

**Zeitschrift:** Revue de linguistique romane  
**Herausgeber:** Société de Linguistique Romane  
**Band:** 37 (1973)  
**Heft:** 145-146

**Artikel:** La dialectométrie dans l'Atlas linguistique de la Gascogne  
**Autor:** Séguy, J.  
**DOI:** <https://doi.org/10.5169/seals-658403>

### **Nutzungsbedingungen**

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. [Siehe Rechtliche Hinweise.](#)

### **Conditions d'utilisation**

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. [Voir Informations légales.](#)

### **Terms of use**

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. [See Legal notice.](#)

**Download PDF:** 22.01.2025

**ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>**

## LA DIALECTOMÉTRIE DANS L'ATLAS LINGUISTIQUE DE LA GASCOGNE

*Qu'il y ait des chiffres, c'est mon régal, en  
des matières où l'on s'obstine à ne pas en mettre.*

Th. LALANNE <sup>1</sup>

C'est sans le moindre scrupule que nous écrivons le néologisme *dialectométrie*. Aux choses nouvelles, des mots nouveaux. Et encore la chose n'est-elle pas si nouvelle : les premiers dialectologues qui ont appliqué à leurs recherches des méthodes numériques ont déjà disparu. Citons, pour la seule France, Adolphe Terracher *Les aires morphologiques des parlers de l'Angoumois* (Paris 1914) <sup>2</sup> ; notre regretté compagnon de travail Théobald Lalanne : nous reviendrons sur ses études, publiées ou non, où se rencontrent des innovations lumineuses ; certaines communications du colloque de Strasbourg 1971 sur les atlas linguistiques de la France par régions (la plus importante, de ce point de vue, étant celle de M. Guiter *Atlas et frontières linguistiques*) ; et notre article paru ici même dans le fascicule 2 de 1971 (p. 335-357) est une réalisation des recherches et procédures analogues actuellement en cours dans divers centres de dialectologie (Montpellier, Nancy). Puisque une *économétrie*, une *sociométrie* et une *jurimétrie* figurent déjà à l'état civil des sciences humaines, rien ne s'oppose au baptême et à l'inscription de la *dialectométrie*.

Nous avons fait subir aux matériaux de l'ALG un traitement mathématique. Mais de quelle mathématique s'agit-il ? Avant tout, de la plus traditionnelle, de la plus familière, celle dont les piliers sont l'addition, la division et la règle de trois, celle qui se constitue en statistique. Rarement, nous avons recouru à la logique mathématique — qu'on appelle, qui sait pourquoi ? *mathématiques nouvelles* —. La logique mathématique, on le sait,

1. Lettre du 23 février 1951. — Les atlas linguistiques qui indiquent l'échelle du fond de cartes sont une infime minorité.

2. V. *Via Domitia*, 1956, p. 40-41.

*Revue de linguistique romane.*

fait merveille dans d'autres branches de la linguistique ; mais il se trouve qu'en dialectométrie, elle n'est utile qu'en quelques occasions, que nous n'avons eu garde de négliger, notamment dans certaines applications de la théorie des ensembles et dans la quantification de la ressemblance.

Le volume VI et dernier de l'*ALG*<sup>1</sup> présente encore des séries de cartes purement descriptives, par exemple le long chapitre consacré au pronom. Mais c'est à la dialectométrie qu'est vouée la part majeure de ce tome, cela dans deux secteurs, dont nous traiterons ici successivement : 1<sup>o</sup> le regroupement, le comptage, et la statistique des faits de toute sorte dispersés dans les cinq volumes précédents ; 2<sup>o</sup> le problème des frontières dialectales, ce secteur se trouvant d'ailleurs sous la dépendance du premier.

La définition de ces deux secteurs va immédiatement évoquer, dans l'esprit du lecteur, le verbe *interpréter* et ses dérivés. Nous refusons cette étiquette. En matière de cartographie linguistique, on en fait un étrange abus : est qualifiée d'interprétative toute carte qui n'est pas un simple report de données point par point, à la manière de Gilliéron. Or, parmi les sens d'*interpréter* définis par Littré, aucun ne correspond à nos activités, si ce n'est le premier, obsolète et remplacé aujourd'hui par *traduire* : *interpréter de l'anglais au français*. Interpréter, c'est dire les causes d'un phénomène, causes réelles ou supposées. Quand nous écrivons, à propos de l'archiphonème /E/ qu'au point 675 N nous avons relevé 54 réalisations [é] fermé, 30 de [è] ouvert et 14 de [e] moyen, sans rien ajouter, où est l'interprétation ? De même quand, nous référant à la carte 1618 du volume V (par M. Allières), nous rangeons le point 645 dans la colonne des localités à verbe fort ? Parfois, nous signalons que tel fait (par exemple les écarts à la finale féminine [o] moyen) donne matière à interprétation ; nous invitons le lecteur à interpréter lui-même, sans lui suggérer quoi que ce soit. On devra aussi se convaincre une fois pour toutes que jamais nous ne procédons à des choix. S'il s'agit d'un échantillon, comme c'est le cas pour le lexique, cet échantillon est procuré par tirage au hasard des numéros des cartes. Pour les autres critères, il n'y a pas davantage choix, puisque l'exploitation est exhaustive. Compter, traduire en chiffres et en graphismes n'est pas interpréter. Le volume VI de l'*ALG* n'enfreint donc pas les normes du genre : nous ne faisons qu'élaborer une matière informe et insaisissable et que mener les données quantitatives jusqu'à leur formalisation terminale et mathématiquement nécessaire.

1. A paraître en 1973.

*Premier secteur.*

Au premier coup d'œil jeté sur un atlas linguistique dans la tradition de Gilliéron, on est frappé par le foisonnement des nuances phonétiques. On admire l'honnêteté et le scrupule des enquêteurs. Mais au long de la fréquentation de l'atlas, une question vient se poser : quelle est la portée de cette minutie ? C'est l'usager qui se pose la question, et non, ou rarement, l'enquêteur-auteur : en son âme et conscience, il a rempli sa mission, qui était de noter objectivement tout ce qui était accessible à son oreille, sans se préoccuper de la fonction linguistique de ces variantes, préoccupation qui pourrait nuire à sa liberté d'esprit. Ayant couché dans les cartes les produits de sa collecte, il s'estime quitte. Il complète sa satisfaction en songeant vaguement que d'autres chercheurs, dans un avenir plus ou moins proche, s'intéresseront à ces milliers de choses minuscules et sauront en tirer parti, en déduire des conclusions.

Nous avons écrit ailleurs <sup>1</sup> que le « réalisme naïf » (Langevin) doit être le seul souci de celui qui décrit les dialectes. Mais cela n'est qu'une étape — une bien longue étape —. Nous ajoutons : « Dans les volumes suivants, nous nous proposons de donner de la langue gasconne une description graduellement abstraite. M. Allières traitera du verbe dans le volume V, en s'efforçant de dégager du réel les structures partielles et globales <sup>2</sup>. Je ferai de même dans le volume VI en ce qui touche les divers aspects de la phonétique, de la phonologie, de la morphologie et de la syntaxe. »

Or, en fait, les riches collections que constituent les atlas linguistiques restent sous-exploitées, comme le serait un herbier dont les consultants n'étudieraient que les caractères macroscopiques des échantillons. Les phénomènes lexicaux, qui sont toujours à l'échelle macroscopique, la phonétique diachronique au niveau du phonème ont donné lieu et donnent lieu à des travaux nombreux, solides et brillants. Mais qui a jamais tenté de faire, à partir des paradigmes verbaux non cartographiés de l'*AIS* et de l'*ALMC*, un travail d'analyse et de synthèse tel que celui de M. Allières ? Qui a pris l'initiative de traiter, au sens informatique du mot, les multitudes de notations micro-phonétiques de l'*ALF* et de sa descendance ? Il

1. *ALG IV, Avant-propos*, p. 4.

2. C'est maintenant chose faite : *ALG*, vol. V, en deux tomes : un atlas de 456 cartes et un livre de commentaires (302 p. et de nombreuses cartes) (Paris, Publications du CNRS 1972.)

est vrai que l'entreprise est peu engageante. Ne parlons pas de la perspective d'une tâche lourde, longue et assidue. Mais ce qui décourage, c'est la formidable dispersion des faits. Les atlas linguistiques sont comme les dictionnaires : leur destination est de recueillir des mots. Si le lecteur ne s'intéresse qu'au lexique, tout va bien : il n'est que de consulter la liste alphabétique des titres de cartes. Mais s'il se propose d'aborder une question de phonétique diachronique, il doit fouiller l'atlas, en suivant des pistes hasardeuses, afin de déceler les vocables porteurs de l'information ; et si certains se laissent facilement appréhender grâce aux lois des correspondances romanes, beaucoup se cachent dans des recoins insoupçonnés. Il existe des atlas où l'on a tenté de remédier à cet inconvénient en mettant au questionnaire une série d'articles spécialement consacrés à des faits phonétiques (*ALPI I*, par exemple). C'est mieux que le chaos habituel. Mais en dépit de l'excellente connaissance que les auteurs ont de leur domaine en général, ils ne peuvent présupposer qu'un nombre de faits limité et de faits déjà plus ou moins connus : l'inconnu qui se révélera sur le terrain s'éparpillera dans la jungle, comme dans les autres atlas. Et alors qu'on croyait un signifié mononyme dans tout le domaine, on a la surprise de rencontrer un autre étymon qui crée une lacune sur une aire plus ou moins étendue.

Dans le premier chapitre du volume VI, nous nous employons à regrouper les données phonétiques éparses dans les volumes I à V. Qui, en effet, connaît mieux que l'auteur d'un atlas les cartes et les caches où se tiennent les mots possédant le phonème primitif dont il faut écrire l'histoire ? Dans l'*ALG*, les voyelles pénultièmes des proparoxytons apocopés, qui donnent les produits les plus divers, souvent mélangés et lexicalisés à la même localité, se trouvent dans quatorze cartes lexicales ; les faits d'assimilation consonantique interne, dans trente-quatre cartes. Ainsi, le lecteur qui serait intéressé par ces questions trouve-t-il le travail tout fait et n'a pas à battre les buissons. On sait que les évolutions phonétiques concordent rarement quand on superpose les aires des mots porteurs : chaque carte marque les limites des traitements phonétiques respectifs. De plus, le lecteur peut toujours contrôler, puisque les sources, à savoir les cartes lexicales exploitées, sont toujours énumérées. Mais fermons vite cette parenthèse dans l'exposé de la dialectométrie de l'*ALG* : car la computation n'y tient aucun rôle, si ce n'est de temps en temps, par des cartes en graphismes qui montrent le gradient géographique de tel produit d'évolution (c'est le cas, justement, des pénultièmes de proparoxytons et des faits d'assimilation et de non-assimilation).

Le deuxième chapitre de l'ALG VI s'intitule *Phonologie* (71 cartes). Là, tout, ou peu s'en faut, repose sur les nombres. Nous nous en tiendrons à citer et à commenter quelques exemples.

Touchant la détermination des phonèmes, nous écrivions il y a dix ans <sup>1</sup> : « Notre premier travail, en transcrivant les phonogrammes, est de noter avec une objectivité maximale toutes les nuances des réalisations audibles. C'est l'écoute répétée d'un son forcément identique chaque fois que la tête magnétique le relit qui nous permet d'arriver à une certitude définitive. Les déterminations des réalisations sont opérées sans préjugés ni visées systématiques. Ensuite, le moment venu de déterminer la nature exacte d'un phonème, nous comptons et classons toutes les réalisations. Si une réalisation ressort avec un pourcentage très important (par exemple de l'ordre de 80 %), c'est cette réalisation qui sera retenue comme phonème. Il se peut que ce phonème vienne se placer dans un schéma parfaitement symétrique ; il se peut aussi qu'il provoque un gauchissement de la figure, un décrochage par rapport à un autre ordre. Mais c'est ce qui nous importe peu. L'essentiel est de décrire les choses telles qu'elles sont, et non selon les plans de notre spéculation ». Ce propos, nous nous sommes efforcé de le réaliser dès que tous les matériaux de l'ALG eurent été cartographiés. La carte 2159 traite du phonème /o/ tonique ; les quatre réalisations [ò] ouvert, [o] moyen, [ó] fermé, [o] légèrement diphtongué sont décomptées à chaque localité. L'opération porte sur 38 cartes lexicales ou morphologiques données en référence, qui fournissent en tout 11 250 occurrences. Partout, sauf à un point, la réalisation [ò] ouvert apparaît en très forte majorité. Nous en déduisons qu'en gascon 1<sup>o</sup> il n'existe qu'un seul phonème /o/, 2<sup>o</sup> que la substance de ce phonème est /ò/ ouvert. D'un autre point de vue, la fréquence du phonème /ò/ tonique varie selon les régions du domaine : le gradient géographique de cette fréquence est figuré dans une carte spéciale.

C'était là un cas simple et facile. Il n'en fut pas de même de l'archiphonème /E/ tonique. La ventilation de trois variétés principales [é], [e], [è] a été faite sur 57 000 occurrences ; elle est présentée en pourcentage dans deux cartes, l'une numérique, l'autre graphique. Mais le résultat est bien moins net que pour /ò/ tonique : ici, le traitement mathématique ne suffisait pas pour déterminer s'il existe, à certaines localités, deux phonèmes /E/

1. *Actes du X<sup>e</sup> Congrès international de linguistique et philologie romanes* (Strasbourg 1962) : *Essai de cartographie phonologique appliqué à l'Atlas linguistique de la Gascogne* (p. 1029-1050).

ou bien un seul <sup>1</sup>. Il a fallu procéder par des moyens plus complexes qui sont détaillés dans une série de cartes : comptage des écarts par rapport aux proto-phonèmes /é/ et /è/, relevé des paires minimales (dans la mesure où la faiblesse du corpus le permet), calcul des coefficients de latitude. Il est résulté de ces recherches où la statistique, la phonétique diachronique et les principes de la phonologie, faisant litière de la distinction des genres, se portent entraînée 1<sup>o</sup> que la très grande majorité des localités gasconnes possèdent deux phonèmes : /é/ et /è/ ; mais que cinq ont le seul phonème /è/ et une le seul phonème /é/ ; 3<sup>o</sup> que l'unique phonème /E/ du « gascon noir » reste bien /è/ ouvert, conformément à son origine historique et cela en dépit du remaniement structural qui a converti en /œ/ le phonème /é/ primitif ; et la latitude de réalisation de cet unique phonème /è/ n'y est ni plus ni moins élevée que celle des deux /é/ ~ /è/ dans la majeure partie du domaine. Le remaniement structural a donc été inutile <sup>2</sup>.

« Il arrive aussi que la réalisation 1 d'un phonème se présente dans une proportion de 56 %, la réalisation 2 dans une proportion de 40 %, sans qu'il y ait le moindre conditionnement d'entourage, de tempo, etc. Dans de telles conditions, il nous paraît abusif de décréter : « la réalisation 1 est le phonème ». Nous dirons que ce phonème est mutant 1/2, ou polymorphe 1/2, ou tout ce qu'on voudra, mais non que les réalisations 2 sont négligeables, relèvent de la variation biologique, etc. Parce que cela n'est pas. Il faudra trouver un moyen quelconque d'intégrer ces réalisations 2, même, répétons-le, si cela doit troubler l'harmonie des représentations graphiques habituelles. » Le cas, par bonheur, ne s'est présenté qu'une fois, mais d'une façon irréductible. Dans les trois quarts du département de la Gironde, l'archiphonème /E/ prétonique est réalisé tantôt [e], tantôt [ə], en proportions égales et suivant des occurrences parfaitement aléatoires. En conséquence de quoi nous avons créé un signe spécial pour ce phonème biface.

Dans les parties centrale et orientale du domaine, la finale féminine apparaît sous forme d'un [o] moyen qui domine les autres timbres /O/ dans des proportions le plus souvent écrasantes (statistique portant sur 30 250 occurrences). Mais le dénombrement des écarts à cet [o] moyen permet des cons-

1. L'existence de trois phonèmes /E/ distincts par leur degré d'aperture est tout à fait exclue du gascon : nulle part ni la commutation ni l'analyse fonctionnelle ne laissent apparaître un tel état.

2. On chercherait en vain cette dernière phrase dans l'une des cartes de l'ALG, puisque son contenu est délibérément interprétatif. Ce qui est exclu d'un atlas linguistique devient légitime dans un article de revue.

tations intéressantes : les timbres « sombres » [u, ó] se groupent au nord et à l'est, les timbres « clairs » [ò], les réalisations de /a/, à l'ouest et au sud (Pyrénées). La carte montre — mais aucun commentaire ne le souligne — que les écarts clairs se rassemblent autour des localités conservatrices qui ont gardé la finale /a/ posttonique primitive. Le lecteur en tirera ou n'en tirera pas des conclusions historiques, à son gré.

On appelait jadis *phonétique descriptive*, ou même *phonologie*, ce qui est devenu l'une des provinces de la science phonologique : les réalisations des phonèmes (*phonèmes* au sens actuel). Dans l'ALG, les réalisations de /s/ implosif, intérieur ou en sandhi, sont étudiées dans dix-huit cartes, selon la nature du phonème suivant. La carte 2185, par exemple, apprend qu'au point 6750 /s/ devant /b/ est réalisé deux fois par un souffle voisé, une fois par une continue bilabiale voisée, une fois par un très bref arrêt du débit, et qu'une fois il disparaît sans laisser de traces, les références étant données aux cartes lexicales d'où ces faits ont été extraits.

Ces échantillons montrent les services que la bonne vieille arithmétique peut rendre à la phonologie, surtout quand on prétend dresser une phonologie diatopique, entreprise à la fois risquée et d'une difficulté peu commune.

La syntaxe des dialectes, on ne le sait que trop, ne se laisse pas capturer au moyen d'un questionnaire. Il faudrait s'aposter à chaque localité des semaines et des mois durant pour y écouter le discours spontané : projet hors de possibilité. Si l'enquêteur demande au témoin de lui donner la traduction d'un énoncé français, il n'obtiendra, le plus souvent, qu'un calque. C'est pourtant à cette manière d'agir que nous avons eu recours, parce que nous ne pouvions nous résigner à passer par pertes et profits la syntaxe du gascon, et parce que cette langue possède une assez large autonomie pour permettre quelque espoir.

Les énoncés français se partagent en deux séries. L'une est continue : ce sont les 274 articles de la morpho-syntaxe du pronom. Grâce au bilinguisme gascon-français, grâce à une batterie de recettes psycho-linguistiques qu'il n'est pas loisible de détailler ici, M. Ravier a obtenu partout, avec une relative facilité, la traduction mécanique et inconsciente de ces phrases, dans des formulations idiomatiques souvent sans le moindre rapport avec le français.

L'autre série est disséminée dans le questionnaire de l'enquête complémentaire. Prenons un exemple concret : « Quand vous voudrez, l'épicier ouvrira une caisse. » On sait d'avance qu'en gascon, ou dans une partie du domaine gascon, le verbe de la subordonnée temporelle, au futur en français, doit être au subjonctif présent, tout comme en castillan. L'informa-

teur vient de traduire à la suite onze petits énoncés contenant des formes du verbe « ouvrir » ; il sait donc que celui de l'épicier est également relatif à « ouvrir ». Effectivement, l'enquêteur obtient le futur de ce verbe ; mais du même coup, le magnétophone enregistre la traduction de « quand vous voudrez », que l'auteur-transcripteur note dans le cahier. Vingt constructions analogues sont disposées dans le questionnaire du verbe. L'auteur n'a plus qu'à compter combien de fois sur vingt le témoin a rendu le futur français par le subjonctif présent gascon. Nous avons pu ainsi dresser une carte en gradient. Le subjonctif est nul dans la partie nord de la Gironde et du Lot-et-Garonne ; il apparaît timidement quand on va vers le sud, puis se renforce, et des têtes à 100 % surgissent dans la Chalosse et dans les Pyrénées, par grappes de trois à quatre localités.

En fait, il s'agit d'un *test de résistance au français*. Les résultats positifs sont acquis. Mais les résultats négatifs restent ambigus : ils signifient ou bien que le tour au subjonctif n'existe pas, ou bien qu'il existe en concurrence avec le futur, ou bien qu'il manque de solidité et qu'il est menacé par le calque français. C'est loin d'être parfait, mais c'est mieux que le néant. De même pour « on » et ses équivalents, etc.

Les résultats de la première série sont également comptabilisés. Les énoncés du genre « il passe sans me voir ; je pars pour le voir ; je m'en vais pour ne pas le voir » sont au nombre de 66. Ils étaient destinés à montrer l'antéposition ou la postposition du pronom régime à l'infinitif prépositionnel. Dans la région pyrénéenne, la postposition (enclise sur l'infinitif) se réalise à 100 % et doit être considérée comme obligatoire ; dans les Landes, elle est nulle ou facultative, mais obligatoire si l'énoncé est négatif. Ailleurs, le régime est antéposé. Pour chacun des types, les occurrences sont chiffrées dans la carte. Mais quand l'infinitif est précédé d'un semi-auxiliaire (« je veux le voir » ; 17 énoncés au minimum), l'enclise est partout facultative dans l'aire positive, les Pyrénées présentant cette fois des chiffres très inférieurs à ceux des Landes.

L'ALG VI contient 34 cartes consacrées à la syntaxe sous forme de statistiques. Dans les cas utiles, les gradients sont figurés en graphismes. Faut-il préciser que les données des cartes syntaxiques sont purement spontanées ? En règle générale, M. Ravier s'est abstenu de solliciter. La suggestion, si féconde quand il s'agit du lexique<sup>1</sup>, est contre-indiquée en matière de syntaxe.

1. ALG IV, *Avant-propos*, p. 11 ss.

Il faut savoir que Lalanne avait réalisé un bon nombre de cartes en champs gradients, fondées sur des statistiques (sous le nom d'*aires échelonnées*<sup>1</sup> ; les têtes des champs gradients étaient des *noyaux d'aires* ; mais les différences d'appellations couvrent les mêmes signifiés).

*Deuxième secteur.*

Il est occupé par la mesure des distances linguistiques entre les localités de l'atlas.

Nous avons exposé dans cette même revue (1971, 2 ; p. 339-342) la méthode que nous avons employée pour mesurer la distance linguistique réduite à sa seule composante lexicale. Résumons. Nous tirons au hasard les numéros de 170 cartes, représentant 10 % de l'effectif total de l'ALG<sup>2</sup>. Ces cartes sont prises au hasard, parce que le classement de l'ALG n'est pas alphabétique, mais logique, c'est-à-dire que les cartes sont groupées par champs sémantiques. Si nous avions pris les cartes de 1 à 100, l'examen du lexique se serait cantonné aux animaux sauvages et aux plantes cultivées. Tandis que dans l'échantillon au hasard, on trouve de tout : la belette, la bergeronnette, égrener le maïs, le pou, chercher, la gyrole, la barrière, la serpe, le tas de foin, se rouler par terre (en parlant des ânes), le cri pour faire avancer les bœufs, une douve de tonneau, maladroit, assommer, changer, cheminée, regarder sans être vu, verrou, printemps, oser, goujon, furoncle, avare, digérer, Toussaint, jambon, gauche, frère, cloche, appeler, non plus, naître, tiède... D'une carte à l'autre, la variété lexicale est dispersée entre 0 (cartes mononymes) et 52 termes différents. Le critère principal de la différence est celui de l'étymologie : 89 % des cas ; 11 % procèdent d'autres critères, parfois discutables et plus ou moins subjectifs. On construit ensuite une matrice, où les lignes correspondent aux 154 localités de l'atlas et où les colonnes — que nous appellerons désormais *items* — représentent les 170 cartes de l'échantillon. Dans chacune des cases ainsi définies, on porte le numéro de code du vocable.

1. Th. Lalanne *L'indépendance des aires linguistiques en Gascogne maritime* (2 vol. ronéotés ; 77 et 112 p. ; Saint-Vincent-de-Paul, 1949).

2. Cent cartes suffisent quel que soit le nombre des cartes d'un atlas : c'est ce qu'a démontré M. Guiter (*Actes du colloque CNRS sur les atlas linguistiques de la France par régions* (Strasbourg 1971) : *Atlas et frontières linguistiques*) ; mais nous ne connaissions pas encore cette communication. Toutefois, pour notre étude sur les distances lexicales dans divers atlas (*RLiR* 1971, 335-357), nous n'avons retenu que cent cartes, y compris dans l'ALG.

Sur ce modèle, nous avons dressé quatre autres matrices : phonétique diachronique, phonologie, morphologie verbale et morpho-syntaxe. La plus nette est la matrice phonologique : chaque item est un phonème observé dans le domaine gascon (*phonème* au sens fonctionnel et structural du mot). Par exemple, les localités qui possèdent le phonème /t/ en position finale sont marquées 1 dans cette colonne ; celles qui ne le connaissent pas sont marquées 0. Les données de cette matrice procèdent des cartes phonologiques dont nous avons parlé dans le premier secteur ; d'autres renvoient directement à des cartes lexicales où l'on voit la répartition géographique du phonème en question. Le paramètre phonologique de l'ALG repose sur une phonologie distributionnelle : on envisage les consonnes dans leurs trois positions initiale, intérieure et finale, et les voyelles suivant l'accent de mot (très important et même pertinent en gascon) : prétoniques, toniques et posttoniques. La phonologie non distributionnelle ne fournirait que des différences très faibles. Cette matrice compte 77 items.

Dans la matrice de phonétique diachronique, ce sont les phénomènes évolutifs qui sont consignés et codés. Par exemple *-n-* intervocalique latin maintenu/disparu ; les divers produits de la diphtongaison de *ò* ouvert conditionnée par *yod* ; le traitement de la finale *-ia*. Il est nécessaire que le fait évolutif soit attesté sur tout le domaine : un fait connu seulement d'une façon lacunaire ne peut être pris comme critère. L'aire marquée peut d'ailleurs être fort petite : c'est le cas, par exemple, de *ò* tonique plus nasale. Il se maintient tel quel à cinq localités pyrénéennes seulement, codées 1 ; il passe à *u* dans les 149 autres localités de l'ALG, codées 0. Les items de cette matrice, au nombre de 67, proviennent des cartes de phonétique historique (premier chapitre de ALG VI).

Nous traiterons ensemble des matrices du verbe et de la morpho-syntaxe, puisqu'elles sont conçues sur le même principe. Tous les faits de caractère lexical ont été laissés de côté, par exemple les tiroirs des verbes « être », « aller », presque entièrement lexicalisés ; les formes du pronom neutre *òk, ik, ek, ak, au, zu* ; ces faits sont cependant très nombreux ; mais le critère lexical était saturé par les 170 items de la matrice afférente. Nous n'avons donc pris en considération que des faits paradigmatiques, ou formant système, ou se présentant comme des règles grammaticales. Exemples : le gérondif organique existe/n'existe pas (1-0) — subjonctif imparfait des classes II et III, vocalisme morphémique : 0 = *u* ; 1 = *e* ; 2 = *u* ; 3 = *i* ; 4 = polymorphe — infinitif de la classe II oxyton/paroxyton (1-0) — le pronom neutre existe/n'existe pas (1-0) — postposition du pronom régime

après les gérondifs : 0 = non, 1 = oui — article devant les noms de personne : 1 en usage, 0 inusité — particule *e* entre le sujet et le verbe : 0 = n'existe pas, 1 = existe, etc. La matrice du verbe (44 items) procède du volume V par M. Allières ; celle de la morpho-syntaxe (68 items) renvoie à diverses divisions du volume VI.

Le total des items des cinq paramètres monte à 426. Dans les matrices, chacun des items porte un numéro : les clés pour le décodage de ces numéros figurent dans le petit volume annexe ; en outre, le décodage de chaque item est assorti des références aux sources du critère : numéros des cartes lexicales, phonologiques, etc.

Pour mesurer la distance lexicale, phonologique, etc. entre deux localités, il n'est que de juxtaposer les deux lignes les concernant <sup>1</sup>. Chaque fois que les deux cases correspondant au même item portent des chiffres différents, on marque un point de distance ; on ne marque rien quand les chiffres sont identiques. On fait le total lorsqu'on a examiné tous les items, et ce total donne la distance lexicale, phonologique, etc. entre la localité A et la localité B. En totalisant les totaux des cinq paramètres, on obtient la distance linguistique, ou distance dialectale. Quant à la signification de cette sixième mesure, elle sera examinée dans la dernière partie de l'article.

Sauf dans la matrice lexicale, l'opération est facilitée par le codage ordinairement binaire (0-1). Toutefois, il ne nous paraît pas que l'homme doive se priver d'utiliser des items ternaires, quaternaires... (voir exemple ci-dessus), pour la raison que l'ordinateur n'est capable de traiter que du binaire.

Cette méthode — qui est une application de la distance de Hamming — paraît avoir les caractéristiques suivantes :

1<sup>o</sup> Elle est intégrale, puisque comme nous l'avons dit dans le paragraphe *Premier secteur*, toute l'information de l'atlas est intégrée, jusque dans le petit détail. Les critères phonologiques, phonétiques et morpho-syntaxiques sont exploités exhaustivement ; quant à l'échantillon lexical, il est incontestablement représentatif de l'ensemble <sup>2</sup>.

2<sup>o</sup> Elle est impersonnelle, puisque le choix n'y a aucune part : les cartes lexicales ont été tirées au hasard, et les autres faits sont pris en compte dans

1. Nous avons écrit les matrices en double exemplaire, au moyen d'un papier carbone ; puis nous avons découpé la copie carbone suivant les lignes horizontales des localités. Il suffisait ensuite de prendre le « ruban » de l'une des deux localités et de le juxtaposer à la ligne de l'autre localité dans la matrice originale.

2. H. Guiter, dans la communication citée (Strasbourg 1971).

leur totalité. L'élimination des faits d'ordre lexical dans les matrices morpho-syntaxiques n'est pas un choix, mais une nécessité statistique.

3<sup>o</sup> Elle permet de calculer les distances entre deux localités quelconques en un temps relativement bref, et sans fatigue, puisqu'on n'a pas à porter de jugement : il suffit d'ouvrir l'œil à l'identité ou à la non-identité des chiffres de code dans la même case de la matrice ; dans ce dernier cas, que ce soit 0-1 ou 17-38 (lexique) revient au même : un point de différence (ou distance).

4<sup>o</sup> Elle donne au lecteur faculté de tout vérifier à tous les niveaux. Supposons un cas. Dans la carte 2524 intitulée *Distances linguistiques en pourcentages*, on lit que cette distance, entre les localités A et B est de 18 %. On va contrôler ce chiffre au moyen des cartes 2518 à 2522 qui donnent le détail des composantes phonétique, phonologique, etc. En consultant les matrices et leurs clés, on voit de quels critères sont faits ces composantes aux deux localités A et B. Si l'un de ces critères est la carte 2126 « -n- intervocalique », on trouvera dans cette carte les références aux cartes lexicales dont elle procède. Ainsi, par récurrence, le lecteur peut-il toujours aller des mesures terminales jusqu'aux sources premières dispersées dans les six volumes de l'atlas.

Mais cette méthode a au moins un inconvénient, et un inconvénient de taille : elle n'est applicable qu'à un atlas du type de l'*ALG*, c'est-à-dire comptant, en plus des cartes lexicales semblables à celles des autres atlas, quelque neuf cents cartes consacrées à la phonétique diachronique, à la phonologie et à la morpho-syntaxe. Or, M. Allières a mis plusieurs années à tirer du chaos les matériaux de l'enquête sur le verbe pour les traduire en 456 cartes occupant tout un volume et accompagnées d'un commentaire de 300 pages. Les opérations mathématiques incombaient à nous seul ; plusieurs fois, nous nous sommes trouvé aux limites des possibilités du calcul manuel, notamment en phonologie ; qu'on songe que pour les seules voyelles, nous avons dû traiter 122 000 occurrences ; le reste à l'avenant. Additionnées, les surfaces des cinq matrices couvrent deux mètres carrés, chaque case (ou carreau du papier quadrillé) mesurant 5 millimètres de côté. Ceci pour un atlas de 154 points d'enquête seulement. On peut donc estimer qu'au delà de 200 points d'enquête, la construction des matrices devient impossible par les moyens ordinaires.

Dès lors, beaucoup de nos lecteurs doivent se dire : « Mais qu'attend ce pauvre homme pour appeler l'informatique à son secours ? » On y a bien pensé : un seul exemple montrera que ce propos ne pouvait être que chimé-

rique. Je prends dans la carte 1181 « rillons », au point 679, le mot *héritys*, et j'en fais l'analyse — analyse dans ce cas relativement facile — destinée au programmeur. Je ne m'occupe pas de la transcription phonétique : transcodage que la perforeuse effectuera elle-même en consultant une table, table d'ailleurs richement garnie. Mais, en plus des numéros de la carte et de la localité, en plus de l'étymologie, je dois prévoir les informations suivantes : FR initial  $>$  *her* ; phonème /h/ réalisé pleinement (en vue d'une carte phonologique où se distingueront les localités à réalisation positive de celles qui présentent une réalisation zéro ; le moment venu, l'ordinateur devra me renseigner sur ce chef pour tous les mots de l'atlas commençant par /h/) ; *e* prétonique fermé (si je ne dégage pas cette information, l'ordinateur ne saura pas la découvrir en lisant l'étymon) ; de même pour *u* tonique final et en syllabe fermée ; nasale finale primitive devant morphème -s ; morphème *s* positif ; suffixe - $\bar{O}$ NE (il se peut que j'aie à connaître le nombre de mots portant ce suffixe dans le corpus). C'est là le minimum : je néglige certaines informations que je pense n'avoir jamais à utiliser, telle la règle phonétique  $\bar{O} >$  *u*, commune à tout l'occitan, ou la position des phonèmes les uns par rapport aux autres (distribution) <sup>1</sup>. Et ainsi de suite pour les centaines de milliers de vocables contenus dans l'ALG (quant à l'analyse des formes du verbe, nous ne nous sentons ni la force ni le courage d'imaginer un seul exemple). En admettant que je vienne à bout de l'analyse, combien de temps faudra-t-il pour réaliser la programmation monstrueuse ? Et combien d'argent ? Certes, l'informatique est capable de performances autrement complexes. Mais les gens étant ce qu'ils sont, les sciences humaines ne peuvent prétendre aux largesses nécessaires, sciences entre lesquelles la dialectologie se fait remarquer par sa rentabilité microscopique.

Il existe une autre méthode dialectométrique : c'est celle qu'expose M. Guiter dans la communication plusieurs fois citée ici. Pour aller vite, nous désignerons par G la méthode de M. Guiter, et la nôtre par S. G et S partent des mêmes matériaux de base : cent cartes lexicales prises au hasard dans un atlas. Les deux méthodes sont impersonnelles. Le principe de la

1. Quelques dialectologues utilisent l'informatique. Mais ils ne mettent en mémoire qu'une quantité dérisoire d'information eu égard aux besoins de la dialectométrie. Contrairement au lexique, les autres composantes du langage, spécialement la phonétique diachronique, ne peuvent se mesurer sur un échantillon au hasard. Et on ne perdra pas de vue que le corpus d'un atlas linguistique n'est lui-même qu'un échantillon assez mince de chaque parler.

mesure par points contigus est également commun à G et à S (réseau de triangulation, v. p. 21) ; la distance lexicale se compte de la même façon : une unité de différence, chaque fois que le mot change. A partir de là, G et S divergent : en G, les matériaux des cent cartes lexicales servent aussi à mesurer les autres composantes de la différence linguistique. La procédure se résume en une phrase : « Chaque fois que l'interpoint coupe une isoglosse, on marque une unité de différence ». Précisons : quelle que soit la nature de cette isoglosse : lexicale, phonétique, phonologique, etc., de sorte que le signifié de la carte lexicale peut être le support de 0 à  $n$  unités de différence à un même interpoint. Une fois les cent cartes dépouillées, on fait la somme des différences à chaque interpoint, et on obtient ainsi la distance linguistique globale de toutes les liaisons entre points contigus.

On aperçoit tout de suite l'avantage écrasant de G sur S : l'énorme machine du premier secteur est inutile et on attaque directement le deuxième secteur, celui de la distance linguistique, de sorte que la méthode G est immédiatement applicable à n'importe quel atlas linguistique.

Mais obtient-on les mêmes chiffres en employant indifféremment G ou S ? Oui, à très peu de chose près. Pourquoi et comment ? Ceci serait une autre histoire. C'est par l'expérimentation que nous avons vérifié cette concordance. Sur deux longs parcours de l'ALG, totalisant 44 interpoints, nous avons refait les calculs avec la méthode G ; nous avons figuré en courbes les valeurs des interpoints dans l'ordre du parcours. Si les valeurs sont rarement identiques aux points représentatifs, elles ne diffèrent que de quelques unités ; et en tout cas, ce qui est l'essentiel, les courbes G et S ont toujours la même allure ; les montées et les descentes ne sont jamais en contradiction. On peut donc présenter G et S comme également fiables. En fin de compte, les lecteurs tentés par la dialectométrie et que cet exposé aurait démoralisés peuvent reprendre courage, grâce en soient rendues à M. Guiter.

Dans ces conditions, on peut s'étonner que nous n'ayons pas nous-même employé la méthode G. La raison est des plus simples : le premier secteur, celui qui exige les calculs les plus laborieux, était au centre de la cible que nous visions depuis des lustres. Du moment que ces calculs étaient faits, il nous a paru tout indiqué de les remployer dans le deuxième secteur <sup>1</sup>.

\*  
\*\*

1. Rappelons que la méthode S permet au lecteur de tout vérifier.

Au long d'un quart de siècle, c'est-à-dire pendant l'élaboration des volumes I à IV de l'ALG, nous nous en sommes tenu volontairement et rigoureusement à collectionner des faits, une multitude de faits, sans autre souci que de les représenter dans les cartes avec toute la fidélité dont nous étions capable. Après quoi nous nous sommes mis à la tâche que nous nous étions toujours assignée : trier, rassembler et compter les éléments de cette poussière<sup>1</sup> ; à cette fin, nous avons usé de procédés mathématiques, et il en est résulté une certaine masse de chiffres. « L'ivresse des chiffres » est bien connue, qui procure au chercheur une euphorie rassurante ; du moment que « les chiffres sont là », notamment la formule synthétique apparue en fin de parcours, tout est dit. Trop commode. A notre avis, il convient alors de critiquer ces chiffres, de rechercher ce qu'ils signifient. Autrement dit, après avoir abstrait le concret, on doit, par une troisième opération, retourner aux sources, confronter l'abstrait au concret, « ce concret intégral et intégré sans le respect duquel il n'est que chimères et bavardages »<sup>2</sup>.

En ce qui concerne le premier secteur, la réponse va de soi, puisque les comptages, les statistiques sont donnés tels quels et comme tels, et que l'interprétation en est à la charge du lecteur : les résultats sont justes et significatifs dans la mesure où nous n'avons pas commis d'erreurs matérielles. Comme il s'agit, le plus souvent, de grands nombres, les erreurs de détail ont une incidence négligeable.

Quand on passe au deuxième secteur — les distances linguistiques — il s'en faut que la réponse se trouve aussi aisément. Ce ne sera pas la même suivant l'utilisation qu'on veut faire des données de ce deuxième secteur : jugement d'identité ou jugement de ressemblance.

Identifier, c'est reconnaître (*déterminer*, dans le langage des naturalistes) un être singulier ; ou bien un être appartenant à un ensemble dont tous les éléments sont censés identiques (notion d'espèce), donc le situer dans cet ensemble. Deux modes de reconnaissance : la reconnaissance naïve (la lune, mon ami Pierrot, un œillet, un dé à coudre) et la reconnaissance savante. Mais en réalité, l'une et l'autre suivant la même démarche : l'analyse en traits qualitatifs, le choix de ceux qui sont pertinents et la saisie de leur agencement. La reconnaissance naïve va au plus économique : ou bien elle

1. Vue au microscope, la poussière des maisons est faite de fragments de matières hétéroclites, qu'on peut, dans un échantillon, classer par catégories et par proportions.

2. ALG IV, *Avant-propos*, p. 4.

est globale <sup>1</sup> (l'ami Pierrot) ou bien analytique, en extension ; bien souvent, elle se fie à un trait unique : on distingue un Européen et un Africain d'après la couleur de leur peau, un haricot flageolet d'un pois sec d'après la forme. La reconnaissance savante est toujours analytique : la détermination d'un animal, d'une plante ou d'une roche repose sur la présence et l'absence des *caractères de diagnose*, parfois extrêmement nombreux et difficiles à discerner ; ces caractères ont été retenus parce qu'ils sont isolables et invariants.

Le rendement de l'identification naïve, globale ou analytique, est immense et universel. Nous en usons tous à chaque instant, et même un entomologiste expérimenté reconnaîtra, dans une poignée d'insectes hétéroclites, les diverses espèces « au premier coup d'œil », sans avoir recours aux clés dichotomiques d'une faune. Il n'est pas exceptionnel que le trait unique soit suffisant : en présence d'un animal doté d'un appendice nasal de deux mètres, mobile et préhensile, point n'est besoin de pousser l'analyse plus loin pour savoir qu'on a affaire à un éléphant <sup>2</sup>. Mais il serait bien long de citer des exemples où l'identification naïve est fautive, mauvaise et même nuisible. Les gens de justice ont constamment à gémir des confrontations entre témoins et suspects, alors que l'anthropométrie est infaillible. Dans la faune populaire, l'orvet est un serpent. En dépit de l'école et des campagnes d'information, on n'arrive pas à vulgariser suffisamment la notion, pourtant simple, d'Amanite phalloïde : les conséquences tragiques de cette erreur se reproduisent partout chaque année.

Est-il possible d'identifier un dialecte, un sous-dialecte ? Certainement. Les identifications naïves existent dans le monde entier, et elles sont dans leur majorité fondées sur un trait unique. « Les Béarnais se moquent des Landais qui disent *yurn* « jour », cependant que les Landais font des gorges chaudes de *diæ* — On légitime, après coup, ce cloisonnement, en s'opposant au voisin par trois ou quatre noms d'objets très usuels : le balai, la chaise, le pot à feu et surtout le maïs : *miļas*, *miļok*, *indun*, *turgét*, *blaspaņæ* <sup>3</sup>... ». Un

1. Dans ce cas, *global* est synonyme d'*intuitif*. Le procès de l'intuition apparaît des plus simples, puisque le résultat est instantané, « fulgurant ». En réalité, il doit se passer dans les arcanes du cerveau des choses d'une complexité effrayante, dans un temps infime, tout comme dans les ordinateurs, qui ne sont, après tout, que la réplique rudimentaire, quoique gigantesque, de la machine cérébrale.

2. Encore que les zoologues, parmi les éléphants actuels, distinguent deux genres, ce qui est encore plus tranché que deux espèces. Presque toujours, l'identification naïve et la savante se tiennent dans deux ordres de choses différents.

3. Lalanne, *op. cit.* I, 7 ; II, 95 ; V. également Bloomfield, *Le Langage*, p. 52, 313.

document remarquable et irrécusable, puisqu'il est enraciné dans le folklore, vient d'être publié <sup>1</sup>. C'est un petit conte recueilli en Gascogne pyrénéenne, un conte de fées. Les fées et leurs enfants parlent *étranger*. La mère fée demande à sa fille, qui vient d'être ébouillantée : « *Qui t'ac a fait ?* (qui te l'a fait ?) — *Mi-madeisha*. — *Salva-t'oc* (soigne-le-toi) ». La différenciation dialectale est marquée par trois traits : deux morpho-phonologiques : languedocien [fajl] vs gascon [hèil], lg. [sɔlbo] vs gasc. [sqɔbo] ; et un lexical : à la fillette qui lui demandait son nom, la méchante vieille a répondu *Mi-madeisha* « moi-même » et la fillette a pris cela pour un nom propre, car en languedocien *madeish* a été évincé depuis longtemps par le gallicisme *mèma* <sup>2</sup>.

Ces identifications naïves existent vigoureusement et elles existent parce qu'elles sont fonctionnelles : les communautés s'y réfèrent, entre autres signes, pour se démarquer. Il est donc assez vain de s'interroger sur leur valeur objective, puisqu'elles n'ont pas, qu'on sache, de conséquences néfastes. Elles doivent être l'objet d'étude comme n'importe quel autre fait social.

Dans l'identification savante des dialectes, ou typologie dialectale, la démarche est la même que dans l'identification naïve : mais l'analyse y est plus poussée, de sorte que la caractérisation est définie par un ensemble de traits nombreux. Ces traits ne sont pas forcément endémiques de ce dialecte : c'est la constellation qu'ils forment qui marque le type. De plus, on valorise certains traits. Cette valorisation est parfois arbitraire, mais pas toujours : ce sont les *isoglosses importantes* de Bloomfield <sup>3</sup>, remarquables par la longueur et la continuité de leur parcours.

Jusqu'ici, l'identification des dialectes et celle des êtres de la nature est la même à ses divers niveaux. Au delà de ce point, on se heurte à une différence radicale entre les deux catégories d'objets. L'espace qu'occupe chacun des êtres de la nature est tranché net : il serait ridicule de se demander où commence et où finit un hanneton. Il n'en va pas de même pour les dialectes (ou même pour des langues apparentées et contiguës <sup>4</sup>). Où finit le domaine territorial du béarnais et où commence celui de l'armagnacais ? L'imbrication des critères <sup>5</sup>, le dégradé des occurrences interdisent de tracer

1. J. C. Dinguirard, *Contribution ethno-linguistique à l'enseignement du gascon langue maternelle (Via Domitia VII, p. 70 ; Univ. Toulouse-Le Mirail 1971)*.

2. On reconnaît évidemment l'un des thèmes du conte-type, *Ulysse et le Cyclope*.

3. *Op. cit.*, p. 321.

4. Bloomfield, *op. cit.*, p. 55.

5. Séguy *Actes Strasbourg 1971*. — Ce brouillage est d'ordre spatial. On le ren-

une frontière. Mettre le problème entre parenthèses, en disant par exemple que le béarnais est le gascon parlé à Pau et à l'armagnacais celui qu'on parle à Nogaro n'est qu'une dérobade, sinon une distorsion de la réalité. La démonstration a été donnée cent fois, et nous n'allons pas la recommencer. Tenons-nous à citer quelques lignes saisissantes de Lalanne, qui, lui, exigeait en bonne logique une définition coextensive <sup>1</sup> de chaque dialecte : « A moins qu'on ne se contente du *dialecte au rabais*, où une seule caractéristique suffit. Décrétons par exemple que la voyelle *æ* tonique, opposée à *é*, suffit à définir le dialecte. N'ayant pu faire cadrer des caractéristiques avec un terroir donné, donnons-nous une caractéristique et attribuons-lui 5 000 km<sup>2</sup> de territoire qu'elle recouvre, et nous sommes ainsi assurés d'avoir un dialecte, en l'espèce, le parler noir. A ce prix, on aura autant de dialectes qu'on voudra. Tant pis si, à ce détail près, tout le reste n'est que divergence et confusion à l'intérieur du territoire dialectal ainsi artificiellement défini... C'est pourquoi toutes les définitions traditionnelles de dialectes dans notre secteur nous ont paru aussi fantaisistes que le découpage de la voûte céleste en constellation. La constellation est peut-être poétique. Il se peut qu'elle soit commode, voire nécessaire, pour les besoins du langage et de l'action, mais l'astronome sait fort bien que ce cloisonnement ne correspond à rien de réel <sup>2</sup>. »

La raison d'être d'un atlas, linguistique ou autre, est de délimiter des surfaces, de tracer des frontières. La chose est impossible s'agissant de territoires dialectaux. Plus exactement, le problème n'a pas de sens, puisque le fait visé n'a pas d'existence objective. Aussi n'en est-il jamais question dans l'ALG. La dernière carte, intitulée *Champ gradient de la gasconité*, ne doit pas faire illusion. Justement parce qu'elle est intitulée « champ gradient » : le nombre de traits que le gascon oppose au reste du gallo-roman <sup>3</sup> y est représenté par la densité de l'encre noire. Cette densité est maximale

contre aussi parfois dans la taxonomie des êtres de la nature : on dit alors que le genre est polymorphe. Dans la flore de la France, c'est le cas des ronces et des épervières. Boulay, dans la *Flore* de Costes, décrit quarante-cinq espèces de ronces, ce qui est déjà beaucoup. Mais tel autre spécialiste du genre en nommait des centaines. A la limite, on trouverait un nom spécifique pour chaque pied de ronce. De même Lalanne avançait qu'il existe autant de dialectes que de villages.

1. Une définition coextensive est celle dont tous les traits sont rigoureusement pertinents ; si l'un est facultatif, ou commutable, la définition n'a plus de valeur.

2. Lalanne II, II, IV. — V. en outre Bloomfield, *ib.* 55, 307, 320. Et tant d'autres.

3. Mais la plupart de ces traits sont communs au gascon et à l'ibéro-roman.

à la tête du gradient, et devient de plus en plus légère à mesure qu'on approche de la périphérie, où les vestiges de noir font brusquement place au blanc pur. Mais ce blanc pur n'est que le reflet de notre ignorance touchant les domaines limitrophes : on peut penser que le noir s'étendra à l'est et au nord quand l'*Atlas du Languedoc occidental* aura paru.

Les calculs que nous avons effectués ne peuvent donc servir à déterminer les contours et la superficie de ce qu'on appelle traditionnellement le béarnais, le médocain, le lectourois, le bigourdan, etc. Cela n'a jamais été notre propos.

Pour Lalanne, l'identification naïve du dialecte n'avait aucune valeur. Il tenta donc d'établir une identification scientifique en intégrant dans sa recherche tous les critères objectifs qu'il avait relevés. Au terme de son travail, il conclut par un constat de faillite : « Il semble désastreux d'employer le mot dialecte quand on veut étudier et opposer des parlers limitrophes : ils sont scientifiquement *inopposables* <sup>1</sup> ». De fait, la démonstration qui aboutit à cette sentence est irréfutable.

La géographie des dialectes est-elle donc impossible ? Elle l'est en effet si on se propose de formuler des jugements d'identité : « l'aire 1 parle le dialecte 1, l'aire 2 parle le dialecte 2 ». Mais il reste une issue si on se contente de porter des jugements de ressemblance <sup>2</sup>.

Le jugement intuitif — ou naïf — de ressemblance est l'une des activités permanentes de l'esprit. Mais contrairement au jugement intuitif d'identité, il montre rarement les preuves de son utilité. On connaît bien les controverses qui s'engagent à ce sujet autour du berceau d'un nouveau-né.

Les sciences de la nature ne négligent pas la ressemblance : les systématiciens savent fort bien énumérer en extension les traits communs à deux, trois catégories d'objets. Les succès de cette démarche se révèlent assez souvent positifs : par exemple, dans la classification botanique, les Papavéracées se placent entre les Renonculacées et les Crucifères, parce que, bien qu'elles forment une famille individualisée, elles ressemblent un peu à celles-ci et un peu à celles-là. N'empêche que les appréciations de cet ordre restent grevées de subjectivité.

Il n'en va plus de même si on met en œuvre la méthode de la distance de Hamming qui, en quantifiant la ressemblance, donne aux résultats le label mathématique et élimine le facteur personnel.

1. Lalanne II, VIII.

2. Lalanne est passé près de cette voie (notamment II, 95). Il l'aurait sans doute explorée s'il avait vécu.

On peut appliquer la méthode de Hamming à la dialectologie, puisque les faits linguistiques de tout ordre (phonétique diachronique, phonologie, lexique, morphosyntaxe) se laissent isoler en unités discrètes, donc dénombrables. Envisageons un problème théorique. Dans le domaine gallo-roman (ou ibéro-roman, ou gascon, ou provençal, etc.), je considère quatre localités A, B, C, D situées à des distances quelconques les unes des autres sur un même méridien. Sur la parallèle de A, je prends une localité A<sub>1</sub> située à dix kilomètres à l'ouest de A ; je prends de même B<sub>1</sub>, C<sub>1</sub>, D<sub>1</sub> sur les parallèles de B, C, D (par convention, je pose que l'intervalle de dix kilomètres définit des points *contigus*). Je fais la même chose à l'est de A, et je marque A<sub>2</sub>, B<sub>2</sub>, C<sub>2</sub>, D<sub>2</sub> <sup>1</sup>. Dans les cinq matrices, je compare successivement les lignes de A et celles de A<sub>1</sub> ; je fais la somme, qui est la *distance linguistique* entre A et A<sub>1</sub>. Même opération pour A en face de A<sub>2</sub>, B en face de B<sub>1</sub>, etc. A la fin, j'ai deux files d'interpoints chiffrés. J'étudie ces chiffres. Là, tout peut arriver. La probabilité la plus faible est qu'on trouve la valeur zéro à tous les interpoints, ou la même valeur à tous les interpoints. L'expérience montre que ce sont des valeurs diverses qui apparaissent. Dès lors, pour pouvoir poursuivre, il faut se donner quelques conventions (nous verrons que les nôtres ne sont pas entièrement arbitraires) : par exemple que les valeurs de 0 à 5 appartiennent à un même ensemble (ou fourchette), qui sera le degré zéro (distance linguistique négligeable) ; celle de 5 à 10 à un autre ensemble, celle de 10 à 15 à un troisième. On relie par un trait jaune les interpoints du deuxième ensemble, qui sera le degré 1 ; par un trait orangé ceux du troisième ensemble, ou degré 2 ; par un trait rouge les interpoints d'un quatrième ensemble (15 à 20, degré 3), et on interprète les résultats. Si tout est jaune, j'en déduis que les quatre points du méridien A-D ressemblent beaucoup à ceux du méridien A<sub>1</sub>-D<sub>1</sub>, et tout autant à ceux du méridien A<sub>2</sub>-D<sub>2</sub> (ce qui ne veut pas dire que A<sub>2</sub>-D<sub>2</sub> ressemble dans la même mesure à A<sub>1</sub>-D<sub>1</sub> : on ne perdra pas de vue que la ressemblance n'est pas transitive). Je peux dire aussi que la différence linguistique entre A-B et A<sub>1</sub>-B<sub>1</sub> est faible. Par contre, si j'ai un trait jaune à l'ouest de A-D et un trait orangé à l'est, je dirai que A-D ressemble davantage à A<sub>1</sub>-D<sub>1</sub> qu'il ne ressemble à A<sub>2</sub>-D<sub>2</sub> (à plus forte raison si le trait à l'est est rouge).

Ce problème était destiné à faire saisir le principe de la distance de Hamming appliqué à la dialectométrie. Mais dans la pratique, une telle

1. Comme les arcs de méridiens sont très courts, le non-parallélisme des méridiens ne causera qu'une différence négligeable entre la distance  $A + A_1 + A_2$  et la distance  $D + D_1 + D_2$ .

procédure serait insuffisante. Il convient non seulement de calculer la distance linguistique entre A et B, B et C, C et D ; A<sub>I</sub> et B<sub>I</sub>, B<sub>I</sub> et C<sub>I</sub>, etc., mais encore d'établir toutes les liaisons possibles entre les localités contiguës de l'atlas, de sorte que le fond de carte rappelle une minute de triangulation géodésique<sup>1</sup>. Ce n'est qu'après avoir calculé les distances de tous les interpoints qu'on joint entre eux ceux qui sont du même degré.

En Gascogne, des lignes jaunes se dessinent un peu partout. Mais quelques lignes orangées apparaissent fort nettement, les unes nord-sud, les autres est-ouest ; elles sont plus nombreuses dans la zone pyrénéenne et dans la région sud-est, où elles délimitent parfois des isolats. Par lignes orangées, il faut entendre des lignes qui ont cette couleur sur la plus grande partie de leur parcours, mais qui présentent çà et là des segments jaunes ou au contraire rouges.

Ce système à quatre tranches présente l'inconvénient habituel : il diminue certaines des valeurs réelles alors que d'autres sont majorées. Mais il est possible de donner une représentation tout à fait fidèle de ces « lignes de clivage », en les projetant en valeurs réelles sur un plan vertical. La frontière apparaît alors comme le profil d'une chaîne de montagnes, où les altitudes des pics et des cols traduisent la distance linguistique exacte entre les localités contiguës.

Les quatre degrés que nous utilisons dans l'ALG ne correspondent pas aux tranches de 5 en 5 du problème théorique, qui seraient purement arbitraires. Nous avons tenté de concilier, dans la mesure du possible, les inconciliables : l'identification naïve du dialecte et la mesure mathématique de la ressemblance. Nous avons pris, comme base expérimentale, la ligne qui sépare les localités parlant le célèbre gascon noir de celles qui parlent le gascon « clair ». Dans l'identification naïve, le critère est unique : c'est la correspondance *é* tonique en gascon clair = *æ* tonique en gascon noir<sup>2</sup> (*la pɛlə də la hɛmnə k ɛʒ nɛgæ* vs *lə pɛlə də lə hɛmnə k æʒ nɛgə* « la jupe de la

1. Le procédé a été exposé pour la première fois, et magistralement, par Guiter (Strasbourg 1971). — Les localités réellement contiguës sont celles du seul atlas exhaustif existant en France, celui des Pyrénées orientales, par M. Guiter ; dans les autres atlas, où les enquêtes sont faites par sondages, les localités ne sont contiguës que par convention. M. Guiter, *ib.*, donne les règles mathématiques qu'il a établies pour passer d'une maille de densité à une autre maille. — Faut-il préciser que dans la pratique, la position des localités sur des méridiens et des parallèles est le dernier souci du « dialectomètre » ?

2. « Il est très courant d'entendre dire que les membres d'une tribu parlent « léger », ou bien qu'ils parlent « lourd », selon qu'ils relèvent de l'un ou de l'autre secteur linguistique » (Elkin, *Les aborigènes australiens* ; Paris 1967, p. 91).

femme est noire »). Dans nos calculs, ce trait vaut un seul point de différence (item n° 5 de la matrice phonologique). Or, la ligne qui se superpose naturellement à cette isoglosse est orangée dans la plus grande partie de son parcours, avec quelques segments rouges (et aussi deux brefs segments jaunes). Ce qui prouve que l'unique critère naïf est en réalité accompagné d'un cortège d'autres faits linguistiques différentiels de tout ordre : rien que pour la phonétique diachronique, nous vérifions 31 items différentiels le long de cette ligne de partage, sur les 67 items que compte la matrice de phonétique diachronique (bien entendu, tous les 31 items ne se vérifient pas à chacun des interpoints de la ligne en question). Comme les valeurs différentielles qui jalonnent la « frontière » du gascon noir sont comprises entre 18 et 22 %, nous avons attribué à cette fourchette le degré 2 ; le degré 3 est celui de la limite d'un isolat bien connu, celui de Labastide-Clairence (6910 ; = 23 %). Les degrés 1 (14-17 %) et 0 ( $\leq$  13 %) ont été fixés arbitrairement : mais ils ont bien moins d'importance que les degrés supérieurs. C'est cette graduation de base quelque peu expérimentale qui a été généralisée à tout le domaine de l'ALG.

D'aucuns, en regardant la carte intitulée *Frontières dialectales du gascon* (ALG VI, 2525) penseront que nous nous sommes livré à bien des contorsions et que nous avons pris bien des détours pour finalement rentrer au bercail de la géo-dialectologie traditionnelle. Il n'en est rien : les « aires dialectales » que les frontières de cette carte semblent circonscrire ne sont que de fausses aires. Si, prenant deux points non contigus à l'intérieur de l'une de ces aires fallacieuses, on en calcule la distance linguistique, il résulte que cette distance atteint et le plus souvent dépasse celle de la frontière. A de rares exceptions près, et bien entendu quand une frontière entoure un isolat fait d'une localité unique, cela se vérifie à chaque sondage. Les conclusions de Lalanne quant à l'inexistence des aires dialectales sont indestructibles. Comme il avait fort bien aperçu les bourrelets d'isoglosses que nous formalisons par des lignes en couleur, nous lui écrivions en 1950 : « Il peut se passer n'importe quoi à l'intérieur du polygone (= la fausse aire) : le bourrelet suffit tout de même à délimiter un grand ou un petit quelque chose, qui tombe sous les sens ». A quoi il nous rétorqua : « En fait, il s'y passe tout ce qu'on veut, sauf la continuité d'un parler unique et homogène. Il ne s'y passe pas *une* petite, ni *une* grande chose, mais *beaucoup trop* de choses, pour quelqu'un qui cherche de l'unité et du défini ». Ni l'un ni l'autre n'avions su voir, à cette époque, la vérité toute simple et toute nue : les bourrelets signifient qu'il existe une différence linguistique notable entre

deux séries de localités contiguës, et ne signifient que cela. Quand nous parlons de « frontière dialectale », c'est dans un sens très restrictif. La désignation est trompeuse ; il faudrait trouver autre chose. Mais il nous paraîtrait peu convenable de proposer plus d'un néologisme par tranche de 20 pages.

C'est dans le dessein de formuler ces jugements de ressemblance, et non d'identification, que nous avons effectué les calculs et dressé les matrices du *deuxième secteur*. But modeste, mais tout de même intéressant.

Que valent nos chiffres ? Après en avoir délimité le champ d'application — champ étroit — la réponse est la même que pour ceux du *premier secteur* : la validité ne dépend plus que de la correction des statistiques. Et à ce propos, on fait deux constatations à première vue inquiétantes : 1<sup>o</sup> ces « frontières » mesurent la fonction démarcative du dialecte<sup>34</sup>. Si on fait la ventilation des paramètres donnant la valeur de chaque interpoint, on voit que le paramètre phonologique y est le plus faible, alors que le paramètre lexical représente à lui seul la valeur des quatre autres réunis. Ceci se concilie mal avec les expériences sur l'intercompréhension faites par M. Ravier<sup>1</sup>, d'où il résulte que la fonction de communication est assurée par le lexique, les autres composantes du langage servant à la démarcation ; en d'autres termes, si le taux de différence lexicale s'élève trop, la communication passe peu et mal, situation intolérable entre des localités contiguës. Mais dans la vie courante, les différences lexicales servent bel et bien à la démarcation tout comme les autres caractéristiques ; la communication n'est pas gênée, parce que les locuteurs sont en quelque sorte bilingues : pour un signifié *x*, A emploie un mot que B n'emploie jamais, mais qu'il comprend fort bien, et réciproquement. Il n'en va pas de même lorsque, par artifice, on met en contact des localités distantes de cent kilomètres : dans de telles conditions, les différences lexicales portent préjudice à la communication, et l'intercompréhension n'est plus que globale<sup>2</sup>. — 2<sup>o</sup> (qui est un corollaire de 1<sup>o</sup>) : dans nos comptes, un trait comme « *é* fermé primitif > *æ* » compte pour un point de différence, pas plus. Ce qui heurte à la fois le sens commun et la notion rationnelle d'isoglosse principale définie par Bloomfield. Tous

1. Séguy, *La fonction minimale du dialecte* (Strasbourg 1971).

2. X. Ravier, *L'incidence maximale du fait dialectal* (Strasbourg 1971). Par le truchement du magnétophone, M. Ravier mettait en rapport des locuteurs gascons séparés par une distance moyenne de cent kilomètres, et qui n'avaient donc pratiquement jamais l'occasion de communiquer dans leurs vernaculaires respectifs.

les critères sont ramenés au même niveau, tout est aplati. Mais il n'en est pas du tout ainsi dans la méthode de M. Guiter : grâce à leur répétition, les traits phonologiques et phonétiques de forte fréquence bénéficient d'une pondération naturelle. Or, au prix d'un changement de constante, les courbes de différence G et S, comme nous l'avons vu, sont très étroitement apparentées : une méthode peut sans inconvénient être substituée à l'autre. Resterait à élucider la cause par laquelle ces deux voies, divergentes sauf en ce qui touche au lexique, conduisent à la même destination.

A moins que les fruits de cette longue quête, expérimentale et même empirique, nullement spéculative, ne se révèlent secs et creux sous la dent de la critique, on peut conclure : 1<sup>o</sup> que la dialectométrie bi-dimensionnelle (alias *aréologie dialectale*) vaut pour des faits singuliers, tels ceux dont nous avons donné des exemples (p. 10-11.) : subjonctif présent = futur ; enclise du pronom régime sur l'infinitif ou le gérondif ; mais que la dialectométrie bi-dimensionnelle fait défaut quand on prétend définir des aires dialectales constituées par l'empilage de faits linguistiques. Elle n'existe pas pour la raison que son objet n'existe pas <sup>1</sup> — 2<sup>o</sup> que, par contre, la dialectométrie linéaire rend compte de ressemblances (ou différences) entre localités contiguës (ou même éloignées entre elles, à condition qu'on les prenne deux par deux), la mesure de ces ressemblances étant la somme d'autant de critères linguistiques qu'il s'en présente sur le terrain.

J. SÉGUY.

1. La dialectométrie bi-dimensionnelle a peut-être un avenir dans les *structures floues*.