



HAL
open science

De l'espace dans la langue ? Distribution de la distance linguistique en breton et incidence de facteurs géographiques

Tanguy Sollicec

► **To cite this version:**

Tanguy Sollicec. De l'espace dans la langue ? Distribution de la distance linguistique en breton et incidence de facteurs géographiques. *La Bretagne Linguistique*, 2022, 24, pp.115-166. 10.4000/lbl.4280 . hal-03835710

HAL Id: hal-03835710

<https://hal.science/hal-03835710v1>

Submitted on 30 Jan 2025

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

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



Distributed under a Creative Commons Attribution 4.0 International License

De l'espace dans la langue ? Distribution de la distance linguistique en breton et incidence de facteurs géographiques

TANGUY SOLLIEC*

À la mémoire de Jean Le Dû et aux discussions
que nous n'aurons pas eues.

Introduction

La question de l'incidence de la géographie sur les structures linguistiques a été renouvelée depuis une quinzaine d'années grâce à la constitution de bases de données importantes et à des analyses construites sur l'emploi d'outils statistiques¹. De façon générale, la géographie est considérée jouer un rôle de premier plan dans la répartition des langues humaines à la surface du globe et de certains faits linguistiques². Les propositions formulées dans ce domaine

* Les travaux de Tanguy Sollicc portent sur la géolinguistique, la dialectologie des langues et leur variation dans l'espace, à partir du cas de la langue bretonne. Il s'intéresse plus particulièrement à la description des phénomènes linguistiques à l'aide de méthodes quantitatives.

1. D. Robert LADD, Seán G. ROBERTS, Dan DEDIU, «Correlational Studies in Typological and Historical Linguistics», *Annual Review of Linguistics*, t. 1/1, 2015, p. 221-241.

2. Johanna NICHOLS, *Linguistic Diversity in Space and Time*. Chicago, University of Chicago Press, 1992. Daniel NETTLE, «Explaining Global Patterns of Language Diversity», *Journal of Anthropological Archaeology*, t. 17/4, 1998, p. 354-374. Daniel NETTLE, *Linguistic Diversity*, Oxford, Oxford University Press, 1999. Johanna NICHOLS, «The vertical archipelago: Adding the third dimension to linguistic geography», dans P. Auer, M. Hilpert, A. Stukenbrock, B. Szmrecsanyi, *Space in Language and Linguistics. Geographical, Interactional, and Cognitive Perspectives*, Berlin, De Gruyter, 2013, p. 38-60. R. DE BUSSE, R. J. LAPOLLA (dir.), *Language Structure and Environment. Social, Cultural, and Natural Factors*, Amsterdam, John Benjamins

de recherche permettent de mieux appréhender l'organisation de la diversité linguistique à l'échelle de la planète et comment l'émergence de certains traits dans les langues résulte des particularités des milieux dans lesquels ces langues sont parlées. En étudiant la possibilité d'une relation entre la présence d'un type de consonne, les éjectives, au sein du système linguistique de certaines langues et l'altitude à laquelle les locuteurs de ces dernières vivent, Caleb Everett a suggéré que les paramètres environnementaux étaient susceptibles d'avoir une influence sur l'organisation des structures des langues³. Ce résultat a ouvert un vif débat qui a interrogé la validité d'un tel lien⁴, amenant l'auteur à réviser ultérieurement sa position⁵. De leur côté, Coupé et Maddieson ont mis en relation la sonorité et un couvert végétal boisé, particulièrement aux tropiques⁶. Ainsi, un ensemble de recherches vise à évaluer dans quelle mesure les langues s'adaptent aux spécificités du milieu géographique dans lequel elles sont utilisées.

Publishing Company, 2015. Antonio BENÍTEZ-BURRACO, Steven MORAN (dir.), *The Adaptive Value of Languages: Non-Linguistic Causes of Language Diversity*. Lausanne, Frontiers Media, 2018. Xia HUA, Simon J. GREENHILL, Marcel CARDILLO, Hilde SCHNEEMANN, Lindell BROMHAM. « The Ecological Drivers of Variation in Global Language Diversity », *Nature Communications*, t. 10/1, 2019, p. 2047.

3. Caleb EVERETT, « Evidence for Direct Geographic Influences on Linguistic Sounds: The Case of Ejectives », *PLOS ONE*, n° 8/6, e65275 [en ligne]. Mis en ligne le 12 juin 2013, URL : <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0065275> (consulté le 26 septembre 2020).

4. Seán G. ROBERTS, « Altitude and ejectives: hypotheses up in the air », *Replicated Typo* (blog). Mis en ligne le 13 juin 2013, URL : <http://www.replicatedtypo.com/altitude-and-ejectives-contact-and-population-size/6357.html> (consulté le 27 septembre 2020). Harald HAMMARSTRÖM, « C. Everett's ejectives/altitude correlation is not significant », *Diversity Linguistics Comment* (blog). Mis en ligne le 17 juin 2013, URL : <https://dlc.hypotheses.org/491> (consulté le 27 septembre 2020).

5. C. EVERETT, « The Sound Systems of Languages Adapt, but to What Extent? Considerations of Typological, Diachronic and Mercurial Data », *Cadernos de Linguística*, vol. 2/1, 2021, p. 1-23, <http://doi.org/10.25189/2675-4916.2021.V2.N1.ID342> (consulté le 27 septembre 2020).

6. Ian MADDIESON, Christophe COUPÉ, « Human spoken language diversity and the acoustic adaptation hypothesis », *The Journal of the Acoustical Society of America*, t. 138/3, 2015, p. 1838-1838. Ian MADDIESON, « Language Adapts to Environment: Sonority and Temperature », *Frontiers in Communication*, n° 3, 2018, p. 28, [en ligne]. Mis en ligne le 24 juillet 2018, URL : <https://www.frontiersin.org/articles/10.3389/fcomm.2018.00028/full> (consulté le 26 septembre 2020).

Bien évidemment, cette approche ne vise nullement à réactualiser naïvement la théorie des climats exprimée par Montesquieu au XVIII^e siècle et par bien d'autres avant lui. L'ambition ici est saisir dans quelle mesure les facteurs géographiques contribuent à façonner les langues. Johanna Nichols, spécialiste de typologie linguistique, considère ainsi que « *[g]eography shapes language*⁷ ». En examinant la relation entre la complexité des structures linguistiques des langues en milieu alpin et les particularités de ce milieu, elle montre qu'une palette de paramètres est à prendre en considération :

« Aspects of the ecological and geological landscape such as vegetation, rivers, soil, length of growing season, climate, altitude, and barriers shape the directionality (literal and sociolinguistic), rate, and frequency of language spreads and constrain the sociolinguistic variables that make for greater vs. lesser complexity. Space in this approach is a matter of physical reality, and it is three-dimensional. »⁸

Or, dans la plupart de ces travaux, les analyses sont menées de langue à langue sans étudier en détail dans quelle mesure les critères à l'œuvre dans la structuration de la diversité linguistique se répercutent au sein même de chaque domaine linguistique. Or, décrire et analyser la variation diatopique, spatiale, au sein d'une langue amène inmanquablement à réfléchir à la place que la géographie joue dans l'organisation des faits étudiés et si les mêmes facteurs y sont partout à l'œuvre. Différentes dimensions de la géographie sont effectivement considérées intervenir dans la distribution des faits dialectaux ou vernaculaires⁹ sans que la nature de celles-ci soit forcément envisagée avec précision¹⁰.

7. J. NICHOLS, *art. cit.*, 2013, p. 38.

8. *Ibidem*, p. 38.

9. Jack K. CHAMBERS, Peter TRUDGILL, *Dialectology*, Cambridge, Cambridge University Press, 1998. John NERBONNE, Peter KLEIWEG, « Toward a Dialectological Yardstick », *Journal of Quantitative Linguistics*, t. 14/2-3, 2007, p. 148-166.

10. David BRITAIN, « Conceptualisations of geographic space in linguistics », dans A. Lameli, R. Kehrein, S. Rababus (dir.), *Language and Space. An International Handbook of Linguistic Variation*, Berlin, De Gruyter Mouton, 2010, p. 69-97. David BRITAIN, « Language and Space: The Variationist Approach », dans P. Auer et J. E. Schmidt (dir.) *Language and Space. An International Handbook of Language Variation*, Berlin, De Gruyter Mouton, 2010, p. 142-162.

Le breton, langue celtique parlée dans l'ouest de la France, constitue un véritable laboratoire pour les linguistes, pour reprendre l'idée de Gary German. Il offre ainsi un terrain d'investigation intéressant pour aborder une telle question. En effet, si les usages écrits du breton sont attestés de longue date, ceux-ci n'ont eu toutefois que peu d'emprise sur les pratiques langagières qui elles, sont restées avant tout marquées par l'oralité. De ce fait, le breton présente un caractère nettement dialectalisé, ce qui permet de réfléchir aux éventuels paramètres géographiques contribuant à une telle situation. Par ailleurs, l'extension géographique relativement limitée du breton (la pointe occidentale de la péninsule bretonne) permet de prendre assez facilement en considération l'ensemble du domaine d'une langue et d'y observer de quelle(s) façon(s) un milieu peut agir sur la différenciation linguistique.

Une telle démarche nécessite d'une part de pouvoir envisager un espace linguistique dans son ensemble et, d'autre part, de disposer de données pouvant être étudiées en détail. C'est à ce titre que nous avons procédé à une analyse des données provenant du *Nouvel atlas linguistique de la Basse-Bretagne* (désormais NALBB) de Jean Le Dù¹¹. Cet ouvrage constitue un corpus particulièrement approprié dans notre perspective car le réseau de points d'enquête sur lequel il s'appuie permet d'établir un maillage relativement fin du domaine bretonnant. En outre, malgré les limites inhérentes à tout atlas linguistique, le NALBB rend compte avec fidélité de plusieurs aspects des variétés de breton en usage à la fin du XX^e siècle¹².

Toutefois, la question de la méthodologie adéquate pour exploiter les données du NALBB, à l'instar de celles issues de tout autre atlas linguistique, se pose. Si le tracé d'isoglosses demeure une approche de choix pour observer la répartition des faits linguistiques dans l'espace, elle ne permet de prendre en compte qu'un nombre limité de phénomènes. D'autre part, les isoglosses ne coïncident que rarement, ce qui peut rendre

11. Jean LE DÛ, *Nouvel Atlas Linguistique de la Basse-Bretagne*, Brest, CRBC, 2001.

12. Francis FAVEREAU, «Jean Le Dù. Nouvel atlas linguistique de la Basse-Bretagne. Brest, CRBC, 2001, 2 vol.», *Bulletin de la Société archéologique du Finistère*, t. 130, 2001, p. 538-543. Elmar TERNES, «Compte rendu Jean Le Dù, Nouvel Atlas Linguistique de la Basse-Bretagne (NALBB). Brest 2001. 2 vols», *Études celtiques*, t. 36, 2008, p. 223-28.

leur représentation graphique difficilement lisible et interprétable¹³. De ce fait, nous avons eu recours à une méthode quantitative pour traiter les données du NALBB, la dialectométrie, approche que nous avons mise en œuvre dans le cadre de notre doctorat¹⁴. Celle-ci vise à quantifier le degré de similarité linguistique (ou de différence, selon le point de vue adopté) qui existe entre les différents points d'enquête d'un atlas linguistique, en comparant les données correspondantes de façon systématique. Ceci permet d'aboutir à un taux de similarité entre chaque point d'enquête d'un atlas linguistique et ses voisins immédiats. À partir de là, il devient possible d'examiner de quelle façon différents facteurs interviennent sur la répartition de la distance linguistique à travers l'espace.

En premier lieu, nous présenterons les données analysées avant de détailler la méthodologie mise en œuvre. Dans un second temps, nous examinerons les résultats obtenus et tout particulièrement comment ceux-ci s'organisent du point de vue spatial. Ensuite, nous évaluerons le rôle de différents paramètres géographiques susceptibles d'intervenir dans la distribution spatiale de la distance linguistique. La section suivante sera consacrée à une discussion des résultats obtenus.

Données étudiées, méthode mise en œuvre, enjeux épistémologiques

Les données ont été analysées selon la méthode de la dialectométrie. Cette approche vise à quantifier et à mesurer la distance linguistique. Cette dernière notion s'entend comme le degré de similarité ou de différence qui s'observe lors de la comparaison de deux variétés linguistiques, tel que le définit Paul Heggarty : « *The term 'linguistic distance' is often used to refer to the degree of similarity/difference between any two language varieties*¹⁵ ».

13. Hans GOEBL, «Dialectometry and quantitative mapping», dans A. Lameli, R. Kehrein, S. Rababus (dir.), *Language and Space. An International Handbook of Linguistic Variation*, Berlin, De Gruyter Mouton, 2010, p. 433-457.

14. Tanguy SOLLIEC, *Distance linguistique et dialectométrie, une application à la langue bretonne. Enjeux, méthodologie et interprétations*, Thèse de doctorat, Brest, Université de Bretagne Occidentale, 2021.

15. Paul HEGGARTY, «Quantifying change over time in phonetics», dans C. Renfrew, A. McMahon, L. Trask, *Time Depth in Historical Linguistics*, Cambridge, McDonald Institute for Archaeological Research, 2000, p. 531-562.

Plusieurs propositions ont été formulées dès le XIX^e siècle pour réaliser des mesures de distance linguistique de façon systématique. Toutefois, c'est Jean Séguy qui le premier a élaboré une méthodologie dans ce but qu'il a nommée *dialectométrie*, ouvrant ainsi un nouveau champ d'étude à l'analyse linguistique¹⁶. Cette nouvelle contribution apporte un renouvellement méthodologique à la dialectologie en lui permettant d'agréger de nombreux faits et de dépasser ainsi l'«atomisme» qui caractérise les faits linguistiques vernaculaires¹⁷. Divers outils computationnels associés à des traitements statistiques et à des représentations cartographiques sont à présent utilisés dans ce domaine¹⁸.

L'intérêt de la dialectométrie est qu'elle permet de mettre au jour des structures qui ne sont pas observables à l'œil nu. La distance linguistique agit souvent comme le révélateur de phénomènes d'organisation anciens dont les faits linguistiques ont enregistré l'écho. Hans Goebel, un des pionniers de la dialectométrie, décrit cet état de fait comme l'«activité basilectale d'Homo loquens» et la dialectométrie permet d'accéder à ce niveau d'organisation des faits linguistiques¹⁹. La distance linguistique constitue ainsi un point de vue heuristique dans la description des phénomènes langagiers.

Nous avons appliqué une telle analyse dialectométrique à des données du breton issues du NALBB. Cet atlas, résultat d'un travail débuté en 1969, couvre le domaine bretonnant de façon dense avec

16. Jean SÉGUY, «La dialectométrie dans l'Atlas linguistique de la Gascogne», *Revue de linguistique romane*, t. 37/145-146, 1973, p. 1-24. H. GOEBL, *art. cit.* Martijn WIELING, John NERBONNE, «Advances in Dialectometry», *Annual Review of Linguistics*, t. 1/1, 2015, p. 243-264.

17. Eugenio COSERIU, *Einführung in die Allgemeine Sprachwissenschaft*, Tübingen, Francke Verlag, 1988, p. 44.

18. John NERBONNE, Wilbert HEERINGA, «Measuring Dialect Differences», dans P. Auer et J. E. Schmidt (dir.), *Language and Space. An International Handbook of Language Variation*, Berlin, De Gruyter Mouton, 2010, p. 550-567. Wilbert HEERINGA, Jelena PROKIĆ, «Computational Dialectology», dans C. Boberg, J. Nerbonne, D. Watt (dir.), *The Handbook of Dialectology*, Oxford, Wiley Blackwell, 2018, p. 330-347.

19. Hans GOEBL, «Regards dialectométriques sur les données de l'Atlas Linguistique de la France (ALF) : relations quantitatives et structures de profondeur», *Estudis Romànics*, t. 25, 2003, p. 59-96.

187 points d'enquête et renouvelle ainsi l'entreprise de Pierre Le Roux, l'*Atlas linguistique de la Basse-Bretagne*, élaboré lui-même dans le sillage de l'*Atlas linguistique de la France* de Gilliéron et Edmont²⁰. Les enquêtes réalisées pour ce dernier portaient avant tout sur les parlers gallo-romans et de ce fait, la Bretagne de langue celtique n'avait pas été étudiée. Le NALBB s'intègre pour sa part au projet du NALF (Nouvel atlas linguistique de la France) dont le but était de couvrir le territoire de la France d'atlas linguistiques régionaux afin qu'ils s'attachent à rendre avec plus d'attention certains faits ethnographiques ou linguistiques locaux ayant pu échapper à un ouvrage d'ampleur nationale²¹.

L'orientation du NALBB est nettement phonétique mais il avait été initialement prévu que cet atlas connaisse une seconde partie consacrée plus spécifiquement à la morphologie et au lexique. Le NALBB comprend ainsi près de 100 000 formes correspondant peu ou prou à 598 notions. Leur retranscription a été assurée par Jean Le Dù lui-même, ce qui a permis d'éviter d'éventuelles isoglosses de transcripteurs. En effet, lorsque plusieurs personnes interviennent au moment de la retranscription du matériel sonore, leurs filtres phonologiques respectifs peuvent les amener à rendre compte différemment d'une même réalité et aboutir à des isoglosses d'enquêteurs²². Le choix d'un retranscripteur unique s'est imposé d'emblée lors de l'élaboration du projet du NALBB²³.

Jean Le Dù a retenu dans ce but une notation fine permettant de distinction quatre degrés d'aperture pour les voyelles moyennes. Il a également différencié trois niveaux de diphtongaison (ei, ej, eⁱ) et a eu recours à un signe qui ne figure pas dans l'alphabet phonétique

20. Edmond EDMONT, Jules GILLIÉRON, *Atlas linguistique de la France*, Paris, Champion, 1902-1910. Pierre LE ROUX, *Atlas linguistique de la Basse-Bretagne*, Rennes, Plihon et Hommay, 1924-1963.

21. Alfred DAUZAT, « Un nouvel Atlas linguistique de la France », *Le français moderne*, t. 7/2, 1939, p. 97-101. Jean SÉGUY, « Les Atlas linguistiques de la France par régions », *Langue française*, t. 18/1, 1973, p. 65-90.

22. Andreas MATHUSSEK, « On the problem of field worker isoglosses », dans M.-H. Côté, R. Knooihuizen, J. Nerbonne (dir.), *The future of dialects: Selected papers from Methods in Dialectology XV*, Language Variation. Berlin : Language Science Press, 2016, p. 99-116.

23. Jean LE DÛ, « Introduction », *La Bretagne Linguistique*, vol. 13, 2004, 5-11.

international, le [f] qu'il décrit, à la suite de Falc'hun, comme « une sorte de [v] qui est prononcé plus énergiquement avec une forte aspiration »²⁴. En outre, il a rendu compte de l'articulation faible de certains segments en les notant en exposant, particulièrement dans le cas des semi-consonnes en position intervocalique, du schwa et de certaines fricatives

Nous avons donc mis en œuvre une analyse systématique du degré de similarité entre les données des différents points du NALBB afin de quantifier la distance linguistique qui les sépare. Nous avons dès lors sélectionné un ensemble de 106 notions en veillant à ce qu'elles soient effectivement comparables du point de vue phonétique²⁵. Le réseau géographique de la figure 1 a constitué le cadre à partir duquel nous avons mis en œuvre ces comparaisons selon le point de vue interponctuel²⁶. Ainsi, chaque point du NALBB a été relié à ses voisins immédiats et dans certains cas, à des localités plus distantes. Les 186 points d'enquête étudiés ont ainsi été regroupés en 534 segments ou couples de localités afin d'en comparer les données respectives²⁷.

Les formes de chaque point d'enquête pour les 106 notions observées ont donc été comparées à celles des points voisins afin d'établir un taux de similarité entre les localités comparées. Ce processus de quantification de la distance linguistique s'est fait grâce à l'utilisation d'un algorithme, la distance de LevenshteinModificationsReturned²⁸. Cet outil est une

24. François FALC'HUN, *Etudes sur la langue bretonne : système consonantique, mutation et accentuation*, Brest, Label LN, 2005, p. 77-79. J. LE DÛ, *op. cit.*, v.

25. Nous avons ainsi écarté les cas de doublets lexicaux comme *sailh/kelorn* ou *mintin/beure* pouvant former des isoglosses nettes mais ne permettant pas une comparaison segment à segment pour l'ensemble du domaine bretonnant.

26. Théobald LALANNE, « Indice de polyonymie. Indice de polyphonie », *Le Français Moderne*, t. 21, 1953, 263-274. E. Bagby ATWOOD, « The phonological division of Belgo-Romance », *Orbis*, t. 4, 1955, p. 365-389. Hans GOEBL, « Parquet polygonal et treillis triangulaire, les deux versants de la dialectométrie interponctuelle », *Revue de linguistique romane*, t. 47, 1983, p. 353-412.

27. Le NALBB compte 187 points d'enquête mais nous avons dû exclure de notre analyse le point 69 Caurel en raison de l'aspect lacunaire des données reportées pour cette localité. Jean Le Dû signale en effet que le breton local a disparu entre les différentes phases des enquêtes réalisées pour cet atlas (J. LE DÛ, *op. cit.*, vi.).

28. Guylaine BRUN-TRIGAUD, « Un usage particulier de l'algorithme de Damerau-Levenshtein dans le domaine occitan », dans F. Diemoz, D. Aquino-Weber, L. Grüner, A. Reusser-Elzingre (dir.), « *Toujours langue varie...* », *Mélanges de linguistique historique du français et de dialectologie galloromane offerts à M. le Professeur Andres Kristol par ses collègues et anciens élèves*, Genève, Droz, p. 127-148.

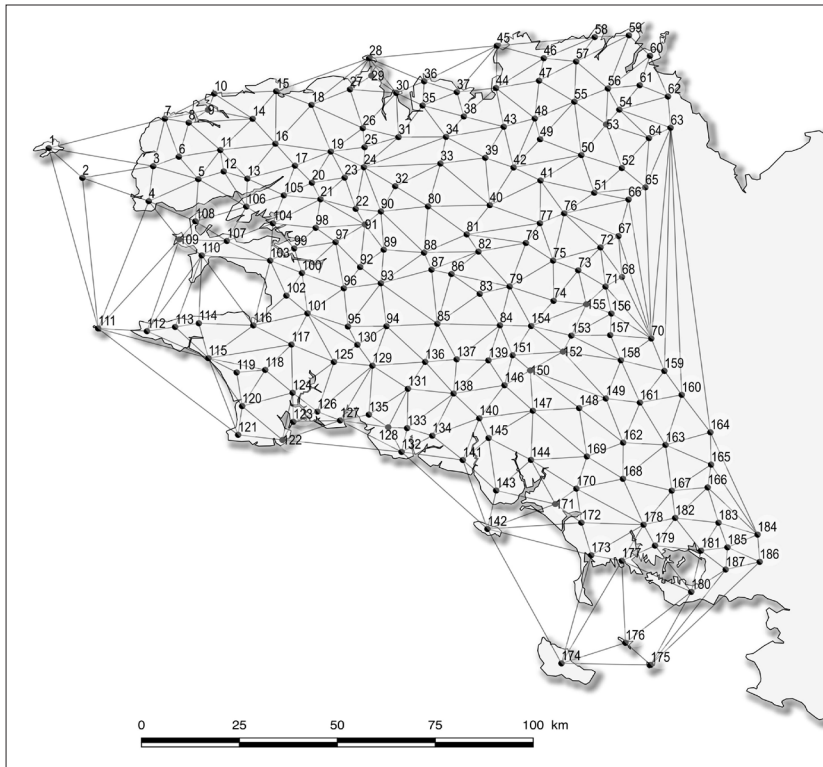


Figure 1. Réseau des points du NALBB étudiés

adaptation de la distance de Levenshtein dont l'objectif est d'évaluer le nombre d'opérations nécessaires pour transformer une chaîne de caractères en une autre. Après avoir aligné les formes comparées, il utilise trois opérations à cet effet :

- le remplacement d'un caractère par un autre
- la suppression d'un caractère
- l'insertion d'un caractère

L'algorithme comptabilise le nombre d'opérations requis et retient la solution la plus économique. À partir de ce décompte, le taux de similarité s'obtient en appliquant la formule suivante²⁹ :

29. Il s'agit ici du calcul de la similarité linguistique. Cette dernière fonctionne en miroir avec la distance linguistique. Ainsi 60 % de similarité équivaut à 40 % de distance.

$$100 - \left(\left(\frac{\text{nombre d'opérations}}{\text{nombre de caractères de la chaîne la plus longue}} \right) \times 100 \right)$$

Si nous prenons le cas de la comparaison suivante :

(1) carte 176 du NALBB DE L'HERBE *geot* :

point 88 Collorec ['jɛwot] ~ point 90 Botmeur ['jɛot]

Le rapprochement entre les deux formes permet d'établir cet alignement :

'j	ɛ	w	o	t	
'j	ɛ		ɔ	t	
1		1	1		(3)

Le passage de la première forme à la seconde nécessite trois opérations, à savoir :

- le remplacement de [o] par [ɔ] ;
- la suppression de [w] ;
- le remplacement de [j] par [j].

À partir de là, le taux de similarité entre les deux formes peut être calculé :

$$100 - \left(\left(\frac{3}{5} \right) \times 100 \right) = 40$$

Le taux de similarité entre les deux formes s'élève par conséquent à 40 %.

En procédant de la sorte pour chacune des 106 notions étudiées, nous avons établi un taux de similarité moyen pour chaque segment en calculant la moyenne des résultats obtenus. Par souci de clarté, nous réservons l'expression « distance linguistique » pour décrire l'espace symbolique plus ou moins important entre deux objets linguistiques (langues, variétés, textes, etc.). Nous retiendrons « similarité linguistique » pour décrire les résultats obtenus et par là, la proximité entre les différents parlers étudiés. La similarité est bien entendue à comprendre comme le miroir de la différence.

Résultats obtenus

Présentation générale

Les résultats obtenus à l'issue de l'analyse dialectométrique s'étendent d'un minimum de 39,15 %, segment 63 (Pléguien) – 160 (Noyal-Naizin) le long de la frontière linguistique à un maximum de 88,47 %, segment 57 (Plougrescant) – 58 (Camlez), sur la façade maritime du Trégor, sur les côtes de la Manche. Le taux moyen de similarité linguistique est de 68,43 % à travers le domaine bretonnant. Le taux médian présente une valeur relativement proche (68,27 %), ce qui indique que les valeurs extrêmes ne pèsent que de façon modérée sur l'ensemble de ces résultats. Le tableau 1 présente les mesures statistiques descriptives réalisées à partir de ces résultats.

nombre de segments examinés	534
moyenne	68,43 %
médiane	68,27 %
maximum	88,47 %
3 ^e quartile	75,63 %
1 ^{er} quartile	62,83 %
minimum	39,15 %
étendue	49,32
écart-type	9,52

Tableau 1. Mesures statistiques descriptives des résultats de similarité linguistique

L'examen de la distribution des résultats de similarité linguistique obtenus à la suite de l'analyse dialectométrique permet d'évaluer la normalité de celle-ci et observer si elle résulte d'un simple hasard. Ainsi, la figure 2 représente les taux de similarité linguistique obtenus dans le cadre d'un graphique de densité et d'un diagramme Quantile-Quantile de normalité.

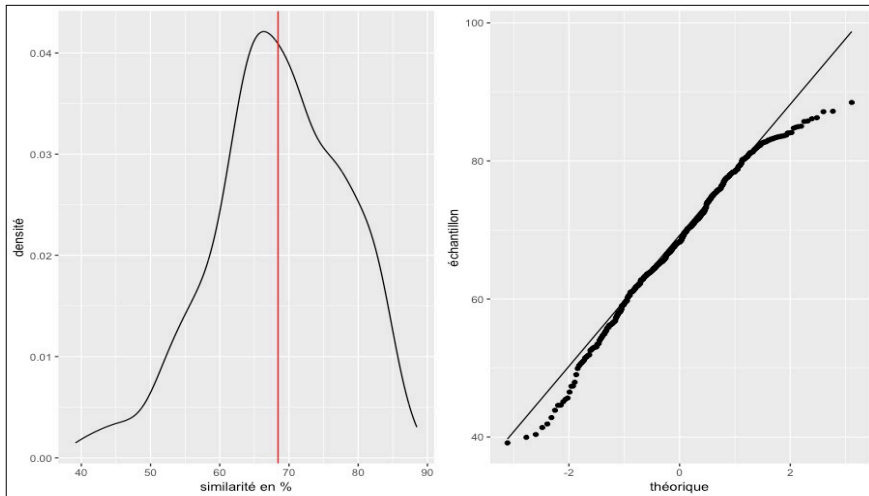


Figure 2. Graphique de densité et diagramme Quantile-Quantile de normalité

Ainsi, le hasard aboutirait à une courbe en cloche parfaite, obéissant à une distribution gaussienne dans le premier cas et à une ligne droite se confondant avec celle donnée en noir dans le second. En outre, le test statistique de Shapiro-Wilk, qui évalue la normalité d'une distribution, permet de confirmer cette première observation. On obtient ainsi : $W = 0,98457$, $p \text{ value} < 0,05$. Ces valeurs confirment que les résultats obtenus ne sont pas la conséquence d'un simple hasard. Ceci invite à penser que des facteurs particuliers viennent modéliser la distribution de la distance linguistique à travers le domaine bretonnant. La figure 3 en donne la représentation cartographique.

La taille et la couleur des segments entre deux points d'enquête est proportionnelle au niveau de similarité linguistique qui les caractérise. Plus celui-ci est important, plus l'épaisseur des segments est grande et plus le bleu sera sombre. Une première inspection visuelle permet de constater que le niveau de similarité linguistique est distribué inégalement à travers l'espace. Une lecture plus attentive conduit toutefois au constat que la répartition de la distance linguistique à travers le domaine bretonnant se structure à plusieurs niveaux.

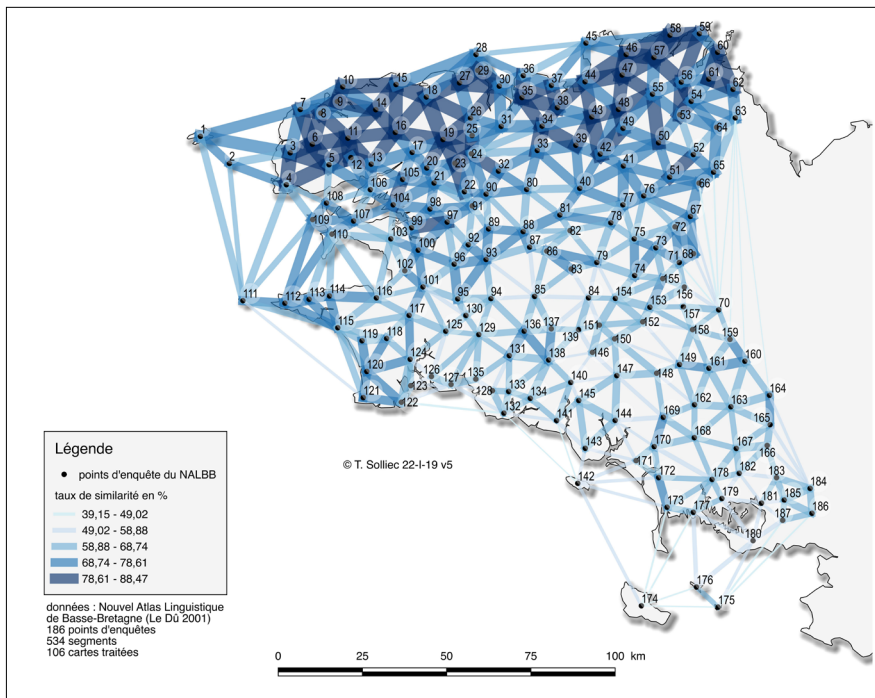


Figure 3. Distribution de la distance linguistique à travers le domaine bretonnant

Des structures organisées sur le plan spatial

Deux niveaux de lecture se distinguent ainsi, s'imbriquant l'un à l'autre et se cumulant. Leur identification se fait visuellement, en les délimitant par proximité spatiale mais également en se basant sur les comportements statistiques observés.

Niveau 1

Le premier niveau est celui d'une opposition qui découpe le domaine bretonnant entre une zone nord-ouest et une zone sud-est, comme l'indique la figure 4.

En allant d'ouest en est, la séparation entre les deux zones suit l'estuaire puis le cours inférieur de l'Odette avant de rejoindre la partie finistérienne de l'Aulne et partage le pays *vFañch*, la région autour de Saint-Nicolas-du-Pélem (22) en deux laissant au Nord-ouest un plateau nommé localement *kroec'haoù* (hauteurs) et au Sud-est, les terres de plaines dites *'n diezaoù*. À l'est, la limite pour les deux zones se confond avec la frontière linguistique.

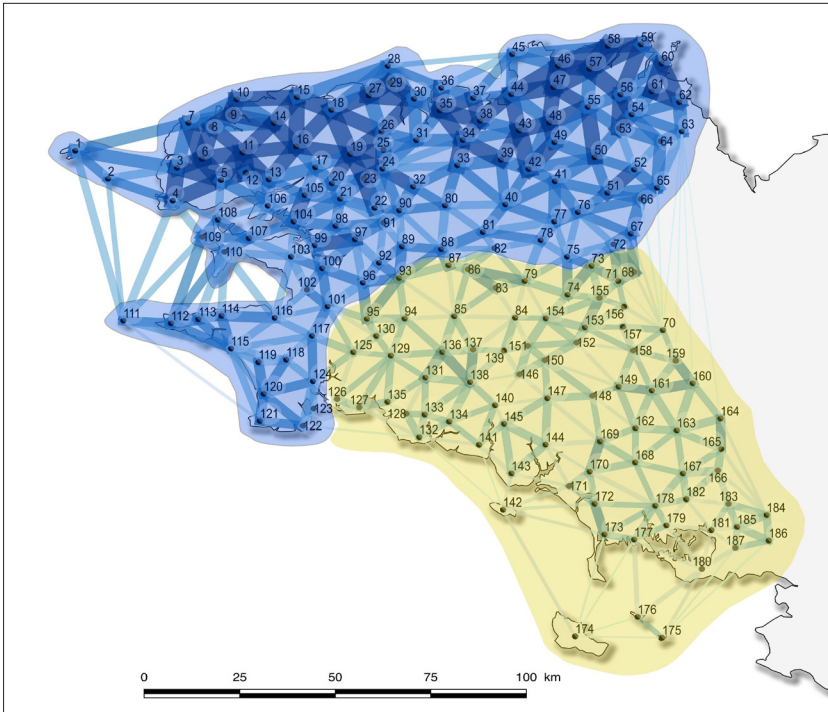


Figure 4. Deux principales aires de Basse-Bretagne en fonction de la similarité linguistique

De plus, chacune de ces zones se caractérise par des valeurs de similarité linguistique distinctes ainsi que par un niveau d'hétérogénéité différent comme le tableau 2 le présente.

	aire nord-ouest	aire sud-est	ensemble de la Basse-Bretagne
nombre de segments	323	211	534
taux moyen de similarité	73,9 %	60,04 %	68,43 %
médiane	74,1 %	61,55 %	68,27 %
maximum	88,47 %	72,39 %	88,47 %
minimum	51,58 %	39,15 %	39,15 %
étendue	36,89	33,24	49,32
coefficient de variation	8,85	11,49	13,91

Tableau 2. Statistiques descriptives pour les zones nord-ouest et sud-est

Ainsi, la zone nord-ouest (en bleu) présente un niveau élevé de similarité linguistique avec un taux moyen de 73,9 %. Il est de 68,43 % pour l'ensemble de la Basse-Bretagne. Les valeurs de similarité relevées dans la zone sud-est (en jaune) sont nettement plus faibles, avec un taux moyen de 60,04 %. On constate également un décalage entre les valeurs extrêmes (minima et maxima) de chacune de ces zones d'une dizaine de points soulignant ainsi le contraste manifeste entre ces deux espaces.

À cette première différence vient s'ajouter une disparité dans l'homogénéité de la distance linguistique pour ces deux territoires. La zone nord-ouest est ainsi relativement homogène du point de vue des valeurs de similarité linguistique, en comparaison de la zone sud-est. En effet, dans le Nord-ouest, la variation de ces valeurs, d'un segment à l'autre, est relativement faible, ce qui n'est pas le cas pour le Sud-est. On relève ainsi pour le premier un coefficient de variation de 8,85 alors qu'il est de 11,49 pour le second. La figure 5 rend compte visuellement de cette différence.

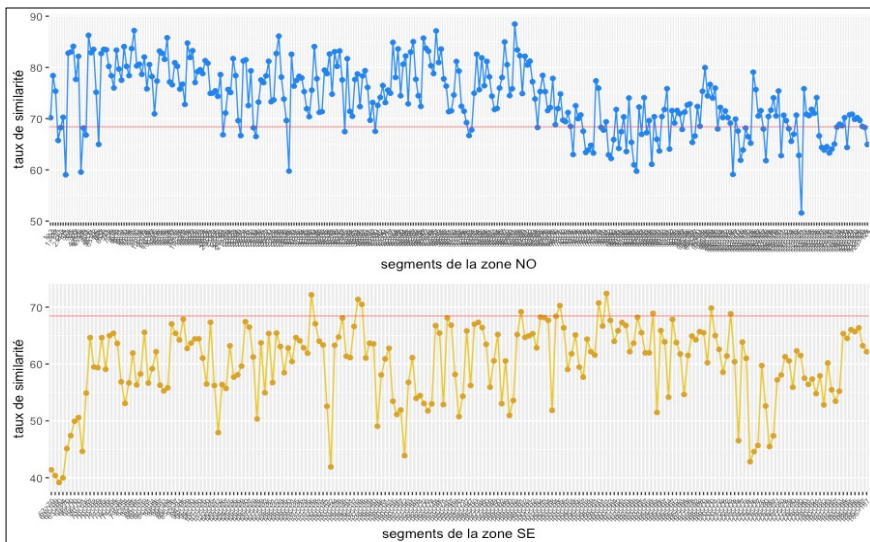


Figure 5. Distribution respective de la similarité linguistique segment par segment dans les zones nord-ouest et sud-est (la ligne rouge indiquant le taux moyen de similarité pour l'ensemble du domaine bretonnant)

L'observation attentive du comportement de la similarité linguistique permet donc d'identifier deux zones géographiques distinctes, la zone nord-ouest et la zone sud-est. Chacune affiche des caractéristiques qui la distinguent de sa voisine.

Niveau 2

D'autre part, au sein de ces deux zones se dessinent des pôles de similarité plus localisés et d'envergure plus limitée. Selon leur appartenance à la zone nord-ouest ou à la zone sud-est, ces pôles présentent des profils différents. Nous avons identifié 10 ensembles au sein du domaine bretonnant. Nous avons attribué à chacun d'entre eux un nom auquel correspond une pastille de couleur plus ou moins étendue comme le montre la figure 6.

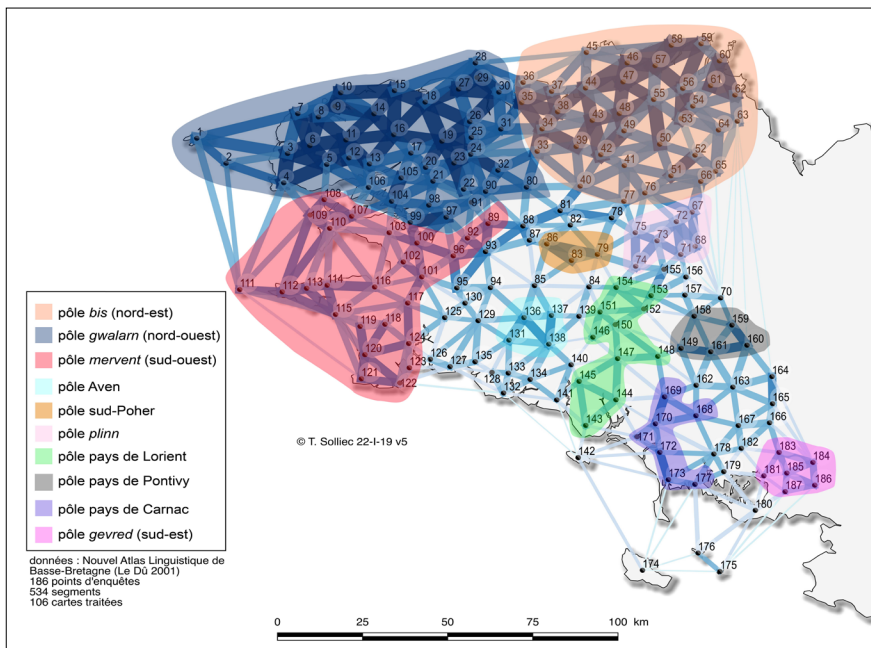


Figure 6. Pôles de similarité linguistique au sein de l'espace bretonnant

La zone nord-ouest est occupée par 3 pôles dont l'emprise tend à recouvrir l'ensemble de la zone. Allant d'est en ouest, on relève ainsi le pôle *bis*, le pôle *gwalarn* et le pôle *mervent*³⁰. Dans ces différents pôles, le taux de similarité entre les segments est élevé voire très élevé (voir tableau 3). Ils se caractérisent également par leur étendue géographique qui est relativement importante au regard des pôles de la zone sud-est.

30. Ces pôles portent le nom breton des points cardinaux qui leur sont associés, littéralement nord-est, nord-ouest et sud-ouest.

La zone sud-est présente en son sein un nombre plus important de pôles, sept au total. Leur étendue géographique est toutefois beaucoup plus restreinte en comparaison de ceux de la zone nord-ouest. D'ouest en est, ont ainsi été identifiés :

- le pôle *Aven* (entre les cours d'eau Odet et Ellé, légèrement au sud-est de Quimper)
- le pôle Sud Poher (légèrement au sud de la ville de Carhaix)
- le pôle *plinn* (à l'est de Carhaix)
- le pôle pays de Lorient
- le pôle pays de Pontivy
- le pôle pays de Carnac
- le pôle *gevred* (immédiatement à l'est de la ville de Vannes)

	Ensemble de la Basse-Bretagne	zone nord-ouest	pôle <i>gwalarn</i>	pôle <i>mervent</i>	pôle <i>bis</i>
nombre de segments	534	323	101	55	90
proportion par rapport à l'ensemble des segments	100 %	60,48 %	18,91 %	10,3 %	16,85 %
moyenne	68,43 %	73,9 %	77,95 %	68,22 %	77,44 %
médiane	68,27 %	74,1 %	78,39 %	69,11 %	77,66 %
maximum	88,47 %	88,47 %	87,19 %	75,86 %	88,47 %
3 ^e quartile	75,63 %	78,76 %	81,58 %	70,67 %	81,21 %
1 ^{er} quartile	62,83 %	69,51 %	75,15 %	65,38 %	74,2 %
minimum	39,15 %	51,58 %	66,68 %	51,57 %	66,69 %
étendue	49,32	36,89	20,51	24,28	21,78
écart-type	9,52	6,54	4,65	4,16	4,92
coefficient de variation	13,91	8,85	5,97	6,09	6,35

Tableau 3. Comparaison des statistiques descriptives pour les différents espaces observés

En raison de la taille limitée de ces pôles et du faible nombre de segments qu'ils contiennent, nous n'avons pas établi de statistiques descriptives, la représentativité de ces dernières aurait été limitée dans ce cas.

Les pôles de similarité linguistique au sein de la zone sud-est se caractérisent en premier lieu par leur faible extension spatiale. En outre, ils affichent des niveaux de similarité linguistique nettement plus faibles que ceux des pôles de la zone nord-ouest. Malgré tout, ces valeurs permettent de contraster les pôles avec les segments qui les entourent qui, eux, exhibent des taux de similarité linguistique encore plus faibles.

Synthèse

La distribution de la distance linguistique se structure donc à plusieurs niveaux. Ces différentes couches se cumulent sans s'exclure. On observe donc en premier lieu une opposition entre une zone nord-ouest et une zone sud-est. Chaque partie se distingue par un comportement différent de la distance linguistique comme nous l'avons montré. De surcroît, au sein de chacun de ces espaces se dessinent des pôles de similarité dont les particularités varient en fonction de la zone à laquelle ils appartiennent. Ainsi, dans la zone nord-ouest les taux de similarité linguistique sont plutôt élevés et stables d'un segment à l'autre. Les trois pôles de similarité linguistique qui y ont été identifiés se caractérisent par une emprise spatiale relativement étendue. Au contraire, dans la zone sud-est, les taux de similarité sont plus bas et varient beaucoup d'un segment à l'autre, produisant un effet général d'hétérogénéité. Les pôles de similarité dans cet espace sont certes plus nombreux (7) mais leur extension s'avère nettement restreinte. De plus, ces pôles présentent des taux de similarité assez élevés mais pas aussi importants que dans la zone nord-ouest. Toutefois, les valeurs relevées de similarité linguistique relevées au sein de ces pôles contrastent avec celles des segments environnants, permettant ainsi de dessiner les contours de ces pôles. Le tableau suivant détaille les différents niveaux d'organisation de la distance linguistique à travers l'espace :

niveau 1	domaine bretonnant	
niveau 2	zone nord-ouest	zone sud-est
niveau 3	3 pôles	7 pôles

Tableau 4. Organisation schématique de l'espace bretonnant du point de vue de la distance linguistique

La distance linguistique est donc organisée sur le plan spatial en deux niveaux distincts mais néanmoins imbriqués l'un dans l'autre, produisant de ce fait un effet de cumul. L'opposition principale repérée à la suite de l'analyse dialectométrique est sans doute à mettre en relation avec les observations formulées par plusieurs auteurs concernant une organisation dialectale ancienne de l'espace bretonnant, antérieure à la structure tripartite actuelle, que ce soit une « dualité dialectale primitive » pour Falc'hun ou bien un espace divisé en deux occupé au septentrion par « *a fundamentally northern dialect* » contrastant avec un espace sudiste pour Jackson³¹. Hélas, aucun de ces auteurs n'a développé sa vision de cette organisation primitive de l'espace dialectal breton. La démarcation précise entre les deux zones repérées par inspection visuelle pourrait être sans aucun doute discutée et son tracé être affiné. À nos yeux, elle correspond à une interface entre deux ensembles et la position actuelle, telle que nous l'avons reconnue, résulte d'une dynamique historique longue dont la chronologie reste à établir dans la mesure du possible.

Il convient à présent de déterminer dans quelle mesure cette organisation de la distance linguistique obéit à des facteurs géographiques et si tel est le cas, leur nature est à préciser.

Quelle est l'incidence des facteurs géographiques sur la distribution de la distance linguistique du breton ?

L'influence de la géographie sur la distribution des faits de langues, que ce soit à travers le cours des rivières ou la présence d'obstacles topographiques à la circulation des êtres humains et des marchandises,

31. François FALC'HUN, *Perspectives nouvelles sur l'histoire de la langue bretonne*, Paris, Union Générale d'Éditions, 1981, pp. 227-231, 525-526. Kenneth H. JACKSON, *A Historical Phonology of Breton*, Dublin, The Dublin Institute for Advanced Studies, 1967, p. 34.

est un phénomène qui a été identifié de longue date et ce, dans différents contextes linguistiques. Cependant, la principale difficulté que rencontre la réflexion géolinguistique dans son analyse des facteurs environnementaux est que le rôle joué par un élément géographique, et son incidence sur les faits linguistiques n'est pas systématique. Par exemple, si certains cours d'eau assurent clairement un rôle démarcateur entre plusieurs formes, tous ne le font pas. Il est dès lors délicat de repérer clairement des relations de cause à effet et de construire en conséquence une modélisation pertinente.

Une influence attendue

L'influence de la géographie sur la distance linguistique nous est apparue dans un travail exploratoire limité au Centre Bretagne à partir des données de 23 points d'enquête du NALBB³². Les premiers résultats obtenus permettent de repérer un lien entre des accidents topographiques et une diminution du niveau de la similarité linguistique comme cela peut s'observer à la figure 7.

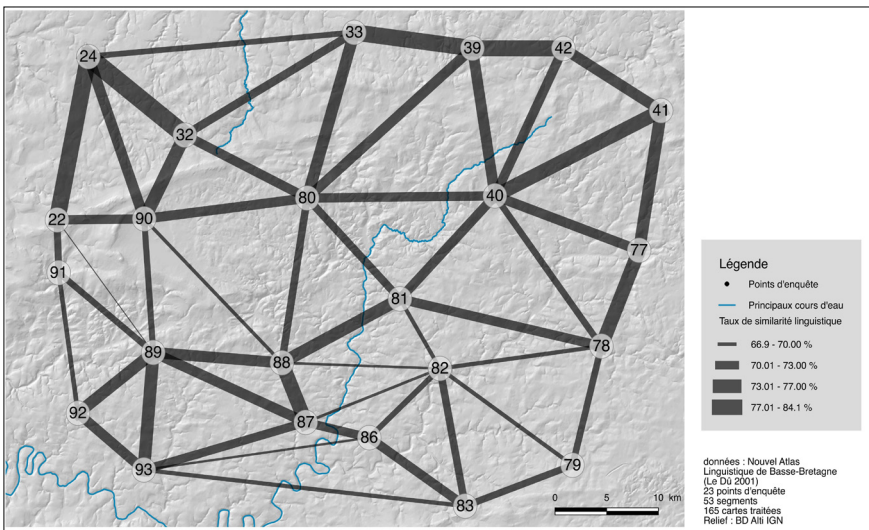


Figure 7. Distribution de la distance linguistique en Centre Bretagne

32. Tanguy SOLLIEC, «Unity in diversity? A fine-grained approach to the linguistic geography of Breton by the means of dialectometry», *Studia Celto-Slavica. Journal of the Learned Association Societas Celto-Slavica*, t. 10, 2019, p. 137-173.

Ainsi, plusieurs espaces, indiqués à la figure 8 par des ensembles grisés clairs, se caractérisent par un taux de similarité linguistique inférieur à la moyenne pour la zone étudiée (74,8 %). Dans la plupart des cas, une correspondance peut être établie avec des particularités topographiques.

Deux cours d'eau y jouent un rôle démarcateur, le Queffleuth dans la partie supérieure de la carte, qui servait également de limite à cet endroit entre les anciens évêchés pré-révolutionnaires du Léon et du Trégor. Dans la partie basse, on observe qu'au cours de l'Aulne, caractérisé par ses nombreux méandres, correspond également une baisse de la similarité linguistique, particulièrement dans sa partie inférieure. En outre, la partie centrale de la carte coïncide avec une cuvette marécageuse dite Yeun Elez³³, réputée pour être l'une des entrées de l'Enfer selon la tradition.

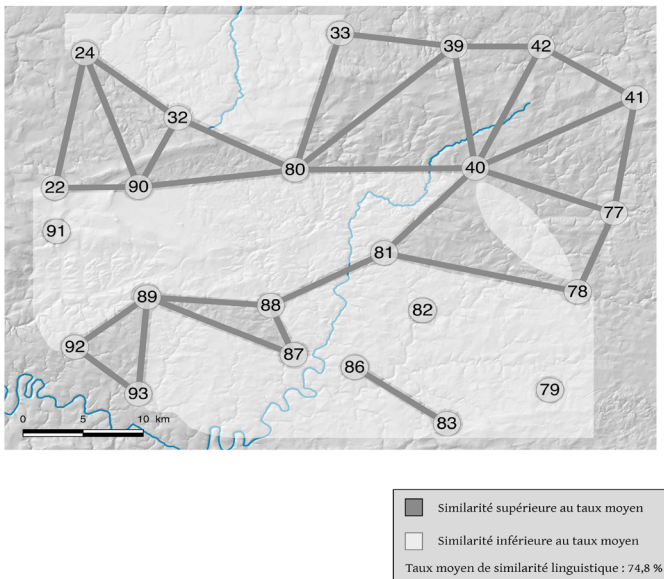


Figure 8. Espaces de divergence linguistique en Centre Bretagne du point de vue de la distance linguistique

33. [jyn] ou [jyn e'les], le marais de l'Elez d'après le nom d'un cours d'eau traversant cet espace.

De ce premier aperçu, on peut s'attendre à ce que dans le cadre d'une étude étendue à l'ensemble du territoire bretonnant, la topographie et les accidents de relief locaux aient une incidence importante sur la distribution de la distance linguistique. Nous évaluerons ce point par la suite

Un modèle sous-jacent ?

Il est intéressant d'éclairer nos constats à la lumière des problématiques actuelles sur la compréhension de la répartition de la diversité linguistique à l'échelle du globe et de réfléchir à la palette des facteurs géographiques susceptibles d'intervenir dans la distribution des langues dans l'espace pour les croiser avec les observations de la dialectologie. Il faut toutefois garder à l'esprit que ces différents facteurs s'associent et forment sans aucun doute un modèle complexe. En effet, il est sans doute naïf de rechercher un facteur unique susceptible d'expliquer à lui seul la distribution de la distance linguistique à travers le domaine bretonnant.

C'est donc un modèle plurifactoriel qu'il est souhaitable d'envisager et d'évaluer de la sorte, le poids qu'y tient chaque paramètre. Une telle approche permet en outre de considérer que l'association de différents facteurs contribue à modéliser la répartition de la distance linguistique.

Quels paramètres prendre en compte ?

Une revue de travaux portant sur la dialectologie, la typologie ou examinant les dynamiques à l'œuvre dans l'organisation de la diversité des langues permet de dresser une liste de facteurs susceptibles d'intervenir dans la structuration spatiale de la distance linguistique³⁴. Dans le cas présent, nous faisons fi de la distinction entre langue et dialecte, et nous considérons ces critères du point de vue de leur influence sur la diversité linguistique, que celle-ci s'observe au sein des langues ou entre elles. Nous considérons que les mêmes processus de différenciation sont à l'œuvre aussi bien entre les langues qu'entre les variétés d'une même langue, seul le niveau d'observation retenu change. Nous posons

34. Voir notamment Harald HAMMARSTRÖM, «Linguistic diversity and language evolution», *Journal of Language Evolution*, t. 1/1, 2016, p. 19-29. Søren WICHMANN, «The Emerging Field of Language Dynamics», *Language and Linguistics Compass*, t. 2/3, 2008, p. 442-455.

facteurs explicatifs	données à corrélér
distance euclidienne (ou spatiale)	distance à vol d'oiseau en kilomètres entre les différents segments progression linéaire selon Séguy ³⁵ <i>Fundamental Dialectology Principle</i> ³⁶
structure dialectale	carte de la structure dialectale de la langue bretonne ³⁷
distribution des innovations linguistiques	réseau routier ancien ³⁸ présence de centres urbains ³⁹
temps de trajet	temps mis pour effectuer le trajet entre les différents points d'enquête constituant les segments étudiés (calcul à partir des temps de trajets actuels) ⁴⁰
facteurs de géographie physique - altitude - milieu côtier - rivières - forêts - obstacle naturel (mer, marais...)	altitude moyenne des points d'enquête, carte de relief (courbes de niveau) ⁴¹ distance de la côte présence de cours d'eau séparateurs grands massifs forestiers (forêt de Quénécan, Coat Quéau, Cragou...) cf. temps de trajet et organisation du réseau routier
densité de la population	nombre d'habitants au km ² par commune dans lesquels sont situés les points d'enquête taille de la population vers 1850/1900 par point d'enquête

Tableau 5. Liste des critères susceptibles de structurer la distribution de la distance linguistique

35. Jean SEGUY, « La relation entre la distance spatiale et la distance lexicale », *Revue de linguistique romane*, t. 35/139-140, 1971, p. 335-57. John NERBONNE, « Measuring the diffusion of linguistic change », *Philosophical Transactions of the Royal Society B: Biological Sciences*, t. 365/1559, 2010, p. 3 821-3 828.

36. J. NERBONNE, P. KLEIWEG, art. cit., p. 154.

37. À titre de comparaison, ce paramètre a été pris en compte dans des travaux sur le néerlandais, l'anglais dialectal d'Angleterre et l'allemand.

38. François FALC'HUN, *L'histoire de la langue bretonne d'après la géographie linguistique*, Quimper, chez l'abbé Nédélec, 1951. François FALC'HUN, *Histoire de la langue bretonne d'après la géographie linguistique*, Paris, PUF, 1965. F. FALC'HUN, *op. cit.*

39. Peter TRUDGILL, « Linguistic change and diffusion : description and explanation in sociolinguistic dialect geography », *Language in Society*, t. 3/2, 1974, p. 215-246.

40. Pour le norvégien : Charlotte GOOSKENS, « Travel time as a predictor of linguistic distance », *Dialectologia et Geolinguistica*, t. 13, 2005, p. 38-62. Pour le sui, langue de Chine : James N. STANFORD, « One size fits all? Dialectometry in a small clan-based indigenous society », *Language Variation and Change*, t. 24/02, 2012, p. 247-278.

41. J. NICHOLS, *op. cit.*, 1992. Johanna NICHOLS, « Modeling Ancient Population Structures and Movement in Linguistics », *Annual Review of Anthropology*, t. 26/1, 1997, p. 359-384. J. NICHOLS, art. cit., 2013.

comme principe que si un facteur favorise la diversité linguistique, il contribue dès lors à la progression de la distance linguistique, tant que l'inverse n'a pas été démontré⁴². Le tableau précédent présente un faisceau de paramètres à explorer afin de déterminer leur incidence sur la structuration de la distance linguistique dans le domaine bretonnant ainsi que les données à corrélérer aux résultats de similarité linguistique obtenus dans le cadre de l'analyse dialectométrique.

Dans certaines situations, il peut être impossible de distinguer l'incidence respective des différents facteurs à l'œuvre. Ainsi, un effet de cumul peut intervenir dans le cas d'une zone périphérique, mal connectée aux centres urbains et caractérisée par des accidents géographiques importants, ce qui a pour conséquence des temps de trajet conséquents.

Examen des différents facteurs

Distance euclidienne

La relation entre la distance euclidienne (ou spatiale) et la distance linguistique est considérée comme étant particulièrement privilégiée, la seconde étant perçue comme étant fonction de la première. Ce constat à amener certains auteurs à formuler un principe, le *Fundamental Dialectology Principle* (désormais FDP), qui énonce que « *geographically proximate varieties tend to be more similar than distant ones* »⁴³. Cette analyse est un avatar de plusieurs observations tant en géographie, la loi d'autocorrélation spatiale de Tobler ou en génétique des populations, sous l'appellation d'« isolation par distance »⁴⁴.

42. Il faut relever qu'un mouvement de différenciation linguistique peut être contrebalancé dans certains contextes par un mouvement de convergence entre les langues ou variétés linguistiques en présence, notamment dans le cas des aires linguistiques.

43. J. NERBONNE, P. KLEIWEG, art. cit., p. 154.

44. Waldo R. TOBLER, « A Computer Movie Simulating Urban Growth in the Detroit Region », *Economic Geography*, t. 46, 1970, 234-240. Sewall WRIGHT, « Isolation by distance », *Genetics*, t. 28, 1943, 114-138. Tobler (1970), (Wright 1943).

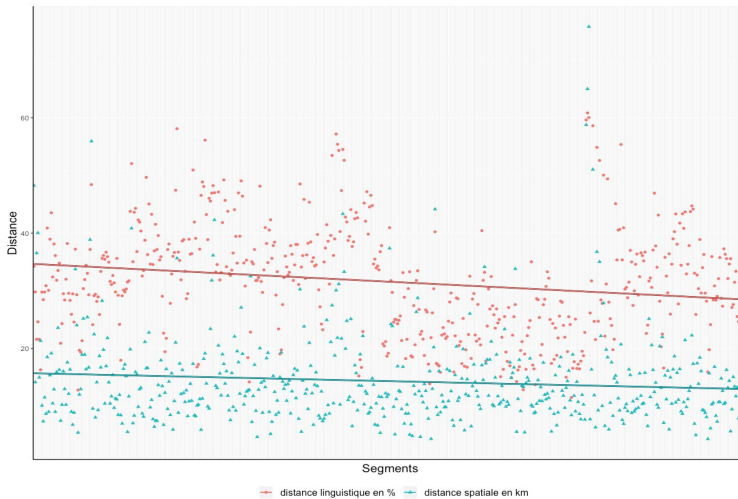


Figure 9. Répartition simultanée des distances linguistique et spatiale par segment

L'évaluation de la corrélation entre les deux variables grâce à la méthode de Pearson (voir figure 9), aboutit au résultat suivant :

$$r = 0,4394714 \text{ (n = 534), } p\text{-value} < 0,01.$$

Dans le cas d'une corrélation parfaite, le coefficient est de 1 ou de -1. Plus il tend à se rapprocher de 0, plus la probabilité d'une corrélation est faible. Un résultat compris entre 0,3 et 0,7 signale une corrélation modérée. Il y a donc une corrélation entre les deux variables observées mais celle-ci se révèle faible.

Le lien entre distance linguistique et distance spatiale étant posé, nous pouvons à présent observer comment la première se déploie dans l'espace. Séguy a identifié à partir de ses travaux sur la distance lexicale dans l'espace roman que la distance linguistique suit une progression sous-linéaire dans l'espace en suivant la formule suivante :

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

Autrement dit, la distance lexicale augmente nettement dans un premier temps au fur et à mesure de la progression dans l'espace⁴⁵. Puis à partir d'un certain seuil, elle se stabilise et n'augmente guère. Nerbonne dans une démarche comparative a constaté qu'une telle relation s'observait

45. J. SÉGUY, *art. cit.*, 1971.

dans le cas de la distance linguistique dans son ensemble d'une part et d'autre part que cette progression sous-linéaire s'observe dans différents contextes linguistiques et a nommé cette progression de la distance linguistique « loi de Séguéy⁴⁶ ».

Dans le cas du breton, on assiste une situation semblable comme le montre la figure 10.

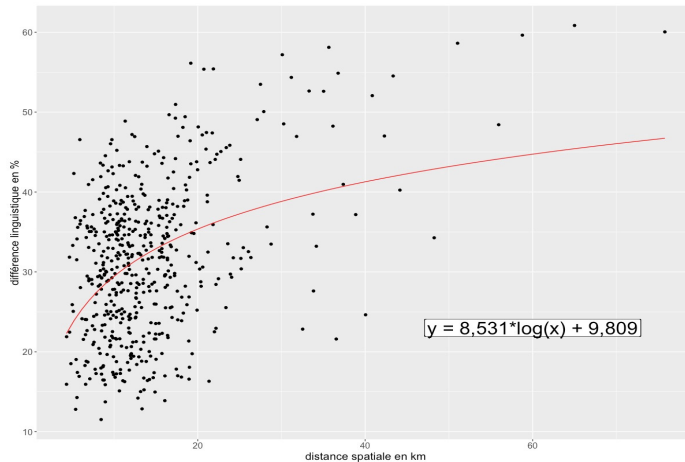


Figure 10. **Progression logarithmique de la différence linguistique en fonction de la distance euclidienne**

La distance linguistique ici aussi est sous-linéaire, elle ne progresse pas aussi vite que la distance spatiale. Dans l'analyse que nous avons retenue, interponctuelle, les différents points d'enquête sont comparés de façon privilégiée à leurs voisins immédiats. Ceci explique pourquoi la majorité des valeurs se concentre entre 4 et 20 kilomètres. La similarité linguistique tend à décroître de façon marquée entre 4 et 35 kms (passage de 78 à 60 %). Le rythme de cette diminution s'atténue au-delà, de 37 à 78 kms, la similarité passe ainsi de 60 à 43,15 %.

Les résultats obtenus se conforment donc à ce qui s'observe dans d'autres domaines linguistiques. Néanmoins, la corrélation entre la distance linguistique et la distance spatiale dans le cas du breton demeure relativement faible, ce qui suggère que d'autres facteurs sont à l'œuvre, et que la place accordée à la distance euclidienne comme facteur explicatif de la distribution de la distance linguistique est peut-être à réévaluer.

46. J. NERBONNE, *art. cit.*, 2010.

Structure dialectale

Le premier facteur qui vient à l'esprit pour expliquer la répartition de la distance linguistique est celui de la structure dialectale du breton. Toutefois, la situation n'est pas aussi simple qu'il le semble. Une première analyse basée sur les variétés du Centre-Bretagne nous a amené à faire le constat que d'autres perspectives étaient à envisager pour appréhender la structuration de la distance linguistique dans l'espace⁴⁷. La prise en compte du territoire bretonnant dans son ensemble confirme ce constat. En effet, la confrontation de la répartition de la distance linguistique avec deux modèles de découpage dialectal aboutit à une absence de correspondance.

L'ancienne vision en vigueur jusqu'au milieu du XX^e siècle, à présent dépassée mais encore largement répandue, reconnaît un dialecte pour chacun des anciens évêchés antérieurs à la Révolution française partageant l'espace bretonnant⁴⁸. La figure 11 présente une telle répartition.

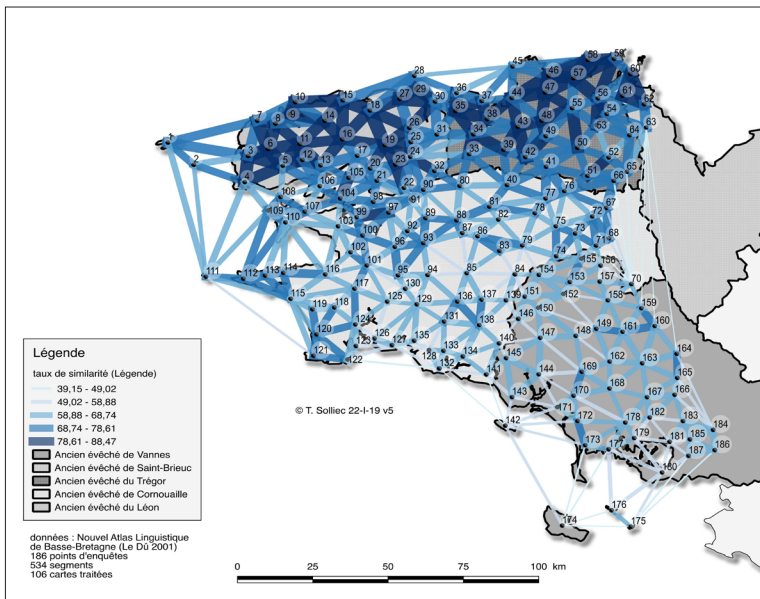


Figure 11. Distribution de la similarité linguistique en relation avec les limites des évêchés pré-révolutionnaires

47. T. SOLLIEC, *art. cit.*

48. Johannes HEINECKE, « 11. Bretonisch », dans T. Roelcke (dir.), *Variationstypologie / Variation Typology. Ein sprachtypologisches Handbuch der europäischen Sprachen in*

L'analyse des données de l'ALBB par Falc'hun a renouvelé la dialectologie du breton et a montré que son espace dialectal se structure de façon tripartite autour d'une vaste bande centrale selon un axe sud-ouest/nord-est, courant de Lannion au sud de Quimper⁴⁹. De nombreuses innovations linguistiques y ont émergé au cours des siècles et s'y sont largement diffusées. On identifie de part et d'autre de cette diagonale deux zones, une au sud-est correspondant à l'ancien Vannetais et l'autre au nord-ouest englobant le Léon, la presqu'île de Crozon et le cap Sizun. Du point de vue linguistique, ces deux espaces se montrent plus conservateurs et moins enclins à l'adoption des nouveautés provenant de l'espace central. La superposition d'un tel découpage avec les résultats de similarité linguistique ne s'avère donc pas plus concluante (voir figure 12).

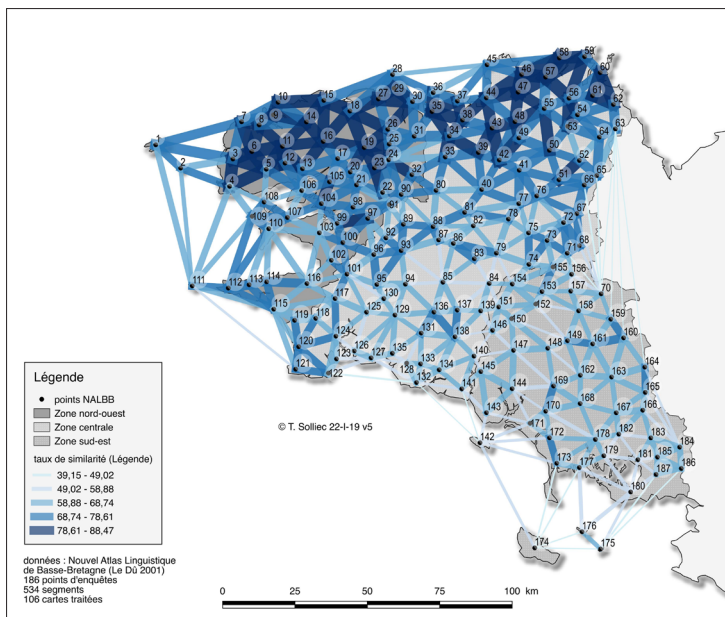


Figure 12. Distribution de la similarité linguistique en relation avec la structuration tripartite de l'espace dialectal bretonnant

Geschichte und Gegenwart / A Typological Handbook of European Languages, Berlin, Walter de Gruyter, 2003, p. 308-324. Fañch BROUDIC, « Bretagne » dans U. Reutner (dir.), *Manuel des Francophonies*, Berlin, De Gruyter, 2017, p. 149-168. Jean SELLIER, *Une histoire des langues et des peuples qui les parlent*, Paris, La Découverte, 2019.

49. F. FALC'HUN, *op. cit.*, 1951. F. FALC'HUN, *op. cit.*, 1965. F. FALC'HUN, *op. cit.*, 1981.

Dans le cas du breton, une disjonction se manifeste donc entre le signal enregistré par la distance linguistique dans le cadre d'une analyse dialectométrique et la structure dialectale de cette langue.

Diffusion des innovations linguistiques

Falc'hun a également constaté que dans le domaine bretonnant, nombre d'innovations linguistiques se sont diffusées dans l'espace central grâce au système routier hérité de l'Antiquité et construit en étoile autour de la ville de Carhaix, l'ancienne Vorgium. La superposition du réseau viarie et de la distribution de la distance linguistique (figure 13) montre une nouvelle fois qu'aucune correspondance ne peut être établie, pas plus que les pôles urbains ne semblent avoir d'influence sur la répartition de la distance linguistique. En effet, la figure 14 indique que la présence d'une ville ne pèse aucunement sur la distribution. Aucune baisse ou augmentation des taux de similarité linguistique ne se dessinent dans le pourtour des principales villes de l'espace bretonnant.

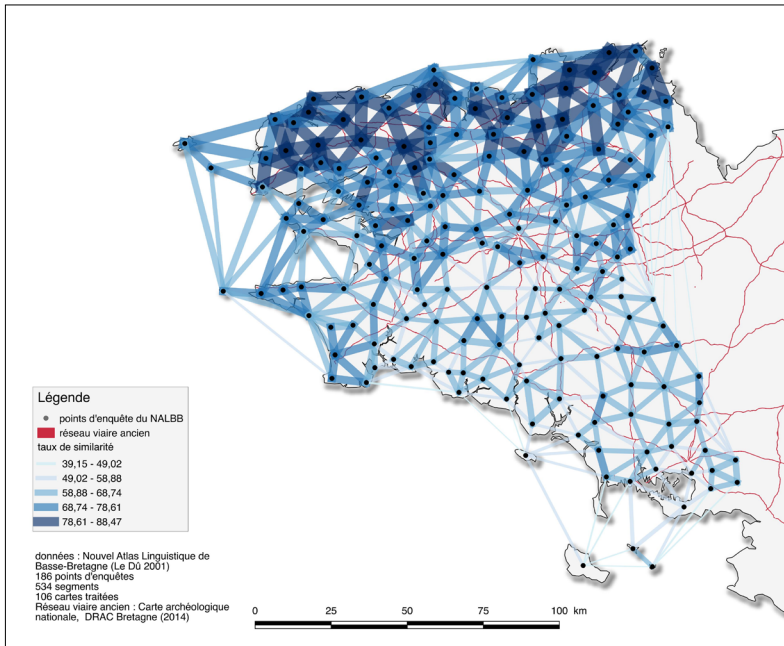


Figure 13. Distribution linguistique et réseau routier ancien, d'avant la période moderne

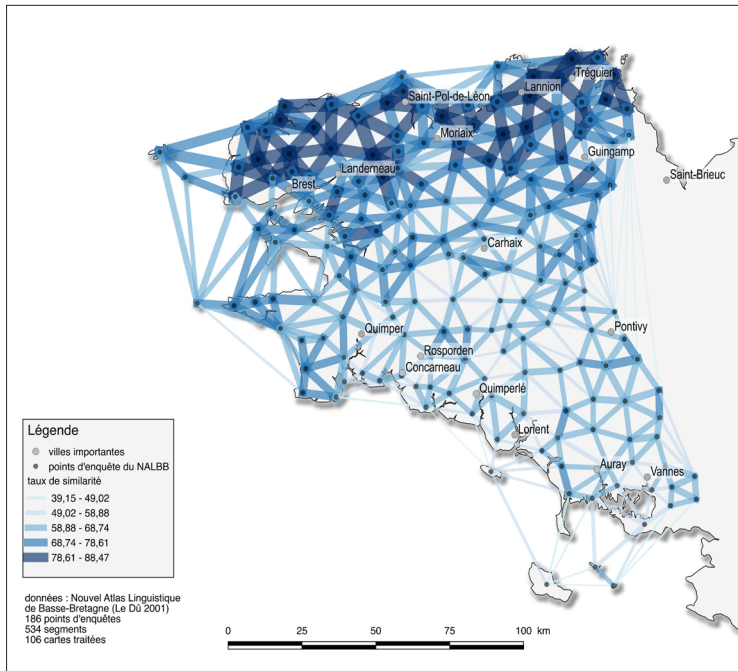


Figure 14. Distribution de la similarité linguistique et situation des villes importantes de Basse-Bretagne

À nouveau, on ne peut que constater l'absence de correspondance entre la distribution de la distance linguistique et les voies de diffusion des innovations linguistiques.

Temps de trajet

Plusieurs travaux ont observé que le temps de trajet entre divers points est fréquemment corrélé à la similarité linguistique, plus le trajet est long, plus la distance linguistique sera importante. Dans le cas du domaine étudié, nous ne disposons pas de données historiques permettant d'établir le temps nécessaire pour se rendre d'un point d'enquête du NALBB à l'autre. Nous nous sommes donc tournés vers un service de calcul de temps de trajet en voiture, ayant bien conscience de l'anachronisme. Nous considérons néanmoins que l'utilisation de moyens de transport moderne permet néanmoins de conserver un ordre de grandeur qui se reflètera dans les résultats obtenus. La confrontation des deux variables aboutit à la figure 15.

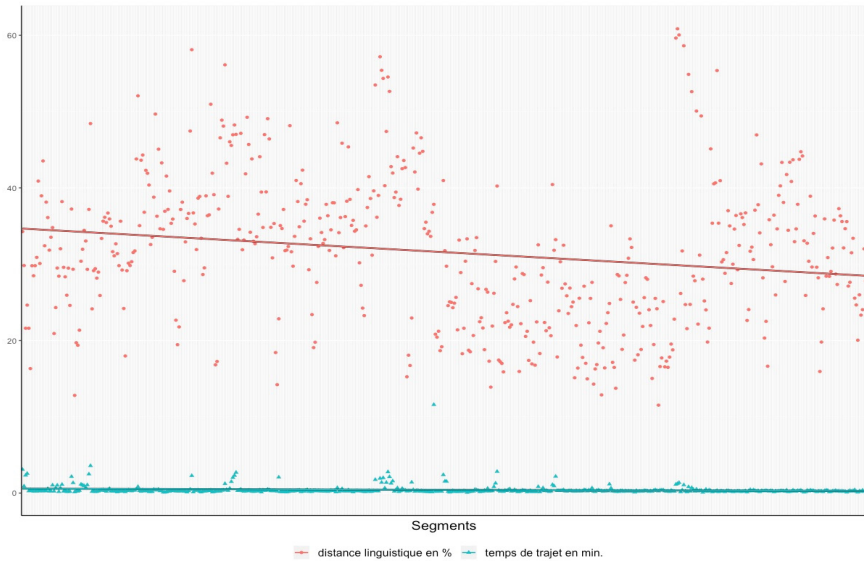


Figure 15. Répartition simultanée de la distance linguistique et du temps de trajet pour chaque segment

La possibilité d'une corrélation entre ces deux variables peut être calculée à partir de leur droite de régression linéaire respective. L'évaluation du lien de corrélation grâce à la méthode de Pearson aboutit au taux suivant :

$$r = 0,3347878 \text{ (n = 534), } p \text{ value} < 0,01$$

Un tel résultat permet d'écarter l'hypothèse nulle d'une absence de corrélation entre les deux variables examinées mais ce lien s'avère faible ici aussi.

Si la corrélation est bien présente, cette association n'en demeure pas moins relativement limitée, ne permettant pas de faire du temps de trajet un critère explicatif majeur pour expliquer la répartition de la distance linguistique dans le domaine bretonnant, à la différence des observations de Gooskens pour le norvégien ou de Szmrecsanyi pour la morphosyntaxe de l'anglais dialectal britannique⁵⁰. Le taux de corrélation entre la distance linguistique et le temps de trajet est ici plus faible que lorsque celle-ci est mise en relation avec la distance spatiale.

50. C. GOOSKENS, *art. cit.* Benedikt SZMRECSANYI, *Grammatical Variation in British English Dialects: A Study in Corpus-Based Dialectometry*, Cambridge, Cambridge University Press, 2013, p. 104.

Les temps de trajet entre les points d'enquête du NALBB sont relativement faibles. S'ils varient ainsi entre 5 minutes et 3 heures 35, la moyenne s'élève à un peu plus de 24 minutes⁵¹. Il est intéressant de s'attarder sur les temps de trajet les plus longs dont nous disposons afin d'observer si l'isolation aboutit à des taux de similarité importants.

Segment	% de similarité linguistique	% de différence linguistique	distance en km	temps de trajet en heure ⁵²
111 – 121	51,58	48,42	55,94	3,58
1 – 111	65,73	34,27	48,24	3,09
28 – 45	59,76	40,24	44,15	2,83
175 – 186	45,48	54,52	43,34	2,77
142 – 174	52,99	47,01	42,3	2,71
1 – 7	75,36	24,64	40,05	2,56
111 – 115	62,83	37,17	38,87	2,49
2 – 111	59,04	40,96	37,38	2,39
1 – 3	78,4	21,6	36,54	2,34
[...]				
20 – 21	74,94	25,06	5,13	0,12
53 – 54	82,58	17,42	5,62	0,12
86 – 87	66,94	33,06	6,97	0,12
95 – 130	64,42	35,58	5,66	0,12
97 – 99	79,97	20,03	14,21	0,12
128 – 132	65,44	34,56	7,76	0,12
128 – 133	63,07	36,93	6,45	0,12
68 – 71	71,96	28,04	5,99	0,10
22 – 91	66,69	33,31	5,07	0,08
En moyenne	68,43	31,57	14,34	0,41

Tableau 6. Valeurs extrêmes de temps de trajet en lien avec la distance linguistique

Un temps de trajet important contribue donc à accentuer la distance linguistique mais cela ne s'avère toutefois pas systématique comme le

51. Il faut garder à l'esprit que ces valeurs correspondent à des temps de déplacement actuels dans un véhicule motorisé.

52. Les valeurs données ici ont été calculées sur une base 100 et non 60 de façon à permettre la comparaison avec les taux de distance linguistique.

montre le cas de Ouessant (point 1 du NALBB). Ce point d'enquête quand il se trouve en relation avec d'autres localités appartenant au pôle *gwalarn* affiche un taux de similarité élevé. Un tel comportement tend donc à relativiser l'isolation comme facteur explicatif et souligne également la logique de pôles à l'œuvre dans la répartition de la distance linguistique. À l'inverse, un temps de trajet court ne correspond pas forcément à un taux de similarité élevé comme cela peut se voir dans le cas du segment entre les points (95) Briec et (130) Landudal.

L'examen des données ci-dessus montre que dans le cas du breton, le temps de trajet entre les points d'enquête du NALBB ne contribue que faiblement à expliquer la répartition de la distance linguistique au sein de cet espace. L'appartenance d'une localité à un pôle tend à favoriser des taux de similarité linguistique importants avec celles qui en font également partie.

Géographie physique

Il est généralement admis dans les études sur la dialectologie des langues que les accidents de la géographie physique (cours d'eau, marais, relief, etc.) concourent à façonner la répartition des faits linguistiques. La prise en compte des cours d'eau principaux et du relief dans l'ouest de la Bretagne permet d'élaborer la figure 16. Les facteurs climatiques

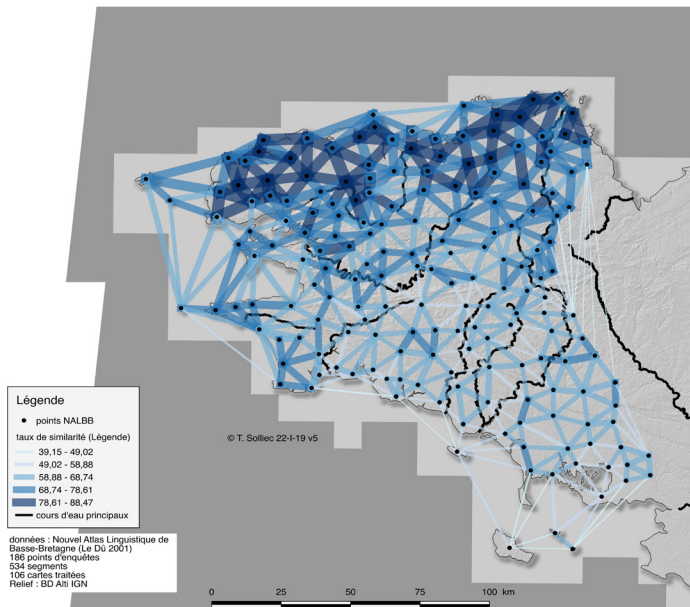


Figure 16.
Distribution
de la similarité
linguistique en
relation avec
la géographie
physique
de l'espace
bretonnant

comme la pluviométrie ou encore la température moyenne tout comme la distance par rapport à la mer n'ont délibérément pas été pris en compte.

Aucune relation tangible ne se dégage à première vue entre les éléments de géographie physique et la répartition de la distance linguistique. Toutefois, une lecture plus attentive permet de repérer de façon ponctuelle des effets de rupture ou de barrière entre certains pôles,

Un relevé des caractéristiques géographiques locales, pôle par pôle permet d'identifier qui introduisent des ruptures dans la répartition de la similarité linguistique. L'inventaire de ces éléments figure au tableau 7.

	cours d'eau	relief	végétation
pôle <i>gwalarn</i>	+ (Aulne, Queffleuth/ rivière de Morlaix)	+ (<i>Yeun Elez</i>)	–
pôle <i>bis</i>	+ (Queffleuth/rivière de Morlaix, Leff/Trieux comme limite interne)	–	–
pôle <i>mervent</i>	+ (Aulne, Steïr, Odet)	–	–
pôle <i>Aven</i>	–	–	–
pôle sud Poher	+ (Aulne)	+ (Montagnes Noires)	–
pôle <i>plinn</i>	–	–	+ (forêt de Quénécan, couvert boisé au sud du pôle)
pôle pays de Lorient	+ (Ellé/Laïta, Scorff)	–	–
pôle pays de Pontivy	–	–	+ (forêt de Quénécan)
pôle pays de Carnac	+ (Blavet, rivière d'Auray)	+ (landes de Lanvaux)	+ (landes de Lanvaux, forêt de Camors)
pôle <i>Gevred</i>	–	–	–
Totaux	6 occurrences sur 10	3 occurrences sur 10	3 occurrences sur 10

Tableau 7. Ensemble des éléments de géographie physique jouant un rôle de barrière dans la distribution de la distance linguistique au sein de l'espace bretonnant

Ainsi, sur les 10 pôles de similarité linguistique identifiés, dans 8 cas, la géographie physique contribue à la répartition de la distance linguistique. D'autre part, le rôle joué par les cours d'eau peut changer d'un pôle à l'autre, étant tantôt vecteur, tantôt barrière. En outre, on constate la coïncidence dans le nord de la Bretagne entre des plateaux géologiques étendus et des taux de similarité élevés dans les pôles *gwalarn* et *bis*. Les massifs forestiers relevés jouent eux un rôle de démarcateur entre différents pôles.

L'influence de la géographie physique sur la répartition de la distance linguistique dans le cas du breton quand elle est présente, opère avant tout localement. L'absence d'effet systématique par les éléments de topographie rend complexe toute modélisation. À l'encontre de notre intuition initiale, la géographie physique n'intervient que localement dans l'organisation de la répartition de la distance linguistique.

Densité de la population

La question d'un lien entre le nombre de locuteurs d'une langue et les caractéristiques structurelles qu'elle présente est débattue depuis la fin des années 2000 par les typologues⁵³. Il a ainsi été proposé que dans le cas d'une population nombreuse, la propension des langues à évoluer rapidement est plus marquée. De ce fait, ce facteur démographique est également susceptible de contribuer à façonner la répartition de la distance linguistique dans l'espace. Les travaux de dialectométrie ayant

53. Eric W. HOLMAN, Søren WICHMANN, *Temporal stability of linguistic typological features*. Munich, Lincom Europa, 2009. Joan BYBEE, «How plausible is the hypothesis that population size and dispersal are related to phoneme inventory size? Introducing and commenting on a debate», *Linguistic Typology*, t. 15/2, 2011, p. 147-53. Mark DONOHUE, Johanna NICHOLS, «Does phoneme inventory size correlate with population size?», *Linguistic Typology*, t. 15/2, 2011, p. 161-170. Steven MORAN, Daniel MCCLOY, Richard WRIGHT, «Revisiting population size vs. phoneme inventory size», *Language*, t. 88/4, 2012, p. 877-893. Simon J. Greenhill, Xia Hua, Caela F. Welsh, Hilde Schneemann, Lindell Bromham, «Population Size and the Rate of Language Evolution: A Test Across Indo-European, Austronesian, and Bantu Languages», *Frontiers in Psychology*, n° 9, 2018, p. 576 [en ligne]. Mis en ligne le 27 avril 2018, URL : <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00576/full> (consulté le 26 juin 2019).

pris en compte le critère démographique ont toutefois abouti à des résultats mitigés⁵⁴.

La prise en compte de la taille de la population à différentes périodes des époques moderne et contemporaine nous a permis d'établir un taux moyen de densité au kilomètre carré pour les points d'enquête du NALBB⁵⁵. Ce premier pas indique une densité moyenne de 73,34 habitants par km². Mais ces valeurs présentent également des disparités importantes entre l'intérieur des terres d'une part et le littoral et les îles de l'autre. Ainsi, Saint-Rivoal (point 91 du NALBB) au cœur des Monts d'Arrée présente une densité de 30,98 habitant au km² alors qu'à l'Île-de-Sein (point 111), on relève un taux de 1282,33 habitants au km²⁵⁶.

À partir des travaux de Trudgill⁵⁷ prenant en compte le critère démographique pour évaluer l'importance et le rôle de différents centres urbains dans la diffusion des innovations linguistiques, nous avons établi un indice *i* de densité humaine afin d'évaluer l'équilibre qui pouvait être établi dans la comparaison des points d'enquête du

54. John NERBONNE, Wilbert HEERINGA, «Geographic distributions of linguistic variation reflect dynamics of differentiation», dans S. Featherston, W. Sternefeld (dir.), *Roots: Linguistics in Search of Its Evidential Base*, Berlin, Mouton de Gruyter, 2007, p. 267-297. Benedikt SZMRECSANYI, «Geography is overrated», dans S. Hansen, C. Schwarz, P. Stoeckle, T. Streck (dir.), *Dialectological and Folk Dialectological Concepts of Space. Current Methods and Perspectives in Sociolinguistic Research on Dialect Change*, Berlin, De Gruyter, 2012, p. 215-231. B. SZMRECSANYI, *op. cit.*, Martijn WIELING, Robert G. SHACKLETON, John. NERBONNE, «Analyzing Phonetic Variation in the Traditional English Dialects: Simultaneously Clustering Dialects and Phonetic Features», *Literary and Linguistic Computing*, t. 28/1, 2013, p. 31-41.

55. Nous nous sommes basés pour cela sur les données de la base de données Cassini élaborée à l'EHESS qui détaille le résultat des recensements à partir de la Révolution française (consultable à l'adresse suivante : <http://cassini.ehess.fr/fr/html/index.htm>). Nous avons retenu quatre dates (1793, 1851, 1906 et 1926) afin de garder une perspective historique. Les données pour les points d'enquête ne constituant pas des communes du point de vue administratif sont celles de leur commune d'appartenance dans ce cas. Il en va de même pour les communes ayant disparu à moins que nous ayons été en mesure de préciser la population de ces dernières après fusion (cas des communes associées).

56. Il s'agit d'une projection pour les besoins de la comparaison, la surface de l'île ne faisant que 0,58 km².

57. P. TRUDGILL, *art. cit.*

NALBB selon les segments établis. Celui-ci s'obtient en appliquant la formule suivante :

$$i = \frac{\text{Densité de population dans la localité A}}{(\text{Densité de la population dans la localité A} + \text{Densité de la population dans la localité B})}$$

Une densité identique dans les deux communes observées aboutit à un résultat de 0,5. Lorsque le résultat est compris entre 0,5 et 1, la densité est plus importante dans la localité A mais une valeur entre 0 et 0,5 indique une population plus dense dans la localité B.

Le calcul d'une éventuelle corrélation entre les deux variables montre que ce lien est extrêmement ténu :

$$r = -0,1187992 \text{ (n = 534), } p\text{-value} < 0,01$$

La mise en relation de la progression de la similarité linguistique avec l'indice de densité humaine permet d'observer la façon dont ce critère est susceptible d'influer sur sa distribution comme cela peut se voir à la figure 16.

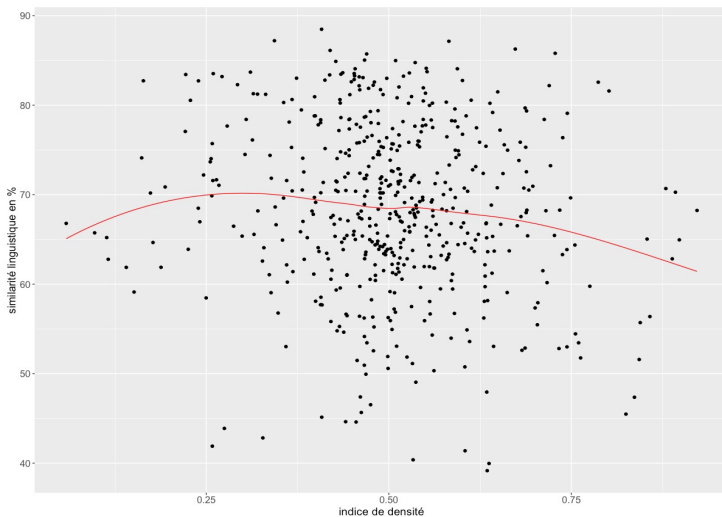


Figure 16. Progression de la similarité linguistique selon l'indice de densité humaine pour les points du NALBB

La tendance de la courbe à s'aplatir au centre du graphique indique que les points d'enquête caractérisés par un taux de densité équivalent tendent à présenter des taux de similarité linguistique assez élevés. Par conséquent, lorsqu'une différence marquée entre le taux de densité des points d'enquête peut être constatée, les taux de similarité linguistique tendent à diminuer. Il semble donc qu'une densité similaire dans les points d'enquête comparés contribue à favoriser des valeurs importantes de similarité linguistique même si cela ne suffit pas à l'expliquer pour autant.

La géographie, un facteur à l'influence variable

L'examen de différents facteurs géographiques susceptibles d'expliquer la répartition de la distance linguistique dans l'espace aboutit à un bilan mitigé dans la mesure où aucun d'entre eux ne permet à lui seul d'expliquer nettement la structuration de cette dernière. Les différents taux de corrélation obtenus se montrent tous faibles, voire très faibles. De plus, aucune correspondance géographique n'a pu être observée de façon claire dans le cas de phénomènes territoriaux. Les différents résultats rassemblés dans le cadre de cette investigation ont été synthétisés dans le tableau 8 ci-dessous :

Paramètres	Corrélation avec la distribution de la distance linguistique et commentaire
distance spatiale	+ (corrélation faible)
aires dialectales	-
structure tripartite	-
modèle épiscopal	-
innovations linguistiques	-
réseau routier ancien	-
distribution des centres urbains importants	-
temps de trajet	+ (corrélation faible)
géographie physique	+ (incidence locale et parfois ponctuelle)
densité humaine	+ (corrélation extrêmement faible)

Tableau 8. Synthèse des divers facteurs géographiques examinés et susceptibles de structurer la distance linguistique

La possibilité pour chacun des paramètres de ce tableau de prédire la distribution de la distance linguistique dans le domaine bretonnant est indiquée ici grâce aux signes + ou -. Aucun facteur ne vient à lui seul expliquer cette répartition. La distance spatiale est la dimension qui explique le mieux cette distribution mais le coefficient de corrélation correspondant demeure faible. Les coefficients relevés pour la corrélation avec le temps de trajet ou avec la densité humaine aboutissent à des résultats encore inférieurs. En outre, on n'observe pas de convergence avec la structure dialectale du breton ni même avec les vecteurs des innovations linguistiques que sont les routes et les villes. La démographie semble avoir une très légère incidence sur la similarité linguistique, celle-ci tend effectivement à augmenter quand les points d'enquête comparés se caractérisent par une densité humaine équivalente.

Certains faits de géographie physique comme le relief ou la présence de cours d'eau peuvent influencer sur la structuration de la distance linguistique mais cela s'observe essentiellement au niveau local et de façon ponctuelle, notamment dans le cas des cours d'eaux. De plus, les plateaux géologiques importants sont associés à des taux de similarité linguistique relativement élevés, particulièrement dans la zone nord-ouest. Les différents paramètres relevés n'agissent pas tous à la même échelle. Leur mise en relation avec les niveaux hiérarchiques repérés précédemment permet d'établir le tableau 9 :

niveau hiérarchique	espaces géographiques		facteurs
niveau 0	domaine bretonnant		typologie et facteurs aréaux (non examinés)
niveau 1	zone nord-ouest	zone sud-est	distance spatiale
niveau 2	3 pôles	7 pôles	topographie densité de la population

Tableau 9. Facteurs à l'œuvre dans la structuration de la distance linguistique aux niveaux 2 et 3

Les facteurs les plus prégnants ont été indiqués pour chaque niveau hiérarchique. Les critères indiqués pour le niveau 1 sont d'ordre général et donnés à titre indicatif. Ils portent avant tout sur les structures de la langue. Les facteurs géographiques interviennent donc aux niveaux 2 et 3. La distance spatiale permet d'expliquer la répartition de la distance linguistique dans les zones nord-ouest et sud-est mais ces deux espaces présentent néanmoins un profil distinct de ce point de vue. Au niveau inférieur, des éléments topographiques contribuent à moduler la distribution de la distance linguistique. Si ces différents facteurs géographiques se cumulent, leur incidence sur la répartition de la distance linguistique demeure toutefois limitée.

Au terme de cet examen, la distance spatiale s'avère être le facteur qui prédit le mieux la distribution de la distance linguistique. Toutefois, le taux de corrélation qu'il présente avec les résultats de similarité linguistique est si faible qu'il ne saurait être retenu comme unique paramètre explicatif. Les autres éléments examinés n'exercent que peu ou pas d'influence dans l'organisation de la distance linguistique. Cet état de fait suggère que, dans le cas du domaine bretonnant, d'autres facteurs interviennent dans la structuration spatiale de la distribution de la distance linguistique. Ce constat montre que la géographie ne contribue que de façon limitée et ponctuelle à l'organisation de l'espace social que la dialectométrie permet d'identifier. Il convient donc de replacer cette dimension dans une perspective plus vaste.

Discussion : ouvrir des perspectives

Les résultats de l'analyse dialectométrique montrent clairement que ceux-ci sont organisés du point de vue spatial. Toutefois, les facteurs géographiques ne contribuent que modérément à la répartition de la distance linguistique dans l'espace. Ce constat paradoxal amène, à partir du cas breton, à reconsidérer la place que tient la géographie dans la structuration des faits linguistiques au sein d'une langue.

La géographie, un rôle à préciser dans l'étude de la variation linguistique

L'observation des différents facteurs géographiques nous a montré que leur influence demeure limitée à l'échelle du domaine bretonnant. Celle-ci s'observe avant tout à un niveau local. De ce fait, il convient

de déterminer avec précision ce que recouvre la notion de géographie lorsqu'elle est convoquée pour expliquer la répartition spatiale des phénomènes linguistiques. En outre, la pertinence des facteurs pris en compte peut varier selon les niveaux observés comme le montre l'exemple de la distance euclidienne ou bien celui de la géographie physique dans le cas du breton. Dès lors, il est important de prendre en compte la géographie dans toute sa complexité au cours de l'analyse des faits de variation linguistique en définissant au mieux les éléments constitutifs de la spatialité (position dans l'espace et dans les réseaux de circulations), en intégrant des données écologiques (climat, végétation, etc.) mais aussi celles relevant de la géographie humaine (type d'habitat, activités professionnelles, densité humaine) afin de dépasser une conception de l'espace limitée à la seule distance euclidienne.

Les observations présentées à la section précédente amènent à se demander dans quelle mesure la situation du breton correspond aux modèles théoriques habituellement retenus pour décrire la variation diatopique. De ce fait, il s'agit de déterminer si le cas breton est une situation courante ou au contraire exceptionnelle. La distance spatiale (ou euclidienne) comme constaté précédemment est ici le paramètre qui prédit le mieux la distribution de la distance linguistique à travers le domaine bretonnant. D'un point de vue théorique, les faits linguistiques sont décrits comme se différenciant dans l'espace de proche en proche, toujours plus au fur et à mesure qu'on s'éloigne d'un point de référence, selon le modèle du continuum dialectal⁵⁸. Ce constat de la variabilité des faits linguistiques dans l'espace a été reformulé par des praticiens de la dialectométrie sous la forme d'un principe, le *Fundamental Dialectology Principle* (FDP) évoqué précédemment. Cette proposition s'inscrit dans le prolongement d'une série de travaux ayant appliqué des analyses dialectométriques à plusieurs ensembles de données dialectales, à travers l'espace néerlandophone⁵⁹ mais également à un niveau plus local en se concentrant sur des variétés de bas-saxon dans le nord-ouest des

58. J. K. CHAMBERS, P. TRUDGILL, *op. cit.*, p. 5.

59. Wilbert HEERINGA, John NERBONNE, «Dialect areas and dialect continua», *Language Variation and Change*, t. 13/3, 2001, p. 375-400.

Pays-Bas⁶⁰. Nerbonne et Siedle observent une situation semblable en analysant des données de l'allemand⁶¹.

Cependant, le faible taux de corrélation entre la distance linguistique et la distance spatiale constaté pour le breton conduit à relativiser l'importance du rôle que joue cette variable. Ce résultat malmène donc le FDP sans pour autant l'infirmier. De plus, cela souligne que plusieurs facteurs s'associent en lien avec ce principe pour façonner la distribution de la distance linguistique dans l'espace. Cette situation n'est pas limitée au cas breton. En se concentrant sur la variabilité morphologique et syntaxique dans un corpus d'anglais britannique vernaculaire, Szmrecsanyi constate que le FDP ne permet d'expliquer que 4 % de la variance au sein de ses résultats. Ceci l'a amené à considérer que la place accordée à la géographie en tant que facteur explicatif dans les travaux de dialectologie tend à être surévaluée⁶² prolongeant ainsi les remarques formulées par Saussure en son temps :

« On objectera que les diversités de milieu, de climat, de configuration du sol, les habitudes spéciales (autres par exemple chez un peuple montagnard et dans une population maritime), peuvent influencer sur la langue et que dans ce cas les variations étudiées ici seraient conditionnées géographiquement. Ces influences sont contestables [...]. La direction du mouvement [impulsé par la différenciation linguistique] est attribuable au milieu ; elle est déterminée par des impondérables agissant dans chaque cas sans qu'on puisse les démontrer ni les décrire. »⁶³

D'autre part, ces observations amènent à s'interroger sur la validité des modèles évoqués plus haut. En effet, ceux-ci, s'ils ne sont pas fondamentalement erronés, s'avèrent peu opérants dans le cas du breton. Leur pouvoir explicatif ne dépasse effectivement guère la généralité énoncée. De ce fait, cela amène à se demander s'il en va de même pour d'autres langues ou bien si le breton demeure un cas particulier malgré tout. Pour ce dernier, un modèle plus complexe prenant en compte un

60. J. NERBONNE, W. HEERINGA, art. cit.

61. John NERBONNE, Christine SIEDLE, « Dialektklassifikation auf der Grundlage aggregierter Ausspracheunterschiede », *Zeitschrift für Dialektologie und Linguistik*, t. 72/2, 2005, p. 129-147.

62. B. SZMRECSANYI, *op. cit.*, p. 159.

63. Ferdinand de SAUSSURE, *Cours de linguistique générale*, Paris, Payot, 5^e édition, 2005 [1913], p. 272.

nombre plus important de variables permettrait sans doute de décrire la distribution des phénomènes linguistiques dans l'espace avec plus de pertinence. Le faible pouvoir explicatif de la géographie observé dans le cadre de cette étude amène ainsi à réinterroger la place accordée à la géographie en dialectologie et d'une certaine façon non pas tant à la minorer qu'à préciser quels sont les paramètres géographiques véritablement pertinents⁶⁴.

La distribution de la distance linguistique : le résultat de dynamiques historiques ?

Les critères géographiques n'expliquant que faiblement la répartition de la distance linguistique à travers l'espace bretonnant, la question des raisons d'une telle organisation se pose toujours. Il se pourrait que celle-ci résulte en premier lieu de dynamiques historiques. Plusieurs éléments invitent à penser dans ce sens. Ainsi, l'opposition initiale Nord-Ouest / Sud-Est, premier niveau de structuration de la distance linguistique, entre en écho avec plusieurs phénomènes historiques dont on constate le cumul comme le résume le tableau 10.

Zone nord-ouest	Zone sud-est
Taux de similarité linguistique élevés et constants	Taux de similarité linguistique faibles et hétérogènes
Niveau de palatalisation faible	Niveau de palatalisation élevé et très élevé
Traits linguistiques sudistes –	Traits linguistiques sudistes +++
Emprunts ponctuels au français dans le vocabulaire de base	Emprunts assez importants au français dans le vocabulaire de base
Noms de famille suivent une distribution nordiste ⁶⁵	Noms de famille suivent une distribution sudiste ⁶⁶

64. B. SZMRECSANYI, *art. cit.*

65. Daniel LE BRIS, « Surnames and geolinguistics in Brittany: a study of concordances », *STUF - Language Typology and Universals*, t. 65/1, 2012, p. 95-109. Daniel LE BRIS, « Concordances géolinguistiques et anthroponymiques en Bretagne », *Corpus*, t. 12, 2013, p. 85-104.

66. *Ibidem.*

Densité importante des noms de lieu en <i>Plou-</i> , <i>Plo-</i> , <i>Plé-</i> , <i>Gui-</i> ⁶⁷	Densité importante des toponymes se terminant en <i>-ac</i>
Civitas gauloise et gallo-romaine des Osismes	Civitas gauloise et gallo-romaine des Vénètes
Transposition plus ou moins dans les limites des évêchés médiévaux de Cornouaille, Léon et Tréguier (KLT)	Evêché médiéval de Vannes
Dialecte occidental ancien ⁶⁸	Dialecte oriental ancien ⁶⁹

Tableau 10. Cumul des faits correspondant à la bipartition ancienne de l'espace bretonnant⁷⁰

Toutefois, c'est au niveau des pôles de similarité linguistique que l'hypothèse historique se caractérise le mieux. En effet, on note des correspondances saisissantes entre ces pôles et la géographie de bassins génétiques identifiés à partir de mutations du gène CFTR responsables de la mucoviscidose. Cette maladie génétique provoque chez les malades un surplus de mucus dans les voies respiratoires et l'appareil digestif. Un enfant ne souffrira de la mucoviscidose que s'il hérite de chacun de ses deux parents d'une mutation du gène CFTR⁷¹.

La prévalence de la mucoviscidose est particulièrement élevée en Bretagne et plus encore dans sa partie occidentale. Afin de comprendre les raisons d'une telle situation dans l'espace breton, Nadine Pellen, socio-démographe, a travaillé sur la généalogie de porteurs sains de la mucoviscidose et en a conclu, du point de vue de la génétique des populations, que la présence de la mucoviscidose prolonge un effet fondateur intervenu à une époque ancienne et lié à un mouvement de population. La réduction de la diversité génétique à la suite de cet épisode a abouti à une présence importante des mutations de la mucoviscidose

67. F. FALC'HUN, *op. cit.*, 1981, figure 54, p. 191.

68. Erwan VALLERIE, *Traité de toponymie historique de la Bretagne*, Ar Releg-Kerhuon [Le Relecq-Kerhuon], An Here, 1995.

69. *Ibidem*.

70. Le détail de ces faits est discuté dans T. SOLLIEC, *op. cit.*

71. On recense à l'heure actuelle plus de 2000 mutations de ce gène, voir Cystic Fibrosis Mutation Database : <http://www.genet.sickkids.on.ca/StatisticsPage.html> (page consultée le 30/05/21).

parmi la population. Cet effet a été amplifié par la position géographique excentrée de la Bretagne, ne facilitant pas l'établissement de nouvelles populations au cours des siècles, susceptibles d'apporter de la diversité génétique. De plus, Pellen a constaté que le marché matrimonial des populations bretonnes était restreint géographiquement, ce qui a favorisé la diffusion de mutations liées au gène CFTR⁷². La mucoviscidose constitue en ce sens un fait social total.

Nadine Pellen a observé ainsi que les ancêtres des porteurs sains identifiés dans le cadre de ses travaux s'inscrivent dans des bassins génétiques privilégiés dont l'ancrage géographique est manifeste, comme cela peut se voir à la figure 17.

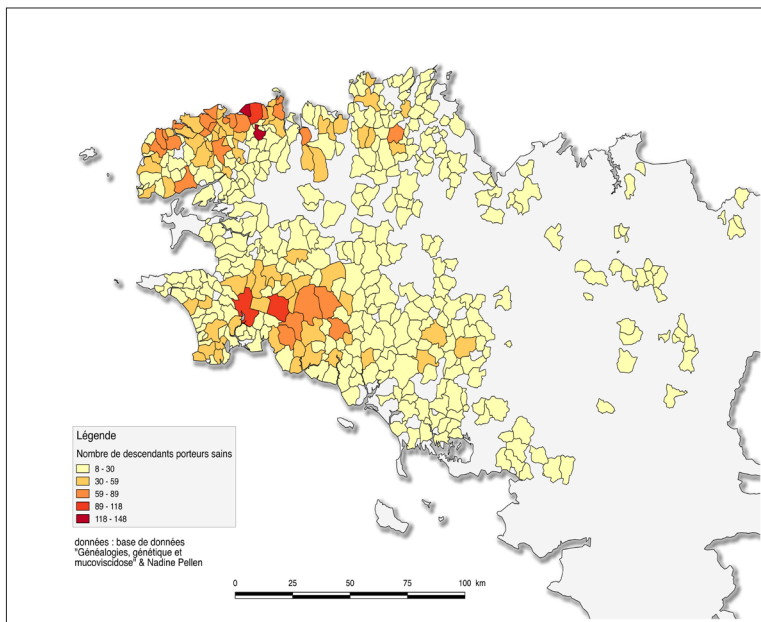


Figure 17. Localisation des descendants de porteurs sains des couples présent dans la généalogie de la base de données « Généalogies, génétique et mucoviscidose »

La dichotomie observée en premier lieu entre une zone nord-ouest et une zone sud-est se retrouve ici. Dans le cas de ce dernier espace, la maladie y moins présente que dans le Nord-Ouest. En outre, une

72. Nadine PELLEN, *La mucoviscidose en héritage*, Paris, INED, 2015.

correspondance s'établit entre les bassins génétiques et plusieurs pôles de similarité linguistique. C'est le cas tout particulièrement pour les pôles *bis*, *gwalarn* et *mervent*. En observant le détail pour 11 mutations du gène CFTR, une récurrence se manifeste clairement entre ces deux dimensions comme le résume le tableau 11 ⁷³.

Nom de la mutation du gène CFTR	<i>gwalarn</i>	<i>bis</i>	<i>mervent</i>	<i>Aven</i>	Pays de Pontivy	<i>gevred</i>
F508del	+	+	+	+	?	-
G551D	+	-	-	+	?	-
1078delT	-	+	+	+	-	?
1717-1G>A	+	+	+	+	-	-
W846XelT	+	+	?	+	+	-
N1303KelT	?	+	+	-	+	+
4005+1G>A	+	+	+	+	-	-
G542X	+	+	+	-	-	-
I507d	-	+	-	-	-	-
G149R	+	?	+	+	-	-
621+1G>T	-	-	-	+	-	-
Nombre de correspondances	7	8	7	8	2	1

Tableau 11. Correspondances entre les principales mutations du gène CFTR et les pôles de similarité linguistique ⁷⁴

Nous proposons de voir dans ce lien entre les deux domaines, le prolongement d'un effet fondateur ayant opéré tant du point de vue génétique que sur le plan linguistique (*Ibidem*). Cette notion est issue de la génétique des populations et désigne une situation dans laquelle un groupe se détache d'une population d'origine, souvent à l'occasion d'un peuplement. On observe alors que la diversité génétique présente dans ce nouveau groupe est moindre et certaines caractéristiques, ici des mutations du gène CFTR, tendent à devenir plus prégnantes

73. Nadine PELLEN, Tanguy SOLLIEC, «Dialectometry and population genetics - when results converge: the case of Western Brittany», *Studia Celto-Slavica - Journal of the Learned Association Societas Celto-Slavica*, t. 12, Brest, CRBC, 2021, p. 63-103.

74. + marque une correspondance, - signale qu'il n'y a pas de correspondance et ? indique que les informations ne sont pas suffisante pour trancher.

en son sein⁷⁵. Salikoko Mufwene a appliqué ce concept aux langues humaines, particulièrement dans le cas des langues créoles et des nouvelles variétés coloniales des langues européennes⁷⁶. Dans ces cas de figure, les nouvelles langues ou les nouveaux dialectes ne sont pas de simples transpositions des idiomes originels mais émergent à la suite d'un processus de restructuration des différentes variétés en présence. Ainsi, certains traits non standards, au statut périphérique dans les langues métropolitaines peuvent acquérir un rôle central au sein de ces nouvelles variétés.

Nous considérons ainsi que les variétés de breton présentes dans les pôles *bis*, *gwalarn*, *mervent*, pays de l'Aven et plus hypothétiquement, pays de Pontivy et *gevred* sont issues d'un effet fondateur qui a conduit à la création d'espaces et de représentations socio-culturels, à des pratiques linguistiques partagées et, *in fine*, à l'émergence de nouveaux dialectes brittoniques⁷⁷. Ce phénomène est ainsi à mettre en lien avec les migrations bretonnes entre les IV^e et VII^e siècles qui ont vu des populations issues de l'île de Bretagne venir s'établir sur le continent et particulièrement notamment la péninsule armoricaine⁷⁸. La documentation historique contemporaine de ce phénomène s'avère fragmentaire et limitée⁷⁹. Toutefois, le croisement des données génétiques et linguistiques permet d'apporter de nouveaux éclairages sur cet épisode.

En ce sens, cela conduit à constater le « primat de l'histoire » comme perspective pour décrire les faits linguistiques, suivant en cela Coseriu

75. Ernst MAYR, *Systematics and the Origin of Species, from the Viewpoint of a Zoologist*. Cambridge, MA, Harvard University Press, 1942. Ernst MAYR, « Change of genetic environment and evolution », dans J. Huxley (dir.), *Evolution as a Process*, Londres, George Allen & Unwin, 1954, p. 157-180. William B. PROVINE, « Ernst Mayr: Genetics and Speciation », *Genetics*, t. 167/3, 2004, p. 1041-1046.

76. Salikoko S. MUFWENE, « The Founder Principle in Creole Genesis », *Diachronica*, t. 13/1, 1996, p. 83-134. . Salikoko S. MUFWENE, *The Ecology of Language Evolution*, Cambridge, Cambridge University Press, 2001.

77. T. SOLLIEC, *op. cit.*

78. N. PELLEN, T. SOLLIEC, *art. cit.*

79. Magali COUMERT, « Le peuplement de l'Armorique : Cornouaille et Domnonée de part et d'autre de la Manche aux premiers siècles du Moyen Age », dans M. Coumert, H. Tétrel (dir.), *Histoires des Breagnes 1. Les mythes fondateurs*, Brest, CRBC, 2010, p. 15-22.

et à reconnaître par là qu'« une langue [est] une tradition historique, de l'activité de parler, c'est-à-dire par définition un objet historique⁸⁰. » Cette perspective permet dans le cas du domaine bretonnant de mieux envisager les relations entre géographie et histoire en tant que paramètres structurants dans la distribution de la distance linguistique. Les deux points de vue y contribuent clairement mais à des échelles différentes et c'est sans doute dans une perspective qui les associe que la répartition de la distance linguistique peut être pleinement comprise.

Envisager des perspectives plus globales ?

La prise en compte de l'influence de la géographie dans l'organisation de la distribution de la distance linguistique nous a amené à relativiser son importance comme facteur explicatif et à lui reconnaître une influence avant tout ponctuelle et locale. La constitution de l'espace linguistique breton tel qu'il se donne à lire à travers une analyse dialectométrique se comprend plus aisément dans une perspective historique.

Or, plus que simplement juxtaposer contraintes géographiques et dynamiques historiques, c'est un ensemble de facteurs qu'il faut prendre en considération pour accéder à une pleine compréhension de l'organisation de la distance linguistique dans l'espace bretonnant. En effet, il est possible d'envisager ce domaine linguistique comme une niche, au sens écologique, et de l'appréhender comme le résultat de plusieurs forces en présence assurant ainsi un ensemble de pressions, le fonctionnement d'un système linguistique en étant une aux côtés des différents facteurs géographiques et des tendances historiques.

Gary Lopyan & Rick Dale, à partir de l'étude de la complexité morphologique plus ou moins importante dans les langues du monde, ont formulé la *linguistic niche hypothesis*. Ils considèrent que les langues tendent ainsi à s'adapter structurellement aux différents environnements dans lesquelles elles sont parlées, sous l'effet conjugué de plusieurs facteurs externes, social, physique et technologique mais également dans le prolongement du fonctionnement interne des

80. Eugenio COSERIU, « Du primat de l'histoire », *Energia*, t. 2, 2010, p. 56-72, p. 57. Traduction française de Eugenio COSERIU, « Vom Primat der Geschichte », *Sprachwissenschaft*, t. 5, 1980, p. 125-145.

structures linguistiques⁸¹. Parmi ces facteurs, ils relèvent notamment que la structure sociale des groupes dans lesquelles ces langues sont parlées peut favoriser l'émergence de certains traits linguistiques⁸². Cette approche prolonge la réflexion d'Haugen qui a proposé de penser et de modéliser les relations des langues entre elles ainsi qu'avec la société où elles se parlent dans le cadre d'une écologie linguistique⁸³.

Un tel point de vue permet de prolonger les observations formulées à partir de l'organisation de la distance linguistique en Basse-Bretagne, en y voyant non pas un état donné, fruit de différents épisodes historiques mais le résultat d'un processus plurifactoriel exerçant des pressions plus ou moins importantes sur les structures linguistiques et les amenant tantôt à converger (similarité linguistique importante) ; tantôt à diverger (similarité linguistique faible). Ce phénomène peut dès lors se comprendre comme une adaptation au milieu et au contexte social dans lesquels ces variétés de breton sont en usage. Cette perspective permettrait de mieux cerner la place qu'a prise la langue bretonne dans l'espace armoricain au tournant de l'Antiquité et du Moyen Âge et de réfléchir aux facteurs qui ont favorisé l'émergence et le développement d'une niche qu'elle a depuis lors occupée.

À partir de là s'ouvre un chantier de recherche appelant à croiser les approches et à multiplier les comparaisons avec plusieurs espaces

81. Gary LUPYAN, Rick DALE, «Language Structure Is Partly Determined by Social Structure», *PLOS ONE* 5, n° 1, 2010, e8559 [en ligne]. Mis en ligne le 20 janvier 2010, URL : <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0008559> (consulté le 1^{er} mai 2017). Rick DALE, Gary LUPYAN, «Understanding the origins of morphological diversity: the linguistic niche hypothesis», *Advances in Complex Systems*, t. 15/03, n° 4, 2012, 1150017. Gary LUPYAN, Rick DALE, «Why Are There Different Languages? The Role of Adaptation in Linguistic Diversity», *Trends in Cognitive Sciences*, t. 20/9, 2016, p. 649-660. James WINTERS, Simon KIRBY, Kenny SMITH, «Languages adapt to their contextual niche», *Language and Cognition*, t. 7, 2015, p. 415-449. Molly LEWIS, Michael C. FRANK, «Linguistic niches emerge from pressures at multiple timescales», dans A. Papafragou, D. Grodner, D. Mirman, J.C. Trueswell, *Proceedings of the 38th Annual Conference of the Cognitive Science Society*, Austin, TX, Cognitive Science Society, 2016, p. 1385-1390. (Lupyan & Dale 2010 ; Dale & Lupyan 2012 ; Lupyan & Dale 2016 ; Winters, Kirby & Smith 2015 ; Lewis & Frank 2016).

82. G. Lupyan, R. Dale, *art. cit.*

83. Einar HAUGEN, *The Ecology of Language*, Stanford, CA, Stanford University Press, 1972. Louis-Jean CALVET, *Pour une écologie des langues du monde*, Paris, Plon, 1999.

linguistiques. Ainsi, l'étude de la distribution de la distance linguistique au sein des parlers gallos permettrait d'examiner la situation au-delà de la frontière linguistique, particulièrement dans la zone dite mixte d'où les parlers bretons ont disparu au cours du Moyen-Âge. De même, des parallèles avec d'autres langues, plus ou moins proches sur les plans généalogiques, sociolinguistique et/ou géographique donneraient l'opportunité de contraster la situation du breton et d'identifier une éventuelle singularité. L'examen de l'incidence des structures sociales sur l'organisation des faits linguistiques, du début du Moyen-Âge jusqu'à nos jours, reste encore à poursuivre en mettant en œuvre les outils de la sociolinguistique historique. Assurément, ces développements ne peuvent être envisagés que dans le cadre d'un dialogue avec les données de l'histoire diachronique du breton.

Conclusion

L'analyse de la distance linguistique, comprise comme une tension entre le même et le différent, nous a permis d'appréhender le domaine bretonnant dans son ensemble et la variation diatopique en son sein. Le point de départ de notre réflexion portait sur l'incidence de la géographie sur la distribution de la distance linguistique dans cet espace.

Nous avons pu constater que dans le cas du breton, la distance linguistique s'agence du point de vue spatial sur plusieurs plans. Tout d'abord, une grande division se dessine entre une zone nord-ouest et une zone sud-est. Cette bipartition correspond à plusieurs phénomènes géohistoriques que plusieurs auteurs ont préalablement relevés. Ensuite, un second niveau se compose de pôles de similarité linguistique, bien localisés et dont l'extension est restreinte voire même très limitée. On constate que le profil de ces pôles varie selon leur appartenance à la zone nord-ouest ou sud-est.

Alors que les facteurs géographiques s'avèrent être des paramètres structurant dans l'organisation de la diversité des langues, la mise en relation de plusieurs d'entre eux, distance spatiale, temps de trajet ou géographie physique, avec la distribution de la distance linguistique montre qu'ils n'influent qu'assez peu sur cette dernière. Lorsque leur effet se manifeste, celui-ci demeure ponctuel et local. L'organisation dialectale du breton ne joue pas plus de rôle dans cette répartition.

Un prolongement de cette analyse permettrait d'identifier du point de vue statistique quelle proportion de la répartition linguistique peut être expliquée par chacun des paramètres analysés. Nous avons constaté que les critères les plus pertinents n'opèrent pas au même niveau. La distance euclidienne agit à l'échelle de l'ensemble du domaine bretonnant. La géographie physique agit avant tout au niveau des pôles et ce, de façon intermittente. Néanmoins, leur influence demeure limitée comme en témoigne les taux de corrélation relativement faibles.

Ceci amène à reconsidérer la place accordée à la géographie comme facteur explicatif dans les études de dialectologie. Plusieurs modèles explicatifs, le *Fundamental Dialectology Principle* et celui du continuum dialectal se montrent décevants dans le cas du breton dans la mesure où ils ne permettent pas de dépasser de grandes généralités. D'autres causes doivent donc être recherchées pour comprendre les raisons d'une telle organisation de la distance linguistique en Basse-Bretagne.

La coïncidence de la répartition des pôles de similarité linguistique avec des bassins génétiques identifiés à partir de la géographie de mutations du gène CFTR responsables de la mucoviscidose nous a permis d'ouvrir une perspective historique et d'identifier là le résultat d'un effet fondateur que nous avons mis en relation avec l'épisode des migrations bretonnes au début du Moyen-Âge. Dans ce cadre se sont mis en place des espaces de communication et de pratiques sociales dont la structuration à terme se donne à lire à travers la distance linguistique d'une part et une géographie locale de la mucoviscidose.

Les faits linguistiques possèdent donc une spatialité et celle-ci n'est que l'inscription d'une temporalité dans un territoire. En ce sens, les phénomènes linguistiques appréhendés dans leur diatopie donnent à voir «du temps dans l'espace» pour reprendre les termes du titre de Guylaine Brun-Trigaud et al.⁸⁴

L'analyse de facteurs géographiques ainsi que l'observation de dynamiques historiques nous a permis de mieux comprendre

84. Guylaine BRUN-TRIGAUD, Yves LE BERRE, Jean LE DÛ, *Lectures de l'Atlas linguistique de la France de Gilliéron et Edmont. Du temps dans l'espace. Essai d'interprétation des cartes de l'«Atlas linguistique de la France» de Jules Gilliéron et Edmond Edmont augmenté de quelques cartes de l'«Atlas linguistique de la Basse-Bretagne»*, Paris, Comité des travaux historiques et scientifiques (CTHS), 2005.

l'organisation de la distribution de la distance linguistique à travers l'espace bretonnant et ce, à plusieurs niveaux. Toutefois, l'adoption d'un point de vue plus global permettra de comprendre la constitution de cet espace en tant que niche linguistique du point de vue de l'écologie des langues. Dans ce cadre, plusieurs forces contribuent à façonner la ou les langue(s) en présence. En complément des paramètres mentionnés, la dimension sociale intervient également dans un tel processus. Comme le signale Haugen : « *Language exists only in the minds of its users, and it only functions in relating these users to one another and to nature, i.e. their social and natural environment.* »⁸⁵. La prise en compte d'une perspective plus vaste permettra selon nous de mieux appréhender les développements historiques du breton à travers les siècles.

La prise en compte de la distance linguistique par la dialectométrie montre son potentiel heuristique. Elle offre ainsi un point de vue supplémentaire pour la description des langues et permet d'accéder notamment à une temporalité invisible de prime abord. Nous avons également constaté que celle-ci interagit à des degrés variables avec plusieurs paramètres géographiques. Ces différentes dimensions extérieures aux structures linguistiques amènent, au-delà du cas breton à interroger la part précise que joue l'arbitraire dans leur organisation et à identifier avec le plus de précision possible l'influence qu'y exerce l'environnement pris dans toute sa complexité.

85. E. HAUGEN, *op. cit.*, p. 325.