

**Zeitschrift:** Revue de linguistique romane  
**Herausgeber:** Société de Linguistique Romane  
**Band:** 35 (1971)  
**Heft:** 139-140

**Artikel:** La relation entre la distance spatiale et la distance lexicale  
**Autor:** Séguy, J.  
**DOI:** <https://doi.org/10.5169/seals-399508>

### **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 RELATION ENTRE LA DISTANCE SPATIALE ET LA DISTANCE LEXICALE

## NOTE IMPORTANTE.

Le présent travail n'aurait pu voir le jour sans l'aide de notre collègue et ami M. Henri Guiter, professeur de linguistique romane à l'Université de Montpellier, agrégé des sciences physiques, docteur ès sciences et docteur ès lettres. Non seulement il a établi les équations des courbes, mais il nous a aussi prodigué des remarques et conseils qui ont guidé et redressé notre démarche.

## PRÉAMBULE.

Il s'agit d'un problème de ressemblance. En dialectologie, on a l'habitude de comparer entre elles des aires phonologiques, phonétiques, lexicales, etc. La validité de ce genre de comparaisons est certaine, mais quand elle porte sur un fait unique, telle évolution phonétique par exemple. On dit alors que les deux aires comparées, si elles sont contiguës, sont séparées par une ligne isoglosse. Mais dès qu'on veut superposer plusieurs faits, c'est-à-dire quand on cherche à comparer deux aires en visant la totalité linguistique, les difficultés apparaissent et croissent suivant le nombre de faits rassemblés. Et encore envisageons-nous le cas où le dialectologue considère vraiment une surface : si on se contente de tracer des isoglosses en avançant de point en point, sans se préoccuper de ce qu'on laisse derrière, il peut arriver que l'extension coïncide avec la réalité, mais le contraire peut aussi se produire. Par exemple, si, dans l'*Atlas linguistique de la France*, on suit le traitement évolutif de *-L*, on rencontre *-u* à toutes les localités gasconnes, puis *-l* conservé de Toulouse jusqu'au méridien de Montpellier, où reparaît *-u*. Dès lors, il faut explorer point par point et dans les deux dimensions l'aire de *-l* conservé, pour la définir par opposition à celle (ou à celles) de

-u. Si on essaie de combiner d'autres faits, alors les ressemblances des aires ne sont plus que partielles, et on en arrive vite à l'imbroglia.

Ces difficultés découlent d'une loi logique : la ressemblance, au contraire de l'identité, n'est pas transitive. Elle ne se vérifie qu'entre deux êtres, constitués en dialectologie par deux points d'enquête. La comparaison irréfutable ne peut être que linéaire et de deux en deux ; dès qu'on infère de l'unidimensionnel au bidimensionnel, sans qu'il soit exclu de tomber juste, on s'expose à des erreurs. Cherche-t-on plus de rigueur en quantifiant les données, par exemple en comptant le nombre de traits différents/semblables qui existent entre deux points ? L'entreprise est sûre en appliquant la méthode des distances de Hamming. Alors, les chiffres sont justes en ce qui concerne les deux points comparés, mais n'ont aucune valeur par rapport à d'autres. Constater que tant d'isoglosses passent entre le point A et le point B est une certitude mathématique ; mais rien ne prouve que les points en deçà de A et ceux au delà de B soient séparés par les mêmes isoglosses. Cela arrive d'ailleurs, et même souvent, mais la démonstration n'est pas faite tant que la comparaison n'a pas été pratiquée entre tous les points du plan pris deux à deux. Comparaison qui pourra bien faire apparaître une aire aux contours nets ; dans le cas contraire, une discontinuité se manifesterait entre les points extrêmes d'un axe ou entre certains points périphériques de la surface. Que de fois, dans les matrices que nous avons établies, tel signifiant codé 1 à l'origine d'un axe reparaît au terminus avec le même chiffre 1, après être passé, chemin faisant, par 2, 3, 4, 5, 6...

Tout ceci est fâcheux pour des gens qui, tels les auteurs d'atlas linguistiques, consacrent leurs facultés à décrire le langage sur deux dimensions, ce qui est la définition même d'une carte. Mais les faits sont têtus, et personne ne peut rien contre la non-transitivité de la ressemblance. Dès qu'ils empilent les faits, les cartographes doivent se résigner à des figurations aux contours plus souvent flous que nets. Ce qui cesse d'être irritant ou mystérieux si on veut bien admettre, comme nous l'avons rappelé et précisé au colloque de Strasbourg (mai 1971), que la fonction du dialecte est double et contradictoire : donner aux communautés voisines la faculté de communiquer tout en se démarquant, ceci par le moyen de l'imbrication des différences.

Il reste deux consolations : accepter que l'aréologie ne soit praticable que pour des faits isolés, ou tout au plus réunis en amalgames légers (alors tout va bien, si on admet que le polymorphisme et les faits d'intercourse déterminent souvent, même dans ces cas, des sutures plus ou moins baveuses) ;

penser que lorsqu'il s'agit non plus de dialectes, mais de langues, la différence bidimensionnelle triomphe : on peut sans remords colorier en vert le domaine basque et en rouge ce qui ne l'est pas ; nous verrons (p. 356) que cette formalisation serait légitime, à condition de ne pas se montrer trop regardant sur le menu détail.

Ceci posé, il convient de retenir deux points : 1<sup>o</sup> la recherche dont nous ferons ici état est d'ordre linéaire, et, dans la mesure du possible, rectiligne ; 2<sup>o</sup> les ressemblances mesurées ne concernent jamais que deux points, qui peuvent d'ailleurs se trouver proches dans l'espace ou bien extrêmement éloignés l'un de l'autre.

#### LE PROBLÈME.

Dans un territoire où sont parlés des dialectes de la même famille — en l'espèce les langues néo-latines — prenons un point d'origine *A* et son point distal *B*. En suivant sur la droite qui les relie les points intermédiaires *a*, *b*, *c*, ...*n*, nous allons mesurer la distance linguistique existant entre *A* et *a*, *A* et *b*, *A* et *c*, ...jusqu'à *A* et *B*. En termes concrets, supposons qu'un habitant d'Arles-sur-Tec (Pyrénées-Orientales), faisant route vers Liège, s'arrête à tous les points de l'*ALF* qui jalonnent ce parcours et essaie de converser dans son vernaculaire avec un habitant de chacune des stations s'exprimant lui-même dans son propre vernaculaire. Qu'en résultera-t-il ?

On ne perdra jamais de vue que cette formulation du problème n'est qu'une fable (v. p. 352) destinée à concrétiser, mais qu'il serait vain et sans doute décevant de vouloir transformer en expérience vécue. Dans notre exposé, ce qui compte, ce sont les chiffres.

#### LA MÉTHODE.

##### *Les documents.*

Ce sont les atlas linguistiques suivants :

- AIS* : Sprach- und Sachatlas Italiens und der Südschweiz, par K. Jaberg et J. Jud ;
- ALAn* : Atlas lingüístico y etnográfico de Andalucía, par M. Alvar ;
- ALC* : Atlas lingüístic de Catalunya, par A. Griera ;
- ALF* : Atlas linguistique de la France, par Gilliéron et Edmont ;
- ALG* : Atlas linguistique et ethnographique de la Gascogne, par J. Séguy ;

- ALLY* : Atlas linguistique et ethnographique du Lyonnais, par P. Gardette ;  
*ALMC* : Atlas linguistique et ethnographique du Massif central, par P. Nauton ;  
*ALPO* : Atlas linguistique des Pyrénées orientales, par H. Guiter ;  
*ALR* : Atlasul lingvistic român (nouă serie).

*L'échantillonnage.*

La totalité linguistique, intégrant tous les paramètres (lexique, phonologie, phonétique historique, morphologie, syntaxe) est pratiquement inaccessible, même puisée dans des atlas, à un chercheur qui ne dispose pas de moyens informatiques et qui, par surcroît, n'a des domaines étrangers au sien qu'une connaissance superficielle. Certes, nous avons pu établir les distances linguistiques intégrales pour tous les points de l'*ALG* — l'exposé occupera une partie du volume VI —, mais au prix d'un travail de longue haleine, qu'il ne pouvait être question de refaire pour tous les atlas ici mis en œuvre ; d'autant, répétons-le, que cette appréhension complète suppose une grande familiarité avec les faits du ou des domaines exploités. Finalement, nous avons retenu un seul critère : le critère lexical, le plus facile à isoler. Et ce mobile de facilité coïncide heureusement avec un motif scientifique. En effet, le problème que nous nous sommes posé n'est pas d'abstraction pure : il ne perd pas de vue la fonction du langage, qui est de communiquer. Or, d'une expérience cruciale faite par M. Ravier en domaine gascon et exposée au colloque de Strasbourg, il ressort que, dans l'intercompréhension, le primat du lexique est écrasant.

Les témoins que M. Ravier a soumis à ses tests n'étaient arrêtés que dans le cas où le phonogramme enregistré à une localité très éloignée de la leur présentait çà et là un lexème inconnu, autre que le leur. Les différences morpho-syntaxiques ne comptaient que comme bruits négligeables, et les récepteurs arrivaient le plus souvent à percevoir à jour les distorsions phonétiques et phonologiques, peut-être grâce à ce que d'aucuns appellent le diastème.

Nous venons de qualifier de cruciales les expériences de M. Ravier. Mais « expérience cruciale » n'est pas synonyme de « loi universelle ». Il se trouve qu'en Gascogne le « bruit » phonologique, pourtant considérable, laisse reconnaître le lexème. Il peut en être différemment dans d'autres domaines : par exemple, un Français ignorant l'anglais ne comprendra absolument rien à un message parlé, en dépit de l'abondance des termes cosmopolites ou

empruntés au français : le bruit phonologique masque le signal lexical. Incontestablement, notre étude serait plus convaincante si nous avions pu appréhender la différence linguistique totale. Il n'en reste pas moins que le lexique est la clé de la compréhension : pour peu qu'on reconnaisse un certain nombre de mots dans une langue qu'on ne sait pas, on comprendra (relativement) l'énoncé ; inversement, si, dans une langue qu'on manie, on rencontre un mot énigmatique, le message présente un trou que le contexte ne permet pas toujours de boucher. L'élève « paresseux » qui, mis en présence d'une version latine, commence par feuilleter le dictionnaire à tour de bras, cherchant à « deviner » au lieu de « construire », suit une démarche réaliste (mais non infaillible, à cause de la polysémie).

Mais pour ce qui touche la typologie d'un ensemble linguistique (disons tout simplement d'une langue), ce sont les traits morpho-syntaxiques qui viennent au premier rang, puis les traits phonologiques et phonétiques ; inversement, le lexique n'y joue qu'un moindre rôle. En conséquence, dans notre travail, il ne sera plus question de « distances linguistiques », mais de « distances lexicales ».

Dans chaque atlas, nous avons suivi deux parcours, les plus longs possibles et se coupant à angle droit ; mais dans les atlas de petite surface, ou représentant un domaine dont l'une des dimensions est bien trop courte par rapport à l'autre, nous n'avons étudié qu'un seul parcours : *ALC*, *ALLy*, *ALMC*, *ALPO*. Il va sans dire que les localités intermédiaires qui jalonnent le parcours *A-B* ne s'alignent pas exactement sur une droite. Mais les distances spatiales ont été mesurées *recta linea* en repartant toujours du point d'origine ; *A-b*, *A-c*, *A-d...A-B* ; *A'-b'*, *A'-c'*, etc. pour le deuxième parcours. Quant aux distances spatiales des segments intermédiaires *A-b*, *b-c*, *c-d...n-B*, elles ont été également mesurées *recta linea*.

Nous avons ensuite établi la matrice lexicale de chacun des parcours, en codant cent cartes de chaque atlas. C'est un code chiffré : dans chaque carte (traitant d'un même signifié), chaque signifiant d'étymon différent a son chiffre propre. Donc, en principe, la ligne affectée au point de départ *A* porte le chiffre 1 pour toutes les cartes. On fait une copie de la ligne *A* et on la confronte successivement à la ligne *a*, puis à la ligne *b*, etc., jusqu'à *B*. Chaque fois que, dans les deux cases correspondant à la même carte, on trouve un chiffre différent, on marque une unité. Puis on fait le total. Ce total donne la distance lexicale entre les deux points considérés (quantification de la ressemblance par la méthode des « distances » de Hamming).

*Cas particuliers.* — Si, à la localité considérée, l'article du questionnaire a été omis par l'enquêteur, on compte comme une identité. Par contre, si l'informateur n'a pu fournir de réponse, on compte comme différence (dénomination zéro, souvent significative de divers points de vue, v. Séguy *Les noms populaires des plantes dans les Pyrénées centrales*, p. 182). S'il y a synonymie, par exemple *A* chiffré 1 et *d* chiffré 1, 5, 7, la présence de 1 aux deux localités annule la distance lexicale (c'est à cause de la synonymie que la ligne de *A* présente parfois des chiffraisons 1, 2 ; 1, 3, 2, etc. à la même case). Dans les cas où l'évolution phonétique rend vraiment méconnaissables deux mots de même étymon, on compte un point de différence. Ces cas sont rares parce que l'expérience de M. Ravier a montré qu'il suffit de deux ou trois phonèmes communs pour que, le contexte et la situation aidant, les témoins identifient le mot sous sa forme « estropiée ». C'est le système binaire, par tout ou rien, de la méthode de Hamming qui nous a contraint d'opter pour ces décisions. Le plus discutable de ces cas est évidemment celui où nous mettons sous plusieurs chiffres les continuateurs d'un même étymon : le procédé est nécessairement entaché de subjectivité. Par bonheur, le fait se rencontre rarement. En ce qui touche l'échantillon de l'ALG, 89 % des chiffres codent des mots d'étymologie différente ; les séparations subjectives ne représentent que 11 %, même en y incluant sans ménagement des cas comme gascon pyrénéen *ézlayète* et bas-gascon *flèu* (emprunté au français) < FLAGELLU. On peut estimer que si la moitié de ces séparations subjectives est mauvaise, l'autre moitié est judicieuse, c'est-à-dire correspond au sentiment des usagers. Les séparations subjectives ne peuvent donc avoir qu'une très faible incidence sur nos calculs.

Dans chacun des atlas, nous avons examiné cent cartes. Au colloque de Strasbourg, M. Guiter a démontré qu'un échantillon de cette taille est suffisant pour mesurer la distance linguistique ; il ne vaut jamais la peine de pousser au delà, parce que les écarts en pourcentage avec la première tranche de cent se montrent d'abord faibles, puis insignifiants à mesure qu'on élargit l'échantillon.

Les cartes de la plupart des atlas sont classées par ordre logique : le ciel et le temps ; les plantes sauvages ; le corps humain, etc. Autrement dit, chaque chapitre correspond à un champ sémantique. Or, il est certain que la variété lexicale est différente suivant les champs : il ne pouvait donc être question d'étudier les cartes n° 1 à 100. Nous avons, dans les 1 000 premières cartes, opéré un tri à 10 % par la méthode des nombres au hasard, utilisant pour cela une table toute faite. Par contre, pour les atlas classés par ordre

alphabétique, suivant les instructions de M. Guiter, nous avons simplement exploré les cent premières cartes, puisque ce classement brouille parfaitement l'ordre logique.

En vertu de ces règles statistiques, nous n'avons pu prendre en compte les atlas qui offrent, dans l'état actuel de leur publication, un échantillon inférieur à cent cartes (ordre alphabétique ou ordre logique).

Toutes les cartes ainsi sélectionnées ont été chiffrées et incorporées aux calculs, même les cartes mononymes (présentant un étymon unique dans tout le domaine ou seulement pour le ou les parcours suivis). Le nombre de ces cartes mononymes varie suivant les atlas. Le maximum de différences, qui est a priori de 100, est ainsi rabaissé, dans tous les atlas, à un maximum théorique inférieur que nous indiquerons ci-dessous. N'ont été exclues que les cartes vraiment incompatibles avec nos fins : cartes présentant de grandes lacunes, cartes morphologiques (par exemple les quelques tiroirs du verbe *avoir* cartographiés dans l'*ALF*, l'*ALPO*, etc.). Dans ces cas, d'ailleurs rares, on passait outre pour les atlas alphabétiques (quitte à dépasser la carte n° 100), et dans les autres, on cherchait le numéro le plus voisin de celui de la table des nombres au hasard. Mais toujours nous avons codé cent cartes.

Les atlas que nous allons comparer entre eux sont-ils comparables ? Il est inutile de se dissimuler qu'ils présentent certaines divergences d'ensemble. Dans les dimensions : le parcours le plus long, *AIS 1*, c'est-à-dire du sud des Pouilles — le fond du « talon » — jusqu'aux Alpes valdôtaines, fait 1 100 kilomètres, alors que celui de l'*ALPO* n'en atteint que 127, de sorte que seul le début du parcours *AIS 1* est comparable au parcours total de l'*ALPO*. Dans l'équidistance des points d'enquête, dont nous donnons les valeurs classées : *ALLy* 15,5 km ; *ALPO* 15,9 ; *ALMC* 22,2 ; *ALAn* 23 ; *ALG* 24,6 ; *ALC* 29 ; *ALF* 34,6 ; *AIS* 47,5 ; *ALR* 52. Ces chiffres ne mesurent ni la densité réelle des atlas, ni les distances réelles des interpoints des parcours : pour chacun des atlas, c'est le quotient de la distance totale du ou des parcours divisée par le nombre d'interpoints (pour l'*ALPO*, atlas, rappelons-le, exhaustif, nous n'avons retenu que dix points sur un trajet W-E). Dans la variété lexicale, en gros inversement proportionnelle au nombre des cartes mononymes, dont voici les valeurs par pourcentages décroissants : *ALPO* 62 ; *AIS 2*, *ALF 2* 31 ; *ALC* 30 ; *ALF 1* 27 ; *ALLy* 24 ; *AIS 1* 23 ; *ALMC* 20 ; *ALR 1* 13 ; *ALR 2* 8 ; *ALG* 8 (pour l'ensemble de l'atlas : compte portant sur les 154 localités) ; *ALAn 2* 6 ; *ALAn 1* 5. Ces pourcentages ne sont d'ailleurs pas forcément en corrélation avec la différence lexi-

cale  $A-B$  (nous verrons de cela un exemple remarquable avec l'*ALPO*). Dans le tableau des mesures, on peut lire la distance réelle entre chaque point.

Les cartes de petit format de l'*ALG*, au lieu d'être classées à leur ordre logique, sont toujours placées en fin des volumes. Or, comme elles sont en petit format justement à cause de leur variété faible, ceci introduit un léger biais dans la statistique, au détriment de l'indice de variété.

Néanmoins, ces divergences ne paraissent pas de nature, dans leur ensemble, à frapper notre entreprise de nullité.

#### *Les éventualités.*

Avant de nous mettre au travail, il était loisible, et même bon, d'envisager les divers résultats qui pouvaient se présenter.

Nous désignerons par  $x$  les distances spatiales et par  $y$  les distances lexicales, celles-ci exprimées en distances de Hamming.

1. Le voyageur qui part d'Arles-sur-Tec pour Liège rencontre un nombre  $n$  de différences à la première station, située à 24 km. Ce nombre restera invariable jusqu'à Liège. La courbe de  $y$  sera une ligne horizontale parallèle à l'axe des abscisses.

2. Le voyageur rencontre, chemin faisant, des difficultés toujours croissantes. De quel ordre ? On peut se donner celui-ci : une unité de différence tous les myriamètres. Ce qui paraît raisonnable : un mot différent tous les 10 kilomètres pour un échantillon de 100 signifiés, cela représente 20 mots pour un vocabulaire courant de 2 000. Cette évaluation n'en reste pas moins arbitraire, et on peut tout aussi bien envisager que la constante multipliant  $x$  soit différente de 1. Mais dans tous les cas, on aurait une fonction linéaire, autrement dit une droite dont la pente changerait selon que la constante  $K$  serait 1, 2...10, 11... ou 0,50, etc.

3 a. Mais l'accroissement de  $y$  peut aussi bien se faire selon une autre loi que  $y = Kx$ . Si  $y$  croît comme le logarithme de  $x$ , on aura la courbe caractéristique. De même si  $y = \sqrt{x}$ .

3 b. On pourrait aussi prévoir que  $y$  est fonction du second degré de  $x$  (la courbe représentative serait une branche de parabole) ou bien que  $y$  serait une fonction exponentielle de  $x$  (la courbe représentative croîtrait plus vite qu'une parabole).

4. Cette éventualité est analogue à la première. Le voyageur rencontre des difficultés qui croissent jusqu'à un point P du parcours, à partir duquel le nombre des différences reste stationnaire jusqu'au terminus. Nous aurons une courbe de même allure qu'en 1.

5. Le voyageur, à la distance de 24 km, rencontre des difficultés qui ne cesseront de s'amenuiser jusqu'à Liège :  $y$  serait alors une fonction décroissante de  $x$ .

6. Le voyageur rencontre des difficultés dont le nombre varie fortement et dans les deux sens quand il passe d'une station à l'autre. La courbe de  $y$  est atypique et ne se prête pas à l'analyse.

7. Pour le parcours Arles-sur-Tec - Liège, nous avons obtenu telle courbe. Passant à l'ALR, nous obtenons des courbes tout à fait différentes, et ainsi de suite, de sorte qu'à la fin de notre travail, nous aurons un ramas de courbes hétéroclites, d'où nous ne pourrions tirer qu'une conclusion : tout va à vau-l'eau, à chaque pays, et même à chaque parcours sa solution.

8. Comparant les courbes, nous voyons que certaines sont apparentées ; nous pouvons ainsi constituer des groupes I, II, III... et tenter de raisonner. Cet état de fait pose à nouveau, pour chacun des groupes, les éventualités 1 à 6.

9. Toutes les courbes sont identiques (formes 1, ou 2, ou 3, etc.), elles se superposent exactement. Une loi stricte gouverne la relation  $y = (f) x$ .

10. Les courbes, sans être exactement superposables, présentent toutes une allure semblable. Dès lors, en éliminant la fluctuation aléatoire, nous devrions parvenir à trouver quelque chose de commun (forme 1, ou 2...).

11. Les courbes présentent des variations brusques d'une localité à l'autre ; mais tous les pics et tous les affaissements, situés aux mêmes repères spatiaux, se correspondent, si bien que la courbe moyenne résultante aura le même aspect tourmenté. Il existe bien une loi ; mais l'analyse mathématique des courbes pose des problèmes ardu, sinon insolubles.

Inutile de marquer d'un astérisque les prévisions hautement improbables, sinon absurdes. Mais l'hypothèse du sens commun, de l'intuition naïve, paraît être la deuxième. C'est ce que nous allons voir.

#### TABLEAU DES MESURES.

Disons tout de suite que pour les seize parcours que nous avons étudiés, ce sont les éventualités 10 et 3 qui se sont réalisées : toutes les courbes croissent selon l'allure caractéristique des fonctions logarithmiques.

Les mesures sont présentées sur cinq colonnes :

1. Numéro de la localité dans l'atlas.
2.  $x$  du point d'origine au point considéré (en kilomètres).
3.  $y$  du point d'origine au point considéré. Du point de vue fonctionnel,

ce paramètre est capital : il constitue l'indice de l'intercompréhension. L'expérience de M. Ravier a montré qu'aux alentours de 50, la compréhension fonctionne encore, mais d'une manière globale, c'est-à-dire plutôt mal que bien. Dans les sens  $W \longleftrightarrow E$ , l'intercompréhension n'était pas exactement bi-univoque : le taux de compréhension était nettement plus faible dans  $W$  écoutant  $E$  que dans  $E$  écoutant  $W$ .

4.  $x$  de l'interpoint :  $A-b$  ;  $b-c$  ;  $c-d$  ; ... $n-B$ .

5.  $y$  de l'interpoint.

Les paramètres 4 et 5 seront utilisés dans la discussion.

*AIS 1* (Pouilles-Alpes)

Le point 523 (Florence) qui se trouvait sur le trajet, a dû être abandonné : enquête trop lacunaire. De plus, la distance lexicale entre 543 et 523 est extrêmement faible.

I	2	3	4	5
749				
738	75	23	75	23
737	135	25	60	16
736	173	28	49	19
727	222	32	55	20
716	272	33	50	20
714	330	35	60	24
668	364	31	38	20
656	435	31	75	15
654	500?	43	65?	29
643	535	44	35	26
633	558	43	26	27
583	623	41	64	17
553	685	45	64	20
543	722	49	51	16
464	824	45	65	19
444	862	50	42	24
413	905	46	44	17
274	972	46	72	22
261	1 000	46	30	17
250	1 035	53	32	22
126	1 075	50	43	29
114	1 100	47	30	20

*AIS 2* (Ligurie-Frioul)

I	2	3	4	5
190				
193	30	24	30	24
185	75	28	45	23
179	155	26	78	23
432	188	36	36	20
413	238	34	47	23
286	264	39	26	17
372	314	44	44	32
354	384	38	71	31
345	407	38	23	18
336	442	36	36	17
318	500	44	52	29

$x$ , dont le maximum devrait être de 100 Mm, a été dépassé de 10 Mm. Mais il n'y a pas d'inconvénient, puisque à 100 Mm,  $y$  est loin d'être saturé.

<i>ALAn 1</i> (W-E)					<i>ALAn 2</i> (S-N)						
	1	2	3	4	5		1	2	3	4	5
H	302					Ca	602				
H	301	29	36	29	36	Ca	600	22	32	22	32
Se	301	62	38	32	32	Ca	400	33	29	12	27
Se	302	90	40	28	32	Ma	303	58	29	23	29
Se	304	109	38	19	32	Ca	202	83	38	22	34
Se	303	133	52	22	32	Se	603	95	36	14	28
Se	400	154	47	21	37	Se	600	115	50	21	35
Se	401	188	45	29	28	Se	404	131	46	14	25
Co	603	205	47	28	30	Se	401	158	45	20	30
Co	604	234	49	24	31	Co	301	182	47	26	31
J	501	262	52	30	44	Co	300	218	48	26	33
J	502	388	56	25	35	Co	102	240	50	24	33
J	503	300	53	14	32	Co	100	266	46	22	33
Gr	300	314	47	15	31						
Gr	401	348	55	36	39						
Gr	402	364	55	17	28						
Gr	404	380	58	18	22						
Al	100	402	55	30	30						
Al	204	440	58	37	25						
Al	205	460	54	21	32						

*ALAn 3* (E-W, *ALAn 1* suivi à rebours)

	1	2	3
Al	205		
Al	204	21	32
Al	100	59	36
Gr	404	84	46
Gr	402	100	42
Gr	401	118	37
Gr	300	147	49
J	503	168	55
J	502	180	61
J	501	204	58
Co	604	234	62
Co	603	254	60
Se	401	278	58
Se	400	316	60
Se	303	339	62
Se	304	360	48
Se	302	370	60
Se	301	410	53
H	301	440	66
H	302	460	54

*ALC*

	1	2	3	4	5
	78				
	79	42	17	42	17
	80	93	17	45	9
	76	124	14	38	13
	77	142	15	24	16
	74	185	16	39	20
	75	221	19	42	11
	70	260	22	38	21
	71	272	23	22	18
	60	328	30	45	21
	61	340	31	13	10
	62	360	35	25	15
	44	372	34	24	16
	45	400	37	24	11
	26	460	40	50	16
	28	480	38	21	14
	10	490	36	18	14
100	500	41	14	10	10
101	520	43	18	14	14

Pour les colonnes 4 et 5, voir *ALAn 1*.

*ALF 1* (Roussillon-Belgique)

	2	3	4	5
796				
795	24	9	24	9
786	51	37	28	32
776	86	40	35	11
757	135	40	51	16
748	182	47	45	14
729	246	45	63	12
813	293	52	49	20
812	318	52	28	11
809	356	56	38	19
808	384	62	30	18
905	415	57	39	17
906	450	55	32	18
907	485	54	26	13
6	505	56	25	16
7	522	56	21	11
16	565	57	44	14
19	588	50	28	12
110	615	54	30	12
122	660	55	40	13
133	690	55	30	13
144	723	51	35	13
156	777	55	51	11
165	805	54	26	10
175	821	51	21	4
176	846	52	19	13
183	871	49	30	15
184	880	56	16	28
190	910	53	28	16
193	950	48	40	22

*ALF 2* (Biarritz-Menton)

	1	2	3	4	5
690					
691		40	22	40	22
685		75	21	35	11
686		110	17	35	12
678		150	27	42	24
679		193	25	42	12
760		221	25	33	8
762		249	23	28	10
763		282	35	35	12
764		314	31	31	14
766		344	31	31	13
757		370	27	28	11
758		390	29	24	9
759		405	25	40	17
862		490	25	60	13
873		530	23	38	11
874		565	31	33	11
885		600	32	31	14
887		650	30	52	12
898		705	38	56	21
899		730	38	24	23

*ALG 1* (SE-NW)

	1	2	3	4	5
790 SE					
790 S		13	30	13	30
790		20	30	6	19
790 NO		30	27	13	32
781 NO		59	39	23	33
688 N		85	44	32	30
668 SO		115	48	30	29
657 S		145	46	30	22
656 SO		175	56	31	30
656 SO		175	56	31	30
645 S		200	56	28	33
664 N		225	56	23	30
653 N		250	53	26	32
641 O		270	54	22	17
650 E		290	55	23	15
548		340	54	48	20

*ALG 3* (N-S : 548 à 6910 gascon  
et à 6910 basque)

	1	2	3	4	5
548					
650 N		27	14	27	14
650		50	25	23	10
662		83	33	32	30
672		117	44	37	39
680		146	44	29	16
680 S		167	41	21	13
681 SE		200	50	33	31
690 E		209	53	9	27
6910 g		217	53	8	22
6910 b		217	90	0	84

*ALG 2* (W-E)

	1	2	3	4	5
681 S					
681 SE		15	15	15	15
683		35	28	20	27
683 E		50	32	21	13
684		70	32	19	16
676 NO		95	42	25	28
667 SE		130	57	35	35
668		155	50	27	23
659		200	58	41	31

<i>ALLy</i> (S-N)					<i>ARL 1</i> (SW-NE)				
I	2	3	4	5	I	2	3	4	5
74					987				
71	13	20	13	20	723	90	29	90	29
66	33	30	21	30	728	126	36	44	33
62	42	37	9	20	182	193	44	67	48
53	55	46	12	22	141	290	55	111	49
50	66	50	12	16	250	320	58	44	45
41	77	50	12	16	284	358	62	40	34
29	98	49	21	23	279	390	54	27	32
10	127	54	29	20	325	420	59	46	29
7	139	55	12	19					
2	152	53	12	14					
1	170	56	18	28					

<i>ALMC</i>					<i>ALR 2</i> (SE-NW)				
I	2	3	4	5	2				
55					836	34	51	34	51
52	30	15	30	15	833	75	57	40	41
50	53	27	21	15	130	110	58	29	42
38	73	41	25	25	228	230	56	122	52
29	110	44	40	22	551	270	60	42	45
26	130	45	16	18	414	310	55	44	46
22	141	51	15	29					
21	158	54	16	22					
3	187	62	27	31					
1	202	54	15	22					

<i>ALPO</i>				
I	2	3	4	5
407				
405	7	14	7	14
378	26	25	18	27
386	39	24	14	11
356	54	30	15	16
327	76	33	18	17
251	89	34	18	21
79	102	33	10	14
64	113	33	12	7
56	127	30	14	9

*Équations des courbes.*

$$\left\{ \begin{array}{l} AIS 1 \\ AIS 2 \end{array} \right\} y = 25 \log (x + 1)$$

$$\left\{ \begin{array}{l} ALLy \end{array} \right\} y = 46 \log (x + 1)$$

$$\left\{ \begin{array}{l} ALC \\ ALR 1 \end{array} \right\} y = 14 [\log (x + 1)]^2$$

$$\left\{ \begin{array}{l} ALR 1 \end{array} \right\} y = 26 [\log (x + 1)]^2$$

$$\left\{ \begin{array}{l} ALG 3^1 \\ ALMC \end{array} \right\} y = 33 \sqrt{[\log (x + 1)]^3}$$

$$\left\{ \begin{array}{l} ALMC \end{array} \right\} y = 40 \sqrt{[\log (x + 1)]^3}$$

$$\left\{ \begin{array}{l} ALAn 1 \\ ALAn 2 \\ ALAn 3 \\ ALF 1 \end{array} \right\} y = 42 \sqrt{\log (x + 1)}$$

$$\left\{ \begin{array}{l} ALF 2 \end{array} \right\} y = 23 \sqrt{\log (x + 1)}$$

$$\left\{ \begin{array}{l} ALG 1 \\ ALG 2 \end{array} \right\} y = 45 \sqrt{\log (x + 1)}$$

$$\left\{ \begin{array}{l} ALPO \end{array} \right\} y = 32 \sqrt{\log (x + 1)}$$

$$\left\{ \begin{array}{l} ALR 2 \end{array} \right\} y = 50 \sqrt{\log (x + 1)}$$

$$\left\{ \begin{array}{l} \text{courbe moyenne} \end{array} \right\} y = 36 \sqrt{\log (x + 1)}$$

## EXPLOITATION DES MESURES.

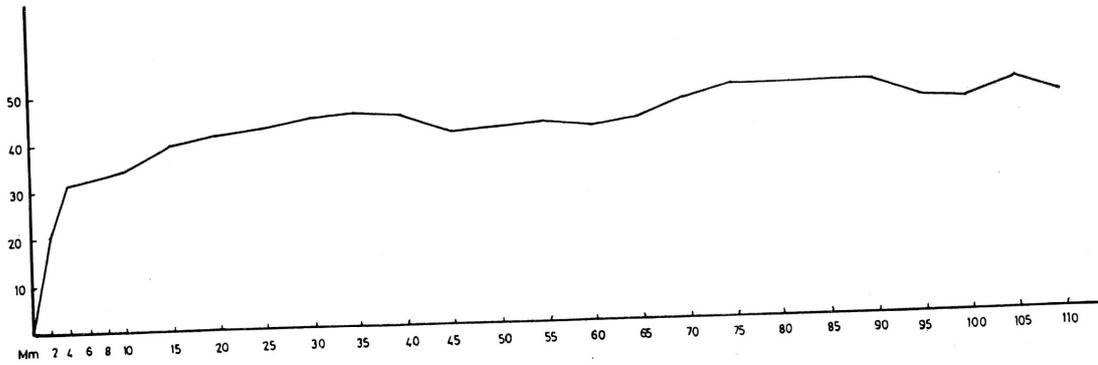
Les courbes sont semblables, mais non identiques (l'éventualité 9 est exclue). Il convient donc de dégager ce qu'elles ont de commun. Le moyen le plus simple — et qui a en outre l'avantage d'éliminer la fluctuation aléatoire — est de réduire les seize courbes à une courbe moyenne. Voici comment nous y avons procédé. Sur chacune des courbes nous avons mesuré directement l'ordonnée à intervalles réguliers : à 2, 4, 6, 8 myriamètres de 0 à 10 Mm ; ensuite de 5 en 5 myriamètres. En effet, il nous a paru que l'analyse devait être plus fine en début de courbe, puisque c'est dans ces débuts que nous observons les allures les plus accidentées.

Ces repères réguliers ne coïncident que rarement avec les stations réelles des parcours : le plus souvent, la valeur de l'ordonnée est extrapolée de la

1. 6910 basque exclu.

COURBE MOYENNE

$$y = 36\sqrt{\log(x+1)}$$



courbe particulière entre deux stations réelles. Bien entendu, cette extrapolation est interdite entre l'origine *A* (dont la mesure est 0) et la première station réelle de chaque courbe. C'est ainsi que *ALR 1*, dont le premier interpoint est de 9 Mm, ne fournit aucune mesure en deçà de 10 Mm. Autre conséquence : il n'y a que 5 parcours sur 16 qui donnent une mesure de *y* à 2 Mm.

Pour chaque repère fixe, les valeurs de l'ordonnée sont additionnées ; la somme est divisée par le nombre de parcours qui entrent en ligne de compte au repère considéré. Ce nombre est de 16 au départ (de moins pour *y* < 10, voir ci-dessus), mais il va s'amenuisant selon la longueur des divers parcours : de 95 à 110 Mm, *AIS 1* reste seul en ligne. A partir de ces valeurs moyennes de l'ordonnée, on obtient la courbe de  $y = (f) x$ .

#### DISCUSSION.

A. L'éventualité 2, celle du bon sens et de l'intuition naïve, était, rappelons-le, que *y* croîtrait proportionnellement à *x* de sorte que les difficultés du voyageur s'accumuleraient selon une proportion arithmétique ; par exemple que *y* connaîtrait partout une augmentation proportionnelle à celle de *x* qui serait la même que celle du premier segment (ce qui donnerait, pour l'*ALC*, 210 % de différences en fin de parcours). L'éventualité 3 *b* — *y* croissant selon d'autres valeurs de *x* :  $y = ax^2$ ,  $y = 2^x$  — n'était elle-même pas absurde a priori, bien que fort improbable. Or, nous constatons que dans tous les cas, c'est l'éventualité 3 *a* qui s'avère : *y* est une fonction logarithmique de *x*.

Dans quatre parcours *AIS 1* et 2, *ALC*, *ALR 1*, le début de la courbe présente une pente assez douce ; M. Guiter observé en outre que dans *AIS* et *ALC* les croissances se font par paliers, chaque saut correspondant à la frontière d'un domaine dialectal italien ou bien à une frontière dialectale du catalan. Dans les douze autres parcours, le début de la courbe est en pente raide. Ensuite, *y* continue de croître, mais par petites valeurs.

B. Les valeurs de *y* interpoints (colonne 5 des tableaux) n'influent que rarement sur les valeurs de *y* *A-B* ; mais il arrive que la corrélation soit certaine : *ALC* ; *ALF* au passage du catalan à l'occitan ; *ALLY* au passage de l'occitan au francoprovençal, etc.

A de rares exceptions près, les valeurs de *y* interpoints sont inférieures à celles de *y* *A-B* considérées aux mêmes stations. Ces exceptions, dans notre corpus, sont au nombre de quatre : *ALG 1* pt 790 NO plus distant

de 5 unités lexicales du point voisin 790 sis à 13 km que du point d'origine sis à 30 km ; *ALC* pt 74 plus distant de 4 unités du point voisin 142 sis à 39 km que du point d'origine sis à 185 km ; *ALR* 1 pt 182 plus distant de 4 unités du point voisin 728 sis à 67 km que du point d'origine sis à 193 km. Un cas semblable se présente dans *ALPO* pt 378 : mais la différence est faible et les distances spatiales très courtes.

Le mieux est de considérer ces exceptions comme des effets de la fluctuation aléatoire.

Tant que le dialecte est fonctionnel, c'est-à-dire sert aux communautés en contact à communiquer tout en se démarquant, la distance lexicale se tient au-dessous du seuil 50, où commence la difficulté de communiquer (v. p. 344). Sa grandeur dépend d'un faisceau de facteurs qu'on n'est pas encore près de mettre en équations : montagne/plaine ; hostilité/amitié ; persistance des divisions tribales/effacement ; unité ethno-culturelle/discontinuité ; autarcie/économie ouverte ; les diverses données historiques, géologiques, climatiques, etc. qui figurent en tête de certains atlas (*ALG*, *ALMC*, *ALPO*) sous la rubrique « cartes auxiliaires » ou « cartes introductives ».

Cependant, deux facteurs privilégiés semblent se dégager : 1° le relief accusé favorise le morcellement lexical ; examiner de ce point de vue la Roumanie carpathique (le relief est figuré dans tous les fonds de cartes), et le début de *ALG* 1 ; 2° de même et plus encore les limites politiques, plus précisément les frontières d'états, à l'intérieur desquelles se forment les vocabulaires nationaux, qu'elles aient été déplacées depuis trois siècles (frontière nord des Pyrénées-Orientales, où  $y$  interpoints fait un bond dans *ALF* 1, bond qui se répercute sur  $y$  *A-B* : c'est la frontière de langue que M. Guiter a mesurée au colloque de Strasbourg) ; encore mieux si cette frontière est immuable : la frontière politique qui sépare le val d'Aran du reste de la Gascogne est responsable de distances lexicales avec les localités françaises contiguës qui battent tous les records, dans la carte de l'*ALG* consacrée à ce paramètre (elle figurera dans le volume VI).

Dans l'*ALLY*, le plus fort interpoint du parcours (30 unités) coïncide avec la frontière objective entre occitan et francoprovençal. Mais sur le même parcours, et à l'intérieur du domaine francoprovençal, il y a tout de même un interpoint de 23 entre les points 41 et 29.

C. Essayons de déceler ce que signifient les valeurs numériques de  $y$  *A-B* référées au concret. Nous avons vu que dans la grande majorité des cas (13 sur 16), la courbe monte en pente raide en début de parcours, étant entendu que la longueur de ce début varie suivant les atlas. On peut dès

lors se poser la question : la différenciation lexicale *A-B* ne serait-elle pas plus forte sur le rayon où le dialecte est fonctionnel, pour s'amortir à partir de la distance spatiale où les gens ne communiquent plus entre eux, soit que la communication vécue n'existe pas, soit que les gens communiquent avec un autre moyen que le dialecte ? Ce qui n'empêche pas qu'à la longue, les petites différences s'accumulant,  $y$  *A-B* n'atteigne ou dépasse le seuil d'intercompréhension (environ 50).

En effet, il ne faut pas oublier que notre histoire de l'habitant d'Arles-sur-Tec cheminant vers Liège pour savoir comment il s'entend avec des gens vivant à 500 ou 950 km de chez lui, n'est qu'une fiction. C'était la réalité au temps où le peuple était unilingue : Savoyards, Catalans, Auvergnats débarquant à Paris, dans les débuts, se débrouillaient comme ils pouvaient. Mais depuis que les Français périphériques sont devenus bilingues, il ne leur viendrait jamais à l'idée d'utiliser leur « patois » pour communiquer au delà d'un certain rayon. Déjà, au temps de l'unilinguisme, les montreurs d'ours du Haut-Couserans (Ariège), qui promenaient leur bête dans toute la France, débitaient leur boniment en français (phonogramme de l'I. E. M. de l'Université de Toulouse-Le Mirail) ; et notre propre mère, âgée de quatre-vingt-douze ans, toulousaine, sait la chanson en français des montreurs de marmottes savoyards. Il en est de même des expériences faites par M. Ravier, à qui il n'a jamais échappé qu'il s'agissait de travaux de laboratoire. Actuellement, jamais un Landais et un Commingeois, distants de 180 km r. l. n'iront se casser la tête à converser dans leurs vernaculaires respectifs, alors qu'il est si commode de recourir au français.

Là-dessus, une autre question vient se greffer, et une question grave : comment se fait-il que, là où le dialecte devient inutile, les choses n'aillent pas au hasard ? Pourquoi  $y$  *A-B* est-il soumis à une détermination mathématique stricte ? Pourquoi, notamment,  $y$  *A-B* ne croît-il plus que faiblement à partir d'une certaine distance ? Généralement parlant, le seul fait que toutes les courbes soient des fonctions logarithmiques nous oblige à nous poser cette question.

Nous pensons qu'il faut considérer les dialectes comme les éléments d'un ensemble. Concrètement : le gallo-roman est un ensemble dont les éléments sont les dialectes (il est d'ailleurs arbitraire de considérer le gallo-roman comme un ensemble clos, surtout en ne retenant que le critère lexical : l'idéal serait de pouvoir opérer sur la Romania en tant qu'ensemble objectif). La typologie ponctuelle de ces dialectes n'est autre que la description du parler de chaque localité ; et la typologie différentielle s'établit par l'opé-

ration des distances de Hamming, comme nous l'avons tenté dans le volume VI de l'*ALG* (en préparation). Dans cette unité supérieure qu'est l'ensemble gallo-roman, ou italo-roman, etc., agit un facteur d'auto-régulation matérialisé par l'équation de la courbe moyenne  $y = K \sqrt{\log(x + 1)}$ , laquelle est un avatar de la loi d'économie toujours présente dans les phénomènes linguistiques. A quoi bon le gallo-roman se serait-il évertué à diversifier son vocabulaire à profusion, puisque cela ne sert à rien ? Le gallo-roman assure son fonctionnement dialectal avec un minimum de bases lexématiques, qu'il agence sur le terrain de façon à assurer à la fois compréhension et contrastes. Rappelons l'une de nos observations. Pour un signifié *S*, le lexème codé 1 n'est pas contrastif entre la localité *A* et la station suivante *a* ; aussi *a* se sert-il du lexème 2. Et ainsi de suite. Mais fort souvent, au terminus du parcours, on voit reparaître le lexème 1. En gros, les localités voisines, où le dialecte est fonctionnel, se tiennent à une distance lexicale forte ; au delà du rayon fonctionnel, la distance lexicale, n'ayant plus d'importance, croît à une allure beaucoup plus faible. Le mécanisme paraît être le suivant : tel ensemble, gallo-roman par exemple, fait resservir les mêmes lexèmes dans la mesure du possible, tout comme un système phonologique fait resservir un nombre très limité de phonèmes pour un nombre de combinaisons théoriquement limité, mais à un chiffre astronomique.

Encore une fable. Imaginons un dialogue entre Pierre et Paul, conseillers municipaux de Rieumes (*ALG* pt 771 NO, 32 km SW de Toulouse) :

*Pierre* — Pour le chien, nous disons *gos*, comme ces vauriens de Toulousains. Il faudrait changer le mot.

*Paul* — Tu as bien raison. Il n'y a qu'à inventer un autre mot.

*Pierre* — Ce n'est même pas la peine. J'en connais un tout prêt : *can*.

*Paul* — Impossible : ces autres vauriens de Roussillonnais disent comme ça.

*Pierre* — Qu'est-ce que ça fait ? Ils sont trop loin, ils ne risquent pas de nous entendre. Et puis, nous dirons comme nos voisins de Lombez [*pt* 679], qui sont de braves types, eux.

A la suite de cette délibération, on a cessé, à Rieumes, de nommer le chien *gos*, mot qu'on a définitivement remplacé par *can* (v. *ALG*, 530).

On objectera que tout cela n'est que fantaisies et billevesées. Selon les lois de la probabilité, là où la fonction dialectale n'existe plus, là où les sujets parlants n'ont plus conscience du fait dialectal, la répartition lexicale ne peut être qu'aléatoire.

La réponse est dans les faits : comment soutenir que seize courbes de

$y$   $A$ - $B$  mesurées sur les parcours les plus divers (Roumanie, Italie, Catalogne, France, etc.) aient toutes la même allure par hasard, et que, réduites à une moyenne, elles obéissent à une loi mathématique stricte ? D'autre part, nous pensons que la liaison langage-conscience est contingente et non nécessaire. La preuve en fut administrée dans une conférence télévisée restée célèbre : M. Lévi-Strauss fit remarquer que le langage génétique, pourtant extrêmement complexe, est dépourvu de conscience. Que la situation dialectale soit réelle (consciente) ou seulement latente (inconsciente), la distance lexicale  $A$ - $B$  est déterminée par une loi mathématique.

La loi d'Estoup-Zipf-Mendelbrot est une loi universelle, qui s'applique aux cas particuliers (encore que cette universalité devienne relative à mesure qu'on la confronte plus minutieusement avec l'expérience : v. R. Michea *BSLP* 1969, 1 ; H. Guiter *ib.* 1970, 1). La loi qui gouverne  $y$   $A$ - $B$  est une loi statistique qui ne s'applique qu'à un groupement de mesures réduites à leur moyenne. Si on considère chaque courbe, on constate qu'il s'agit toujours d'une fonction logarithmique ; mais l'équation n'est pas exactement la même, ne serait-ce que par la constante  $K$ . Cependant nous constatons que 9 sur 16 de ces courbes se groupent sous l'équation générale :  $y = K \sqrt{\log(x+1)}$ , et que parmi ces neuf, quatre ont la même constante :  $y = 42 \sqrt{\log(x+1)}$  (v. p. 348).

2. Quand deux atlas couvrent le même domaine, et c'est le cas de l'*ALF* et des atlas régionaux français, les mesures que nous avons relevées doivent permettre de faire certaines comparaisons.

Examinons le parcours *ALF* 2 (Biarritz-Menton). La courbe  $y$  est étrangement plus basse que celle de *ALF* 1. Après un pic à la localité 763 (Revel), la courbe redescend fortement jusqu'à Mm 53 (Eygalières, Vaucluse), où elle retombe presque à la valeur de l'interpoint initial 690-691 — il ne s'en faut que d'une unité — pour remonter jusqu'à son maximum qui coïncide avec le terminus, sis à 73 Mm.

De cela naît un dilemme : ou bien l'Occitanie, en ce qui touche la répartition de son lexique, est un drôle de pays ; ou bien l'*ALF* ne donne de cette répartition qu'une image peu fiable.

C'est cette dernière éventualité qui s'est avérée, au vu d'une part des mesures effectuées sur l'*ALG*, d'autre part sur l'*ALF* aux localités communes qui jalonnent le parcours *ALF* 2 jusqu'aux limites de l'*ALG*. On a disputé, savoir si Dauzat avait eu raison de nous imposer d'enquêter à nouveau tous les points de l'*ALF*. Voici au moins un cas où ce principe trouve de l'utilité.

Tableau des mesures.

LOCALITÉS	$x$ (km)	$y$ ALF	$y$ ALG
690			
691	40	22	39
685	75	21	40
686	110	17	54
678	150	27	48
679	193	25	58
760	221	25	61
762	249	23	63

La comparaison des colonnes *ALF* et *ALG* se passe de commentaire.

L'*ALMC* ne reprend pas les localités de l'*ALF*. Cependant, il existe dans l'*ALF* un trajet 735-817 qui coïncide assez bien avec celui que nous avons choisi dans l'*ALMC*. Voici les valeurs de  $y$  prises sur les deux courbes à des intervalles de 5 Mm.

$x$ (en Mm)	$y$ ALF	$y$ ALMC
5	15	27
10	20	44
15	25	50
20	31 ?	55 (après un point culminant de 62 sis à 18,8 Mm)

La convergence avec la confrontation *ALF-ALG* se passe également de commentaire. Rappelons que *ALG* et *ALMC* sont en domaine de langue d'oc. N'était ce constat numérique, nous aurions surmonté notre inquiétude touchant les données occitanes de l'*ALF* : notre démarche étant délibérément empirique, nous nous tenons prêt à enregistrer n'importe quel fait objectif.

4. Nous avons calculé les  $y$  propres aux interpoints (paramètre 5 du tableau des mesures). Observe-t-on de fortes valeurs aux endroits où se situent les frontières de langues ? Pour ainsi dire nulle part : la valeur des interpoints qui coïncident avec les frontières italo-roman/francoprovençal, italo-roman/rhêto-roman, occitan/francoprovençal (v. cependant p. 351), francoprovençal/langue d'oïl ne se distinguent pas des interpoints internes. C'est que le lexique, capital dans l'intercompréhension, n'est pas la carac-

téristique principale des langues : la morpho-syntaxe et en second lieu la phonétique historique et la phonologie constituent les vrais critères de la typologie et assurent le sentiment d'une communauté linguistique (p. 339). Une exception : la frontière lexicale entre le catalan et l'occitan, que l'on aperçoit clairement aussi bien dans l'*ALF* que dans l'*ALPO* (et peut-être les frontières des anciens états de l'Italie ?). Mais nous avons dit plus haut notre sentiment à ce sujet (p. 351) : c'est l'influence d'une frontière politique qui agit ici.

Ces remarques ne valent qu'à propos de langues de même famille : elles sont démenties dans les cas contraires. On peut en faire l'expérience au moyen de l'enquête basque de Labastide-Clairence, village où cohabitent une communauté gasconne majoritaire et une communauté basque minoritaire. L'enquête gasconne et l'enquête basque figurent dans l'*ALG* au même point 691 O. Prenons un trajet N-S de 548 à 691 O (*ALG* 3 dans le tableau des mesures) ;  $y$  croît de 548 à 691 O à la même allure que dans les autres cas. Il atteint 53 à 691 O gascon, mais passe verticalement à 90 pour 691 O basque (90 et non 100 parce que dans notre échantillon, le basque présente 10 % d'emprunts au gascon). Donc, le rideau qui, à Labastide-Clairence, séparerait un habitant de Saint-Vivien-de-Médoc (548) de son interlocuteur basque serait opaque à 90 % ; celui de la morpho-syntaxe l'est a priori à 100 % (celui de la phonologie est bien moins épais, mais dans la circonstance, le taux est dépourvu de la moindre utilité). Dans ces conditions, aucune communication n'est possible. Il en va d'ailleurs de même pour un Gascon et pour un Basque de Labastide-Clairence :  $y = 84$ , et la morpho-syntaxe 100 %.

5. Nous avons mentionné incidemment (p. 341-2) que la grandeur moyenne de  $y$  sur un parcours donné n'était pas forcément en corrélation avec le taux général de variété lexicale d'un atlas. Pour l'*ALC*, il semble bien que la faiblesse relative de  $y$  dépende de l'uniformité du vocabulaire catalan ; de toute façon, cette langue est remarquablement homogène et présente un état dialectal peu différencié, encore que les divisions dialectales du catalan retentissent sur  $y$  *A-B* (p. 350). Nous avons rencontré un exemple saisissant de non-corrélation dans l'*ALPO*, également en domaine catalan. Durant que nous établissions la matrice, les cartes mononymes (en totalité ou bien pour le parcours retenu) s'amoncelaient en nombre décourageant ; finalement, nous en comptâmes 62, ce qui limitait à 38 le maximum théorique de  $y$ . Or, en mesurant les distances de Hamming, nous avons constaté que l'*ALPO* se comportait tout aussi brillamment que les autres atlas régionaux de

France en ce qui touche les distances lexicales. Les 38 lexèmes disponibles sont agencés, de la paroisse andorrane d'Ordino jusqu'à Banyuls-sur-Mer, de telle sorte que le démarquage est parfaitement assuré, du moins jusqu'aux régions basses du Roussillon et de l'Ampourdan. Un bel exemple d'économie, dans tous les sens du mot. Il est vrai aussi que, sur ces parcours, les grands massifs montagneux de l'ouest jouent un rôle diviseur important, ainsi que l'observait M. Guiter au colloque de Strasbourg.

6. Nous tenons à insister sur le caractère résolument empirique et positiviste de notre entreprise. Parmi les éventualités que nous avons envisagées, deux seulement — qui au fond se ramènent à une — se sont réalisées. Mais nous n'avons étudié que seize parcours. Qu'est-ce qui prouve qu'il n'existe pas, quelque part dans le monde, un pays où  $y$   $A-B$  marcherait à la façon des écrevisses (éventualité 5) ? Notre base statistique est bien trop faible pour que nous puissions chiffrer une probabilité ; et si un dialectologue découvrait un jour l'occurrence 5, la nouvelle serait accueillie joyeusement. Cela démontrerait tout simplement que la loi de corrélation entre distance spatiale et distance lexicale n'est pas un « universal », mais une loi régionale confinée à une partie de l'Europe.

7. Pour finir, *quattro passi sui nuvoli*. Supposons que l'histoire du monsieur qui va d'Arles-sur-Tec à Liège soit une histoire vécue. Il trouve des difficultés croissantes jusqu'à un maximum qui se situe au cœur du domaine franco-provençal. A partir de là, si ce n'est pas plus facile, du moins n'est-ce pas sensiblement plus difficile. Alors, le mécanisme d'auto-régulation, qui limite la croissance de  $y$ , agirait à la manière des répéteurs jalonnant une ligne de télécommunication par fil, qui redressent de proche en proche la déperdition de l'information et permettent ainsi aux gens de se comprendre en dépit de l'éloignement.

J. SÉGUY.