Des pages d'un atlas linguistique à une fenêtre sur l'histoire, le cas du breton. Résultats d'une analyse dialectométrique appliquée au *Nouvel Atlas linguistique de la Basse-Bretagne*

A Linguistic Atlas as History, Breton Words as a Window Opened onto the Past. Insights from a Dialectometrical Study of the Nouvel Atlas linguistique de la Basse-Bretagne

Tanguy Sollicec

Édition électronique

URL : <https://journals.openedition.org/geolinguistique/5990>

DOI : 10.4000/geolinguistique.5990

ISSN : 2650-8176

Éditeur

UGA Éditions/Université Grenoble Alpes

Édition imprimée

ISBN : 978-2-37747-303-8

ISSN : 0761-9081

Référence électronique

Tanguy Sollicec, « Des pages d'un atlas linguistique à une fenêtre sur l'histoire, le cas du breton. Résultats d'une analyse dialectométrique appliquée au *Nouvel Atlas linguistique de la Basse-Bretagne* », *Géolinguistique* [En ligne], 21 | 2021, mis en ligne le 06 décembre 2021, consulté le 19 décembre 2021. URL : <http://journals.openedition.org/geolinguistique/5990> ; DOI : <https://doi.org/10.4000/geolinguistique.5990>

Ce document a été généré automatiquement le 19 décembre 2021.

Géolinguistique

Des pages d'un atlas linguistique à une fenêtre sur l'histoire, le cas du breton. Résultats d'une analyse dialectométrique appliquée au *Nouvel Atlas linguistique de la Basse-Bretagne*

A Linguistic Atlas as History, Breton Words as a Window Opened onto the Past. Insights from a Dialectometrical Study of the Nouvel Atlas linguistique de la Basse-Bretagne

Tanguy Sollicec

1. Introduction

- 1 La notion de langue permet de désigner intuitivement et avec facilité un ensemble de pratiques linguistiques qui pourtant, lorsqu'on les analyse en détail peuvent parfois s'avérer contradictoires ou divergentes. Il en va de même pour la langue bretonne dont le domaine à première vue, peut sembler former un ensemble monolithique à la périphérie de l'espace roman. Pourtant une observation attentive montre un espace travaillé par la variation amenant ainsi à s'interroger sur les points de convergence permet au breton de se constituer pleinement comme langue. En ce sens, le cas breton illustre la « question de Babel » appliqué au cas d'une langue (Steiner, 1998 : 18). George Steiner en réfléchissant aux enjeux de la traduction a relevé un paradoxe de la condition humaine. Bien que l'espèce soit unique de par son aptitude au langage, cette caractéristique ne se réalise toutefois que dans la multitude, en une diversité de langues alors qu'il serait *a priori* plus économique de ne disposer que d'un seul idiome. Ce paradoxe, qui inscrit ainsi l'être humain dans un antagonisme entre unité et

différence, opère également au sein des langues humaines. En effet, les usages que font les locuteurs d'une langue privilégient généralement une variété donnée qui en côtoie plusieurs autres. L'étude de la variation linguistique dans sa dimension diatopique permet donc d'accéder pleinement au fonctionnement des langues humaines.

1.2. Une approche quantitative au croisement de plusieurs problématiques

- 2 Dans le cadre de cet article, nous nous proposons d'examiner dans quelle mesure les structures linguistiques d'une langue, ici le breton, sont affectées par une telle tension entre le semblable et le différent. Cette perspective offre un point de vue inédit sur l'organisation géolinguistique de la langue bretonne. Pour y parvenir, nous mobilisons une notion, la distance linguistique et sa transposition méthodologique, la dialectométrie. L'approche mise en œuvre ici est quantitative dans la mesure où les données linguistiques étudiées ont été quantifiées, transposées en valeurs numériques selon une procédure déterminée. Cette démarche permet réaliser des comparaisons qu'il n'aurait pas été possible d'établir autrement et ainsi, d'évaluer leur degré de similarité. L'étape de la quantification permet donc de ne pas se restreindre aux seuls critères « identique » ou « différent » dans le cadre d'une catégorisation binaire et d'obtenir des résultats plus nuancés. Il s'agit là d'une procédure d'objectivisation dans l'analyse des données observées.
- 3 Sur le plan méthodologique, cette démarche constitue la traduction d'une notion plus générale de distance linguistique. Celle-ci englobe symboliquement l'ensemble des différences linguistiques séparant deux objets linguistiques plus ou moins circonscrits (parlers, dialectes, langues). La distance linguistique s'avère être un outil heuristique pertinent pour appréhender des données linguistiques qui se caractérisent en premier lieu par leur variabilité. L'application de cette notion à des données issues d'atlas linguistiques a abouti, au début des années 1970, à l'émergence d'une nouvelle méthode, la dialectométrie, autour de la figure de Jean Séguy (1971, 1973). L'apparition de cette méthode s'inscrit dans le contexte plus vaste de recherches conduites au sein de plusieurs courants de la linguistique en vue de l'élaboration d'outils de mesure de la variation linguistique. L'enjeu d'une telle approche est de repérer à travers la grande variabilité des données observées si des phénomènes, invisibles au premier abord, sont identifiables grâce à la médiation d'un traitement quantitatif. Dans un second temps, il s'agit d'en comprendre la structuration et éventuellement, le caractère inédit.
- 4 La recherche que nous avons entreprise se base sur la conviction qu'un atlas linguistique constitue un corpus linguistique à part entière et qu'à ce titre, il offre un point de vue sur la langue décrite. Les données que contient un atlas linguistique se caractérisent par leur aspect limité, souvent dépourvu de leur contexte mais elles ne sont pas moins systématiques et géolocalisées. Évidemment, l'entreprise que constitue un atlas linguistique n'est pas exempte de critiques (Canobbio, 2004), mais elle offre un point de vue de premier ordre pour envisager les langues dans leur spatialité. En outre, nous faisons nôtre la proposition de Séguy (1973b) de considérer les atlas linguistiques comme de véritables laboratoires au sein desquels peuvent s'observer les solutions et créations linguistiques que l'être humain est capable d'élaborer tant pour communiquer que pour se différencier. Une telle étude est donc susceptible de contribuer à la linguistique générale.

- 5 Un atlas linguistique demeure un petit corpus en comparaison des outils élaborés pour les grandes langues modernes de communication. Vraisemblablement, la grande majorité des langues humaines ne bénéficiera sans doute jamais de tels outils et leur description s'appuiera avant tout sur des corpus relativement limités. Pour autant, Evangelia Adamou invite à ne pas renoncer à leur appliquer des analyses quantitatives semblables à celles déployés sur des collections de données gigantesques. Elle a effectivement observé que : « *Despite the small size of the corpora, a systematic and quantitative analysis of the spontaneous speech is full of unexpected and novel findings, confirming or challenging qualitatively-produced observations*¹. » (2016 : 222)
- 6 L'enjeu dès lors est d'intégrer un maximum de données provenant d'un atlas linguistique, de façon à considérer un tel corpus dans sa globalité et ainsi, ne pas se limiter aux seules données qui peuvent sembler intéressantes ou pertinentes à première vue. La méthodologie mise en œuvre dans le cadre de ce travail, la dialectométrie, vise un tel objectif. En ce sens, cette démarche montre de quelle façon les données d'un atlas linguistique peuvent être analysées de façon systématique et les apports qu'il est possible d'en retirer.

2. La langue bretonne, un laboratoire idéal pour une telle démarche

- 7 Conduire une telle analyse sur la langue bretonne s'avère intéressant à plusieurs titres. En effet, cet espace linguistique, de par son extension géographique limitée (équivalent à la surface d'environ deux département français), peut être considéré comme un véritable laboratoire linguistique au sein duquel les hypothèses de recherche peuvent être éprouvées et ce faisant, de réfléchir aux dynamiques à l'œuvre dans les systèmes des langues telles qu'elles se donnent à voir par le prisme des usages linguistiques². En outre, la faible prégnance de l'écrit sur les structures de la langue, du fait d'une situation de diglossie, font du breton à l'heure actuelle une langue essentiellement orale et nettement dialectalisée. Cette situation permet donc de réfléchir aux facteurs qui, à l'échelle d'une langue, favorisent ce processus de différenciation. Le contexte breton a fait ainsi l'objet de travaux géolinguistiques tant dans la description du domaine par le biais d'un atlas de première génération, l'*Atlas linguistique de la Basse-Bretagne* (désormais ALBB) réalisé par Le Roux (1924-1963). L'analyse approfondie des données de ce dernier par Falc'hun (1951, 1965, 1981) a permis de renouveler la compréhension de l'espace dialectal du breton. Le *Nouvel Atlas linguistique de la Basse-Bretagne* (Le Dû, 2001, désormais NALBB) a été conçu dans le but de renouveler et d'affiner les données de son prédécesseur, dès lors, on peut espérer qu'il dessine un portrait plus nuancé de l'espace dialectal du breton.
- 8 La prédominance des usages oraux du breton a été soulignée, mais cela ne doit pas occulter une documentation historique, plus ou moins disparate selon les périodes, qui permet d'envisager cette langue dans sa profondeur chronologique. L'histoire du breton demeure toutefois une histoire en pointillés en raison du caractère sommaire de notre connaissance des états de la langue pour certaines périodes. De ce point de vue, l'étude des parlers vernaculaires dans leur dimension géolinguistique peut apporter des éclairages sur les développements historiques d'une langue, comme nous le verrons par la suite. C'est dans ce contexte que nous avons consacré notre doctorat à l'exploitation dialectométrique des données du NALBB et à l'analyse des résultats obtenus (Sollic, 2016).

2021). Cette perspective nous a offert l'opportunité d'examiner une langue dans sa spatialité et d'observer tant les processus œuvrant à la différenciation linguistique que les phénomènes centripètes suivant en cela la proposition de Séguy (1973c : 87-88) de considérer les atlas linguistiques comme de véritables concentrés de langue.

2.1. Une méthodologie, la dialectométrie

- 9 Selon la typologie des travaux de géolinguistique établie par Jean-Philippe Dalbera (2002), notre démarche appartient ainsi la troisième génération de travaux géolinguistiques. Ceux-ci se caractérisent par leur emploi à des degrés divers de différents outils informatiques dans l'analyse des données étudiées. C'est dans ce contexte que la méthodologie de la dialectométrie s'est constituée et développée à la suite de son élaboration par Séguy (1973a) participant au mouvement plus général de l'intégration des outils informatiques dans les sciences humaines, paradigme à présent connu sous l'appellation d'humanités numériques (Goebl, 1984, 2010 ; Nerbonne & Heeringa, 2010 ; Nerbonne & Kretzschmar, 2013).
- 10 En tant que méthode quantitative, la dialectométrie vise à mettre au jour parmi les données d'un atlas linguistiques des « structures profondes » (Goebl, 2002), des modes d'organisation inédits des faits vernaculaires, indécélables de prime abord. C'est dans cette perspective que nous avons mobilisé cette méthodologie, prolongeant ainsi le projet initial du NALBB d'actualiser et d'affiner les données de son prédécesseur, l'ALBB (Le Dû, 1972). De plus, la dialectométrie, telle que nous l'envisageons, offre la possibilité d'exploiter les données d'un atlas linguistique dans leur ensemble, en les établissant comme un corpus à part entière et non pas comme la juxtaposition de formes ou de variables (Nerbonne, 2008).
- 11 Dans un premier temps, nous nous attacherons à présenter le NALBB, les données étudiées et les notions mobilisées dans le cadre de cette étude. Nous préciserons ensuite la méthodologie qui a été mise en place afin de réaliser l'analyse dialectométrique du corpus étudié. Puis, nous exposerons les résultats obtenus avant d'en proposer une lecture explicative au regard des analyses de nos devanciers.

3. Contextualisation

3.1. La langue bretonne

- 12 Le breton est une langue de la famille celtique parlée dans l'ouest de la France, dans la péninsule bretonne. Le domaine bretonnant recouvre le département du Finistère et la partie occidentale des départements des Côtes-d'Armor et du Morbihan. Attestée à l'écrit dans des gloses médiévales dès les VIII^e et IX^e siècles de notre ère, puis dans des textes longs lors de la période dite du moyen breton (XII^e-1659), le breton demeure avant tout une langue à tradition orale, parlée dans les milieux populaires, dans la mesure où la majorité de ses locuteurs actuels l'ont acquise dans un contexte familial et informel, n'ayant généralement de contact qu'épisodiques avec ses différentes formes écrites.
- 13 D'autre part, les usages vernaculaires du breton tendent à s'effacer au sein de la société locale en raison d'une rupture de la transmission familiale en faveur du français dans les décennies qui ont suivi la seconde guerre mondiale. Ce sont donc avant tout les

générations les plus âgées qui parlent le breton le plus spontanément (Broudic, 2009). En contrepoint, une présence limitée dans l'enseignement et les médias s'est développée, aboutissant à la formation de nouveaux locuteurs. Cette situation a conduit l'Unesco à cataloguer le breton comme langue « sérieusement en danger » (Moseley, 2010)³.

3.2. La notion de distance linguistique

- 14 L'approche mise en œuvre dans le cadre de cette étude, la dialectométrie, constitue la facette méthodologique d'une notion plus large, la distance linguistique, appliquée à l'échelle d'une langue. Paul Heggarty donne à cette notion de distance linguistique la définition suivante : « *The term 'linguistic distance' is often used to refer to the degree of similarity/difference between any two language varieties*⁴. » (2000 : 531)
- 15 Ce concept constitue un outil heuristique efficace, particulièrement lorsque la réflexion en linguistique porte sur des données caractérisées par leur variabilité. La distance linguistique en tant que notion rend possible l'exploration des différents éléments impliqués dans une comparaison. La distance linguistique correspond ainsi un écart au sens que lui donne François Jullien (2016) et permet ainsi de dépasser le simple constat de l'identique ou du différent dans l'analyse d'une comparaison en explorant l'espace (métaphorique) qui sépare les éléments mis en balance. En ce sens, la notion de distance linguistique peut être mobilisée pour appréhender l'activité basilectale d'*Homo loquens* (Goebel, 2003) et permet par-là, d'accéder à une dimension du fonctionnement des langues.
- 16 Diverses propositions ont été formulées au cours du XIX^e siècle pour mobiliser la distance linguistique dans le cadre d'une réflexion scientifique, notamment dans le sillage de la Société de géographie de Paris. Toutefois, ce n'est que dans la seconde moitié du XX^e siècle que la transcription méthodologique a réellement aboutie, notamment suite à différentes démarches visant à mesurer divers phénomènes linguistiques avec plus ou moins de succès (variable sociolinguistique, lexicostatistique, dialectométrie, etc.) (Sollic, 2021).
- 17 La dialectométrie, telle qu'élaborée par Jean Séguy (1971, 1973a), à partir de son expérience dans le domaine gascon et plus largement, roman, permet d'appréhender deux ensembles de données appartenant chacun à des localités précises. Ces informations sont généralement issues d'atlas linguistiques et la comparaison met en relation deux points d'enquête plus ou moins distants. L'enjeu de l'analyse dialectométrique est de quantifier la distance linguistique qui sépare de tels ensembles en relevant de façon systématique les différences qui les distinguent l'un de l'autre. Ces opérations sont répétées autant de fois que nécessaire en vue de couvrir l'ensemble du corpus étudié et du territoire exploré.

3.3. La géolinguistique du breton et les données étudiées

- 18 La compréhension géolinguistique du domaine bretonnant s'est profondément renouvelée au cours du XX^e siècle. Pendant longtemps, il a été considéré qu'à chacun des anciens évêchés antérieurs à la Révolution française correspondait un dialecte. Bien que cette idée ne corresponde pas à la réalité des parlers locaux, elle demeure encore présente à l'esprit de nombre de locuteurs.

Figure 1. – Carte des évêchés de langue bretonne et points d'enquête du NALBB.

Carte élaborée par l'auteur.

- 19 En s'appuyant sur les données présentes dans l'ALBB, et plus particulièrement sur la diffusion des innovations linguistiques, François Falc'hun a montré qu'en réalité le domaine de la langue bretonne se structure du point de vue dialectal en trois zones distinctes. Ainsi, un espace central orienté selon un axe nord-est/sud-ouest, allant de Lannion à Pont-l'Abbé, au-delà de Quimper, contraste avec les espaces qui lui sont contigus au nord-ouest et au sud-est. Les innovations linguistiques se sont propagées dans cette zone centrale au cours des siècles grâce au rôle joué par la ville de Carhaix située en son centre. Cette dernière a assuré un rôle de nœud routier vers lequel convergeaient de nombreuses voies romaines à l'époque antique et propice aux échanges. Les espaces périphériques, véritables zones latérales, au nord-ouest et au sud-est se caractérisent aussi bien par la rétention de réflexes anciens ou bien par des évolutions locales qui leur sont propres (Falc'hun, ouvr. cités). À ce titre, le cumul de ces différences amène à distinguer nettement les parlers du Vannetais du reste du domaine bretonnant (Le Dû & Brun-Trigaud, 2003).

Figure 2. – Carte de la structuration tripartite du domaine bretonnant.

Carte élaborée par l'auteur.

- 20 Au-delà de cette dynamique profonde qui s'est mise en place au Moyen Âge, des aires locales se sont formées, à des périodes plus récentes, autour de pôles urbains comme Morlaix, Landerneau ou Quimper, exerçant une influence plus ou moins profonde sur les campagnes environnantes.
- 21 Les données que nous avons analysées dans le cadre de cette étude proviennent toutes du NALBB. Cette entreprise a débuté dès 1969 à l'initiative de Jean Le Dû, récemment disparu, dans le but d'actualiser et d'affiner les données de l'ALBB qui le précédait. C'est ainsi qu'un réseau dense de points d'enquête, au nombre de 187, a été constitué couvrant un tiers des communes de la Bretagne bretonnante. Cet atlas porte sur un ensemble de notions courantes appartenant plusieurs champs sémantiques. Son orientation est clairement phonétique mais le NALBB sous sa forme actuelle ne constituait que la première phase du projet initial, de plus grande ampleur. Un deuxième volet devait explorer plus en profondeur des questions liées à la morphologie et au vocabulaire. Cependant, par manque de moyens et de temps, celui-ci n'a pu voir le jour (Brun-Trigaud & Le Dû, 2016).

Figure 3. – Points d'enquête du NALBB.

Carte élaborée par l'auteur.

- 22 Le NALBB contient plus de 100 000 formes réparties en 598 cartes. Tout le matériel recueilli lors des enquêtes a été retranscrit par Jean Le Dû les faisant passer au crible d'un filtre phonologique unique et évitant ainsi des distorsions liées à la présence de plusieurs transpositeurs (Mathussek, 2016). Aux dires de plusieurs observateurs, le NALBB offre un bon aperçu des parlers bretons de la fin du xx^e siècle (Favereau, 2001 ; Ternes, 2008).
- 23 Dans le cadre de l'analyse dialectométrique réalisée pour cette étude, 106 cartes ont été retenues en excluant celles pour lesquelles plusieurs lemmes expriment une même notion. Par exemple, la carte NALBB 310 UN SEAU présente des formes correspondant aux lemmes *sailh* et *kelorn*. L'outil utilisé pour réaliser des mesures de distance linguistique doit établir des comparaisons entre des formes d'un même lemme pour que le résultat obtenu ait une quelconque pertinence. Le fonctionnement de cet algorithme sera détaillé par la suite. Ainsi, 17,7 % des cartes du NALBB ont été intégrées à l'analyse. Au total, près de 19 500 formes ont été étudiées dans le cadre de 55 120 comparaisons. Le seuil empirique établi par Guiter (1973) de 100 cartes en dessous duquel une analyse dialectométrique ne saurait être valide est donc atteint⁵.
- 24 Notre analyse porte avant tout au niveau phonétique. Ce niveau est généralement considéré comme celui étant le plus sensible à la variation et au changement linguistique. Son analyse constitue dès lors une voie privilégiée pour contraster les données du NALBB avec celles de l'ALBB. Cette dimension phonétique est celle que le projet NALBB a privilégié et qui par conséquent, offre un matériel sans pareil pour observer la variation linguistique à ce niveau. D'autre part, comme le signale Paul Heggarty, d'un point de vue méthodologique, le processus de quantification peut être

mis aisément en œuvre dans le cas de comparaisons d'ordre phonétique. Si nous utilisons par extension le terme de distance linguistique, nous gardons à l'esprit que nous avons traité les données étudiées avant tout dans leur dimension phonétique.

- 25 Le domaine breton a déjà fait l'objet d'études de dialectométrie, basées essentiellement sur les données de l'ALBB. Gary German (1984, 1993) a cherché à comprendre comment le parler de Saint-Yvi, légèrement au sud-est de Quimper, s'intégrait dans le réseau dialectal breton. Pour cela, il a calculé les distances qui séparent chaque point d'enquête de l'ALBB avec le breton de cette commune. En exploitant les résultats obtenus grâce à un algorithme d'analyse par répartition⁶, il constate que la distribution de la distance linguistique dans le domaine bretonnant recouvre des structures socio-historiques anciennes et montre ainsi que l'étude systématique des données vernaculaires permet d'accéder à une profondeur temporelle importante prenant à revers les doutes exprimés par des chercheurs appartenant à d'autres générations (Jackson, 1960). Denis Costaouec (1998) s'est engagé dans une démarche similaire à partir de son travail de description du breton de La Forêt-Fouesnant, située à quelques kilomètres de Saint-Yvi. Il s'est concentré sur la répartition fine de la distribution linguistique autour de cette commune et a constaté que des changements relativement faibles dans les niveaux de similarité linguistique correspondent à certains changements dans les pratiques sociales (répartition géographique des réseaux d'entraide traditionnels, choix préférentiel du conjoint dans une zone de similarité linguistique importante) montrant que certaines limites dialectales mises au jour par la dialectométrie s'inscrivent de façon structurante dans le paysage social (Costaouec, 2012).
- 26 De tels résultats obtenus à partir des données de l'ALBB invitent à penser que l'étude systématique du contenu du NALBB ouvrira de nouvelles perspectives sur l'organisation géolinguistique du domaine bretonnant. À notre connaissance, le NALBB n'a été exploité pleinement comme corpus que par la seule analyse de Brun-Trigaud et Le Dû (art. cité) illustrant par la même occasion les possibilités offertes par l'application web Gabmap (Leinonen, Çöltekin & Nerbonne, 2015). D'autres recherches s'y réfèrent à divers degrés pour l'analyse de variables dialectales ou l'étude de l'évolution des réflexes diachroniques (Brun-Trigaud, 2004 ; Plourin, 2005 : 298 ; Wmffre, 2007 : 527-566 ; Jouitteau, 2021).

4. Méthodologie mise en œuvre

- 27 La dialectométrie contribue à construire un regard distancié sur les données analysées, mais elle souffre du point de vue méthodologique de ce que nous avons nommé « effet boîte noire » dans la mesure où le lecteur ne saisit pas toujours aisément les différentes étapes du processus de quantification qui permet de transformer des données linguistiques en valeurs numériques. Pour ce faire, nous avons eu recours à une version de la distance de Levenshtein (1955) qui évalue le nombre d'opérations nécessaires pour transformer une suite de caractères A en une suite de caractères B. Pour y parvenir, cet algorithme procède dans un premier temps à un alignement des formes comparées puis, peut mettre en œuvre 3 types d'opérations pour effectuer cette transformation :
- la suppression d'un caractère ;
 - l'ajout d'un caractère ;

- le remplacement d'un caractère par un autre.

- 28 Son objectif premier est d'identifier la procédure la plus économique en nombre d'opérations pour effectuer cette transformation. Si nous prenons pour exemple des formes suivantes issues du NALBB, carte 165 DE LA TERRE *douar*, on aboutit à la comparaison suivante :

Tableau 1. – Détail des opérations effectuées par la distance de Levenshtein dans la comparaison de deux formes phonétiques : point NALBB 88 Collorec ['du^wəɛ] ~ point NALBB 90 Botmeur ['du:ar]?

point 88	D	U	^w	ə	ɛ
point 90	D	u:		A	r
Nature de l'opération réalisée		Rempl. de u par u:	Suppr. de ^w	Rempl. de ə par a	Rempl. de ɛ par r
Nombre d'opérations		1	1	1	1

- 29 On note ainsi un total de 4 différences entre les deux formes comparées. Dès lors, un pourcentage de similarité peut être établi entre les deux formes en appliquant la formule suivante⁸ :

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

- 30 Ceci permet d'obtenir un résultat de 20 % de similarité entre les deux formes comparées⁹ :

$$100 - (400/5)$$

- 31 La distance de Levenshtein également connue sous la dénomination de distance d'édition (*edit distance* en anglais) a été appliquée à la dialectométrie à partir des années 1990, dans un premier temps par Brett Kessler (1995) pour analyser des données de l'irlandais. Par la suite, John Nerbonne et ses collaborateurs l'ont adopté et en ont affiné l'exploitation afin de procéder à des mesures de distance linguistique, initialement dans le domaine néerlandophone (Nerbonne *et al.*, 1996 ; Nerbonne & Heeringa, 1997 ; Nerbonne, Heeringa & Kleiweg, 1999). Wilbert Heeringa (2004) a évalué en profondeur l'efficacité et la pertinence de cet outil pour l'analyse de données dialectales. Il a également identifié les différents paramètres techniques sur lesquels il est possible d'intervenir dans le but d'obtenir une meilleure granularité des résultats. Il en a validé l'utilisation en l'appliquant à plusieurs domaines linguistiques.
- 32 La distance de Levenshtein présente plusieurs avantages dans l'optique d'une analyse dialectométrique. Elle permet d'obtenir un résultat nuancé, dépassant le couple « différent » / « identique » tout en s'avérant synthétique et en concentrant la variabilité des faits linguistiques en une valeur numérique. De plus, cet algorithme est symétrique c'est-à-dire que le résultat de similarité obtenu est valable dans les deux sens de la comparaison. Son usage est relativement aisé et en outre, certains paramètres sont modulables selon les besoins de l'analyse. D'autre part, cet outil ne nécessite pas d'être utilisé sur un corpus important pour identifier la structuration

dialectale d'une langue dans la mesure où une liste de mots est suffisante (Heeringa, Nerbonne & Kleiweg, 2002).

- 33 La version de la distance de Levenshtein que nous avons retenue dans le cadre de ce travail est celle élaborée par les soins de Brun-Trigaud (2014), Levenshtein ModificationsReturned, qui permet de retourner les opérations effectuées par l'algorithme. Si nous revenons à l'exemple précédent, cela signifie que l'algorithme fournit à l'utilisateur le détail des opérations qu'il a réalisées afin d'obtenir un taux de similarité. Dans le cas présent, il s'agit :
- du remplacement de u par u: ;
 - de la suppression de ^w ;
 - du remplacement de ə par a ;
 - du remplacement de ʁ par r ;
- 34 Ces informations ne sont généralement pas renvoyées à l'utilisateur qui se contente donc le plus souvent de la valeur numérique qu'est le taux de similarité. À la suite d'un premier travail exploratoire (Brun-Trigaud, Sollicec & Le Dû, 2016 ; Sollicec, 2019), nous avons formulé l'hypothèse que les opérations retournées de la sorte par l'algorithme permettent d'accéder au contenu de la distance linguistique, aux éléments qui la constituent. Dès lors, il devient envisageable d'identifier les phénomènes linguistiques qui y contribuent et si ceux-ci varient dans l'espace.
- 35 Cette approche permet de dépasser ce que nous avons nommé l'effet « boîte noire » qui est souvent associé à cette méthodologie. En effet, il est fréquent, de prime abord, que l'observateur ne saisisse pas de quelle façon les faits linguistiques ont été transformés en valeurs numériques et, *in fine*, d'où proviennent les nombres donnés comme résultats. En outre, d'un point de vue plus technique, cette approche permet de contrôler et de superviser les opérations réalisées lors de l'analyse dialectométrique et de s'assurer que celles-ci soient justifiées sur le plan linguistique. Afin de comparer les données de chaque point d'enquête avec celles de ses voisins immédiats, un réseau de segments a été élaboré. À partir des 186 points d'enquête¹⁰, 534 segments ou couples de points ont été établis selon une perspective interponctuelle comme le montre la figure 4.

Figure 4. – Réseau des segments étudiés.

Carte élaborée par l'auteur.

- 36 Ainsi, pour chaque segment, les données provenant des 106 cartes sélectionnées ont été comparées de façon systématique. Dans un second temps, un taux moyen de similarité linguistique a été établi à partir des taux obtenus pour chacune des comparaisons effectuées.

5. Résultats

- 37 La représentation cartographique des taux de similarité obtenus pour chaque segment étudié permet d'aboutir à la figure 5 :

Figure 5. – Distribution de la distance linguistique à travers le domaine bretonnant.

Similarité phonétique en Basse-Bretagne

Carte élaborée par l'auteur.

- 38 Cette carte est à lire de la façon suivante : plus la similarité linguistique entre deux points d'enquête est importante, plus le trait qui les relie sera épais et plus foncée sera la nuance de bleu qui y est associée. À travers l'ensemble du domaine bretonnant, on relève un taux moyen de similarité linguistique de 68,43 %, le taux médian est de 68,27 % ce qui indique que l'influence des valeurs extrêmes demeure modérée. Les résultats obtenus s'étendent d'un minimum de 39,15 % (pour le segment des points 63 – Pléguen – et 160 – Noyal-Naizin)¹¹ à un maximum à 88,47 % (entre les points 57 – Plougrescant – et 58 – Camlez)¹².
- 39 Une première évaluation statistique en appliquant le test de Shapiro-Wilk à ces résultats permet de constater que la distribution obtenue n'est pas le fruit du hasard. Celui-ci permet de déterminer dans quelle mesure une distribution est normale, autrement dit correspond à une représentation en cloche (loi de Gauss). L'examen des résultats obtenus à l'issue de l'analyse statistique aboutit aux valeurs suivantes :

$$W = 0,98457, p\text{-value} = 0,00001952$$

Avec une *p-value* nettement inférieure à 0,05, l'hypothèse d'une distribution normale peut être écartée. Ce premier constat étant posé, il en résulte que des facteurs, encore à identifier, viennent structurer cette distribution.

5.1. Une organisation spatiale

- 40 Les résultats s'avèrent être agencés sur le plan spatial (Pellen & Sollicec, 2021; Sollicec, à paraître). Deux niveaux de lectures se distinguent. En premier lieu, une bipartition claire apparaît entre une aire nord-ouest et une aire sud-est, telle qu'elle se donne à voir à la figure 6.

Figure 6. – Bipartition de l'espace bretonnant entre une aire nord-ouest et une aire sud-est.

Carte élaborée par l'auteur.

- 41 Au-delà de la proximité géographique, on observe avant tout que la distance linguistique se comporte différemment dans chacune de ces deux aires. Les segments appartenant à l'aire nord-ouest présentent en moyenne un de taux de similarité relativement important, s'élevant en moyenne à 73,9 %. Au contraire, dans l'aire sud-est, le niveau de similarité linguistique y est plus faible, 60,04 % en moyenne. En outre, on constate qu'au nord-ouest le degré de similarité linguistique est plus homogène (coefficient de variation de 8,85) alors que dans l'aire sud-est, le niveau de similarité linguistique se montre beaucoup plus versatile d'un segment à l'autre (coefficient de variation de 11,49). Le tableau 2 résume ces observations.

Tableau 2. – Statistiques descriptives pour les différentes espaces étudiées.

	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

- 42 Une observation plus approfondie permet de repérer au sein de ces aires des pôles de similarité. Ceux-ci regroupent des segments dont les valeurs sont particulièrement importantes. Nous en avons relevé 10 au sein de l'actuel espace bretonnant, 3 pour l'aire nord-ouest et 7 pour l'aire sud-est comme le montre la figure 7 :

Figure 7. – Cartes des pôles de similarité linguistique.

Carte élaborée par l'auteur.

- 43 Chacun de ces pôles a été nommé d'après la dénomination des points cardinaux en breton, de celui d'une ville ou du terroir local, de façon à les distinguer du niveau précédent. Au sein de l'aire nord-ouest, trois espaces géographiques, relativement étendus, peuvent être ainsi distingués. En revanche, dans l'aire sud-est, le nombre des pôles est plus important. Toutefois, leur extension est nettement plus restreinte, contrastant donc nettement avec leurs pendants de l'aire nord-ouest. Le tableau 3 donne certains éléments permettant de les décrire d'un point de vue statistique.

Tableau 3. – Statistiques descriptives pour les différents pôles de similarité.

Nom du pôle	Taux de similarité moyen	Taux de similarité maximum	Taux de similarité minimum	Écart-type	Nombre de segments
Gwalarn	77,95 %	87,19 %	66,68 %	4,65	101
Bis	77,44 %	88,47 %	66,69 %	4,92	90
Mervent	68,22 %	75,86 %	51,57 %	4,16	55
Aven ¹³	69,92 %	73,28 %	66,6 %	2,99	5
Sud Poher	69,99 %	72,27 %	65,9 %	2,99	3
Plinn	71,71 %	77,85 %	68,48 %	2,68	11
Pays de Lorient	66,34 %	68,37 %	62,83 %	1,65	17
Pays de Pontivy	66,55 %	72,39 %	61,55 %	3,93	7
Pays de Carnac	65,89 %	69,83 %	62,57 %	2,59	8
Gevred	64,76 %	66,36 %	62,16 %	1,56	7

- 44 Ainsi, selon l'aire à laquelle ces pôles appartiennent ils présentent un profil différent. Dans l'aire nord-ouest, ils connaissent une extension importante et un taux moyen de similarité relativement élevé. Au contraire, les pôles de l'aire sud-est sont beaucoup moins étendus dans l'espace, organisés autour de quelques points d'enquête. Les niveaux de similarité linguistique y sont toutefois nettement plus bas. Ces pôles n'apparaissant parfois que par contraste avec leur environnement immédiat dans lequel la similarité est encore plus faible. Si la structuration spatiale de la distribution de la distance linguistique apparaît clairement, l'incidence de la géographie sur cette organisation s'avère relativement limitée (Sollic, à paraître). En effet, seule la topographie et certains cours d'eau jouent ponctuellement un rôle démarcateur, à l'incidence avant tout locale.

5.2. Des structures organisées (patterns) dans l'espace, écho de phénomènes historiques anciens

- 45 Les résultats obtenus diffèrent de l'organisation tripartite de l'espace bretonnant décrite par Falc'hun. Toutefois, chaque niveau de lecture repéré précédemment entre en écho avec plusieurs phénomènes socio-historiques, qui parfois s'inscrivent dans une longue durée. La distribution de la distance linguistique est donc à envisager comme le fruit de dynamiques historiques. L'idée d'une bipartition ancienne de l'espace dialectal breton a ainsi déjà été évoquée par plusieurs auteurs sans que ceux-ci ne s'attachent à en décrire les contours. Falc'hun (1981 : 227-231, 525-526) s'est ainsi brièvement référé à une « dualité dialectale primitive ». Jackson mentionne que : « *In Tréguier in particular we should see a fundamentally northern dialect affected to a greater or lesser by "Central"*

*influences*¹⁴. » (1967 : 34) Cette structuration ancienne entre deux aires dialectales a été explorée par Erwan Vallerie (1995) en se basant sur l'étude des formes anciennes des toponymes d'origine bretonne passés en français à diverses périodes, notamment dans la zone où le breton a été présent historiquement et en les mettant en parallèle avec les équivalents modernes. Tout comme l'opposition majeure entre les parlers du Vannetais et les variétés de breton dit KLT¹⁵, ces découpages ne correspondent pas dans le détail à nos observations sans pour autant les contredire, ni les invalider. Une correspondance entre les faits linguistiques et des structures socio-historiques, particulièrement l'ancienne organisation du territoire des cités gauloises puis gallo-romaines a été relevée (German, 1993 ; Le Bris, 2013). Dans le cas du breton, cela concerne avant tout l'interface entre le territoire des Osismes (au nord-ouest) et des Vénètes (au sud-est) comme l'indique la figure 8 (a-b) :

Figure 8a. – Limites du territoire des tribus gauloises dans l'ouest de la péninsule armoricaine.

Carte élaborée par l'auteur.

Figure 8b. – Limites du territoire des tribus gauloises avec la superposition des aires nord-ouest et sud-est (cf. fig. 6).

Carte élaborée par l'auteur.

- 46 Une telle convergence suggère la possible continuité de certaines limites territoriales, ce qui aurait ainsi contribué à façonner la distribution de la distance linguistique. Il est à noter que cette bipartition s'observe également du point de vue dialectométrique dans certaines cartes élaborées par Brun-Trigaud et Le Dû (2016 : fig. 2, 7a, 9b et 9c)
- 47 Les profils différenciés que présente la distribution de la distance linguistique dans ces différents espaces laisse supposer que l'aire nord-ouest s'est constituée aux dépens de l'aire sud-est dont la formation est elle-même sans doute plus ancienne. En ce sens, cette situation illustre à un autre niveau de granularité l'affirmation de Sapir : « *The greater the degree of linguistic differentiation within a stock, the greater is the time that must be assumed for the development of such differentiations*¹⁶. » (1916 : 76). Dès lors s'établit une chronologie relative. Ceci amène à considérer cette bipartition comme un phénomène dynamique inscrit dans une temporalité. Cet aspect expliquerait donc pourquoi nous n'avons pas relevé de concordance exacte entre nos propres observations avec les conclusions de Vallerie ou avec celles des historiens de l'espace armoricain. Par certains aspects, ce phénomène d'expansion de l'aire nord-ouest aux dépens de l'aire sud-est se prolonge toujours à l'heure actuelle. En effet, Le Pipec (2004) relève une tendance dans certains parlers vannetais à réorganiser leur système d'accentuation en vue de l'aligner sur celui des variétés dites KLT.
- 48 La présence de différents pôles linguistiques disséminés à travers l'espace bretonnant est en partie inédite et apporte une perspective différente des conclusions de Falc'hun quant à l'organisation dialectale tripartite du domaine de la langue bretonne¹⁷. Un premier regard sur les pôles repérés au sein de l'aire nord-ouest (pôles *bis*, *gwalarn* et *mervent*) invite à reconnaître l'ancien découpage des évêchés prérévolutionnaires du Léon, du Trégor et de la Cornouaille dans sa partie occidentale (voir fig. 1) sans pour

autant qu'il y ait, ici aussi, de correspondance totale. Ces pôles de l'aire nord-ouest ont pour point commun d'occuper des plateaux présentant peu d'obstacle d'ordre géographique. En outre, les variétés de breton parlées au sein de ces trois pôles se singularisent à l'heure actuelle par des traits linguistiques distinctifs¹⁸. Ceux ne transparaissent dans les documents écrits qu'à partir de la fin du XVI^e siècle et encore plus nettement au XVII^e siècle. En raison d'une documentation historique particulièrement pauvre avant le XV^e siècle, la question de l'ancienneté de ces particularités en diachronie fait l'objet de débat tout autant que la question de savoir ce que les écrits en moyen breton dévoilent des pratiques vernaculaires de l'époque. La singularité des parlers associés à des pôles de similarité linguistique ne se limite pas à l'aire nord-ouest. Elle s'observe également dans les pôles Aven, pays de Pontivy et *gevred* au sein de l'aire sud-est. Le tableau 4 présente la typologie des différents pôles de similarité identifiés.

Tableau 4. – Typologie des différents pôles de similarité dans l'espace bretonnant.

Catégorie 1 Pôles dont le taux moyen de similarité linguistique est élevé, dont l'étendue géographique importante et bien individualisés sur le plan linguistique	<i>bis, gwalarn, mervent</i>
Catégorie 2 Pôles dont le taux moyen de similarité linguistique est assez élevé, dont l'étendue géographique restreint et assez bien voire bien individualisés sur le plan linguistique	Aven, pays de Pontivy, <i>gevred</i>
Catégorie 3 Pôles dont le taux moyen de similarité linguistique est peu élevé, dont l'étendue géographique restreint et sont faiblement individualisés sur le plan linguistique	Sud Poher, <i>plinn</i> , pays de Lorient, pays de Carnac

- 49 L'individualité des pôles pour les catégories 1 et 2 invite à les considérer comme des espaces symboliques dont les contours ont été façonnés par des usages sociaux et des références en communs au fil du temps et qui se sont singularisés sur le plan linguistique par des faits dialectaux plus ou moins récents. Ces observations suggèrent ainsi que l'émergence de ces différents ensembles de pôles de similarité linguistique résulte de dynamiques différentes.
- 50 À ce titre, il est assez surprenant de constater que plusieurs des pôles de similarité linguistique identifiés à l'issue de l'analyse dialectométrique trouvent un écho dans la géographie locale d'une maladie génétique, la mucoviscidose. Sa prévalence en Bretagne y est élevée, particulièrement dans la partie occidentale de la péninsule (Pellen, 2015 ; Bellis, Dehillotte & Lemonnier, 2019 : 10)¹⁹. Cette maladie est provoquée par une mutation du gène CFTR provoquant l'accumulation d'un surplus de mucus dans les organes. À ce jour, plus de 2 000 mutations différentes de ce gène ont été répertoriées²⁰. Pour qu'un enfant soit victime de la mucoviscidose, il faut qu'il ait reçu de chacun de ses parents une mutation de ce gène. Autrement dit, il faut que chaque parent soit lui-même « porteur sain ». De ce fait, les mutations susceptibles de

provoquer la mucoviscidose se transmettent de génération en génération. Les mutations du gène CFTR peuvent être datées par les généticiens apportant ainsi des éclairages supplémentaires sur l'histoire longue des populations étudiées (Heyer, 2020). Nadine Pellen (2015) en cherchant à comprendre les raisons d'une forte prévalence de la maladie en Bretagne, particulièrement dans la zone bretonnante et en réfléchissant aux pratiques sociales sous-jacentes associées à cette maladie, s'est rendue compte que les différentes mutations du gène CFTR présentes en Bretagne tendent à se regrouper dans des bassins génétiques dont l'ancrage géographique est manifeste. Elle est parvenue à de telles conclusions en construisant une analyse basée sur les informations d'une base de données généalogiques recueillies auprès de personnes souffrant de cette maladie. Il s'avère que cette organisation en bassins génétiques recoupe la distribution de plusieurs des pôles identifiés précédemment comme le montre le tableau suivant dans lequel sont mises en relation les principales mutations du gène CFTR et les différents pôles de similarité linguistique (Pellen & Sollic, 2021) :

Tableau 5. – Correspondance entre les mutations du gène CFTR et les pôles de similarité linguistique⁵¹.

Type de mutation du gène CFTR	Gwalarn	Bis	Mervent	Aven	Pays de Pontivy	Gevred
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

51 Les correspondances s'avèrent donc fortes dans le cas des pôles de similarité linguistique nettement singularisés à l'heure actuelle sur le plan linguistique. L'opposition entre l'aire nord-ouest et l'aire sud-est se manifeste à nouveau ici, car les bassins génétiques apparaissent clairement dans la première alors que dans la seconde,

à l'exception notable du pôle Aven, dont le profil est bien particulier, ces bassins génétiques ressortent avec une consistance moindre.

- 52 Nadine Pellen (2015) voit dans la prévalence élevée de la mucoviscidose en Bretagne le résultat d'un effet fondateur, c'est-à-dire l'implantation d'un groupe humain dont la diversité génétique s'est retrouvée réduite à un moment de son histoire, généralement à la suite d'une migration et favorisant ainsi l'émergence de certains traits. Les mutations génétiques du gène CFTR constituent un véritable marqueur dont la diffusion au sein d'une population a été amplifiée par des contraintes culturelles et historiques auxquelles elle a dû faire face comme l'absence de mouvements migratoires ultérieurs susceptibles de renouveler le stock génétique des individus ou encore un marché matrimonial à l'espace réduit. La distribution de la distance linguistique et sa structuration à travers l'espace bretonnant apparaissent donc être le fruit de dynamiques historiques longues. Elles constituent l'écho d'un effet fondateur sur le plan génétique que nous mettons en relation avec l'épisode historique dit des migrations bretonnes qui a vu au début du Moyen Âge (entre les ^v^e et ^{vii}^e siècles) l'implantation dans la péninsule armoricaine de populations issues de l'île de Bretagne (Pellen & Sollic, 2021).

6. L'étude des faits vernaculaires contemporains, une perspective sur l'histoire des langues ?

- 53 Malgré un travail de compilation des sources important (Fleuriot, 1980), les contours et les modalités de ce phénomène historique demeurent encore assez mal connues en l'absence de sources contemporaines permettant de le documenter, au point où une spécialiste de la période résume la situation de la sorte : « La naissance de la Bretagne continentale est une énigme. » (Coumert, 2010 : 15) En effet, une part importante de la connaissance de la péninsule armoricaine aux premiers temps du Moyen Âge s'appuie avant tout sur des sources ultérieures, vies de saints, anthroponymes ou toponymes. L'exploitation des résultats présentés précédemment permet d'apporter des éléments supplémentaires et de mieux appréhender les modalités de ce déplacement de populations d'une part et d'autre part, d'apporter une nouvelle perspective sur l'histoire de la langue bretonne et de ses dialectes.
- 54 Fleuriot (1980) a reconnu dans ce mouvement migratoire deux temps, une phase militaire aux ^{iv}^e et ^v^e siècles dans le cadre de la réorganisation de la politique défensive de l'Empire romain sur les côtes de la Manche et ensuite, une phase civile sous l'égide d'autorités civiles et religieuses au ^{vi}^e siècle (Chédeville & Guillotel, 1984). Cette interprétation a été discutée récemment en raison du peu de traces matérielles qu'il est possible d'observer d'un point de vue archéologique (Brett, 2011, 2019). Le cas breton est à envisager dans le contexte général de la réorganisation politique et territoriale de l'Europe occidentale au début du Moyen Âge, à un moment où des groupes humains entiers pouvaient migrer, amenant dans certains cas une population autochtone à adopter une langue nouvelle (Guy, 2014). L'organisation bipartite de la distance linguistique relevée précédemment peut être lue dans cette perspective, voyant ainsi une zone nouvelle, l'aire nord-ouest qui s'est constituée aux dépens de l'aire sud-est, initialement plus vaste. Cette dernière, de ce fait, possède un ancrage chronologique plus ancien qui se reflète dans la distribution hétérogène de la distance linguistique. Par ailleurs, l'origine des spécificités des parlers vannetais actuels est souvent

envisagée par le fait que cet espace était déjà plus peuplé au temps des migrations bretonnes et avait été fortement romanisé au cours de l'Antiquité contrastant avec l'aire nord-ouest ou l'ancien territoire des Osismes marqué par une déprise démographique à la suite de la crise du III^e siècle à travers l'empire romain (Fleuriot, 1980 ; Vallerie, 2015).

- 55 L'émergence des pôles de similarité linguistique auxquels sont associés des variétés linguistiques bien singularisées à l'heure actuelle s'inscrit dans le prolongement des migrations bretonnes. Ceux-ci peuvent être considérés comme des occurrences de nouveaux dialectes sur le modèle des variétés coloniales des langues européennes ou des langues créoles. Salikoko Mufwene (1996a, 1996b, 1997, 1998, 2001) décrit ce processus de formation de nouvelles variétés en reprenant la notion d'effet fondateur, mais en l'appliquant aux langues. Dans un tel cas de figure, des populations d'origine différente se regroupent dans un nouvel espace, de façon volontaire ou forcée, entraînant un brassage des différentes variétés ou langues en présence au sein de ce que Mufwene (2002) a appelé métaphoriquement un bassin de traits linguistiques²². À l'issue de ce processus de restructuration, une série de caractéristiques s'imposent et contribuent à définir les nouveaux dialectes. Nous supposons qu'un tel processus d'effet fondateur a opéré en plusieurs points de l'espace bretonnant au début du Moyen Âge, particulièrement pour les pôles de similarité *bis*, *gwalarn*, *mervent*, Aven, pays de Pontivy et *gevred*. Leurs caractéristiques phonétiques et morphologiques actuelles constituent ainsi l'aboutissement d'un tel processus de sélection dont les prémices se sont constituées entre les V^e et VII^e siècles. Ce processus a abouti selon nous à la formation de nouveaux dialectes brittoniques²³. L'analyse de la distribution de la distance linguistique dans le domaine bretonnant permet ainsi d'accéder à une strate historique ancienne différente de celles mises au jour par Falc'hun (1981) dans son analyse des données de l'ALBB.
- 56 Toutefois, l'interprétation que nous proposons ne va pas à l'encontre des conclusions de ce dernier quant à une structuration tripartite de l'espace bretonnant. En effet, nous considérons qu'à la suite de l'émergence de ces nouveaux dialectes brittoniques s'est mis en place un mouvement de convergence entre ces derniers et les variétés de latin tardif présentes sur le territoire et peut-être même de gaulois tardif²⁴. C'est dans ce contexte que le breton s'individualise et se forme comme langue autonome au sein de l'espace brittonique aboutissant à une distinction claire d'avec le gallois au XII^e siècle et d'avec le cornique au XIV^e siècle. Ce phénomène de convergence a été porté par la ville de Carhaix, l'antique *Vorgium* et ainsi aboutit au cours du Moyen Âge à la formation de la zone centrale de l'espace dialectal breton axée autour de cette ville. Ainsi, à la structuration bipartite ancienne du domaine bretonnant succède une organisation en trois zones, phénomène relativement fréquent en dialectologie (Nichols, 2013 : 53-54). Toutefois, ce mouvement de convergence de variétés les unes vers les autres voit ses effets contrebalancés au XVII^e siècle, dans le sillage de la Contre-Réforme et l'élaboration de standards ecclésiastiques de référence basés sur les parlers périphériques du nord-ouest (Léon) et du sud-est (Vannetais). Le scénario proposé ici amène donc à réinterroger l'idée des développements historiques du breton considérés comme un morcellement progressif d'un ensemble initialement uniforme, en faveur d'un modèle plus complexe. L'étude des données vernaculaires telles qu'elles sont documentées dans le NALBB permet d'ouvrir des perspectives historiques sur un temps long. En ce sens, le signal que la dialectométrie permet de capter entre en

contradictions avec les affirmations de Jackson (1960) qui voyait dans les données d'un atlas linguistique la possibilité d'accéder uniquement à l'histoire récente d'une langue.

7. Conclusion

- 57 L'exploitation de données phonétiques du NALBB dans le cadre d'une analyse dialectométrique nous a permis d'observer la répartition de la distance linguistique à travers le domaine de la langue bretonne. Cette distribution se structure dans l'espace et se partage ainsi entre une aire nord-ouest et une aire sud-est, et entre en écho avec des structures socio-historiques présentes à plusieurs périodes. Un second niveau de lecture nous a permis d'identifier à travers l'espace bretonnant 10 pôles de similarité. La typologie de ces derniers varie, mais une contrepartie d'ordre génétique peut s'observer pour plusieurs d'entre eux, dans la mesure où leur répartition correspond à celle de bassins génétiques que dessinent les mutations du gène CFTR responsables de la mucoviscidose.
- 58 La mise en relation de ces observations avec l'histoire linguistique du breton et de ses dialectes nous a permis d'esquisser une nouvelle perspective sur les développements de cette langue à travers le temps en identifiant un effet fondateur à l'origine de plusieurs des pôles linguistiques repérés dans le cadre d'une analyse dialectométrique. Une datation au début du Moyen Âge, entre les ^v^e et ^{vii}^e siècles, est envisagée pour ce phénomène, au moment des migrations bretonnes de l'île de Bretagne vers le continent. La notion de distance linguistique, dont la dialectométrie constitue le pendant méthodologique, s'avère être un outil heuristique probant qui permet d'appréhender la variation diatopique des langues et d'accéder à un niveau d'analyse généralement peu connu et peu analysé. Celui-ci résulte de l'accumulation de diverses strates temporelles dans les données linguistiques. La distance linguistique constitue ainsi un prisme pour envisager les langues là où elles se situent, à la croisée du temps et de l'espace.
- 59 Les atlas linguistiques permettent d'envisager les langues du point de vue de leur spatialité. Ils s'avèrent constituer des corpus à part entière contribuant à la description des langues. Nous avons vu, en prenant appui, sur le NALBB, qu'il est possible d'envisager une exploitation d'une masse importante de données qu'un atlas contient en faisant appel à la dialectométrie. Généralement, cette approche est utilisée pour identifier des zones dialectales, notamment en faisant appel à la distance de Levenshtein (Heeringa, 2004) et de ce fait, elle a été appliquée à différents contextes linguistiques²⁵. De façon surprenante, dans le cas de cette étude, les résultats obtenus ne correspondent qu'assez peu aux divisions dialectales les plus courantes du breton. Le contenu de la distance linguistique et des phénomènes qui la produisent ont pu être identifiés dans leurs grandes lignes mais nous ne les avons pas présentés par souci de brièveté²⁶.
- 60 L'étude de la distance linguistique permet donc d'accéder à un signal distinct de celui sur lequel les répartitions dialectales sont formées. Si cela s'observe pour le breton, la probabilité est importante pour qu'il en soit de même dans le cas d'autres langues. La question de la définition de ce signal, comprendre la place qu'il occupe au sein des données linguistiques et son rôle sur le plan théorique demeure à cette heure un chantier à engager. La perspective que nous a ouvert l'analyse dialectométrique du NALBB amène à réfléchir plus largement à la nature de la géolinguistique et si, malgré

sa dénomination, elle n'a pas un rôle à jouer comme une science auxiliaire de l'histoire voire souscrire dans une démarche géohistorique, dans la mesure où elle permet d'observer l'inscription du temps dans l'espace et par là, l'histoire des langues et des hommes tel que l'avait formulé Jean Le Dû (2008). Une telle approche s'appuie sur l'affirmation singulière que le présent éclaire le passé dans la mesure où c'est à partir de données contemporaines que se construisent cette perspective sur le passé, traversant une suite de strates temporelles accumulées dans les données linguistiques.

- 61 L'analyse des données bretonnes telles qu'elles se trouvent dans le NALBB nous a permis d'ouvrir une telle fenêtre sur le passé, bénéficiant en outre d'éléments d'ordre socio-génétique pour enrichir notre analyse. À présent, il convient de déterminer dans quelle mesure le cas breton est original et si une démarche semblable pourrait aboutir à des résultats similaires dans d'autres domaines linguistiques. Sans doute que l'histoire du breton est à lire à la lumière d'une interprétation écologique afin d'en comprendre l'émergence et le développement à travers les siècles. Les résultats obtenus dans le cadre de l'analyse dialectométrique du NALBB invite à l'élaboration d'un nouveau modèle pour décrire au mieux les développements historiques du breton en les croisant avec la diachronie de la langue. En effet, c'est une autre perspective, plus complexe, que celle de la dislocation toujours continue d'un ensemble uniforme qui se dessine à présent. C'est à ce titre qu'un travail similaire sur les variétés romanes parlée au-delà de la frontière linguistique serait le bienvenu afin d'y rechercher d'éventuels pôles de similarité linguistique.

BIBLIOGRAPHIE

- ADAMOU Evangelia, 2016, *A Corpus-Driven Approach to Language Contact. Endangered Languages in a Comparative Perspective* (Language Contact and Bilingualism, 12), Berlin, Boston, Walter de Gruyter.
- BRETT Caroline, 2011, « Soldiers, Saints, and States? The Breton Migrations Revisited », *Cambrian Medieval Celtic Studies*, n° 61, p. 1-56.
- BRETT Caroline, 2019, « An Invisible Migration? The Origins of Brittany », dans G. Pigeon et G. Hily (éds), *Migrations et territoires celtiques : mouvement spatial et mutations culturelles / Divroañ hag enbroañ ar Gelted: dilec'hiañ ha cheñch sevenadur* [2016], Longueil / Rennes, Tir / L'instant même, p. 11-26.
- BROUDIC Fañch, 2009, *Parler breton au XXI^e siècle. Le nouveau sondage de TMO Régions*, Brest, Emgleo Breiz.
- BRUN-TRIGAUD Guylaine, 2004, « Le breton : un éclairage sur l'histoire des parlers gallo-romans ? », *La Bretagne linguistique*, n° 13, p. 53-71.
- BRUN-TRIGAUD Guylaine, 2014, « Un usage particulier de l'algorithme de Damerau-Levenshtein dans le domaine occitan », dans F. Diemoz et al. (éds), *"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*, Université de Neuchâtel, Genève, Droz, p. 127-148.

BRUN-TRIGAUD Guylaine & LE DÛ Jean, 2016, « Petite leçon de dialectométrie à partir des données du *Nouvel Atlas linguistique de la Basse-Bretagne* », *Lapurdum*, n° 19, p. 133-151.

BRUN-TRIGAUD Guylaine, SOLLIEC Tanguy & LE DÛ Jean, 2016, « A New Dialectometric Approach Applied to the Breton Language », dans M.-H. Côté, R. Knooihuizen et J. Nerbonne (éds), *The Future of Dialects: Selected Papers from Methods in Dialectology*, n° xv, Berlin, Language Science Press. p. 135-154. <<https://langsci-press.org/catalog/view/81/147/377-1>> [consulté le 29 septembre 2021].

CANOBBIO Sabina, 2004, « L'atlas linguistique comme outil de recherche ? À propos de quelques expériences italiennes », *La Bretagne linguistique*, n° 13, p. 281-312.

CHEDEVILLE André & GUILLOT Hubert, 1984, *La Bretagne des Saints et des Rois, V^e-X^e siècle*, Rennes, Ouest-France.

COSTAUEC Denis, 1998, *Le breton parlé à la forêt-Fouesnant (Finistère-sud). Pratique actuelle. Problèmes de phonologie et de syntaxe*, thèse de doctorat, Paris V – René Descartes.

Costaouec Denis, 2012, « Linguistic Geography of Breton and Sociocultural Motivations », *STUF – Language Typology and Universals*, vol. 65, no 1, p. 47-64. <<https://doi.org/10.1524/stuf.2012.0004>>.

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

DALBERA Jean-Philippe, 2002, « Géolinguistique : un nouveau souffle ? », *Revue belge de philologie et d'histoire*, t. 80, fasc. 3, p. 831-849. <<https://doi.org/10.3406/rbph.2002.4643>>.

FALC'HUN François, 1951, *L'histoire de la langue bretonne d'après la géographie linguistique*, Quimper, chez l'abbé Nédélec.

FALC'HUN François, 1965, *Histoire de la langue bretonne d'après la géographie linguistique*, 2 vol., Paris, Presses universitaires de France.

FALC'HUN François, 1981, *Perspectives nouvelles sur l'histoire de la langue bretonne*, Paris, Union Générale d'Éditions, coll. « La nation en question ».

FAVEREAU Francis & LE DÛ Jean, 2001, « *Nouvel Atlas linguistique de la Basse-Bretagne*. Brest », C.R.B.C., 2 vol., *Bulletin de la Société archéologique du Finistère*, t. CXXX, p. 538-543.

FLEURIOT Léon, 1980, *Les origines de la Bretagne. L'émigration*, Paris, Payot.

GERMAN Gary, 1984, *Une étude linguistique sur le breton de Saint-Yvi*, thèse de doctorat de 3^e cycle, Université de Bretagne occidentale, Brest.

GERMAN Gary, 1993, « Une méthode dialectométrique (assistée par ordinateur) pour l'analyse des atlas linguistiques », *La Bretagne linguistique*, vol. 7, p. 177-213.

GOEBL Hans, 1984, *Dialektometrische Studien: Anhand italo-romanischer, rätoromanischer und galloromanischer Sprachmaterialien aus AIS und ALF*, 3 vol., Tübingen, Max Niemeyer Verlag.

GOEBL Hans, 2002, « Analyse dialectométrique des structures de profondeur de l'ALF », *Revue de linguistique romane*, n° 66, p. 5-63.

GOEBL Hans, 2003, « Regards dialectométriques sur les données de l'*Atlas linguistique de la France* (ALF) : relations quantitatives et structures de profondeur », *Estudis Romànics*, n° 25, p. 59-96.

- GOEBL Hans, 2010, « Dialectometry and Quantitative Mapping », dans A. Lameli, R. Kehrein et S. Rababus (éds), *Language and Space. An International Handbook of Linguistic Variation*, Berlin / Boston, De Gruyter Mouton, p. 433-457. <<https://doi.org/10.1515/9783110219166.1.433>>.
- GOEBL Hans, 2012, « Introduction aux problèmes et méthodes de l'« École dialectométrique de Salzbourg » (avec des exemples gallo-, italo- et ibéroromans) », dans X. A. Álvarez Pérez et al. (éds), *Proceedings of the International Symposium on Limits and Areas in Dialectology (LimiAr)* (Lisbonne, 2011), Lisbonne, Centro de Linguística da Universidade de Lisboa, p. 117-166.
- GUITER Henri, 1973, « Atlas et frontières linguistiques », dans G. Straka (éd.), *Les dialectes romans de France à la lumière des atlas régionaux*, Paris, Éditions du CNRS, p. 61-109.
- GUY Ben, 2014, « The Breton Migration: A New Synthesis », *Zeitschrift für celtische Philologie*, vol. 61, n° 1, p. 101-156.
- HEERINGA Wilbert, 2004, *Measuring Dialect Pronunciation Differences Using Levenshtein Distance*, Ph.D. dissertation, University of Groningen, Groningen. <www.rug.nl/research/portal/files/9800656/thesis.pdf> (consulté le 29 septembre 2021).
- HEERINGA Wilbert, NERBONNE John & KLEIWEG Peter, 2002, « Validating Dialect Comparison Methods », dans W. Gaul et G. Ritter (éds), *Classification, Automation, and New Media : Proceedings of the 24th Annual Conference of the Gesellschaft für Klassifikation* (University of Passau, 15-17 mars 2000), Berlin / Heidelberg, Springer, p. 445-452. <http://dx.doi.org/10.1007/978-3-642-55991-4_48>.
- HEGGARTY Paul, 2000, « Quantifying Change over Time in Phonetics », dans C. Renfrew, A. McMahon et L. Trask (éds), *Time Depth in Historical Linguistics*, Cambridge, McDonald Institute for Archaeological Research, p. 531-562.
- HEYER Évelyne, 2020, *L'odyssée des gènes*, Paris, Flammarion.
- JACKSON Kenneth, 1960, « Linguistic Geography and the History of the Breton Language », *Zeitschrift für celtische Philologie (ZcP)*, vol. 28, n° 1, p. 272-293. <<https://doi.org/10.1515/zcph.1961.28.1.272>>.
- JACKSON Kenneth, 1967, *A Historical Phonology of Breton*, Dublin, The Dublin Institute for Advanced Studies.
- JOUITTEAU Mélanie, 2021, « The Politeness Systems of Address, Variations across Breton Dialects », *Roczniki Humanistyczne*, vol. 69, n° 11, p. 107-127.
- JULLIEN François, 2016, *Il n'y a pas d'identité culturelle. Mais nous défendons les ressources d'une culture*, Paris, L'Herne.
- KESSLER Brett, 1995, « Computational Dialectology in Irish Gaelic », dans *Proceedings of the Seventh Conference on European Chapter of the Association for Computational Linguistics (EACL '95)*, San Francisco (USA), Morgan Kaufmann Publishers Inc, p. 60-66. <<https://doi.org/10.3115/976973.976983>>.
- LE DÛ Jean, 1972, « Le Nouvel Atlas linguistique de la Basse-Bretagne », *Études celtiques*, vol. 13, fasc. 1, p. 332-345.
- LE DÛ Jean, 2001, *Nouvel Atlas linguistique de la Basse-Bretagne*, Brest, CRBC.
- LE DÛ Jean, 2008, « Les atlas linguistiques : une fenêtre sur le passé des langues », dans G. Raimondi et L. Revelli (éds), *La dialectologie aujourd'hui : Atti del convegno internazionale "Dove va*

la dialettologia ? (Saint-Vincent, Aosta, Cogne, 21-24 septembre 2006), Alexandrie, Edizioni dell'Orso.

LE DÛ Jean & BRUN-TRIGAUD Guylaine, 2003, « La bipartition de la langue bretonne : quelques questions sur la zone vannetaise », dans R. Caprini (éd.), *Parole Romanze. Scritti per Michel Contini*, Alexandrie, p. 245-260.

LE PIPEC Erwan, 2004, « Lecture critique du Nouvel Atlas linguistique de la Basse-Bretagne », *La Bretagne linguistique*, n° 13, p. 157-175.

LE ROUX Pierre, 1924, *Atlas linguistique de la Basse-Bretagne*, Rennes, Plihon et Hommay.

LEINONEN Thérèse, ÇÖLTEKIN Çağrı & NERBONNE John, 2015, « Using Gabmap », *Lingua*, vol. 178, p. 71-83. <<https://doi.org/10.1016/j.lingua.2015.02.004>>.

MATHUSSEK Andreas, 2016, « On the Problem of Field Worker Isoglosses », dans M.-H. Côté, R. Knooihuizen et J. Nerbonne (éds), *The Future of Dialects: Selected Papers from Methods in Dialectology XV* (Groningue, 11-15 août 2014), Berlin, Language Science Press, p. 99-116. <<https://langsci-press.org/catalog/view/81/145/375-1>> [consulté le 29 septembre 2021].

MOSELEY Christopher (éd.), 2010, *Atlas of the World's Languages in Danger*, Paris, Unesco.

MUFWENE Salikoko S., 1996a, « The Founder Principle in Creole Genesis », *Diachronica*, vol. 13, n° 1, p. 83-134. <<https://doi.org/10.1075/dia.13.1.05muf>>.

MUFWENE Salikoko S., 1996b, « The Development of American Englishes: Some Questions from a Creole Genesis Perspective », dans E. W. Schneider (éd.), *Focus on the USA*, Amsterdam / Philadelphie, John Benjamins Publishing Company, p. 231-264.

MUFWENE Salikoko S., 1997, « The Legitimate and Illegitimate Offspring of English », dans L. E. Smith et M. L. Noonan (éds), *World Englishes 2000*, Honolulu, University of Hawai'i Press, p. 182-203.

MUFWENE Salikoko S., 1998, « What Research on Creole Genesis Can Contribute to Historical Linguistics », dans M. S. Schmid et al. (éds), *13th International Conference on Historical Linguistics* (Düsseldorf, 10-17 août 1997), Amsterdam / Philadelphie, John Benjamins Publishing Company, p. 315-338.

MUFWENE Salikoko S., 2001, *The Ecology of Language Evolution*, Cambridge, Cambridge University Press.

MUFWENE Salikoko S., 2002, « Competition and Selection in Language Evolution », *Selection*, vol. 3, no 1, p. 45-56. <<https://doi.org/10.1556/select.3.2002.1.5>>.

NERBONNE John, 2008, « Variation in the Aggregate: An Alternative Perspective for Variationist Linguistics », dans K. Dekker et al. (éds), *Northern Voices: Essays on Old Germanic and Related Topics Offered to Professor Tette Hofstra*, Louvain, Peeters, p. 365-382.

NERBONNE John & HEERINGA Wilbert, 1997, « Measuring Dialect Distance Phonetically », dans J. Coleman (éd.), *Workshop on Computational Phonology*, Madrid, Association for Computational Linguistics, p. 11-18. <<https://aclanthology.org/W97-1102.pdf>> (consulté le 29 septembre 2021).

NERBONNE John & HEERINGA Wilbert, 2010, « Measuring Dialect Differences », dans P. Auer et J. E. Schmidt (éds), *Language and Space. An International Handbook of Language Variation*, vol. 1 : *Theories and Methods* (Handbücher zur Sprach- Und Kommunikationswissenschaft, 30), Berlin, De Gruyter Mouton, p. 550-567.

NERBONNE John, HEERINGA Wilbert, HOUT Erik VAN DEN, KOOI Peter VAN DER, OTTEN Simone & VIS Willem VAN DE, 1996, « Phonetic Distance between Dutch Dialects », dans G. Durieux et al. (éds),

CLIN VI, *Papers from the Sixth CLIN Meeting*, Antwerp, University of Antwerp, Center for Dutch Language and Speech, p. 185-202.

NERBONNE John, HEERINGA Wilbert & KLEIWE Peter, 1999 [1983], « Edit Distance and Dialect Proximity », dans D. Sankoff et J. B. Kruskal (éds), *Time Warps, String Edits and Macromolecules: The Theory and Practice of Sequence Comparison* (2^e éd.), Stanford, CSLI Publications, p. v-xv.

NERBONNE John & KRETZSCHMAR William A. Jr., 2013, « Dialectometry++ », *Literary and Linguistic Computing*, vol. 28, n° 1, p. 2-12. <<https://doi.org/10.1093/lc/fqs062>>.

NICHOLS Johanna, 2013, « The Vertical Archipelago: Adding the Third Dimension to Linguistic Geography », dans P. Auer, M. Hilpert, A. Stukenbrock et B. Szmrecsanyi (éds), *Space in Language and Linguistics: Geographical, Interactional, and Cognitive Perspectives*, Berlin / Boston, De Gruyter, p. 38-60. <<https://doi.org/10.1515/9783110312027.38>>.

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

PELLEN Nadine & SOLLIEC Tanguy, (2021), « Dialectometry and Population Genetics – When Results Converge: The Case of Western Brittany », *Studia Celto-Slavica, Journal of the Learned Association Societas Celto-Slavica*, no 12, p. 63. <<https://doi.org/10.54586/NJGT6315>>.

PLAIN Jean-Yves, 2005, *Tammoù Gwaskin. Au cœur du breton légitime*, Brest, Éditions Armeline.

SAPIR Edward, 1916, *Time Perspective in Aboriginal American Culture: A Study in Method*, Geological Survey Memoir 90, n° 13, Anthropological Series, Ottawa.

SÉGUY Jean, 1971, « La relation entre la distance spatiale et la distance lexicale », *Revue de linguistique romane*, n° 35, p. 335-357.

SÉGUY Jean, 1973a, « La dialectométrie dans l'Atlas linguistique de la Gascogne », *Revue de linguistique romane*, n° 37, p. 1-24.

SÉGUY Jean, 1973b, « La fonction minimale du dialecte », dans G. Straka (éd.), *Les dialectes romans de la France à la lumière des atlas régionaux* (Strasbourg, 24-28 mai 1971), Paris, Éditions du CNRS, p. 27-37.

SÉGUY Jean, 1973c, « Les atlas linguistiques de la France par régions », *Langue française*, n° 18, p. 65-90. <<https://doi.org/10.3406/lfr.1973.5631>>.

SOLLIEC Tanguy, 2019, « 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*, no 10, p. 137-173. <<https://doi.org/10.54586/BLHO8775>>.

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

SOLLIEC Tanguy, (à paraître), *De l'espace dans la langue ? Distribution de la distance linguistique en breton et incidence de facteurs géographiques*.

STEINER George, 1998, *Après Babel. Une poétique du dire et de la traduction* (3^e éd.), Paris, Albin Michel.

TERNES Elmar, 2008, *Compte rendu « Jean Le Dû, Nouvel Atlas linguistique de la Basse-Bretagne (NALBB), Brest, 2001, 2 vol. »*, *Études celtiques*, vol. 36, p. 223-228.

VALLERIE Erwan, 1995, *Traité de toponymie historique de la Bretagne*, Le Relecq-Kerhuon, An Here.

WMFFRE Iwan, 2007, *Breton Orthographies and Dialects. The Twentieth-Century Orthography War in Brittany* (Contemporary Studies in Descriptive Linguistics, 19), 2 vol., Oxford, Bern, Berlin, Peter Lang.

NOTES

1. Bien que ces corpus soient de petite taille, une analyse quantitative systématique de la parole spontanée apporte beaucoup de résultats inattendus et inédits, confirmant ainsi des conclusions provenant d'approches qualitatives ou les remettant en cause.
2. Je reprends cette idée à Gary German, un de mes directeurs de thèse et pionnier de la dialectométrie bretonne. Qu'il y voie le témoignage de toute ma gratitude.
3. Les données de cet atlas sont à présent disponibles en ligne. Le cas du breton peut être consulté à l'adresse suivante : <<http://www.unesco.org/culture/languages-atlas/fr/atlasmap/language-id-336.html>> (consulté le 24 septembre 2021).
4. La formule « distance linguistique » est souvent employée pour désigner le degré de similarité/différence entre deux variétés linguistiques, quelles qu'elles soient.
5. Toutefois, à la suite d'une discussion avec Guylaine Brun-Trigaud, il s'avère que cette question du nombre minimal de cartes à traiter semble devoir être envisagée autrement afin de gagner en pertinence. Quelle est l'incidence du nombre de cartes traitées sur les résultats obtenus ? Comment le signal capté par l'analyse évolue en fonction du nombre de cartes traitées qu'elles soient au nombre de 25, 250 ou 500 ?
6. *Cluster analysis* en anglais.
7. Jean Le Dû dans le NALBB distingue pour certains sons ([w] ou [ə] par exemple) la force avec laquelle ces sons ont été réalisés lors des enquêtes. Il a rendu graphiquement les réalisations atténuées par des signes en exposant.
8. L'algorithme se contente de relever les différences graphiques entre les chaînes de caractères et ne prend en considération leur nature. Ceci impose une préparation des données en les encodant grâce à une police de caractères adéquates de façon à ce que les signes diacritiques d'un caractère phonétique ne soit pas traité comme des unités graphiques autonomes. Pour notre part, nous avons eu recours à la police de caractères Times Unicode CD de l'université de Neuchâtel.
9. Il va de soi que 20 % de « similarité » sont équivalents à 80 % de différence. Nous avons privilégié le terme similarité que nous trouvons plus intuitif pour décrire les résultats obtenus et avons fait usage de la distance linguistique en tant que notion générale afin d'appréhender un domaine linguistique.
10. Le point 69 Caurel a été laissé de côté en raison des données lacunaires pour cette commune. Le breton local y a en effet disparu au cours des enquêtes.
11. À l'est du domaine, le long de la frontière linguistique, à l'interface entre les domaines celtique et roman.
12. Au nord-ouest dans le Trégor. Petit clin d'œil, Plougrescant est la commune dont Jean Le Dû s'est attaché à décrire le breton local au long de sa vie et c'est à présent là où il repose.
13. Malgré le faible nombre de segments pour ce pôle et les suivants, nous donnons néanmoins les valeurs à titre de comparaison. Ces valeurs sont donc à manipuler avec précaution.
14. Le trégorrois particulièrement doit être vu comme un dialecte foncièrement nordiste influencé à des degrés divers par la zone centrale.
15. Une opposition dialectale fondamentale en breton distingue les parlers bretons vannetais de ceux en usage dans les autres parties de l'espace bretonnant. Ces derniers sont souvent qualifiés de KLT, abréviation pour *Kerne, Leon, Treger*, nom breton des anciens évêchés de Cornouaille,

Léon, Trégor. Le breton du vannetais a conservé une accentuation finale des mots alors que celle-ci est remontée en position pénultième au XI^e siècle. De ce fait, certaines réalisations anciennes se sont maintenues contrastant, dès lors, avec les évolutions des parlers KLT (Le Dù & Brun-Trigaud, 2003).

16. Plus le niveau de la différenciation linguistique au sein d'une famille de langues est important, plus la durée de temps nécessaire pour aboutir au développement de telles différenciations doit être envisagé comme longue.

17. Conclusions que nous estimons pour notre part, tout à fait pertinentes. Nous pensons que nous sommes parvenus à des résultats différents parce que notre objet d'analyse, la distance linguistique, n'est pas le même que celui de Falc'hun, la diffusion des innovations linguistiques avant tout.

18. Par souci de concision, nous renvoyons à notre thèse pour le détail de ces traits linguistiques (Sollic, 2021 : 10).

19. Le registre indique entre 12 et 17 patients pour 100 000 habitants. Il s'agit de la tranche la plus élevée du relevé.

20. Source : <<http://www.genet.sickkids.on.ca/StatisticsPage.html>> (consulté le 29 septembre 2021). Une mutation désigne une modification de la séquence d'ADN du gène en question.

21. Seuls les pôles bien caractérisés par des traits linguistiques ont été indiqués dans le tableau. Nous n'avons pas relevé de correspondance dans le cas des autres pôles. Pour plus de détails, se référer à Pellen et Sollic (2021).

22. *Pool of features* en anglais.

23. Brittonique désigne au sein de la famille des langues celtiques le sous-ensemble regroupant le gallois, le cornique et le breton. On entend également par brittonique l'état historique commun à ces langues entre les I^{er} av. J.-C. et VII^e siècles apr. J.-C.

24. Un débat existe sur la date de l'extinction du gaulois entre les tenants d'une extinction précoce et les partisans d'un maintien tardif, notamment dans les campagnes isolées.

25. Le chapitre 6 de notre thèse donne un aperçu de la variété des langues étudiées de ce point de vue (Sollic, 2021).

26. Ces éléments sont présentés au chapitre 8 de notre thèse (*ibid.*).

RÉSUMÉS

Cet article présente les résultats d'une analyse portant sur la langue bretonne par le prisme de la notion de distance linguistique, à partir de données issues du *Nouvel Atlas linguistique de la Basse-Bretagne* de Jean Le Dù (2001). La méthodologie mise en œuvre est celle de la dialectométrie, approche quantitative qui permet de réaliser des mesures de similarité linguistique. Les résultats obtenus montrent que la distance linguistique est organisée dans l'espace et se fait l'écho de structures socio-historiques appartenant à plusieurs strates temporelles. On identifie au sein du domaine bretonnant la présence de plusieurs pôles de similarité linguistique. Leur répartition reflète la géographie locale des mutations d'une maladie génétique, la mucoviscidose. Ces observations apportent des éclairages sur les premiers temps de la Bretagne au Moyen Âge et sur l'émergence de la langue bretonne.

This article analyses the Breton phonetic data provided by Jean Le Dù's *Nouvel Atlas linguistique de la Basse-Bretagne*. It is studied from the standpoint of a concept called "linguistic distance".

To this end, I used a quantitative approach known as dialectometry, the aim of which is to measure linguistic similarity between the points of a linguistic atlas. The results obtained demonstrate that linguistic distance is spatially-structured in Celtic-speaking Brittany and appears to reflect certain socio-historical patterns which are linked to the history of Breton. The poles of linguistic similarity, which vary considerably in size, are also identified across Western Brittany. Their distribution matches the geography of various cystic fibrosis mutations, a genetic disease common in the area. These observations shed light on the history of Brittany in the early Middle Ages and on the development of the Breton language formation.

INDEX

Mots-clés : dialectométrie, breton (langue), géolinguistique, dialectologie, histoire des langues, atlas linguistiques

Keywords : dialectometry, Breton language, geolinguistics, dialectology, history of languages, linguistic atlases

AUTEUR

TANGUY SOLLIEC

Centre de recherche bretonne et celtique (CRBC), Université de Bretagne occidentale (UBO, Brest)
tanguy.solliec@gmail.com