leur copie, certainement par la voie hiérarchique comme le sceau de l'Inspection de Pau sur les documents le démontre pour les versions basques du moins. Ce travail ne fut donc pas celui d'enseignants simplement volontaires mais de fonctionnaires disciplinés.

La conséquence en est pour nous très intéressante: il n'est presque pas de commune dont la version n'ait pas été retournée. Bourciez devait faire parvenir à

[2] Nous tenons à remercier particulièrement:
- Monsieur Paul-Henri Allioux, Responsable de la Bibliothèque Universitaire de l'Université Michel de Montaigne Bordeaux 3, qui nous a permis de travailler sur le manuscrit Bourciez en nous aidant à résoudre tous les problèmes techniques ou administratifs qui auraient pu nous ralentir.
- Bernard Oyharçabal, directeur du laboratoire de recherches du CNRS Iker 5478, sis à Bayonne, et qui a mis les ressources du centre à notre disposition.
- Nous ne pouvons que joindre dans ces remerciements la mémoire du Professeur Jacques Allières, qui fut notre maître, et l'un des premiers à avoir découvert l'intérêt du recueil Bourciez sur lesquels il publia des articles importants (Allières 1977 et 1986).

©Universitat de Barcelona

chaque instituteur le texte de la parabole en français pour que chaque instituteur en fît ou en fît faire la traduction. Grâce à cette collecte, Bourciez réunit un réseau extrêmement serré sur un grand espace et il pensait à juste raison que bien des objets linguistiques seraient saisis dans les mailles de ces nombreuses versions. Par ricochet en quelque sorte, ce souci synchronique fait notre bonheur en domaine basque; nous rappelons que nos textes basques ne constituent qu'une faible partie du matériau recueilli, 150 sur un total de près de 4500 mais c'est déjà un riche butin

Ces manuscrits sont conservés à la Bibliothèque Universitaire de l'Université Michel de Montaigne Bordeaux 3.

2.1. Les manuscrits

Les 150 versions en langue basque couvrent l'ensemble du Pays Basque dans le département des Pyrénées-Atlantiques actuelles (alors département des Basses-Pyrénées). Tous les textes recueillis en Pays Basque historique ne sont pas en langue basque: c'est ainsi que les communes suivantes ont fourni des versions occitanes (gascon) à Bayonne (1188-1191), Anglet (1197), Bassussary (1199), Biarritz (1200), Urt (1239), Labets-Biscay (1289), Osserain (1296), Montory (1356). Dans le canton d'Oloron, Esquiule a fourni un texte en basque. Quelques villages n'ont fourni aucun texte, comme à Saint-Martin d'Arrossa, Behorléguy, Gamarthe, Ithorrotz-Olhaiby ou Berrogain-Laruns. Dans quelques villages, ce sont deux textes qui ont été recueillis, avec des variantes orthographiques entre les deux textes ou bien des différences plus importantes. En tout ce sont 138 villages qui ont fourni des textes pour le recueil Bourciez.

Ce recueil est d'une importance capitale pour la dialectologie basque et la philologie basque. En effet, on ne possède pas beaucoup de séries comparables de textes recueillis presque dans chaque commune. Même s'il s'agit d'une transcription écrite et donc sujette à caution au regard des enquêtes plus récentes faites au magnétophone, nous défendons la valeur du témoignage ainsi recueilli: ces instituteurs, sachant ou non le basque, ont généralement fait conserver la qualité de leur traduction par un locuteur connaisseur de la variété de la commune.

On sait que la collecte des données a progressé depuis les premières tentatives menées par Georg Wenker en Allemagne qui obtint une célèbre collecte à partir d'une

©Universitat de Barcelona

enquête par correspondance. Envoyant son questionnaire par la poste, il ne pouvait à l'époque procéder à l'enquête directe comme on le fait aujourd'hui. La collecte du recueil Bourciez est donc tributaire de son époque et des méthodologies alors en vigueur. Elle reste de par sa densité et son réseau serré une source de grande valeur.

On connaît les dangers et les risques d'un enquête menée en l'absence du linguiste: en particulier, chaque instituteur n'a pas poussé au même degré le souci de refléter les particularités du parler de la commune et certains ont été plus laxistes sur ce point en produisant des formes moins endémiques.

Parfois la chose est claire comme dans le cas de la commune de Pagolle: l'instituteur n'a absolument pas suivi le parler de la commune et a fourni un texte marqué comme bien plus occidental, peut-être en raison du prestige prêté aux parlers plus éloignés du souletin.

Un autre cas est à constater: le copiage. Il y eut des instituteurs qui eurent moins de scrupules et même l'un ou l'autre peut être comparé au mauvais élève qui compte sur ses camarades pour ne pas faire le travail donné: c'est ainsi que nous expliquons par exemple que les deux versions 1268 et 1269 soient presque identiques, le premier texte étant signé par l'instituteur Maigret et le second par Monsieur Aguer-Achal de Saint-Palais. Certes le lecteur attentif relève quelques différences (*mithila*/*muthila*, *pounitu*/*punitu*, *vesta*/*vestat*, *edo*/*eddo* mais on est certain que les deux textes ont le même auteur, qui utilise certains traits endémiques intéressants comme le futur en *hartiko* (et non pas **hartuko*) ou la verbe *ezari* (et non pas **ezarri*). De toute façon, l'inconvénient est minime puisque les deux textes sont portés comme étant représentatifs l'un et l'autre de Saint-Palais.

Le cas de Bunus (1303) et Hosta (1305) est plus piquant, comme nous l'avons montré ailleurs (Videgain 1989, 35-38). La solidarité des anciens de l'Ecole Normale aidant, il semble que Sagaspe instituteur de Bunus ait prêté son texte à l'instituteur Loustau de Hosta, puisqu'on peut s'apercevoir – Loustau était béarnais sans doute- que c'est Loustau qui a copié, non sans faire des erreurs de copie, ce qui laisse entendre qu'il ne comprenait pas le texte qu'il copiait, les fautes s'accumulant en fin de texte, signe d'une attention moins vive: *lehin* pour *lehen*, *ehaztin* pour *ehaztun*, *bezta* pour *besta*. Ici encore le plagiat est sans grande conséquence puisqu'il touche deux communes limitrophes.

©Universitat de Barcelona

2.2. Dactylographie des manuscrits

Du point de vue technique, nous avons voulu passer directement des microfiches de la Bibliothèque à l'écran de l'ordinateur. Mais le mauvais état de certaines fiches ou leur degré d'illisibilité nous ont poussés à choisir une autre solution: la photographie pure et simple des pages du manuscrit à partir d'un appareil photo numérique. Nous avons pu pour ce faire obtenir les autorisations nécessaires du responsable de la Bibliothèque Universitaire de Bordeaux qui a bien voulu mettre le manuscrit à notre disposition. Une bonne journée de travail intense nous a permis de photographier l'ensemble des textes basques.

Ces documents ont été ensuite formatés en pdf et publiés sous cette forme dans le CD qui accompagne notre ouvrage. On peut aussi les consulter sur Internet à l'adresse www.armiarma.com

Cette phase assez longue qui consiste à saisir sur support informatique l'ensemble ds manuscrits était nécessaire. L'un d'entre nous a saisi chaque texte commune par commune puis l'équipe de travail a corrigé la première version dactylographiée sur Word. La qualité de la copie conditionnait la suite du travail et nous avons essayé d'éliminer le plus possible les erreurs de transcription. Un processeur de textes a ensuite de réunir toutes les versions dans un document unique. Le résultat en a été le premier volume (G. Aurrekoetxea & X. Videgain, 2004).

3. Traitement géolinguistique

3.1. Du texte à la base de données

Comme indiqué ci-dessus, les données originales sont organisées commune par commune. Le texte informatisé a été exploité de manière systématique en utilisant des ressources automatisées.

©Universitat de Barcelona

3.1.1. Isolement des mots-texte

L'ensemble du document ayant été intégré dans un processeur de textes, nous avons procédé à la mise en évidence de chaque mot. Un logiciel appelé 'Simple Concordance' permet d'isoler toute chaine de caractères allant d'un blanc à un autre, chaque chaîne étant isolée et étiquetée. Le programme indique pour chaque séquence dans quel texte elle apparaît, combien de fois et à quel endroit du texte.

Nous avons donc appliqué la procédure autant de fois qu'il y avait de textes, soit 150 fois. Tout n'a pu être fait de façon automatique, certains signes graphiques n'ayant pu être discriminés. Il a fallu avant chaque texte dresser la liste de tous les signes graphiques utilisés par les auteurs.

Nous avons ensuite réuni les documents ainsi traités sous le format txt. C'est à partir de ces documents que nous avons poursuivi le traitement du corpus.

3.1.2. Introduction des textes dans une base de données

Une fois les textes traités, chaque mot-texte a été introduit dans une base, chaque mot étant codé avec le texte auquel il appartient.

Cette opération est assez longue et a soulevé quelques difficultés auxquelles il fallait trouver solution. La coupure ou mécoupure des mots a été une des difficultés principales. En effet nous avons conservé dans le corpus informatisé les caractéristiques des mots telles que données par les manuscrits.

Mais cependant, si la table des caractéristiques du corpus original a été conservée dans la base, une autre table a été élaborée dans laquelle les « erreurs » ont été lissées pour faciliter la lecture et le regroupement des mots-texte à notre avis semblables.

Le tout a fourni une base comprenant 59.902 registres. Ces registres représentent le total des mots composant le corpus original. A partir de ces 59.902 registres nous avons obtenu 12.499 item. Bien que le texte original de la parabole soit de dimensions réduites, la base de données constitue un ensemble qui permet une série de traitements significatifs.

©Universitat de Barcelona

4. Lemmatisations

Notre intention étant celle de procéder à une étude géolinguistique des parlers du Pays Basque Nord, nous avons eu recours à une lemmatisation, étape indispensable dans le traitement de tout corpus linguistique. Sans lemmatisation en effet, le corpus est lourd et d'utilisation pénible. Nous avons construit deux lemmatisations, l'une orthographique, l'autre sémantique ou onomasiologique. L'une est relative à la forme des mots et à leurs variantes, l'autre est relative au signifié.

4.1. Lemmatisation orthographique

Nous avons décidé de créer un programme auxiliaire pour la lemmatisation, les logiciels existants ne pouvant nous fournir de solution satisfaisante. Ce programme donne tout d'abord la liste des mots à lemmatiser (l'illustration n° 1)

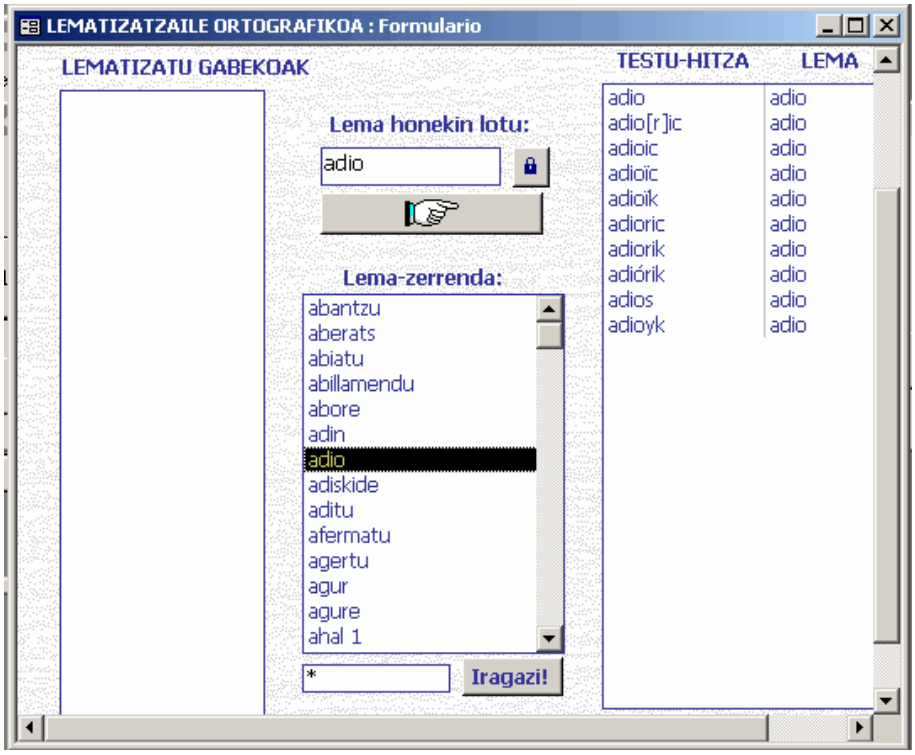


Illustration 1: Écran du programme de lemmatisation.

©Universitat de Barcelona

Les mots-texte sont donnés dans la colonne de gauche dite '*lematizatu gabekoak / non lemmatisés*'. Le chercheur commence par choisir un des mots du texte et lui assigne le lemme qui lui correspond dans le cartouche intitulé '*lema honekin lotu / attacher à ce lemme'*. Ensuite, après avoir sélectionné par un bouton le lemme voulu, les deux termes mis en relation apparaissent dans le cartouche de droite, avec à gauche le mot du texte (*adio*, *adio[r]ic*, *adioic*) et à droite le lemme choisi (*adio*). La classification se fait jusqu'à épuisement de la liste des mots du texte.

Une vérification ultérieure est possible pour s'assurer de la classification adéquate de tous les mots du texte par lemme et il est possible de changer le lemme correspondant à un mot donné du texte. Au milieu de l'écran, se situe la colonne intilulée '*lema-zerrenda / liste des lemmes*'. Il suffit de sélectionner l'un des lemmes de la liste pour voir dans la colonne de droite la liste des mots-texte classifiés sous ce lemme. Dans cet exemple précis dix mots-texte ont été classifiés sous le lemme *adio*. Les mots comme *adioric* portant une marque grammaticale (cas partitif ici) peuvent être ainsi classés avec les formes nues.

La "lemmatisation orthographique" permet de fournir le lexique de tout le corpus. Dans ce lexique, chaque lemme constitue une entrée indiquant le nombre d'occurrences du lemme, ses variantes et le code de chaque commune dans laquelle se trouve chaque variante. Il s'agit dont d'une procédure lexicographique qui ramène bon nombre d'items différents à des lemmes considérés comme entrée d'un dictionnaire. Au total, nous sommes passés de 12499 items à 1199 lemmes distincts servant d'entrée de dictionnaire.

4.2. Lemmatisation onomasiologique

Après cette lemmatisation consacrée aux variantes orthographiques ou grammaticales, il fallait procéder à la lemmatisation sémantique. Nous sommes partis du texte source, soit la version française de la *Parabole de l'enfant prodigue* telle que Bourciez la fit parvenir aux instituteurs priés de donner une version en basque. Nous y avons donc distingué un certain nombre de signifiés, au nombre de 151. Nous n'avons pas utilisé tous les signifiés possibles mais seulement ceux dont la distribution nous intéressait et pour chaque signifié, nous avons examiné comment le terme français avait été traduit par un signifiant basque trouvé dans la liste de la lemmatisation orthographique indiquée ci-dessous.

©Universitat de Barcelona

Ici encore, les difficultés n'ont pas manqué. Il nous a fallu faire des choix ou procéder à des exclusions qui peuvent être discutés ou critiqués. Par exemple pour le signifiant 'tomber' les lemmes suivants sont apparus dans la liste: *erori*, majoritaire avec 97 occurrences, *eori* avec 24 occurrences offrant la chute de la vibrante intervocalique, mais aussi d'autres termes minoritaires comme *jarri*, *etzatu*, *paratu*, *iraili*, et *aurtiki*, termes fort éloignés en eux-mêmes de la notion de 'tomber'. Comme le contexte du texte était 'il se laissa tomber sur un escabeau', les solutions des divers traducteurs ont été très variées et pas toujours faciles à être comparées entre elles. De même un terme comme 'traiter' dans le contexte 'traitez-moi comme le dernier de vos valets' a fourni maintes solutions comme *tratatu*, emprunt manifeste qui apparaît 106 fois, *ibili* et son factitif *erabili*, six fois chacun, *atxiki* deux fois, etc… Dans d'autres cas heureusement, et le plus souvent, la surface sémantique du signifié est nettement limitée et la lemmatisation est aisée: 'canard', 'chemin', 'naître'…

Cette lemmatisation a bien entendu conditionné dans une grande mesure les étapes suivantes du travail. Mais d'une part elle a supposé un classement lexicographique, d'autre part, elle est marquée par le fait même que l'on travaille à partir d'une traduction. Cette difficulté est bien connue des atlas linguistiques dans lesquels les artefacts provenant de la traduction sont monnaie courante.

Nous pouvons donc dire que nous avons ainsi constitué plusieurs bases de données:

a) une base organisée sur une lemmatisation orthographique qui rassemble toutes les données premières. Dans cette base, sont réunis tous les mots du texte et toutes les unités grammaticales;

b) une base organisée sur la dite lemmatisation sémantique et qui fournit, prêtes à l'analyse, les données d'ordre lexical;

c) une base grammaticale réunissant les données phonologiques, morphonologiques et syntaxiques, serait à ajouter mais c'est une base que nous n'avons pas encore réalisée.

Le traitement dialectologique des textes passe par la construction de ces bases de données. L'information distribuée dans le corpus est ainsi organisée pour l'analyse de manière à ce que le chercheur puisse en faire l'étude.

Dans les lignes qui suivent nous procédons à l'analyse du lexique pour produire une cartographie. Nous avions choisi 227 signifiants mais nous nous sommes rendu

©Universitat de Barcelona

compte que, dans un nombre important de versions, certains d'entre eux n'étaient pas traduits. Cela nous a poussé à éliminer un certain nombre d'entre eux. Les signifiants trop ténus ou trop vastes, ou bien les réponses rendues trop fragmentés par des choix grammaticaux très éclatés ont été mis en marge. Comme indiqué plus haut, finalement, nous nous sommes limités à un total de 151 termes

Nous avons aussi procédé à une réduction du nombre de textes quand une commune a fourni plus d'une version. De 150 textes, nous sommes descendu à 137 puisque les communes suivantes présentaient deux versions: Saint-Jean de Luz, Ustaritz, Espelette, Mauléon, Saint-Palais, Arberats-Sillègue, Béguios, Iholdy, Ostabat, Saint-Etienne de Baïgorry, Saint-Jean-Pied-de-Port et Tardets-Sorholuz. Ce choix nous a été imposé par l'appareil informatique qui n'admet qu'une réponse par commune.

La base lexicale, touchant 151 signifiés dans 137 villages, nous a fourni donc un ensemble de 20.838 items ou *taxats* pour reprendre la nomenclature de H. Goebl.

Ce nombre nous amène à formuler une remarque importante. Tiré d'un texte assez court, ce total 151 signifiants représente le total le plus élevé qui ait jamais été utilisé à ce jour en dialectologie basque pour opérer un traitement géolinguistique. De plus, il est appliqué seulement à une partie du Pays Basque et non pas à tout le domaine de la langue. Ces analyses sont premières et devront être reprises, améliorées et notre groupe de travail a l'intention de s'atteler à cette tâche dans les années qui viennent afin de fournir rapidement des résultats encore plus fiables.

Nous ne sommes pas les premiers cependant à travailler le corpus Bourciez et nous avons cité déjà Jacques Allières qui avait publié un « Petit atlas linguistique basque français, Bourciez 1 » (Allières, 1978). Ce document présentant 20 cartes commentées et avait l'immense mérite, après Vinson, de signaler l'intérêt pour les dialectologues des textes recueillis par Bourciez.

Dans le traitement géolinguistique que nous proposons, trois axes de travail, parmi d'autres, peuvent être ouverts:

- des études basées sur les points de vue habituels et traditionnels de la dialectologie

- un traitement géolinguistique proche des méthodes de J. Allières.

- un traitement dialectométrique selon les perspectives de Hans Goebl.

©Universitat de Barcelona

4.3. Cartes lexicales

Nous avons d'abord procédé à un traitement géolinguistique inspiré des atlas linguistiques traditionnels en publiant des cartes signifié par signifié. Dans le premier volume d'analyse nous avons publié 96 cartes. La raison de ce nombre limité est donnée plus loin.

Sur les 151 signifiés que nous avions tirés de la *Parabole*, nous avons constaté que dans 48 cas la carte obtenue aurait été mononyme: autrement dit dans ces 48 cas, la carte ne montrerait aucune variation puisqu'à un seul signifiant français correspond un seul lexème en basque. Quand, de façon très minoritaire, dans quelques communes, on trouve une réponse divergente due à une traduction éloignée, telle que le fait de traduire 'doigt' par *esku* 'main' dans une commune, nous avons considéré que la carte était mononyme si *esku* est la seule réponse partout ailleurs. Nous avons aussi neutralisé les marques morphologiques et considéré que par exemple *gazteago* litt 'plus jeune' ne se distinguait pas de *gazte* 'jeune'. Quand la différence est morphologique nous avons repoussé l'analyse au troisième volume de notre trilogie.

4.4. Cartes effectivement réalisées

Nous avons classé les 96 cartes dans l'ordre alphabétique à partir du français et non pas de l'ordre d'apparition dans le texte, pour de simples raisons de commodité.

Pour chaque carte nous donnons le nom du signifiant français avec son contexte puis:

a) la liste des mots-texte, par commune et en conservant l'orthographe originale tirée du manuscrit.

b) la carte elle-même, en couleurs, bâtie sur les polygones de Thiessen, le titre de la carte reprenant le signifiant français.

c) la légende fournit la liste des lemmes sous lesquels ont été classées les réponses du texte. A chaque lemme est assignée une couleur dont on trouve le répondant sur la carte.

d) Les observations nécessaires à la compréhension de certains choix de cartographie ou de lemmatisation.

©Universitat de Barcelona

Le fond de cartes reprend le territoire de Soule, Basse-Navarre et Labourd mais nous n'avons pas porté les communes dans lesquelles Bourciez n'a pas recueilli de version basque. Le fond de cartes est le même dans tout le travail, construit sur 138 communes dont chacune est représenté par un polygone. Il va de soi que ce polygone ne suit pas les limites géographiques réelles de chaque commune sur le terrain.

Les cartes de ce type ont une tradition bien connue en géolinguistique, surtout depuis les travaux fondateurs de Hans Goebl. L'atlas linguistique du Pays Basque sur lequel nous travaillons au sein d'Euskaltzaindia repose sur la même cartographie (dans cet atlas, toutes les communes bascophones n'ont pas été enquêtées). Pour le tracé des polygones, on a suivi la méthode inaugurée par Voronoï et Thiessen. Hans Goebl donne toute information sur ce mode de représentation (Goebl 1981, 1992). Que l'on sache cependant que dans un tel polygone, tout point appartenant à ce polygone est plus près du centre de son polygone que du centre de tout autre polygone.

Sur la carte, chaque polygone porte un numéro, qui est celui de la commune dont la liste apparaît sur la page de gauche de notre ouvrage. Ceci permet d'identifier assez aisément chaque commune, la taille de chaque polygone étant cependant variable, plus grande en Haute Soule qu'autour de Saint-Palais ou Saint-Jean-Pied de Port.

4.4.1. Les légendes

Chaque carte a sa légende. La légende est construite d'après la lemmatisation, elle fournit la liste des lemmes dans lesquels toutes les réponses des textes ont été distribuées. La couleur de chaque lemme est bien entendu celle que l'on retrouve sur la carte. Il faut ajouter quelques remarques sur la légende et donc la lemmatisation:

a) il peut arriver que quelque lemme ne corrresponde pas à une traduction exacte du signifiant français. Dans la carte 'ressembler (se)', des réponses comme *heldu* et *jin* (qui signifient en principe 'venir') peuvent apparaître incompréhensibles en dehors du contexte. Il faut se reporter au texte lui-même qui éclaire sur ces termes imprécis;

b) il existe des signifiés pour lesquels certaines traductions n'offrent aucun équivalent. Nous nous retrouvons dans la situation lacunaire bien connue en géolinguistique, celle des réponses zéro. On leur assigne la dernière position dans les lemmes de la légende et aucune couleur –sinon le blanc- n'est donnée sur la carte pour ces réponses zéro;

©Universitat de Barcelona

c) chaque lemme est suivi d'un nombre qui indique le nombre de parlers ou versions dans lequel il apparaît. Il donne déjà en soi une idée de la distribution géolinguistique;

d) le chiffre qui précède le lemme indique l'ordre dans lequel ces lemmes sont donnés mais ce chiffre n'à aucun intérêt géolinguistique;

e) Cependant, nous avons tenté de regrouper les lemmes en allant de ceux ayant le plus d'occurrences vers ceux les moins représentés, donc des lemmes couvrant un espace étendu vers ceux ayant une aire exigüe.

4.4.2. Liste des mots-texte

L'information donnée sur la carte n'est pas toujours exactement la même que la liste des formes attestées dans les textes. Comme nous l'avons déjà dit, les formes du corpus n'entrent dans la carte qu'après avoir été lemmatisées. Aussi, sur la page de gauche, en face de la carte correspondante, nous avons donné la liste des formes attestées dans les documents. Ces formes sont données dans la graphie du texte du manuscrit.

Sur cette page on retrouve donc:

- le titre de la carte avec la phrase dans laquelle se trouve le mot traduit, celui-ci mis en italique;
- le code du texte et de la commune;
- le nom de la commune en basque;
- le mot du texte, dans sa graphie originale.

4.4.3. La liste des lemmes

A la suite des cartes, nous avons fourni dans l'ouvrage la liste des lemmes ayant servi à la construction de ces cartes. Les lemmes sont donnés dans l'ordre alphabétique et chacun d'eux est suivi du numéro de la carte dans lequel il apparaît. S'il apparaît dans plus d'une carte, le lemme est répété.

©Universitat de Barcelona

4.4. Les cartes grammaticales

Les cartes grammaticales sont réunies dans le troisième volume consacré aux données Bourciez (Aurrekoetxea, Videgain & Iglesias (2005) [2008]). Nous avons lu les textes en essayant d'y trouver une série d'objets linguistiques intéressant la morphosyntaxe, tout en y ajoutant des questions d'ordre phonologique. Le texte étant d'une longueur limitée, bien des objets linguistiques sont restés en dehors du champ de cette étude.

Nous avons organisé ces 101 objets linguistiques de la façon suivante:

a) les cartes 1 à 11 sont consacrées à la déclinaison (génitif, sociatif, ablatif, datif pluriel, partitif) à partir des exemples proposés par le corpus;

b) dans un deuxième groupe, celui des cartes 12-28, nous avons réuni essentiellement déterminants, pronoms et adverbes de lieu (possessif, démonstratif, indéfini…);

d) le groupe suivant, dans les cartes 29-57 est consacré à la morphologie verbale, depuis le verbe *izan*, 'être' monovalent jusqu'aux formes tripersonnelles puis les formes allocutives;

e) un groupe consacré à 'plus', 'très', 'même', 'car', sans grande unité (58-61);

f) un autre groupe (62-75) réunit les questions de syntaxe, surtout les propositions subordonnées;

g) enfin le dernier groupe (75-101) réunit les cartes phonologiques. Bien entendu elles fournissent aussi une information lexicale dans ces cartes monolexématiques dans laquelle la seule variation est phonologique, comme pour les deux seules variantes *zeru* ou *zelü* pour 'ciel';

4.4.1. Méthodologie

Nous avons suivi la même méthodologie que pour les cartes lexicales mais la lemmatisation a été faite selon d'autres critères. Nous y reviendrons en une autre occasion.

©Universitat de Barcelona

4.4.2. Un dernier mot sur ces données

Avant d'aborder le traitement dialectométrique proposé plus loin, il nous apparaît que ces trois volumes fournissent une information à prendre en compte pour la distribution géolinguistique des objets linguistiques repérés sur le territoire concerné en Soule, Basse-Navarre et Labourd. Nous avons conscience de ne pas avoir épuisé le corpus. Il mérite encore approfondissement et l'ouverture de divers chantiers à venir est souhaitable. Par exemple, une étude pourrait mesurer l'influence des langues voisines, Par exemple, du point de vue diachronique, Par exemple, on peut procéder à une interprétation plus fine de ces données que dans nos cartes qui fournissent un point de vue à gros traits. Nous avons constaté que deux articles ont ouvert cette voie, l'un sur les apports lexicaux du recueil (Ormaetxea 2005), l'autre sur un point de déclinaison (Iglesias 2005)...

5. Dialectométrie

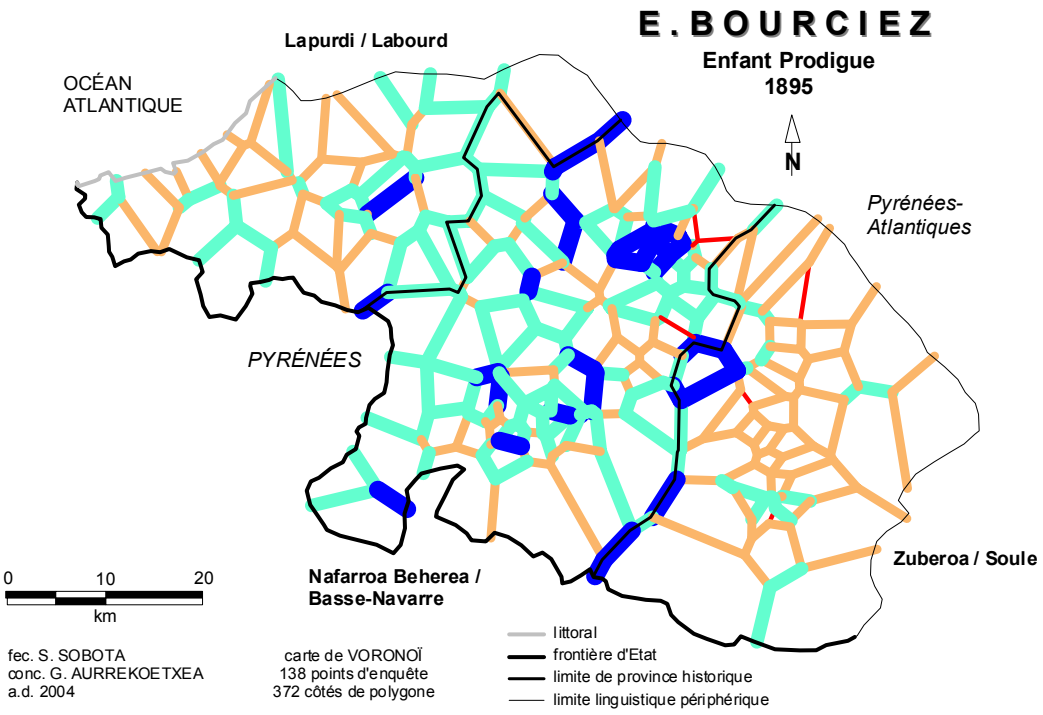
5.1. La distance linguistique entre les parlers ou DM interponctuelle

La dialectométrie interponctuelle est une branche de la discipline qui sert à mesurer la distance linguistique entre deux communes du territoire examiné. La dialectométrie peut s'intéresser soit à la similarité entre les deux parlers étudiés, soit à leur distance linguistique. Dans une direction on étudie ce qui réunit les deux parlers, et dans l'autre ce qui les distingue.

Pour ce faire, nous entrons dans une procédure globalisante. Nous n'utilisons pas les données d'une seule carte, d'un seul item sémasiologique dans notre cas mais nous prenons en compte l'ensemble des données fournies par la base.

Le carte 1 est parmi celles qui offrent le moins de complexité. Le traitement statistique y est réduit. On y montre les similarités et différences tirées des données des 151 cartes, en prenant les communes deux à deux.

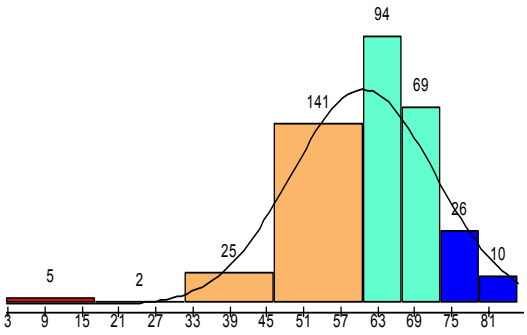
©Universitat de Barcelona



Algorithme de regroupement
MINMWMAX 4 groupes

[1] de 2.65 à 31.65 (n = 7)
[2] à 60.64 (n = 166)
[3] à 73.59 (n = 163)
[4] à 86.55 (n = 36)
Total des éléments: 372

distribution des fréquences de similarité
MINMWMAX 8 groupes



Carte 1: carte interponctuelle montrant la fonction distinctive des parlers (isoglosses synthétiques et quantitatifs)
Corpus: 151 cartes lexicales. Indice de distance. IPD(1)$_{jk}$ ou GIW. Algorithme de regroupement: MINMWMAX 4 groupes

5.2. Distance linguistique entre deux communes contiguës

Nous signalons que le terme 'isoglosse' n'a pas tout à fait le même sens que dans la dialectologie traditionnelle. En effet, nous pouvons traiter ici d'isoglosse synthétisant, dans uns vision globalisante des données. Il ne s'agit pas donc d'un isoglosse concernant un seul objet linguistique mais une ligne imaginaire qui met en

©Universitat de Barcelona

évidence la différence graduelle entre deux parlers en se basant sur une somme de données.

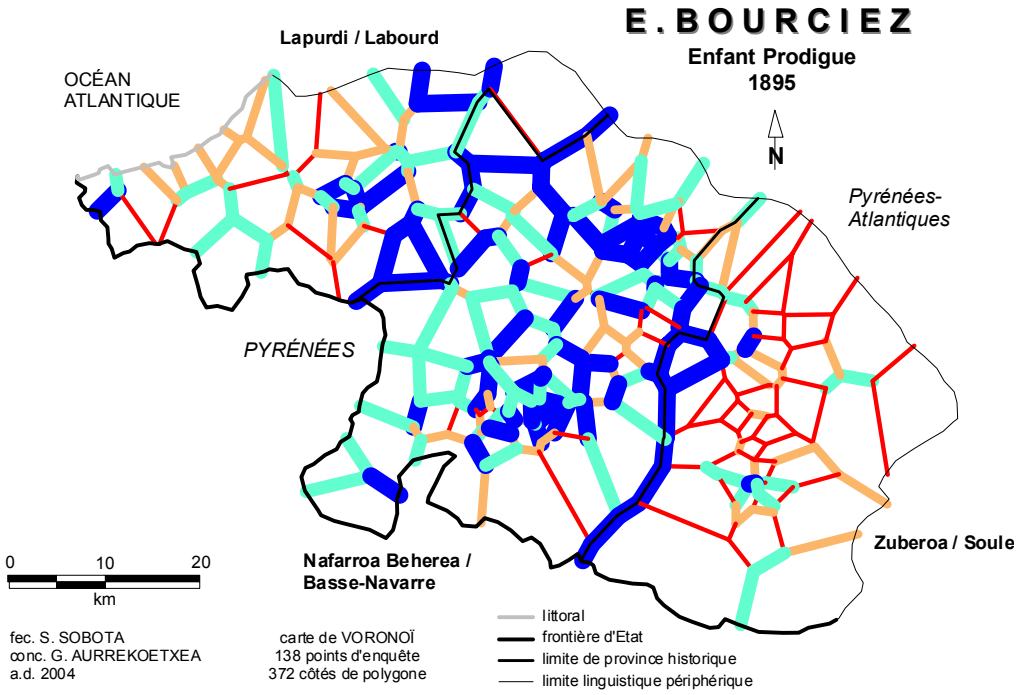
Du point de vue technique, cet isoglosse suit les bords des polygones placés de part et d'autre de l'isoglosse. Cet isoglosse est donc matérialisé par une succession de segments rectilignes et non pas par une ligne ondulante suivant la limite géographique des communes concernées comme cela se fait en cartographie traditionnelle.

Dans la dialectométrie interponctuelle, on ne traite que les communes contigües. La deuxième carte montre bien ce que nous voulons dire: cette carte est réalisée à partir de l'indice de similarité GIW. L'algorithme de regroupement est MINMWMAX, ce qui facilite l'apparition de données réparties en quatre groupes et donc de quatre couleurs sur la carte. Aux couleurs chaudes (rouge et ocre) correspondent les groupes allant de la moyenne faible (2,65) à la moyenne arithmétique [60,64) et aux couleurs rouges les groupes allant de la moyenne arithmétiques à la moyenne forte (86,55).

Il faut comprendre sur cette carte qu'elle visualise la relation linguistique qu'entretient chaque parler. La couleur rouge et ocre signale avec quels parlers un parler donné à une grande relation linguistique, autrement dit une similarité assez élevée. Ce type de relation apparaît surtout en Soule et Labourd. Bien qu'on trouve en Basse-Navarre un certain nombre de ces relations de niveau élevé, les types de relations faibles ou mauvaises sont plus nombreux. Il est significatif de voir comme il existe un haut degré de mauvaises relations entre Basse-Navarre et Soule. La frontière de province est très semblable à la distance linguistique.

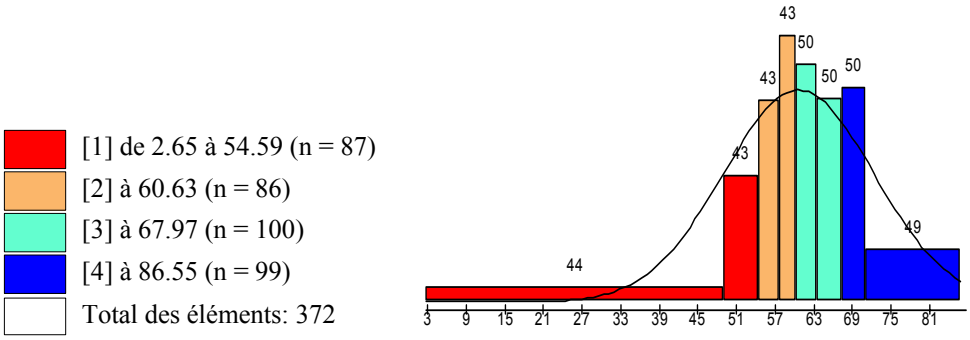
Dans la carte n° 2, bien qu'on utilise le même indice de similarité, GIW, l'algorithme de regroupement est différent; il s'agit de MEDMW qui donne aussi quatre groupes. Les données sont les mêmes que ci-dessus en carte 3. Les différences visibles sur la carte sont donc à imputer au changement d'algorithme.

©Universitat de Barcelona



Algorithme de regroupement
MedMw 4 groupes

distribution des fréquences de similarité
MedMw 8 groupes



Carte 2: carte interponctuelle montrant la fonction distinctive des parlers (isoglosses synthétiques et quantitatifs)

Corpus: 151 cartes lexicales. Indice de distance. $IPD(1)_{jk}$ ou GIW. Algorithme de regroupement MedMw 8 groupes

Il faut ajouter qu'outre la nature des couleurs chaudes ou froides qui renseignent sur la qualité des relations linguistiques, l'épaisseur du trait qui relie les parlers est significative. Plus le trait est épais, plus la distance linguistique est grande; plus le trait est fin, plus la distance linguistique est réduite.

Dans la carte 2, la plupart des parlers qui entretiennent de bonnes relations sont groupés sur deux espaces, l'un très nettemet en Soule, l'autre, moins nettement en

©Universitat de Barcelona

Labourd. Autour de la frontière entre Labourd et Basse-Navarre, la frontière linguistique se détache, en couleurs bleu, clair ou bleu sombre. Il n'apparaît pas en Basse-Navarre un espace à forte cohésion: les isoglosses bleu clair ou bleu foncé y sont nombreux, bien qu'on y trouve quelques isoglosses aux couleurs ocre ou rouges. Cette carte signale donc que la Basse-Navarre ne montre pas un territoire linguistique composé de parlers ayant entre eux une très grande proximité ni donc une grande homogénéité.

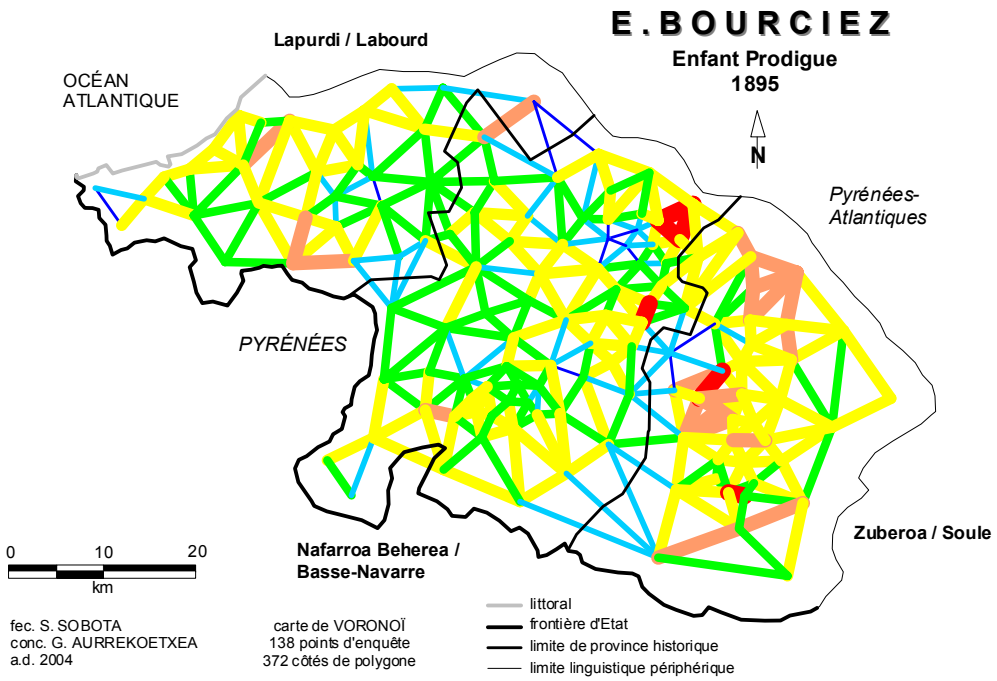
Des isoglosses de couleur bleue séparent Basse-Navarre et Soule, davantage au Sud qu'au Nord. Cette rupture est évidente et montre une faille importante entre les deux domaines. Il s'agit d'un phénomène de grande importance géolinguistique.

La Soule est l'espace présentant la plus forte proximité linguistique entre ses parlers, et la plus grande homogénéité, sauf à Pagolle dans le recueil Bourciez. Mais l'instituteur de Pagolle a manifestement rédigé le texte dans un parler basque éloigné de celui de la commune. C'est ce qui explique que le parler de Pagolle dans ce recueil semble jouer une partition propre.

La carte 2, obtenue avec l'algorithme MEDMW, montre des espaces linguistiques plus visibles. En effet cet algorithme augmente les contrastes en rapprochant les proximités et en éloignant les oppositions. Ici encore, c'est la Soule qui est l'espace le plus cohérent avec une multitude d'isoglosses rouges. Avec cet algorithme, la frontière entre Labourd et Basse-Navarre est plus marquée tout comme celle entre Basse-Navarre et Soule. Une zone de transition importante s'étend entre Basse-Navarre et Labourd, tandis que la rupture entre la Basse-Navarre et la Soule est très nette, surtout au Sud en zone de montagne.

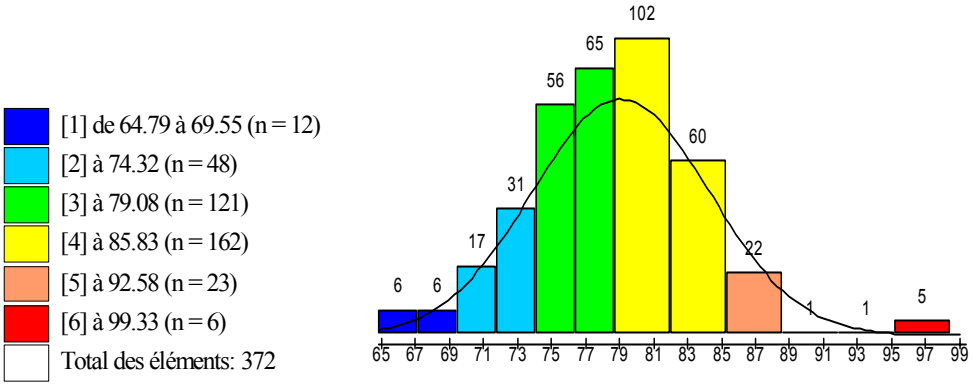
©Universitat de Barcelona

5.3. Carte de similarité: interprétation géolinguistique de la carte 3



Algorithme de regroupement
MINMWMAX 6 groupes

Distribution de la fréquence de similarité
MINMWMAX 12 groupes



Carte 3: carte ponctuelle montrant la fonction communicative entre les parlers.
Corpus: 151 cartes lexicales. Indice de distance. IRI. Algorithme de regroupement MINMWMAX. Carte réalisée avec la triangulation de Delaunay.

Cette carte basée sur la triangulation de Delaunay montre la fonction communicative des parlers, c'est-à-dire la similarité ou ressemblance entre les parlers. Comme les cartes 1 et 2 cette carte ne compare que des parlers contigus, soit chaque parler avec ses premiers voisins. Cependant, malgré ces limites, elle nous intéresse du

©Universitat de Barcelona

point de vue géolinguistique. Elle sert à détecter les noyaux dialectaux homogènes; plus le trait est épais, et plus il est 'chaud' et donc la proximité linguistique est importante

L'un de ces noyaux le plus net est celui visible en Soule qui va jusqu'aux limites du pays de Mixe. Les parlers qui le composent ont de bonnes ou très bonnes relations entre eux. Ils dessinent un espace ferme, à la grande homogénéité. Par contre, en Basse-Navarre, les parlers ayant de bonnes relations ne dessinent pas un espace marqué. On peut en trouver quelques-uns par vallées comme en Ostabarret (1306-Ibarrole, 1308-Juxue, 1310-Larceveau, 1311-Ostabat auxquels s'ajoutent les voisins 1307-Irissarry, 1309-Lantabat. De même une homogénéité se manifeste entre Les Aldudes, Saint-Etienne de Baïgorry, Arnéguy, Uhart-Cize et Lasse. Dans l'ensemble, la Basse-Navarre est avare d'espaces très unis.

La situation est assez comparable en Labourd. On notera en Labourd oriental que des parlers ont de très mauvaises relations avec les voisins de Basse-Navarre. Les chiffres sont bien en dessous de la moyenne arithmétique.

Parmi les parlers ayant de très mauvaises relations avec les voisins, on retrouve Pagolle (1297), en raison d'une 'imposture' de l'instituteur local. Plus curieusement Luxe-Sunberraute (1292) a aussi de mauvaises relations. On y verra aussi, moins nettement qu'à Pagolle que l'instituteur a eu recours à bon nombre de tours très occidentaux. Reste le cas de Louhossoa (1224) et Amorots (1272) isolés de leurs voisins.

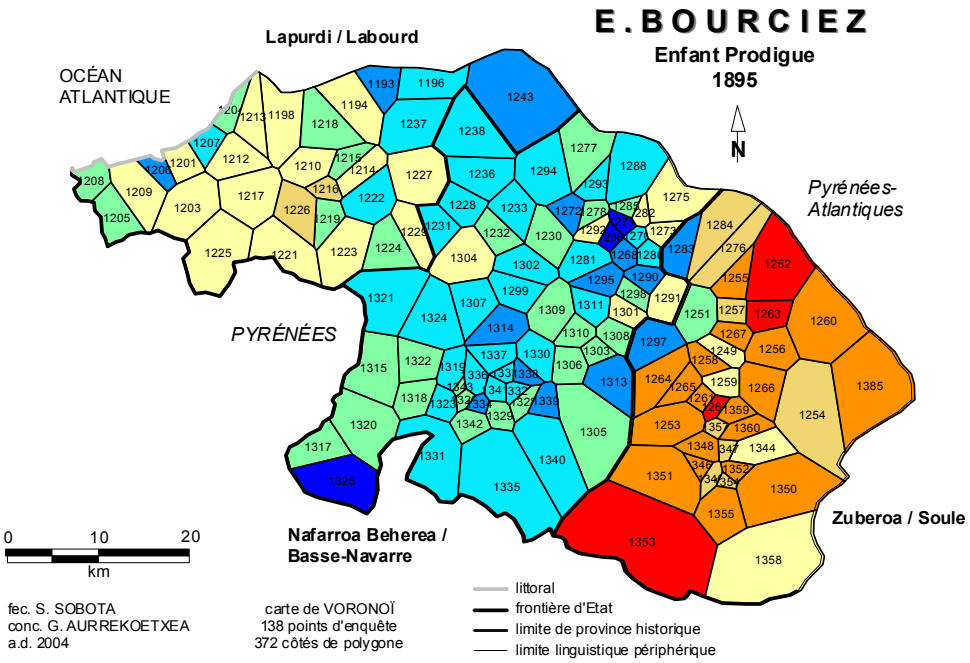
Dans l'ensemble, on peut dire qu'à l'exception de la Soule, le territoire examiné ne présente pas un ensemble dialectal très homogène s'il faut en croire le matériau que nous avons utilisé.

5.4. Présentation de la carte 4: carte des écarts-types

L'écart-type est un outil classique de la statistique descriptive (Goebl 2002, 24). La synopse des écarts-type sert surtout à repérer les parlers-pont ou les zones de transition. A définir ce qu'est une zone de transition ou un parler-pont, nous dirions qu'il s'agit des parlers intermédiaires situés entre les noyaux dialectaux signalés ci-dessus. La caractéristique de ces parlers est de posséder peu de caractères propres et d'offrir peu à peu le 'chemin' ou 'pont' qui permet de passer d'un système linguistique à un autre.

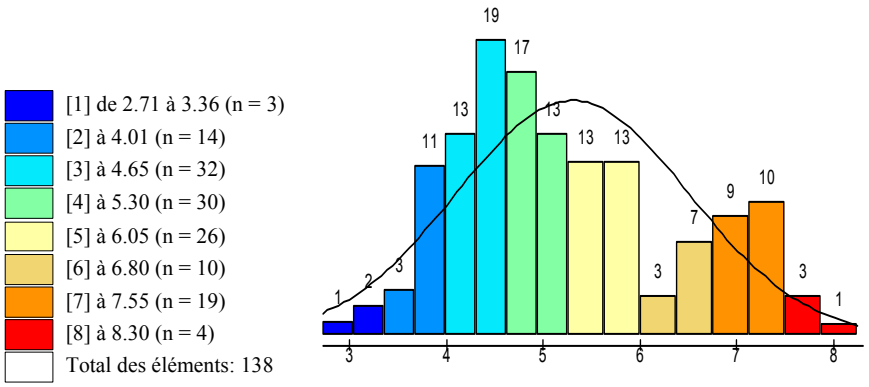
©Universitat de Barcelona

L'écart-type permet d'isoler et de mettre en évidence ces parlers situés entre noyaux dialectaux à forte cohésion.



Algorithme de regroupement
MINMWMAX 8 groupes

Distribution de la fréquence de similarité
MINMWMAX 16 groupes



Carte 4: carte des écarts-type: desbideratze tipikoaren mapa.
Corpus: lexique des 151 cartes. Indice de similarité IRI_{jk}. Algorithme de regroupement MINMWMAX-8 groupes

Dans la carte 4, deux noyaux dialectaux se dessinent de part et d'autre de l'espace considéré, en Labourd et en Soule. Entre ces deux noyaus assez nets, on trouve les parlers de transition (colorés en bleu ou jaune) qui font transition. L'ensemble des parlers bas-navarrais serait concerné s et constituerait une vaste zone de transition. On notera que cette zone de transition connaît tout demême une rupture plus forte avec la

©Universitat de Barcelona

Soule qu'avec le Labourd. Cette structuration de l'espace linguistique révélée par cette carte ne manque pas d'importance puisque synthèse des données de 151 cartes lexicales de 137 parlers.

5.5. Synopse des coefficients d'asymétrie de Fisher (carte 5)

Nous n'avons pas insisté jusqu'ici sur le graphique de distribution des fréquences qui apparaît en bas à droite de nos cartes. Un simple regard permet de voir que cette distribution varie d'une carte à l'autre. Le fait que cette courbe soit symétrique ou asymétrique est une donnée d'importance.

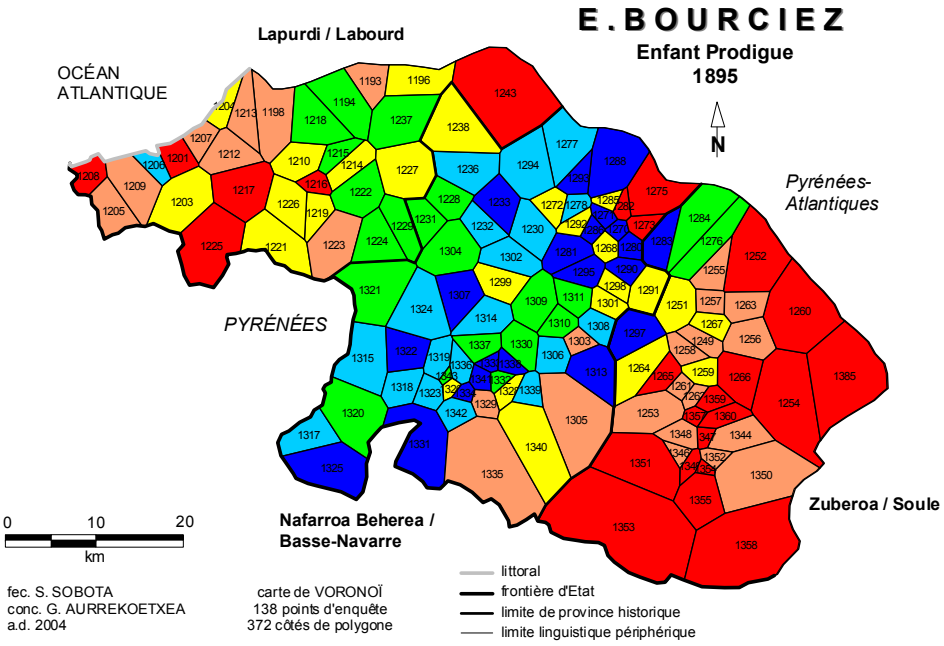
L'un des outils les plus connus pour mesurer ceci est appelé coefficient d'asymétrie de Fisher. On dit que la distribution des fréquences est symétrique quand la distribution des fréquences est également répartie de part et d'autre de la moyenne arithmétique. Au contraire l'asymétrie apparaît si l'un des côtés est nettement plus important, L'asymétrie apparaîtra à droite (valeurs positives) quand le nombre de fréquences au-dessus de la moyenne arithmétique est plus élevé et l'asymétrie apparaîtra à gauche (valeurs négatives) quand le nombre de fréquences au-dessous de la moyenne arithmétique est plus élevé. Le coefficient est 0 quand la courbe est symétrique.

On considére généralement que d'un point de vue géolinguistique, les parlers qui ont obtenu des fréquences supérieures à la moyenne arithmétique (valeurs positives) sont des parlers conservateurs et parallélement que les parlers qui ont obtenu des valeurs négatives ont de bonnes compétences communicatives avec les autres parlers.

Dans la carte 5 deux aires conservatrices apparaissent, en Labourd et Soule, de couleur rouge, aux deux extrémités de la carte.D'un point de vue géolinguistique, ces aires ont de mauvaises relations linguistiques avec les parlers voisins et conservent bon nombre de caractéristiques propres. Inversement, en jaune et bleu, les parlers aux valeurs négatives dont de bonnes relations avec les autres parlers et en même temps peu de caractéristiques propres. Apparaît ici un point de vue antagoniste à celui dégagé par la carte 10: alors que la carte 4 montrait des parlers de transition, la carte 4 montre surtout des parlers qui ont peu évolué.

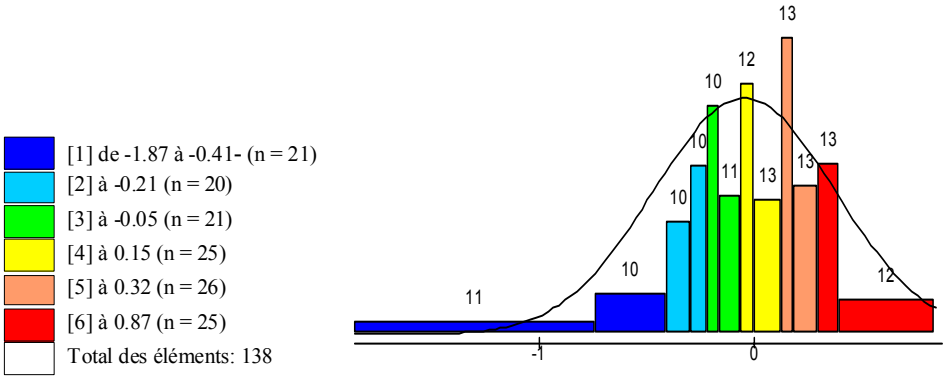
©Universitat de Barcelona

Le coefficient de Fischer permettrait aussi une analyse diachronique comme l'a fait Goebl à partir des données de l'Atlas Linguistique de la France en étudiant les aires d'oc et d'oil? Nous ne le suivrons pas sur ce point dans l'article présent.



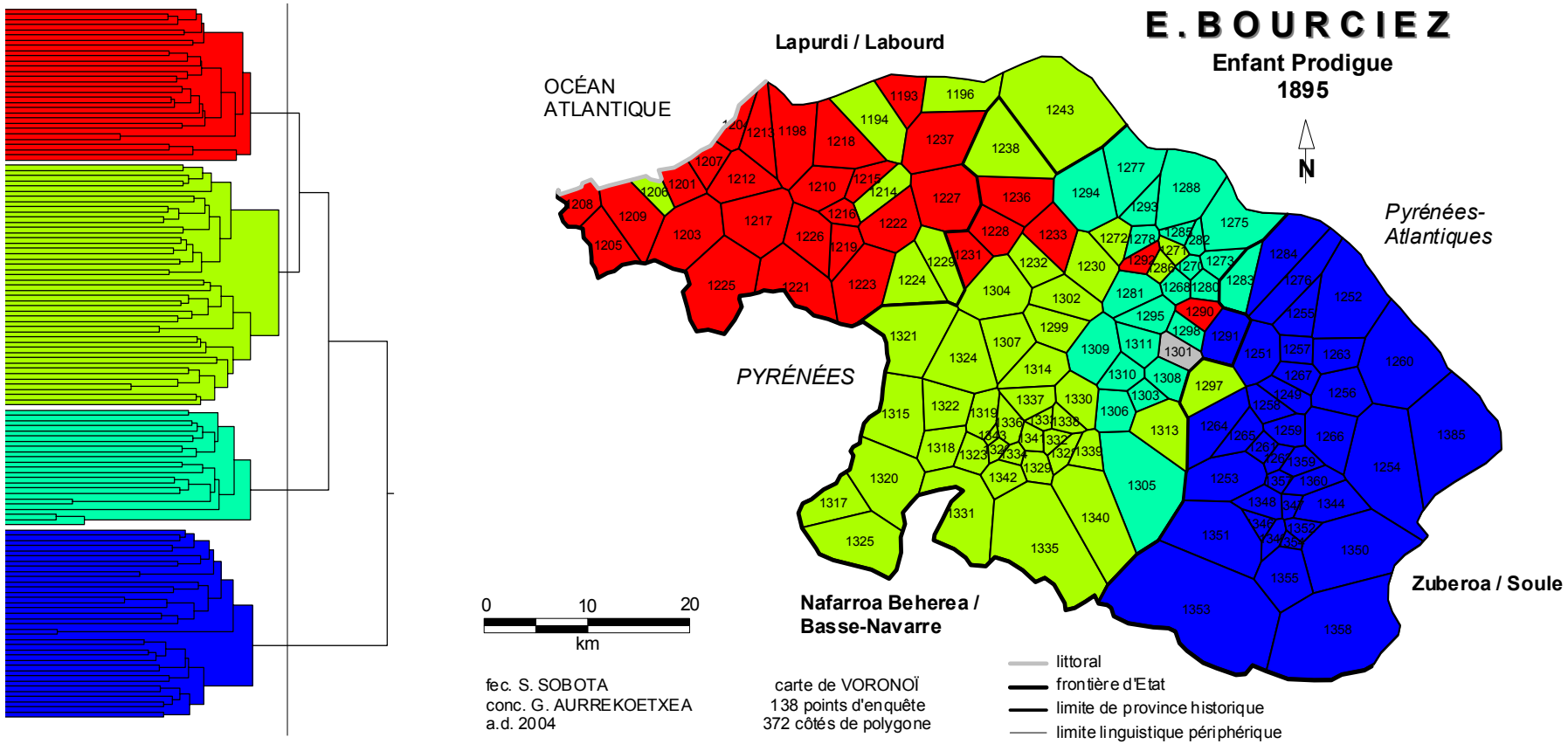
Algorithme de regroupement
MedMw 6 groupes

Similaritate maiztasunen distribuzioa
MedMw 12 groupes



Carte 5: carte choroplèthe de synopsie du coefficient d'asymétrie de Fischer.
Corpus: lexique des 151 cartes. Indice de similarité IRI_{jk}. Algorithme de regroupement MEDMW -6 groupes.

©Universitat de Barcelona



Carte 6. La structuration géolinguistique des données du dendrogramme
Algorithme Ward.
Indice de similarité IRI_{jk}.

5.6. Dialectométrie multidimensionnelle ou dendographique (cartes 6 et 7)

La dialectométrie multidimensionnelle permet de franchir un pas supplémentaire dans l'étude de la structuration géolinguistique. Elle s'appuie sur la classification ascendante hiérarchique (CAH). Parmi les outils que nous avons choisi pour la classification automatique, nous utilisons les dendogrammes qui sont représentés sous forme d'un arbre logique.

Dans cette classification, tous les parlers sont pris en compte. Sur la carte 6 tous ces dendrogrammes sont placés sur la gauche au début de la ligne. Ces parlers sont situés en fonction de leur proximité linguistique. Chaque groupe rassemble l'ensemble des parlers qui en font partie.

Cette procédure de classification est bien connue en sciences sociales (sociologie...) et en linguistique (glottochronologie) et la dialectologie et géolinguistique y ont recours: outre Goebl (1992) d'autres travaux l'ont utilisée (Aurrekoetxea 1995).

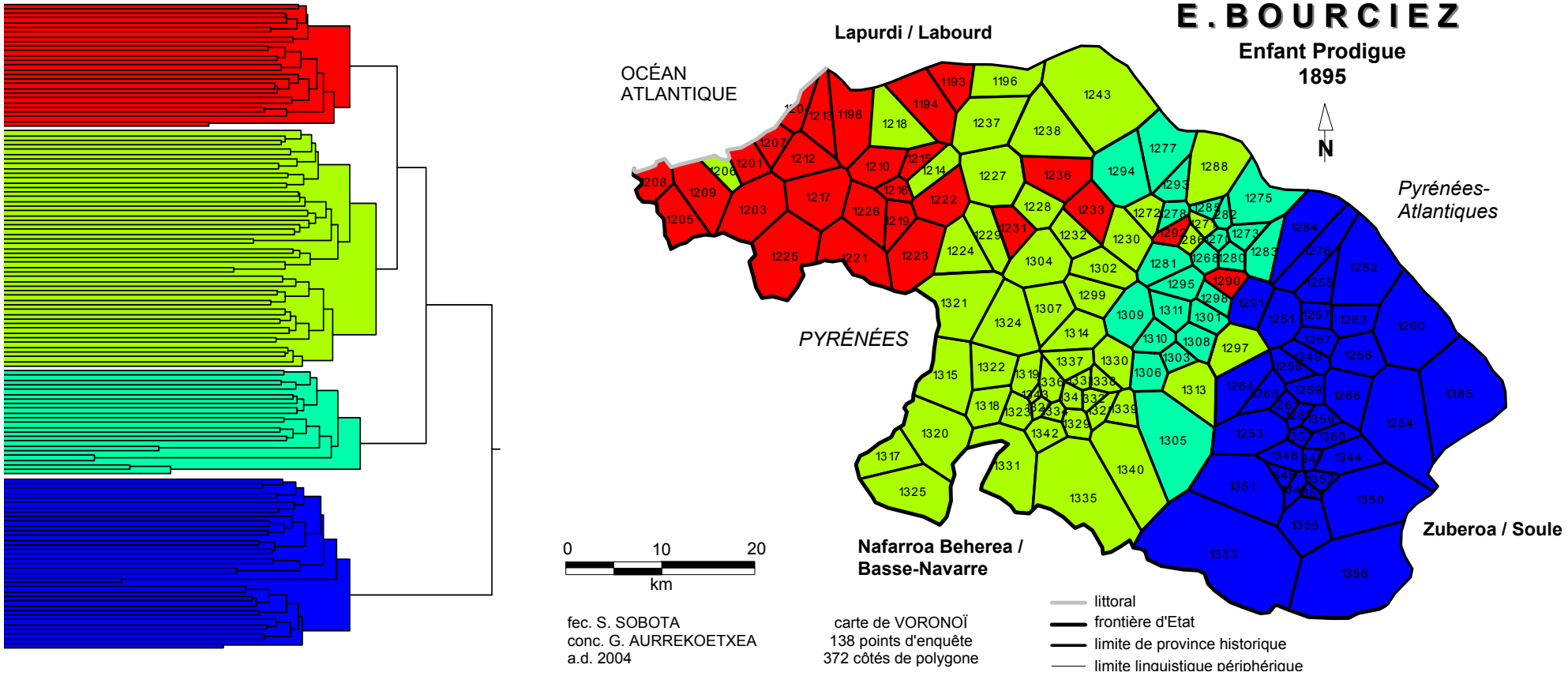
La carte 7 présente la carte et le dendrogramme. La carte n'est rien d'autre que la projection géolinguistique du dendrogramme. Ici encore, on procède à un regroupement des parlers en quatre groupes aux couleurs distinctes. La couleur de chaque groupe est celle portée sur la carte.

Sur la carte, quatre groupes cohérents apparaissent; sauf une exception:

a) le groupe de couleur rouge regroupe la plupart des parlers du Labourd et quelques parlers de Basse-Navarre. L'aire est assez cohérente, mêle si on y trouve Larribar (1290) et Luxe-Sumberraute (1292. Et cette aire comprend en vert des parlers qui sont regroupés dans le groupe suivant: il s'agit de Ciboure (1206), Mouguerre (1194) et Jatxou (1215);

b) le groupe de couleur verte regroupe les parlers de Basse-Navarre occidentale. Il constitue une aire dense et cohérente si on élimine les trois communes citées ci-dessus appartenant au Labourd. Un autre groupe discontinu au nord est formé pr Urcuit (1196), Bardos (1194) eta Isturitz (1238). On trouve deux autre communes isolées en Mixe, Amendeuix (1271) et Garris (1286). Enfin Saint-Just-Ibarre (1313) et Pagolle (1297) forment une autre micro-aire;

©Universitat de Barcelona



Carte 7. La structuration géolinguistique des données du dendrogramme
Algorithme Ward.
Indice de similarité $IPI(1)_{jk}$

c) le groupe de couleur bleu clair regroupe les parlers de Basse-Navarre orientale et Domezain-Berraute (1283) de Soule. 23 parlers sont ainsi regroupés, essentiellement situés en pays de Mixe plus Lantabat, Domezain et 7 communes du pays d'Ostabarret. Cette aire est très cohérente qui n'a pas de parler hors de son aire continue, même si elle enclave deux communes regroupées ici dans le premier groupe (rouge);

e) le groupe de couleur bleu sombre regroupe tous les parlers souletin, sauf Domezain qui 'marche' avec le pays de Mixe, et Pagolle dont nous avons déjà signalé le peu de pertinence dans le recueil Bourciez. La cohérence de ce groupe est étonnante et très supérieure à celle des trois autre groupes.

Le dendrogramme présente ici quatre groupes, nous l'avons dit. C'est le choix du chercheur. Mais celui-ci peut changer le nombre de groupements. Il a choisi ici quatre groupes parce que c'est un niveau d'analyse qui montre une répartition géolinguistique assez proche de la fameuse carte des dialectes basque élaborée par Louis-Lucien Bonaparte à une date à peine plus ancienne que les données du recueil Bourcies que nous utilisons ici. Le chercheur peut donc jouer sur le nombre de regroupements dans le dendrogramme selon les phénomènes qu'il veut étudier.

La carte 7 ne présente que quelques différences: elle utilise l'algorithme $IPI(1)_{jk}$ qui pondère les données. L'aire rouge et l'aire verte sont les plus concernées, très peu l'aire bleu clair et pas du tout l'aire bleu foncé.

6. Une première comparaison entre les cartes obtenues par procédure classificatoires et la carte de Louis-Lucien Bonaparte

Nous pouvons esquisser une première ébauche de comparaison entre les cartes produites ici et leurs résultats et la fameuse carte des dialectes basques proposée à Londres par Louis-Lucien Bonaparte au 19ème siècle.

a) d'une part nous faisons mine d'ignorer ici toutes les discussions sur la nature des dialectes, sur leur existence même;

b) d'autre part, la démarche de Louis-Lucien Bonaparte est totalement différente de celle suivie ici. Quant aux données elles-mêmes, les nôtres ne reposent pour l'instant ici que sur des faits lexicaux alors qu'on sait que L.L. Bonaparte privilégia les données d'ordre morphologique. Travaillant 'à la main', sans concours d'aucun outil

©Universitat de Barcelona

mathématique ou de classification, et pour cause, L.L. Bonaparte avait sélectionné un nombre d'objets linguistiques assez réduits à partir desquels il distinguait les dialectes et sous-dialectes;

c) les procédures classificatoires ne laissent préjuger de rien quant à la structuration géolinguistique. Nous avons ici fourni les données de 151 cartes et 151 parlers sans chercher à l'avance qu'elles nous indiquent une ou autre distribution dans l'espace;

d) nous avons cependant utilisé deux indices de similarité qui fournissent une structuration légèrement différente, l'indice pondéré contribuant à fournir une carte dans laquelle les faits linguistiques minoritaires sont surreprésentés, ce que l'autre indice de similarité ne fait pas.

On sait comment Bonaparte en vint peu à peu à diviser les parlers des trois provinces de Soule, Basse-Navarre et Labourd en quatre dialectes: labourdin, bas-navarrais occidental, bas-navarrais oriental et souletin. Pour le dialecte labourdin, il proposait trois sous dialectes: le labourdin propre (avec trois variantes, Saint-Jean de Luz, Sare et Ainhoa), et le labourdin hybride (Arcangues). Pour le bas-navarrais occidental il proposait trois sous-groupes, ainsi que pour le bas-navarrais oriental. Pour le souletin, il proposait deux sous-dialectes: le souletin proprement dit et le roncalais (celui-ci en dehors de notre champ d'études ici puisque situé en Navarre).

Une comparaison entre la fragmentation dialectale selon Bonaparte et la nôtre en carte 7 montre quelques changements, qui ne sont pas cependant d'une ampleur considérable.

Labourdin: les communes qui limitent le labourdin sont Lahonce 1193, Mouguerre 1194, Jatxou 1215, Cambo 1222 et Itxassou 1223. A l'intérieur du domaine labourdin, deux communes semblent montrer une plus grande proximité avec le bas-navarrais occidental: Ciboure 1206 et Villefranque 1218.

Alors que Bonaparte classe Lahonce et Mouguerre comme bas-navarraises, notre carte les signale comme labourdines.

Alors que Bonaparte classe la variété d'Ustaritz comme bas-navarrais occidental notre carte la donne comme labourdine.

Bas-navarrais occidental. Les communes limitrophes à l'Ouest sont Urcuit 1196, Briscous 1238, Hasparren 1227, Macaye 1229, Louhossoa 1224, Bidarray 1231. A l'est les communes limitrophes sont Bardos 1243, Amorots 1272, Meharin 1230,

©Universitat de Barcelona

Armendaritz 1302, Iholdy 1299, Suhescun 1314, Ainhice-Mongelos 1330, Lecumberry 1339 et Mendive 1340.

Ici, la différence entre la carte Bonaparte et la nôtre réside dans le fait que bon nombre de parlers que Bonaparte situait dans l'un ou l'autre bas-navarrais apparaissent dans notre carte comme faisant partie du bas-navarrais occidental. Autrement dit, dans notre carte, le bas-navararrais occidental couvre une aire plus étendue que dans la carte de Bonaparte.

Bas-navarrais oriental; cette aire recouvre les parlers du pays de Mixe et les communes de Lantabat 1309, Larceveau 1310, Ibarrole 1306 et Hosta 1310. A part ces quatre parlers, la carte de Bonaparte et notre carte 14 montrent la même aire quant au bas-navarrais oriental.

Souletin. A part le cas de Pagolle (pas fiable dans nos données Bourciez) et celui de Domezain 1283, la carte de Bonaparte et notre carte diffèrent sur un parler, celui de Lohitzun. Bonaparte le donne comme bas-navarrais oriental, notre carte le porte comme souletin.

En conclusion, nous espérons avoir montré l'intérêt du recueil Bourciez et de l'avoir traité selon diverses méthodes pour montrer la structuration géolinguistique des parlers basques de Basse-Navarre, Soule et Labourd, en introduisant dans la dialectologie basque les méthodes de la dialectométrie.

7. Bibliographie

ALLIÈRES, J. (1978) "Petit Atlas linguistique basque français Bourciez 1, (20 cartes commentées)", *Fontes Linguae Vasconum*, 27, 353-386.

AURREKOETXEA, G. (1995) *Bizkaieraren egituraketa geolinguistikoa [la structuration géolinguistique du bizcayen],* Bilbao: Université du Pays Basque.

AURREKOETXEA, G. (2009) "Iparraldeko hizkeren sailkapena (1): lexikoa", *ASJU,* XXXVIII-1 (2004), 287-331.

AURREKOETXEA, G. & Ch. VIDEGAIN (eds.) (2004) *Haur prodigoaren parabola Ipar Euskal Herriko 150 bersiotan* [La parabole de l'enfant prodique en 150 versions basques], Anuario del Seminario de Filología Vasca "Julio de Urquijo", annexe XLIX, Bilbao: Service de publications de l'Université du Pays Basque.

AURREKOETXEA, G. / A. IGLESIAS / Ch. VIDEGAIN (2007) *Bourciez Bildumako Euskal Atlasa (BBEA-1): 1. Lexikoa. [El atlas lingüístico Bourciez: 1. Léxiko]*, ASJU 38:2 (2004).

©Universitat de Barcelona

AURREKOETXEA, G. / A. IGLESIAS / Ch. VIDEGAIN (2008) *Bourciez Bildumako Euskal Atlasa (BBEA-2): 2. Gramatika. [El atlas lingüístico Bourciez: 2. Gramática]*, ASJU 39-1 (2005).

AZKUE, R.M. (1922) *Cancionero popular vasco*, 2 vol., Barcelona: Boileau & Bernusconi.

BOURCIEZ, E. (1895) "Introduction au Recueil des Idiomes de la Région Gasconne", *Revue des Pyrénées et de la France méridionale*, 321-324.

GOEBL, H. (1992) "Problèmes et méthodes de la dialectométrie actuelle (avec application à l'AIS)", in Aurrekoetxea, G. & Ch. Videgain (eds.), *Actes du congress international de dialectologie,* Bilbao: Euskaltzaindia, 429-475.

GOEBL, H. (2003) "Regards dialectométriques sur les données de l'Atlas linguistique de la France (ALD): relations quantitatives et structures de profondeur", *Estudis Romànics* XXV, 59-96 +24 cartes.

GOEBL, H. (2005) "La dialectométrie corrélative: un nouvel outil pour l'étude de l'aménagement dialectal de l'espace par l'homme", *RLiR* 69, 321-367.

GOEBL, H. (2006) "Sur le changement macrolinguistique survenu entre 1300 et 1900 dans le domaine d'oïl. Une étude diachronique d'inspiration dialectométrique", *Linguistica* XLVI, 1-37.

VIAUT, A. (1995) "Bourciez et les études gasconnes", in *L'ethnologie à Bordeaux. Hommage à Pierre Métais*. Actes du colloque du 10 mars 1994, Bordeaux: Département d'Anthropologie Sociale-Ethnologie / Centre d'Etudes et de Recherches Ethnologiques / Université de Bordeaux II, 109-128.

VIDEGAIN, Ch. (1989) *Le vocabulaire de l'élevage en pays d'Oztibarre. Contribution aux archives de l'oralité basque*, (dir. J Haritschelhar & J.L. Fossat), Université de Bordeaux III, n.p., 2178 p.

©Universitat de Barcelona