HAL
open science

Explorer la variation en breton grâce à la dialectométrie : la Basse-Bretagne considérée par la distance linguistique

Tanguy Solliec

To cite this version:

Tanguy Solliec. Explorer la variation en breton grâce à la dialectométrie : la Basse-Bretagne considérée par la distance linguistique. La Bretagne Linguistique, 2017, 21, pp.29-61. hal-02930224

HAL Id: hal-02930224

https://hal.univ-brest.fr/hal-02930224v1

Submitted on 4 Sep 2020

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.

La Bretagne Linguistique

21 | 2017
Varia

Explorer la variation en breton grâce à la dialectométrie : la Basse-Bretagne considérée par la distance linguistique

Exploring variation in Breton through dialectometry: Lower Brittany considered from the linguistic distance perspective

Tanguy Solliec

Édition électronique
URL : http://journals.openedition.org/lbl/307

Éditeur
Université de Bretagne Occidentale – UBO

Édition imprimée
Date de publication : 1 mai 2017
Pagination : 29-61
ISBN : 979-10-92331-31-8
ISSN : 1270-2412

Référence électronique
Tanguy Solliec, « Explorer la variation en breton grâce à la dialectométrie : la Basse-Bretagne considérée par la distance linguistique », *La Bretagne Linguistique* [En ligne], 21 | 2017, mis en ligne le 01 mai 2020, consulté le 25 juin 2020. URL : http://journals.openedition.org/lbl/307

La Bretagne Linguistique est mise à disposition selon les termes de la Licence Creative Commons Attribution 4.0 International.

Tanguy SOLLIEC*

Explorer la variation en breton grâce à la dialectométrie : la Basse-Bretagne considérée par la distance linguistique

La variation linguistique en breton est attestée de longue date et constatée par les locuteurs mais aussi les voyageurs au XIXe[1]. Elle a été enregistrée et décrite par plusieurs atlas, *l'Atlas Linguistique de la Basse-Bretagne* de Pierre Le Roux[2] et le *Nouvel Atlas Linguistique de la Basse-Bretagne*[3] de Jean Le Dû, repères importants dans l'étude de la langue, dans le sillage des avancées de la dialectologie contemporaine. Ce champ disciplinaire a joué au XIXe siècle un rôle important en linguistique, dans la mesure où elle venait apporter des données inédites et pouvait remettre en cause certaines affirmations théoriques d'alors en les confrontant à la réalité des faits et des pratiques. Cependant, son appareil critique et ses méthodes ne se sont que peu renouvelés

* Doctorant en linguistique, UBO/CRBC (EA 4451), tanguy.solliec@gmail.com

1. Fañch BROUDIC, « La perception de la variation diatopique à travers le temps », *La Bretagne linguistique*, 13, 2004, p. 31-52.
2. Pierre LE ROUX, *Atlas linguistique de Basse-Bretagne*, Rennes & Paris, Plihon-Hommay, 1924-1963, 6 volumes, 77 points d'enquête. 600 cartes. [Désormais abrégé en ALBB]
3. Jean LE DÛ, *Nouvel Atlas Linguistique de la Basse-Bretagne*, Brest, CRBC, 2001, 2 volumes. 187 points d'enquête, 600 cartes. [Désormais abrégé en NALBB]

depuis son apparition en tant que champ de recherche autonome au commencement du XIXe siècle [4].

Le développement de l'informatique et de nouvelles réflexions théoriques sur la variation linguistique dans les langues du monde ont contribué à la mise en œuvre de nouvelles approches en dialectologie. Ainsi, l'article que nous proposons à la lecture reflète les premiers pas de notre recherche doctorale consacrée à un traitement quantitatif de données géolinguistiques du breton par le biais d'une méthode informatique dite « dialectométrie ». Elle vise à établir une distance linguistique entre les points du NALBB puis à en considérer sa structure et son agencement.

Notre propos portera sur la prise en compte de cette distance linguistique dans le traitement de données dialectales du breton, sur la façon de la mesurer et sur l'observation des éléments qui la constituent. Nous revenons ici plus particulièrement sur des résultats obtenus lors d'une première démarche exploratoire dont les développements techniques ont été présentés par ailleurs [5]. L'analyse de ces résultats inédits sera pour nous l'occasion de développer une réflexion heuristique sur la nature des catégories d'analyse à mettre en œuvre dans une approche alliant description qualitative et outils quantitatifs.

Dans un premier temps, nous expliciterons les principes et les enjeux de l'analyse dialectométrique avant d'expliciter la méthodologie retenue dans le cadre de cette étude. Ensuite, nous présenterons les premiers résultats auxquels nous sommes parvenus et nous en étudierons la structure. Puis, nous porterons notre attention sur les diverses possibilités d'améliorations techniques avant de réfléchir aux implications potentielles de cette approche sur la description de faits de langue.

---

4. Jack K. CHAMBERS & Peter TRUDGILL, *Dialectology*, Cambridge, Cambridge University Press, 2e édition, 1998.
5. Guylaine BRUN-TRIGAUD, Tanguy SOLLIEC, Jean LE DÛ, « A new dialectometric approach applied to the Breton language », dans J. Nerbonne, M.-H. Côté, R. Knooihuizen (éd.), *The future of dialects: Selected papers from Methods in Dialectology XV*, Berlin, Language Science Press, 2016. Accessible à l'adresse suivante http://langsci-press.org/catalog/book/81

Enjeux et méthodologie de la dialectométrie

La distance linguistique à la source de la dialectométrie

Au cœur de la démarche de la dialectométrie se trouve le concept de distance linguistique. Cette distance ici n'est pas géographique mais d'ordre symbolique et s'inscrit sur une échelle de valeur abstraite plus ou moins importante selon la ressemblance entre les éléments comparés. Séguy est le premier à avoir élaboré une telle méthodologie[6] afin de comparer les données recueillies dans le cadre de *l'Atlas Linguistique de la Gascogne*[7]. La démarche de la dialectométrie s'appuie sur des données principalement issues d'atlas linguistiques. Ce travail privilégie l'utilisation d'outils conceptuels provenant de plusieurs disciplines en vue de constituer une approche interprétative des données recueillies[8] et prolonge ainsi l'analyse des cartes dialectales.

En établissant une mesure de la distance linguistique entre deux points d'un atlas linguistique, on établit une comparaison entre deux formes linguistiques pour lesquelles il est possible d'établir un taux de similarité (ou au contraire, de différenciation). La multiplication de telles mesures accompagnées de tris statistiques permet de dessiner des espaces caractérisés par des taux de différenciation identique et formant des zones dans lesquelles la proximité linguistique est forte.

L'analyse dialectométrique part du postulat que traiter une grande quantité d'informations permet d'une part, d'objectiver l'analyse conduite et, d'autre part, de s'assurer de la représentativité des faits linguistiques décrits.

La principale difficulté à laquelle se confrontèrent les premiers chercheurs qui ont investi ce domaine fut d'élaborer les outils et les méthodologies pour atteindre un tel objectif mais également de rendre compte de leurs investigations. Les premières recherches ont porté sur

---

6. Jean SÉGUY, « La relation entre la distance spatiale et la distance lexicale », *Revue de Linguistique Romane*, 35, 1971, p. 335-357 et « La dialectométrie dans l'Atlas linguistique de la Gascogne », *Revue de Linguistique Romane*, 37, 1973, p. 1-24.
7. Jean SÉGUY *et alii*, *Atlas linguistique et ethnographique de la Gascogne*, Toulouse & Paris, Éditions du CNRS, 1954-1973, 6 volumes, 2 531 cartes, désormais abrégé en ALG.
8. Jean-Philippe DALBERA, « Géolinguistique : un nouveau souffle ? », *Revue Belge de Philologie et d'Histoire*, 80, 2002, p. 831-849.

l'élaboration de mesures à même d'évaluer la distance linguistique [9]. En travaillant sur les degrés de concordance ou de divergence de plusieurs faits de différente nature (morphologique, phonologique, lexical) dans l'espace linguistique gascon, Séguy a été en mesure d'établir un « gradient de gasconité » (carte 2525 de l'ALG) révélant sa structure générale. Sa méthode consistait à comparer sur un point précis de lingustique les réponses apportées dans les différents lieux d'enquête et à comptabiliser les différences qui pouvaient s'établir.

Les développements de la dialectométrie grâce à l'informatique

Hans Goebl en Autriche a repris les problématiques développées par Séguy et dès ses premiers travaux [10], il a fait entrer la dialectométrie dans l'ère de l'informatique et de son automatisation par le biais d'algorithmes. Au fil de son travail sur plusieurs atlas et différentes langues [11], il a affiné la méthodologie de la discipline. Il a notamment développé une réflexion sur la représentation cartographique des analyses dialectométriques, a mis à profit des outils issus des sciences du vivant (dendrogrammes) pour analyser et représenter les résultats de ses analyses. Il a inscrit pleinement la dialectométrie dans une optique pluridisciplinaire.

Sa démarche reflète une ambition théorique importante qui se donne pour objectif, grâce à l'examen de données issues d'atlas linguistiques, de repérer les principes de la gestion basilectale et dialectale de l'espace géographique par *Homo loquens* [12].

La dialectométrie propose un renouvellement de la dialectologie en faisant appel à des techniques et à des outils provenant de plusieurs dis-

---

9. Brett KESSLER, « Computational Dialectology in Irish Gaelic », *Proceedings of the European ACL*, Dublin, Association for Computational Linguistics, 1995.
10. Hans GOEBL, *Dialektometrie. Prinzipien und Methoden des Einsatzes der Numerischen Taxonomie im Bereich der Dialektgeographie*, Vienne, Denkschriften der Österreichischen Akademie der Wissenschaften, 1982.
11. H. GOEBL, « Dialectometry and quantitative mapping », dans A. Lameli, R. Kehrein et S. Rabanus (éd.), *Language and Space. An International Handbook of Linguistic Variation, volume 2: Language Mapping*, Berlin, de Gruyter, 2010.
12. H. GOEBL, « La dialectométrie corrélative. Un nouvel outil pour l'étude de l'aménagement dialectal de l'espace par l'homme », *Revue de linguistique romane*, 69, 2005, p. 321-367.

ciplines. L'instrument de synthèse usuel de la géographie linguistique est principalement l'isoglosse, une ligne imaginaire créée par l'analyste et séparant dans l'espace une zone marquée par un fait linguistique d'une autre zone d'où ce fait est absent ou différent. Il s'agit certes d'un outil de lecture utile mais ses faiblesses d'un point de vue théorique ont déjà été soulignées de longue date[13]. La dialectométrie propose une ouverture pour dépasser les limites que pose la notion d'isoglosse. En effet, plutôt que d'observer si pour un point d'enquête précis correspondent une ou plusieurs isoglosses, les données sont agrégées et représentées sous forme cartographique en gardant la possibilité de considérer la relation de chaque point d'enquête à tous les autres de l'atlas étudié selon la distance linguistique qui les sépare.

Par la suite, Kessler, en prenant appui sur l'algorithme dit '*edit distance*' ou distance de Levenshtein a ouvert une nouvelle voie à la méthodologie de la dialectométrie[14]. Cet outil a été élaboré initialement dans le but de comparer des séries de caractères ou des suites de données dans le champ de l'informatique. Il permet également de calculer la distance séparant différentes suites de signes. L'auteur a montré qu'il était possible d'utiliser cet outil pour traiter des données dialectales de façon automatisée. L'algorithme procède à une série de manipulations pour passer d'une chaîne de caractères à une seconde. En considérant que ces opérations constituent autant de différences entre ces deux suites, le nombre de ces changements de caractères équivaut à un score de distance phonétique.

Heeringa a validé cette approche en comparant dans sa thèse[15] différentes méthodes de mesure mathématique appliquée à des formes phonétiques et a validé l'utilisation qu'il est possible de faire de cette approche en traitant des formes provenant d'atlas linguistiques de plusieurs langues (néerlandais, norvégien). Il est parvenu également à intégrer un grand nombre de données et à automatiser leur comparaison.

---

13. J. K. CHAMBERS & P. TRUDGILL, *op. cit.*, p. 7.
14. B. KESSLER, art cit.
15. Wilbert HEERINGA, *Measuring dialect pronunciation differences using Levenshtein distance*, Groningue, University of Groningen, PhD thesis, 2004.

La dialectométrie et les langues celtiques

Le domaine des langues celtiques ne constitue pas une *terra incognita* pour les analyses dialectométrique. En effet, dès 1986, Elsie[16] propose de tester la validité de la structuration dialectale des langues goïdéliques modernes grâce un traitement quantitatif sur la base de 184 unités lexicales et aboutit à la conclusion que si la distinction entre les trois langues se manifeste, la répartition dialectale ne correspond pas à la réalité géographique.

Kessler, afin de traiter de données de l'irlandais, a introduit une innovation méthodologique en faisant appel à la distance de Levenshtein comme nous l'avons mentionné précédemment. Grâce à ce nouvel outil, il a confirmé les analyses précédentes d'Elsie et a montré tout l'intérêt que représente cet algorithme pour la dialectométrie. Il a précisé ainsi la situation particulière de l'Ulster dans l'espace gaélophone et la façon dont se distinguent gaéliques irlandais, mannois et écossais. Nous présenterons par la suite les travaux ayant porté sur le breton.

Bilan de la dialectométrie

Toutefois, si la dialectométrie a connu des développements importants depuis une quinzaine d'années, il n'en demeure pas moins que la réflexion a essentiellement porté sur l'intérêt de prendre en compte un grand nombre de faits d'un point de vue quantitatif[17] d'une part et d'autre part sur les possibilités de repérer dans ces corpus des schémas d'organisation[18]. L'approche classificatoire proposée par la dialectométrie découle des possibilités du traitement computationnel des données et de la capacité à les trier automatiquement[19]. Cela contribue donc à renouveler l'approche de problématiques importantes en dialectolo-

16. Robert ELSIE, *Dialect Relationships in Goidelic, a study in Celtic dialectology with 8 figures*, Hambourg, Helmut Buske Verlag, 1986.
17. John NERBONNE, « Data Driven dialectology », *Language and Linguistics Compass*, 3 (1), 2009, p. 175-198.
18. John NERBONNE, « Identifying Linguistic Structure in Aggregate Comparison », *Literary and Linguistic Computing*, 21, 2006, p. 463-476.
19. W. HEERINGA, *op. cit.*

gie comme la structuration des *continua* linguistiques et le repérage des limites dialectales [20] au-delà de l'individuation des isoglosses.

Mais, une tendance récente dans les travaux de dialectométrie vise également à analyser les faits dialectaux plus finement. Il est ainsi possible de développer des mesures de distance linguistique en travaillant sur la différenciation entre voyelles ou bien en analysant les règles phonologiques à l'œuvre dans un dialecte précis afin de cerner les faits à la source de la distance linguistique. Ces derniers développements [21] accompagnent le développement d'outils accessibles en ligne pour réaliser des analyses dialectométriques comme DiaTech [22] ou Gabmap [23].

Le breton et la dialectométrie

Premiers travaux

Dans le domaine bretonnant, l'essentiel des analyses géolinguistiques est basé sur l'ALBB de Pierre Le Roux. Même s'il n'a pas pratiqué la dialectométrie, la figure du chanoine Falc'hun demeure toutefois incontournable dans ce domaine car il est le premier à avoir conduit une exploitation minutieuse de l'ALBB et à avoir décrit la structuration géolinguistique de la Basse-Bretagne en trois espaces distincts. Il a également souligné le rôle majeur que les villes et les axes routiers ont joué dans la diffusion des innovations linguistiques lexicales ou morphologiques [24].

German est le premier à avoir initié de tels travaux [25] dans le cadre de sa thèse consacrée à la description du parler breton de Saint-Yvi

---

20. Wilbert HEERINGA & John NERBONNE, « Dialect areas and dialect continua », *Language Variation and Change*, 13(3), 2001, p. 375-400.
21. Martijn WIELING & John NERBONNE, « Advances in dialectometry », *Annual Review of Linguistics*, 1, 2015, p. 243-264, doi: 10.1146/annurev-linguist-030514-124930
22. Gotzon AURREKOETXEA, Karmele FERNANDEZ-AGUIRRE, Jesús RUBIO, Borja RUIZ, Jon SÁNCHEZ, « 'DiaTech': a new tool for dialectology », *Literary and Linguistic Computing*, 28, 2013, p. 23–30.
23. Therese LEINONEN, Çağrı ÇÖLTEKIN & John NERBONNE, « Using Gabmap », *Lingua*, sous presse. doi:10.1016/j.lingua.2015.02.004
24. François FALC'HUN, *Perspectives nouvelles sur l'histoire de la langue bretonne*, Paris, UGF, 1981.
25. Gary GERMAN, « Une méthode dialectométrique (assistée par ordinateur) pour l'analyse des atlas linguistiques », *La Bretagne linguistique*, 7, 1991, p. 177-213.

(14 km à l'est de Quimper)[26]. Pour observer la façon dont le parler de Saint-Yvi s'intègre dans l'espace dialectal du breton, German a sélectionné une série de faits phonétiques à partir des données de l'ALBB et a établi 20 variables. Pour chacune d'entre elles, il a examiné les réponses de chaque point de l'ALBB avec les faits correspondants dans le breton de Saint-Yvi. Selon la similarité des données examinées, il leur a attribué un indice compris entre 0,00 et 1,00, établissant ainsi la distance dialectale entre Saint-Yvi et les autres points de l'ALBB.

Une fois les calculs établis, il lui a été possible de découper la Basse-Bretagne en différents secteurs en fonction de leur proximité avec Saint-Yvi. En utilisant le programme Lerman de traitement de données, il a établi une classification arborescente de celles-ci en constituant des classes plus ou moins nombreuses (en 4, 65 ou 49 catégories). La comparaison de ces zones dialectales avec d'autres types de pratiques culturelles ou sociales ayant une profondeur historique plus ou moins importante, montre des concordances intéressantes. Cette approche stimulante se base avant tout sur l'examen de la variation portant sur les consonnes qui manifestent une stabilité relativement importante en diachronie. À partir de l'étude de données synchroniques, German est parvenu à accéder à une temporalité longue et à éclairer la structuration symbolique d'un territoire donné.

Cette première approche a été prolongée par Costaouec[27] qui, également à la suite de sa thèse traitant du parler breton de La Forêt-Fouesnant[28] s'est engagé dans une démarche identique. Il en a affiné l'analyse en travaillant avec la notion de gradient et en prenant pour centre La Forêt-Fouesnant. À partir de données phonétiques, il a considéré les différentes catégories du gradient selon l'importance de la distance linguistique entre le parler des points de l'ALBB et La Forêt-Fouesnant. Il a fait à l'instar de German le constat que, d'une part, la différenciation linguistique est corrélée à la distance géographique et, d'autre part, que les pratiques sociales coïncident avec la répartition de

26. G. German, *Une étude linguistique sur le breton de Saint-Yvi*, Brest, UBO, thèse de doctorat, 1984, n.p.
27. Denis Costaouec, « Linguistic geography of Breton and sociocultural motivations », *STUF - Language Typology and Universals*, 65(1), 2012, p. 47-64.
28. Denis Costaouec, *Le breton parlé à La Forêt-Fouesnant (Finistère). Pratique actuelle, problèmes de phonologie et de syntaxe*, Paris, Université Paris 5, thèse de doctorat, 1998, n.p.

la distance linguistique. En outre, Costaouec en examinant le pourtour immédiat de La Forêt-Fouesnant a repéré une rupture importante dans le continuum dialectal malgré la proximité géographique. En enquêtant plus précisément sur la limite entre les gradients 100 et 350 dans la commune d'Elliant, il a remarqué que des rivières venaient établir une forte différenciation dans les pratiques linguistiques et la structuration des réseaux de sociabilité, influencées plus par la géographie physique que les divisions administratives. Costaouec montre ici qu'une limite dialectale, au-delà d'être une simple isoglosse, peut posséder une profondeur sociale et historique.

Ces études montrent donc tout l'intérêt que représente une approche dialectométrique des données du breton. Ces travaux se sont appuyés tous deux sur les données de l'ALBB en les comparant à celles d'endroits non-pris directement en compte par cet atlas et, en outre, relativement proches l'un de l'autre. Ces deux études ont porté sur des faits phonétiques, German ayant fait usage de variables préétablies, utilisées pour leur valeur diachronique. Tous les deux ont fait le choix de concentrer leur analyse uniquement sur la variation touchant les consonnes, délaissant ainsi les voyelles beaucoup plus instables. Ces premiers travaux menés à partir de l'ALBB peuvent être poursuivis et approfondis grâce à des données nouvelles, celles du NALBB et à des approches techniques renouvelées, l'algorithme de Levenshtein.

Données et zone étudiées

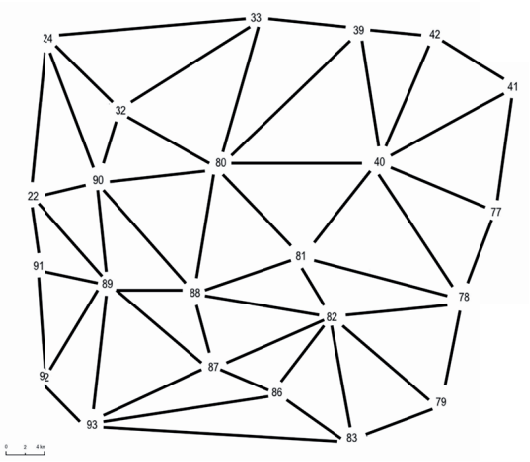
Le NALBB porté par Jean Le Dû depuis 1969 et paru en 2001 couvre l'ensemble de la Basse-Bretagne. Près d'une commune sur 3 y figure par l'intermédiaire des 187 points présents. Les six cents cartes contiennent environ 100 000 formes linguistiques. L'ensemble des informations présentes a été transcrit par Jean Le Dû lui-même de façon à éviter les biais phonologiques qu'induiraient plusieurs transcripteurs. Néanmoins, les notations sont rendues avec une grande finesse, permettant d'envisager des analyses détaillées.

Nous avons retenu pour cette étude 23 points issus du NALBB et situés au sein d'un quadrilatère approximatif compris entre Guimiliau (point 24) au nord-ouest, Plougonver au nord-est (point 41), Paule (point 79) au sud-est et Pleyben (point 92) au sud-ouest.

Carte de la zone étudiée [29]

Chacun des points est rassemblé à ses voisins dans le cadre d'un réseau constitué de 52 segments distincts en vue de la comparaison ultérieure de chaque forme avec celle des points voisins, comme le montre le schéma suivant :

Réseau de la zone étudiée



29. Carte réalisée grâce au site de dialectométrie Gabmap, www.gabmap.nl

Pour réaliser ce premier essai de dialectométrie, nous avons retenu 165 cartes différentes sélectionnées sur le critère qu'elles ne comportaient chacune qu'un seul et unique lexème pour exprimer la notion recherchée, dans la zone concernée, nous concentrant ainsi sur la variation phonétique. Ce principe a été le seul retenu pour établir notre corpus et nous n'avons pas tenu compte de variables linguistiques précises.

Notre objectif a été toutefois de rassembler suffisamment de données pour qu'il n'y ait pas surreprésentation de certains phénomènes.

Une nouvelle approche dialectométrique du breton

Objectifs

Notre démarche vise à appliquer au domaine bretonnant les avancées techniques apparues dans le champ de la dialectométrie et à traiter d'un point de vue quantitatif les données du NALBB en prenant en compte la distance linguistique entre les différents points de l'atlas. Nous souhaitons également repérer dans le cadre d'une telle analyse dialectométrique, les faits linguistiques à la source de la variation diatopique en breton.

L'utilisation de cet outil a donc porté sur une zone restreinte de la Basse-Bretagne pour le tester d'une part et l'améliorer par la suite mais également, pour sonder les phénomènes phonétiques à l'œuvre dans le domaine [30]. Ce premier essai nous a permis de mettre à l'épreuve nos premières hypothèses et d'affiner notre approche.

Méthodologie

Nous avons ainsi examiné 3795 formes différentes en établissant un total de 8745 comparaisons grâce à l'algorithme de Levenshtein. La version que nous avons déployée, dite algorithme de Damerau-Levenshtein, a été revue par Guylaine Brun-Trigaud de façon à ce que celui-ci nous renvoie la nature des opérations qu'il a effectuées.

L'algorithme compare deux chaînes de caractères et propose une succession d'opérations pour passer d'un ensemble à l'autre. Chaque opération engendre un coût qui s'ajoute à celui des opérations déjà

30. G. Brun-Trigaud *et alii*, art. cit.

réalisées auparavant. Au terme de la procédure, la somme des différents changements conduit donc à une valeur plus ou moins importante, dans notre cas, il s'agit de la distance linguistique. La distance de Levenshtein ne procède qu'avec trois opérations (remplacement d'un caractère par un autre, insertion d'un caractère ou suppression d'un caractère) qui peuvent être cumulées ou délaissées selon les besoins. Dans le cas suivant, nous comparons deux formes issues de la carte 163 NALBB 'le soleil' AN HEOL.

(77) Saint-Servais ən ˈɛwəl ~ (78) Locarn əm̥ ˈhɛo̤l

(77) Saint-Servais	ə	n			ɛ	w	ə	l
		remplacement de n par m̥		insertion de h		suppression de w	Remplacement de ə par o̤	
(78) Locarn	ə	m̥		h	ɛ		o̤	l
		1		1		1	1	
	total de 4 changements							

Le nombre de syllabes n'ayant pas varié, l'accent ne s'étant pas déplacé, nous comptons 4 changements ce qui nous permet d'établir un taux de différenciation de 66,6 %.

L'algorithme qui a été intégré dans une base de données comportant les informations du NALBB compare de façon automatique les données voulues, caractère par caractère, soit dans notre cas de figure, les différents sons constitutifs des réponses de l'atlas. Cependant cette opération ne va toujours sans heurt comme indiqué par ailleurs [31]. Nous avons donc été amenés à éditer manuellement un nombre important d'éléments en raison de ces difficultés.

Le déploiement de l'algorithme dans un tableur Excel nous a permis, une fois l'ensemble des formes comparées, de recueillir un total de 11 949 modifications de caractères que nous désignerons à présent par le terme d'alternance. Nous avons été amenés à regrouper ces alternances par types. En effet, si pour un segment reliant un point *a* à un point *b* nous observons une modification de [a] en [ə] et pour

31. *Ibidem*, p. 140 et suiv.

un segment d'un point *b* vers un point *c*, la modification de [ə] en [a], il s'agit d'un seul et même type d'alternance que nous notons [a]/[ə]. Chaque type d'alternance possède à l'issue de notre analyse un nombre plus ou moins important d'occurrences parmi l'ensemble des 11 949 modifications relevées.

L'intérêt d'utiliser cet outil est que non seulement il offre une mesure de la distance linguistique entre deux mots mais qu'il procède en outre à une commutation automatisée des différents sons. Cette approche permet d'obtenir une vue contrastive entre les prononciations examinées en obtenant les éléments qui les distinguent. Procéder ainsi permet en outre de contrôler la façon dont la distance linguistique se calcule et de s'assurer du bon fonctionnement de l'algorithme. Cette nouvelle approche nous permet à la fois de quantifier la variation linguistique mais également de repérer les faits qui la produisent et *in fine* de la qualifier.

Résultats

Afin de procéder à l'analyse des résultats obtenus, nous examinerons attentivement leur structuration, les types d'alternance relevés, ainsi que leur fréquence respective. Si les variables présentes en breton sont bien connues de la dialectologie, nous ne nous y référerons pas pour le moment. Nous nous proposons de rechercher des catégories d'analyse pertinentes pour rendre compte des résultats obtenus et les comprendre. Cette démarche revêt un caractère exploratoire. En effet, alors que souvent les résultats issus de la dialectométrie traitent essentiellement de l'importance de la distance linguistique, nous traiterons, quant à nous, de la nature de celle-ci. Et dans ce cas de figure, nous ne disposons pas de point de comparaison.

Structuration des données

Une première vue d'ensemble nous fait constater que les 11 949 alternances relevées se répartissent entre 442 types différents avec des intensités variables. Ainsi, une seule alternance (ʀ/ʁ) représente plus de 5 % des occurrences recueillies (5,9 %). La seconde (a/ə) compte elle pour moins de 5 % de celles-ci (4,4 %). L'importance des autres

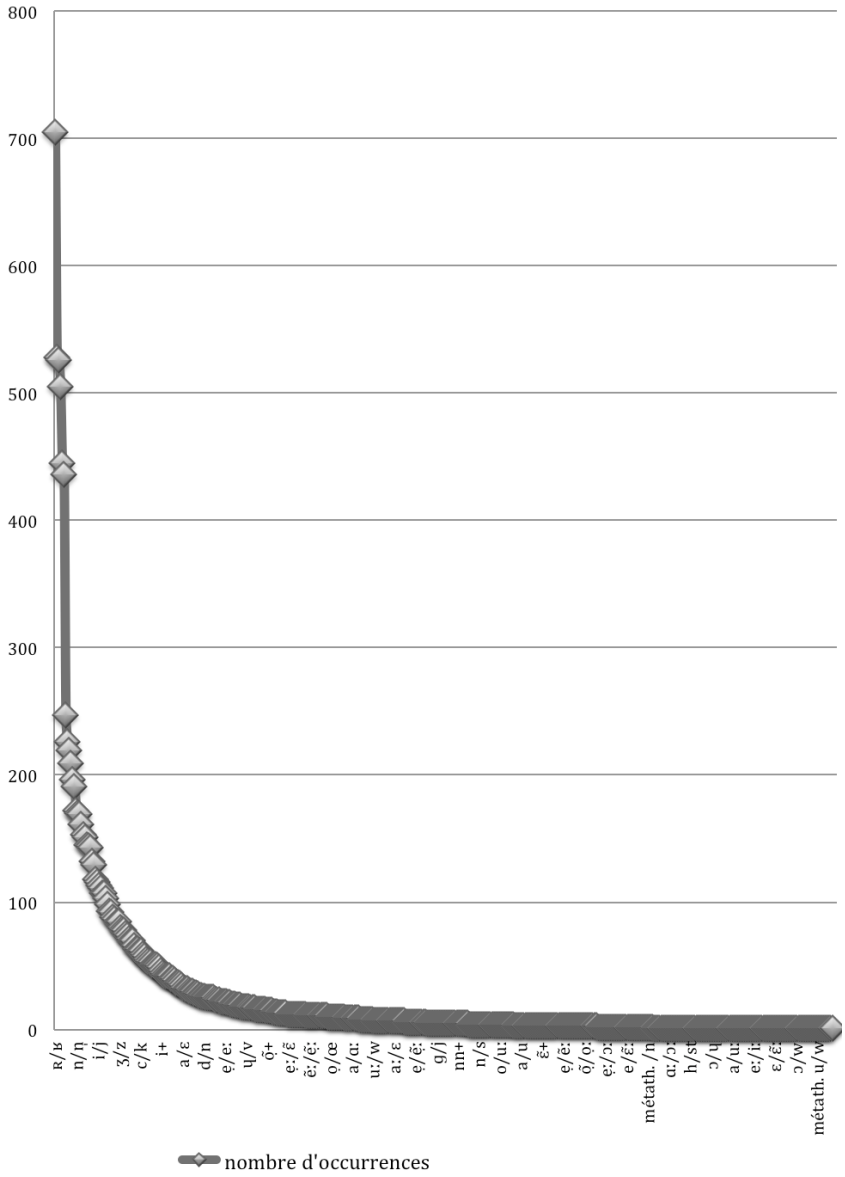
alternances va de façon décroissante et finalement les 25 premiers types recouvrent 53 % des occurrences relevées. Une telle répartition suggère d'emblée un éclatement des résultats qui ne ferait que refléter une situation de continuum linguistique. Dans cette situation, les données de chaque point d'enquête ne se distinguent de celles des localités voisines que par quelques traits fugaces. La distribution des alternances de façon décroissante est représentée dans la figure 1 ci-contre.

La répartition hiérarchique de ces données prend la forme d'une courbe de distribution de fréquence (dite courbe de Gauss) scindée en deux dont on n'aurait gardé que la partie droite. La courbe démarre ainsi par la valeur la plus élevée de la courbe et décroît très vite avant de se stabiliser à une valeur très faible pour un très grand nombre des types d'alternance relevés dans le corpus. Cette distribution non-linéaire indique donc que la majorité des occurrences que nous avons relevées se répartit entre quelques types d'alternance. Ainsi, d'un point de vue quantitatif, les 25 premières alternances représentent avec 6 336 occurrences, 53 % de l'ensemble des alternances relevées. Et inversement, la majorité des différents types d'alternance ne comportent qu'un nombre limité d'occurrences.

Il est intéressant de relever que ce type de distribution s'observe régulièrement lors de l'examen de la fréquence de faits linguistiques selon Kretzschmar [32]. Ce dernier prend appui sur l'analyse de données lexicales et phonétiques recueillies pour un atlas linguistique américain, le LAMSAS [33] et montre que la variation linguistique ne saurait être une simple alternance entre deux catégories mais qu'elle s'établit entre quelques variables qui possèdent une grande fréquence d'occurrences et un grand nombre d'autres variables qui elles, au contraire, n'ont qu'une très faible fréquence de réalisation. Kretzschmar, qui considère le langage comme un système complexe dynamique, estime que ce type de distribution obéit à la loi de Zipf [34] qui veut que dans un texte, la

32. William A. Kretzschmar, Jr., *Language and Complex Systems*, Cambridge, Cambridge University Press, 2015, p. 21-35.
33. *Linguistic Atlas of the Middle and the South Atlantic States* accessible à l'adresse suivante http://us.english.uga.edu/lamsas/ consulté le 10 mars 2015.
34. W. A. Kretzschmar, *op. cit.*, p. 83-86.

Fig. 1 : **Répartition des types d'alternances selon leur fréquence***



* Par souci de lisibilité, le nom de chaque type d'alternance n'a pas été indiqué sur l'axe des abscisses. Un caractère suivi du signe plus (+) correspond à l'insertion ou à la suppression du son représenté. Deux caractères séparés par une barre oblique (*a*/*b*) indique les cas où le son *a* est remplacé par le son *b* ou le cas contraire. Métath. indique une métathèse entre deux sons.

fréquence d'occurrence d'un mot est inversement proportionnelle à son rang. Une autre caractéristique de la distribution des faits linguistiques que Kretzschmar souligne est qu'elle suit le principe de Pareto qui veut que *grosso modo* 80 % des effets relevés pour un phénomène donné soient causés par 20 % de l'ensemble des facteurs susceptibles de le produire.

Tableau 1 : **Les 25 premiers types d'alternance**

Type d'alternance	Nombre d'occurrences	Proportion au sein du corpus en %
ʀ/ʁ	705	5,90
a/ə	528	4,42
ʀ/r	526	4,40
ʁ/r	505	4,22
e/ẹ	445	3,72
ə+	436	3,65
e/ə	247	2,07
ʁ+	226	1,89
ẹ/ə	219	1,83
o/ọ	209	1,75
i/iː	196	1,64
h/x	191	1,60
n/ṇ	172	1,44
r+	170	1,42
e/ɛ	169	1,41
u/w	161	1,35
ẹ/ẹː	153	1,28
ẹ/i	151	1,26
a+	145	1,21
a/aː	144	1,20
w+	143	1,20
ọ/u	132	1,10
ɥ/w	129	1,08
h+	118	0,98
i/j	116	0,98
	6 336	53

Ces éclairages nous permettent donc de constater que la distribution des données à l'issue de notre traitement dialectométrique se structure selon des modes d'organisation connus. De ce point de vue, nos résultats ne témoignent pas d'une dispersion particulière. En outre, la variation phonétique ne s'ordonne pas de façon catégorielle selon un principe binaire mais plutôt en termes de fréquence avec quelques phénomènes peu nombreux qui prédominent (les 20 % des types d'alternance qui comptent pour 80 % des occurrences). Cette particularité rend possible l'observation d'un continuum de potentialités allant de certaines très fréquentes à d'autres bien plus périphériques et marginales.

L'examen de la fréquence d'occurrence des types d'alternance vient confirmer cette caractéristique. Ainsi, les 25 types d'alternance les plus importants quantitativement comptent pour 53 % des occurrences de notre corpus bien que ne représentant que 5,6 % de l'ensemble des types alternances[35]. Les 5 premiers types d'alternance correspondent eux-mêmes, à un peu plus d'un cinquième des occurrences du corpus (cf. tableau 1).

La grande dispersion des résultats que nous avons relevée dans un premier temps s'avère donc structurée et répond à la loi statistique de Zipf, à l'instar d'autres faits linguistiques lorsqu'ils sont analysés par le biais de leur fréquence.

Le regroupement des différents types d'alternance par catégories larges permet de voir se dessiner des tendances au sein du corpus. Pour y parvenir, nous avons procédé à des rapprochements en examinant les faits phonétiques impliqués dans les différentes alternances. En effet, les premières étapes du traitement des données suggèrent une nette tripartition des alternances entre d'une part, celles portant sur les voyelles, celles concernant les consonnes et une autre à l'intérieur de cette catégorie, impliquant uniquement les rhotiques. L'établissement d'un tableau de synthèse montre la répartition suivante :

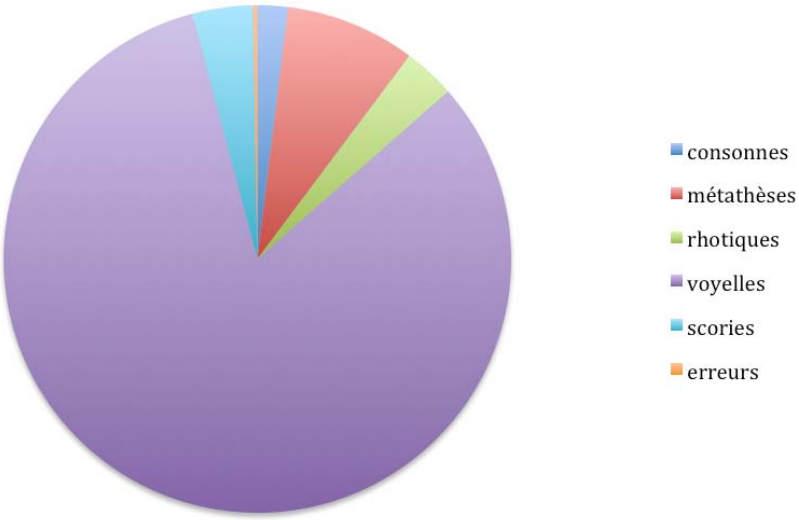
35. Chacun de ces 442 types d'alternance représentant 0,22 % de l'ensemble.

Tableau 2 : **Répartition des types d'alternance par catégorie**

	consonnes	rhotiques	voyelles	métathèses	scories	éléments erronés	totaux
nb des différents types d'alternance	87	12	298	30	14	1	442
nb d'occurrences cumulées	2 600	2 266	6 950	85	26	22	11 949
% de l'ens. des occ.	21,8	19	58,1	0,7	0,2	0,2	100 %

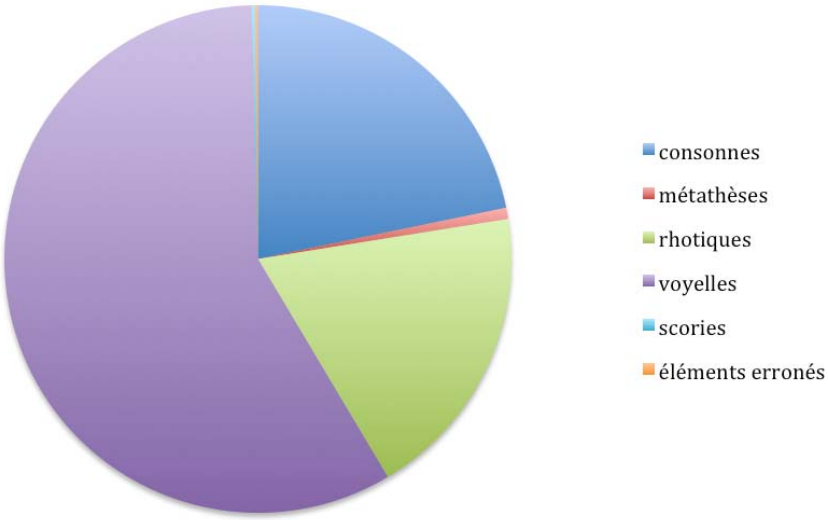
Nous désignons par 'scories' un ensemble d'alternances que nous n'avons pu classer ou dont nous nous sommes rendu compte par la suite qu'elles découlaient manifestement d'erreurs commises lors du traitement algorithmique. Par 'éléments erronés' sont entendues des erreurs indiquées comme telles par l'algorithme et n'ayant pu être éditées ultérieurement par nos soins. Nous commenterons plus tard ces éléments. La représentation de ce tri aboutit au graphique suivant :

Fig. 2 : **Répartition des types d'alternance par catégories**



La variété des alternances regroupées par catégories y est représentée indépendamment de leur importance numérique respective. On y note la très grande prédominance des alternances portant sur les voyelles. La catégorie des métathèses d'une importance bien plus faible se distingue cependant nettement. Enfin, consonnes et rhotiques sont d'importance équivalente, ce qui justifie la distinction que nous avons établie. Cette répartition par type d'alternance ne reflète pas forcément l'importance numérique de chacun d'entre eux comme le montre la figure 3.

Fig. 3 : **Répartition des catégories d'alternances selon leur importance numérique**



La plus grande part des alternances affecte donc les voyelles. La distinction opérée entre les rhotiques d'une part et les consonnes non-rhotiques d'autre part, se visualise nettement, confirmant ainsi la pertinence de les distinguer. Une analyse plus détaillée de chaque catégorie nous permettra de saisir dans quelle mesure ce type de répartition peut être éclairé par des facteurs linguistiques.

Analyse des types d'alternance par catégories

Chacune des catégories établies présente un comportement particulier et le choix d'une telle répartition se justifie par des arguments linguistiques.

a. Les voyelles

Dans le corpus examiné, la catégorie des voyelles est constituée de 298 types d'alternance distincts regroupant 6 950 occurrences, soit 58,1 % d'entre elles. D'une manière générale, les données portant sur les voyelles présentent une distribution semblable à celle que nous avons présentée pour l'ensemble des données. La majorité des occurrences se partage encore une fois entre quelques types d'alternance comme le montre le tableau suivant :

Tableau 3 : **Les 15 premières alternances vocaliques**

Type d'alternance	Nombre d'occurrences	Proportion p.r. à l'ensemble des occurrences en %
a/ə	528	4,42
ə+	436	3,65
e/ə	247	2,07
ẹ/ə	219	1,83
ə/ɛ	111	0,93
ə/ǫ̃	107	0,90
ə/i	74	0,62
ə/o	58	0,49
ə/u	40	0,33
ã/ə	33	0,27
ə/ọ	25	0,21
ə/ɔ	23	0,20
æ/ə	22	0,18
ə/y	10	0,08
eː/ə	8	0,06
Totaux	1 941	16,24

Les faits examinés montrent donc que dans la zone sur laquelle nous travaillons, les voyelles sont particulièrement sensibles à la varia-

tion et par là possèdent une propension à créer de la distance linguistique.

Les voyelles, même si le phénomène ne semble pas avoir été traité à sa juste mesure dans la littérature, constituent une catégorie de sons particulièrement sensible à la variation notamment d'un point de vue diachronique comme dans le cas du *Great Vowel Shift* anglais intervenu entre le XIVe et le XVIIe siècle.

Plusieurs facteurs contribuent à expliquer pourquoi les voyelles sont aussi enclines à la variation. Le fait qu'elles ne possèdent pas de point d'articulation fixe et qu'elles mettent en œuvre des organes labiles peut y contribuer. De plus, Ladefoged[36] fait remarquer que la description des voyelles peut être problématique dans la mesure où il n'y a pas de coupure claire entre chaque type de voyelle et qu'elles s'inscrivent en fait le long d'un continuum. En outre, les voyelles ont une grande propension dans les langues du monde à s'adapter à leur entourage consonantique[37] du fait de leur position préférentielle au centre de la syllabe[38]. Ces éléments expliquent donc pourquoi la catégorie des voyelles dans ce corpus est aussi variée et importante d'un point de vue numérique.

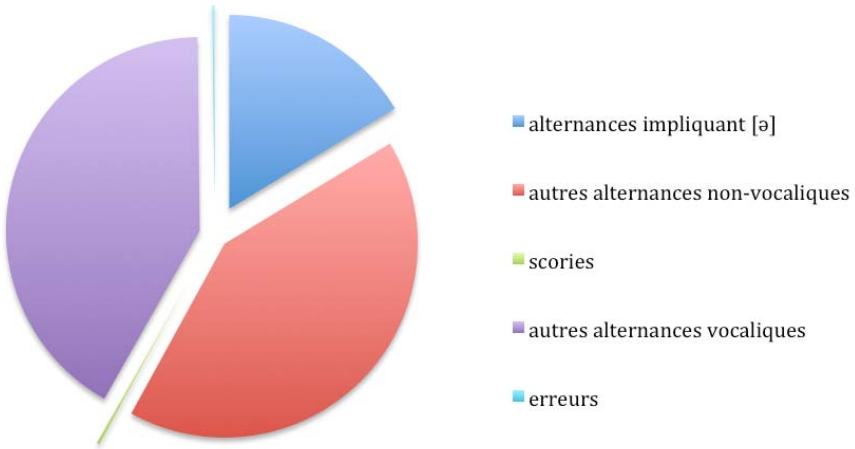
Une tendance de fond se dessine au sein de cette catégorie. La part la plus importante de la variation qui y est à l'œuvre implique le schwa essentiellement dans le cas d'alternances entre ce son et une voyelle antérieure ou postérieure. Cela concerne 22 des types d'alternance de la catégorie recouvrant un ensemble de 1 960 occurrences (soit 16,4 % d'entre elles).

Ce sous-ensemble d'alternances s'individualise ici clairement. Cependant, il faut garder à l'esprit que plusieurs phénomènes distincts peuvent être désignés par cette étiquette (centralisation, réduction vocalique en contexte posttonique, insertion d'une voyelle épenthé-

36. Peter LADEFOGED, *A Course in Phonetics*, Boston, Thomson/Wadsworth, 5e édition, 2006, p. 85.
37. Timothy RIESE, « Permian », dans D. Abondolo (éd.), *The Uralic Languages*, 249-275, Londres, Routledge, 1998, p. 261.
38. Ian MADDIESON, « Vowel Quality Inventories », dans M.S. Dryer & M. Haspelmath (éd.), *The World Atlas of Language Structures Online*, Leipzig, Max Planck Institute for Evolutionary Anthropology, 2013. (accessible en ligne à http://wals.info/chapter/2, consulté le 08-03-2016)

tique). En outre, l'importance numérique de ce sous-ensemble pourrait découler d'une surreprésentation dans les données traitées des articles définis. Nous avons précédemment examiné les contextes dans lesquels l'alternance [a]/[ə] était susceptible d'intervenir[39].

Fig. 4 : **Part des alternances impliquant [ə] par nombre d'occurrences**



L'examen de cette première catégorie nous permet de souligner l'importance de la variabilité que manifestent les voyelles dans la zone étudiée, tant en nombre que dans leur variété. Cette caractéristique peut s'expliquer par les propriétés acoustiques et phonologiques des voyelles. Au-delà, la question des interactions des voyelles avec les autres catégories que nous avons relevées demeure ouverte.

b. Les consonnes

La seconde grande catégorie de types d'alternance qui se distingue concerne celle qui regroupe les consonnes, à l'exclusion des rhotiques. Dans le corpus étudié, nous relevons 87 types distincts d'alternance pour 2 300 occurrences, soit 21,8 % de l'ensemble d'entre elles. Les 11 premières catégories constituent près de 11 % des occurrences du corpus étudié. Nous constatons encore une fois une distribution non-linéaire de ces données.

39. G. Brun-Trigaud *et al.*, art. cit. p. 144 et suiv.

Tableau 4 : **Principaux types d'alternance consonantique**

Type d'alternance	Nombre d'occurrences	Proportion p.r. à l'ensemble des occurrences en %
h/x	191	1,60
n/ṇ	172	1,44
w+	143	1,20
ɥ/w	129	1,08
h+	118	0,98
j+	107	0,89
t+	103	0,86
d+	93	0,78
n+	92	0,77
l+	87	0,73
ʒ/z	85	0,71
Totaux	1 320	11,04

La lecture de ce tableau amène à formuler plusieurs constats. Tout d'abord, on note l'importance de l'insertion ou de la suppression de sons dans les unités linguistiques examinées et ce, à une fréquence relativement importante. Dans le cas de [h], il est possible d'y voir la réalisation ou pas d'un h initial non-étymologique. Ensuite, l'alternance entre la fricative glottale et son équivalent vélaire prédomine dans le corpus étudié. Si ce phénomène est fréquent dans le domaine bretonnant et sujet à commentaire comme l'indique Favereau[40], il est ici délicat de statuer sur le statut phonologique de cette alternance par manque d'information sur la position de ce son dans les mots étudiés. L'analyse informatique en effet n'avait pas pris en compte ce critère.

Autre alternance intéressante, la syllabisation de [n] qui se présente sous la forme [ṇ]. Ce cas de figure implique la réalisation de deux autres phénomènes qui ne sont pas indiqués ici, premièrement la disparition d'une voyelle qui a pour conséquence le transfert de la syllabicité au son suivant. Ce fait découle de l'intensité du schéma accentuel, phénomène relevé précédemment au sujet des réductions vocaliques. Le contexte phonologique précis demeure encore à définir.

40. Francis FAVEREAU, *Grammaire du breton contemporain*, Morlaix, Skol Vreizh, 1997, p. 159.

On remarque également le cas des palatalisations entre [z] et [ʒ], ce qui est aussi bien connu dans le domaine. L'implication des approximantes dans la variation linguistique est également importante avec des types d'alternance mettant en jeu la réalisation ou pas de [l], [w], [j] mais aussi une alternance entre [w] et [ɥ]. Dans le cas des consonnes, pouvoir identifier la place où se réalise cette alternance au sein des unités linguistiques permettrait d'enrichir l'analyse et d'examiner ces alternances particulièrement à la lumière de la dérivation morphologique.

Les alternances consonantiques que nous avons relevées sont certes peu nombreuses mais il faut garder à l'esprit l'importance de leur incidence sur la structuration des unités linguistiques. *A contrario* des voyelles qui semblent être particulièrement sensibles à la variation, les consonnes manifestent plus de stabilité comme en témoigne leur faible importance numérique au sein des résultats que nous avons obtenus. Par conséquent, les modifications qui portent sur les consonnes sont d'une importance toute autre que celles qui touchent aux voyelles, tout particulièrement dans une perspective diachronique comme le montrent bien des cas traités dans les manuels de linguistique historique. D'un point de vue diatopique, cela explique pourquoi German [41] et Costaouec [42] n'ont pris en compte que les consonnes dans leurs analyses. Par conséquent, la question se pose de savoir s'il serait judicieux d'établir une distinction analytique entre chacune de ces catégories car leurs implications sur l'agencement phonologique des mots ne sont pas les mêmes. Toutefois les modalités d'un tel choix n'en demeurent pas moins à établir.

c. Les rhotiques

La dernière catégorie que nous soumettons à l'analyse est celle des rhotiques. Du point de vue de sa distribution statistique, elle paraît clairement individualisée comme le montre la fig. 3. Cependant, si les critères pour établir la classe des rhotiques sont difficiles à établir en termes phonétiques, les sons qu'elle recouvre se caractérisent par une tendance à se comporter de façon semblable du point de vue phono-

41. G. German, art. cit.
42. D. Costaouec, art. cit.

logique[43]. En outre, cette catégorie de sons voit une variation intervenir fréquemment entre les différents éléments qui la constituent[44]. Par rhotiques, nous entendons les différentes réalisations de /ɾ/ relevées dans notre corpus :

[ʀ] soit une vibrante uvulaire voisée
[ʀ̥] la même mais sans la vibration des cordes vocaliques
[ʁ] une fricative uvulaire voisée
[r] une vibrante alvéolaire avec une variante géminée [rr]

Pour cette catégorie, nous relevons 12 types d'alternance différents pour un ensemble de 2 266 occurrences soit près de 19 % de l'ensemble des modifications constatées.

Tableau 5 : **Les différents types d'alternance entre rhotiques**

Type d'alternance	Nombre d'occurrences	Proportion p.r. à l'ensemble des occurrences en %
ʀ/ʁ	705	5,90
ʀ/ɾ	526	4,40
ʁ/ɾ	507	4,24
ʁ+	226	1,89
ɾ+	170	1,42
ʀ+	113	0,94
ɾɾ/ʀ	5	0,04
ʀ̥/ʀ	5	0,04
ʀ̥+	4	0,03
ɾ/ʀ̥	3	0,02
ʀ̥/ʁ	1	0,008
ɾɾ/ɾ	1	0,008
Totaux	2 266	18,936

43. Mona LINDAU, « The story of /ɾ/ », dans V. A. Fromkin (éd.), *Phonetic Linguistics: Essays in honor of Peter Ladefoged*, Orlando, FA, Academic Press, 1985, p. 157-168. Peter LAGOFELD & Ian MADDIESON, *The Sounds of the World's Languages*, Oxford, Blackwell Publishers, 1996, p. 215-216. Richard WIESE, « The Representation of Rhotics », edited by M. van Oostendorp, C. Ewen, E. Hume & K. Rice, *The Blackwell Companion to Phonology*, Malden, MA & Oxford, Wiley-Blackwell, 2011, p. 711-729.
44. James M. SCOBBIE, « (R) as a Variable », dans K. Brown (éd.), *Encyclopedia of Language & Linguistics*, volume 10, Oxford, Elsevier, 2e édition, 2006, p. 337-344.

L'examen du tableau précédent montre d'une part que la majorité des occurrences se concentrent entre 6 alternances distinctes qui correspondent à 18,6 % de l'ensemble des occurrences du corpus. Ce constat étant fait, une dichotomie numérique s'établit entre les remplacements d'un /r/ par un autre d'une part et les insertions et suppressions de ces mêmes /r/.

Certaines remarques que nous avons formulées précédemment peuvent être ici répétées. Premièrement, l'importance de cette catégorie pourrait découler de faits de variation liés à la prise en compte des articles définis et indéfinis du breton dans lesquels une forme de /r/ peut être fréquemment réalisée selon les contextes (**ar** ou bien -**e**-**ur**) ? L'absence d'information sur la position respective de chacun de ces sons ne permet aucunement de préciser leur statut respectif.

L'examen de la distribution de ces sons dans les parlers considérés laisse par ailleurs entrevoir une situation complexe. La réalisation d'un type de /r/ ne semble pas devoir exclure au sein d'un même parler d'autres possibilités comme le montrent les formes suivantes :

Tableau 6 : **Exemples de réponses comportant un son rhotique**

	(22) Saint-Cadou	(89) Lannédern	(40) Plourac'h
carte 202 VOLEUR	lɛ:**r**	'lae**ʀ**	lɛ:**ʀ**
carte 223 UN CORBEAU	'ə v**r**ãn	ọ v**ʀ**ãn	øʁ v**ʀ**ã:n
carte 7 PAROISSE	'pa**rr**es	'pa:**ʀ**əs	'pa:**r**us
carte 231 GRENOUILLE	'glẹsk**ər**	'glẹsk**əʁ**	'glẹskɛ**ʁ**

Cet échantillon indique que, pour chaque point, une réalisation préférentielle semble prédominer mais d'autres réalisations ne sont pas exclues, ce qui invite à envisager les réalisations en fonction de leur fréquence.

La grande variabilité des rhotiques ne se manifeste pas uniquement d'un point à un autre du NALBB mais également au sein de chaque topolecte, ce qui suggère que l'intensité de la réalisation de chaque rhotique se déploie le long d'un continuum.

Ces oscillations dans l'emploi des rhotiques expliquent pourquoi ces différentes réalisations sont susceptibles d'être connotées positivement ou de façon négative dans les communautés linguistiques comme

en témoignent les travaux de Labov sur la structuration sociolinguistique de l'anglais à New York[45] en observant les différentes réalisations de /r/ dans les grands magasins de la ville. Dans le domaine bretonnant, Jean Le Dû décrit dans son introduction au NALBB le comportement d'un informateur passant d'une réalisation vélaire à une réalisation apicale au fur et à mesure de l'avancée de l'entretien, la confiance s'instaurant.

Il est possible de signaler la convergence entre les voyelles et les rhotiques. En effet, ces deux catégories sont particulièrement sensibles à la variation comme nous avons pu le constater dans notre corpus. Plusieurs travaux mentionnent les rapports particuliers qu'entretiennent ces deux catégories de sons. Ladefoged & Maddieson indiquent que les rhotiques influencent la couleur des voyelles qui les entourent et ont tendance à se fondre avec elles[46]. Cette proximité s'explique par une caractéristique acoustique commune aux deux catégories.

Validité de l'outil employé et développements ultérieurs

Les analyses issues du traitement du corpus étudié montrent que des phénomènes courants dans les langues du monde sont à l'œuvre également en breton. Toutefois, la question de l'implication de la méthode d'analyse retenue sur la structuration des faits peut se poser. Par conséquent, nous proposons tout d'abord d'observer dans quelle mesure notre traitement dialectométrique rend compte des faits examinés. Ceci permettra de mieux cerner les phénomènes qui ont posé problème lors du traitement des données et qui ont été délaissés par la suite. Cette démarche vise à tester la validité de notre approche et de déceler les aspects qu'il est souhaitable d'affiner en vue d'approfondir nos analyses.

Évaluation de l'usage de l'algorithme de Damerau-Levenshtein

À l'issue du traitement computationnel, nous avons relevé 442 types d'alternance distincts. Cependant, comme nous l'avons indiqué

45. William Labov, *The Social Stratification of English in New York City*, Cambridge, Cambridge University Press, 2006 [1966[1]], p. 40-57.
46. P. Ladefoged & I. Maddieson, *op. cit.*, p. 216.

(tableau 1), 48 alternances correspondent à des cas de figure que nous n'avons pas été en mesure d'expliquer (26), ou de comprendre quand il ne s'agissait pas d'erreurs issues du traitement algorithmique (22).

Ainsi, dans le premier groupe sont présents 14 cas que nous avons été en peine d'associer à une catégorie précise comme dans le cas des types ʁ/ɛ ou ǫ/t. Ensuite, les métathèses regroupent 30 types d'alternance pour 85 occurrences. Nous n'avons pas analysé cette catégorie car les choix que nous avions retenus pour le traitement informatique ne sont pas satisfaisants. En effet, nous avions pris le parti d'indiquer une inversion de caractères dans ce cas de figure mais cela ne rend pas compte du phénomène de métathèse. Il implique ici une consonne et une voyelle alors qu'il nous semble qu'il s'agisse avant tout du déplacement de la position d'une consonne.

Autre cas de figure que nous avons laissé de côté, l'alternance procédant d'erreurs dues à l'algorithme (22 occurrences). En effet, nous avons déjà relevé précédemment que l'algorithme de Damerau-Levenshtein pouvait être mis en difficulté par les données du breton [47].

La comparaison de ces différents éléments à l'échelle du corpus étudié montre que nous n'avons pas été en mesure de prendre en compte 45 types d'alternance (10,6 % de l'ensemble de ceux-ci) correspondant à 133 occurrences (1,1 % des modifications relevées dans le corpus). Ce résultat nous paraît satisfaisant dans la mesure où nous sommes parvenus à analyser près de 98,9 % des modifications relevées. Cependant, si l'importance numérique de ces scories paraît faible, leur représentativité parmi l'ensemble des types d'alternance demeure encore élevée. De ce fait, plusieurs améliorations peuvent être conduites d'un point de vue technique afin de diminuer ce bruit.

Perspectives

Premièrement il convient de garder en mémoire certaines propriétés de l'algorithme de Damerau-Levenshtein qui éclairent certaines des difficultés rencontrées à l'issue du traitement informatique. En effet, l'algorithme cherche le moyen de passer d'une chaîne de caractères

47. G. Brun-Trigaud *et alii*, art. cit.

à l'autre mais en utilisant la voie la plus économique c'est-à-dire en procédant avec le moins d'opérations possibles. Ceci amènera donc l'algorithme à privilégier des modifications qui n'auront pas forcément de justification linguistique comme dans le cas suivant par exemple pour la carte 189 DE L'ORAGE du NALBB pour les points 89 Lannédern et 93 Lennon, l'analyse brute donne les transformations suivantes :

(89) Lannédern	ˈgyːʁn̩		
	ˈg yː ʁ n̩		
modifications erronées		R. de n̩ par ᵊ	Ins. de n
		Ins. de ᵊ	R. de n̩ par n
	ˈg y ʁ		n
(93 Lennon)	ˈgyʁᵊn		
	total de 2 changements		

Dans ce cas, nous pouvons constater que l'algorithme a procédé à des modifications en allant de la droite vers la gauche sans tenir compte de la nature des caractères mais en considérant uniquement l'ordre dans lequel ils se présentaient à lui. Nous avons indiqué à la troisième ligne les rectifications apportées lors de l'édition de ces opérations.

En outre, des difficultés pourraient être évitées en travaillant sur le codage des signes phonétiques. En effet, la présence de signes diacritiques venant préciser un caractère peut s'avérer problématique dans la mesure où l'algorithme le lit comme étant deux caractères distincts. Il nous faut donc trouver le moyen de les coder par le biais d'un seul caractère de façon à ce que l'algorithme ne se trompe pas. Cette difficulté se répète pour les diphtongues, les consonnes affriquées et dans une moindre mesure, les consonnes géminées. Le cas des semi-consonnes (j, w, ɥ) dans cette perspective est délicat car elles interviennent dans différents contextes que ce soit au sein d'une diphtongue ou bien en position intervocalique. De fait, une alternative à un codage automatique confondant ces différents cas s'avère nécessaire.

Une autre difficulté à laquelle l'algorithme doit faire face porte sur le traitement à réserver aux cas des métathèses. Nous avons fait le choix de le considérer comme le déplacement d'une consonne, bien souvent

une liquide et/ou une nasale qui peuvent permuter d'une syllabe à l'autre ou au sein d'une même syllabe. Plutôt que de se limiter à la simple suppression d'un signe et à son insertion à une autre position dans le même mot, cette opération doit pouvoir être identifiée *a posteriori* afin de la distinguer de simples opérations de commutation. Il en va de même pour les épenthèses, essentiellement vocalique. L'enjeu, dans ces deux situations, est de parvenir à distinguer ces phénomènes de simples alternances pour ne pas confondre plusieurs phénomènes différents.

Notre analyse s'est basée dans un tout premier temps sur l'examen des alternances entre différents sons par le biais de la comparaison des signes qui les représentent. Or, comme nous avons pu l'évoquer à plusieurs reprises, l'accentuation intervenait dans plusieurs situations et par conséquent, elle doit être prise en compte dans la mesure de la distance dialectale. Pour y parvenir, un dispositif technique doit pouvoir comparer entre plusieurs unités la syllabe accentuée au sein d'un mot, et ce d'autant plus que les données de la zone d'accentuation pénultième débordant le Vannetais seront intégrées dans nos analyses à venir. Les variations portant sur la structure syllabique qui peut découler de la force de l'accentuation et entraîner l'ajout ou la perte d'une syllabe seront à prendre également en compte.

Nous avons également constaté d'un point de vue quantitatif la grande place qu'occupent dans le corpus les alternances impliquant les voyelles et les rhotiques. L'une des raisons possibles que nous avons à l'esprit pour expliquer cet état de fait est qu'il pourrait découler d'une surreprésentation de ces types de sons en raison de la structuration de notre corpus. En effet, sur les 165 cartes analysées, 18 comportaient un article défini ou indéfini, chacun de ces mots se composant d'une voyelle et souvent d'une rhotique mais se réduisant fréquemment à l'un ou l'autre. La possibilité de distinguer au sein de l'ensemble analysé également les modifications issues de la comparaison de ces articles permettrait de prendre la pleine mesure de cette possible influence.

Autre cas problème à résoudre, celui de la position des consonnes au sein des réponses fournies par le NALBB. La façon dont nous avons procédé jusqu'alors nous a conduit à rassembler sous une même alternance consonantique des modifications de consonnes indépendam-

ment du lieu où elles étaient réalisées dans le mot, que ce soit en position initiale, intervocalique ou finale. Ceci a pour conséquence que le nombre de types d'alternance consonantique soit relativement faible (87) comparé à celui des voyelles (298). Cependant, cela n'aurait aucune incidence sur le nombre total d'occurrences. Toutefois, cela demeure un préalable nécessaire pour initier une réflexion phonologique à partir des données analysées.

Enfin, l'un des aspects que notre analyse a souligné est la difficulté que nous avons à repérer les phénomènes susceptibles de s'impliquer mutuellement. En effet, dans ce cas de figure, plusieurs catégories distinctes peuvent être impliquées sans qu'il soit possible de les comparer en raison de leur nature différente (son, accent, syllabe). Sans nul doute, une analyse statistique approfondie permettra de valider certains rapprochements.

Ces pistes de travail dessinent autant de possibilités de parvenir à un traitement plus fin des données et à permettre de faciliter l'observation des phénomènes à l'œuvre dans la variation phonétique en breton. Mais elles indiquent également en écho plusieurs phénomènes prégnants dans la variation linguistique en breton.

Que soulignent les résultats observés ?

L'examen de faits analysés amène à plusieurs remarques portant sur la nature de l'objet que nous analysons et sur la façon de le décrire.

Tout d'abord, dans la zone observée et pour les données analysées, la variation sur le plan phonétique est présente et ne saurait constituer un phénomène exceptionnel. La question de fond qui nous anime est de repérer les éléments qui la composent et de déterminer le(s) moyen(s) de la mesurer. La surprise qui a été la nôtre dans un premier temps a été de constater un éparpillement des données ne laissant comme possibilité d'analyse que celle de faire le constat d'un continuum dialectal. Cette notion forgée d'après Bloomfield[48] décrit les transitions graduelles dans l'espace géographique d'une langue amenant à la différenciation de celle-ci au fur et à mesure de l'éloignement d'un point initial[49].

48. Leonard BLOOMFIELD, *Language*, New York, Holt, Rhinehart and Winston, 1933, p. 51-54.
49. J. CHAMBERS & P. TRUDGILL, *op. cit.*, p. 7

Or, dans son ouvrage, Bloomfield réserve le terme de continuum, à la description de la suite des sons dans la chaîne parlée[50] et à leur variabilité intrinsèque. Ainsi, certaines catégories de sons s'inscrivent le long d'un continuum du point de vue acoustique[51]. Il est intéressant à partir de nos données de prolonger la métaphore et de considérer que les différents types d'alternance que nous avons relevés s'inscrivent eux aussi dans un continuum de possibilités, seule la fréquence d'occurrences de certaines en fait des types pertinents à distinguer d'autres plus périphériques. Par conséquent, l'objet que nous décrivons semble s'inscrire dans un continuum pluridimensionnel dont les lignes de structure et de rupture demeurent à préciser. Cette vue rejoint des descriptions d'autres auteurs sur des faits de langue différents[52].

Par conséquent, la fréquence de réalisation nous paraît être un outil intéressant pour éprouver les catégories descriptives avec lesquelles nous travaillons et ainsi parvenir à les préciser. À plusieurs reprises, nous avons souligné l'intérêt d'adjoindre la fréquence à une approche dialectométrique afin de repérer l'intensité des phénomènes à l'œuvre dans la production de la distance linguistique. La méthodologie que nous avons retenue procède essentiellement par contrastes et ne souligne que les différences lors des comparaisons de prononciations. Une telle démarche rejoint celle de Séguy dans son approche des données du gascon par le biais d'un gradient[53] et s'inscrit dans une réflexion actuelle[54]. En effet, l'emploi de cet outil ne va pas sans poser de question sur la construction d'un fait linguistique et sur sa pertinence[55].

---

50. L. Bloomfield, *op. cit.*, p. 76.
51. P. Ladefoged & I. Maddieson, *op. cit.*
52. Cf. le continuum stylistique labovien, le continuum créole selon Bickerton. À titre de réflexion : Hansjakob Seiler, « Le continuum linguistique et la relation entre un invariant et ses variantes », *Bulletin de la Société de linguistique de Paris*, CV(1), 2010, p. 7-33.
53. Cf. Jean-Louis Massourre, *Le Gascon, les mots et le système*, Paris, Honoré Champion, 2012, p. 32-36.
54. Sylvain Loiseau, « Les faits statistiques comme objectivation ou comme interprétation : statistiques et modèles basés sur l'usage », *Travaux de linguistique*, 62(1), p. 59-78. doi 10.3917/tl.062.0059
55. William Labov, « Qu'est-ce qu'un fait linguistique ? », *Marges Linguistiques*, 1, 2001, p. 25-68. Accessible à l'adresse suivante http://www.revue-texto.net/Parutions/Marges/artml0000_ml.pdf

Nous sommes donc face à une situation où nous devons mettre en œuvre une double démarche d'analyse quantitative mais aussi qualitative. Ceci est d'autant plus intéressant que cela nous offre un aperçu sur le fonctionnement d'un système linguistique par le biais du breton et amène donc des questions sur le fonctionnement d'une langue. L'enjeu est de parvenir à élaborer des catégories pertinentes afin de décrire la complexité des faits analysés.

Conclusion

La présentation du traitement exploratoire des données du NALBB que nous venons de mener nous a permis de présenter la méthode de la dialectométrie et les enjeux afférant à la mesure de la distance linguistique entre les points d'un atlas linguistique par le biais de la distance de Levenshtein. Nous avons signalé l'originalité de notre approche visant à associer une analyse qualitative à un traitement quantitatif. Ceci nous a permis de présenter des résultats inédits que nous avons explorés en testant par là des catégories d'analyses tant du point de vue quantitatif que linguistique.

S'il nous semble avoir établi les premiers facteurs expliquant la structuration des résultats que nous avons obtenus, nous avons également dégagé plusieurs voies afin d'affiner nos analyses.

La distance d'édition utilisée comme nous l'avons fait s'avère donc un outil efficace pour rendre compte de la variation linguistique en breton. Cependant, des améliorations techniques et l'association avec une démarche statistique approfondie permettront de rendre compte de phénomènes plus complexes faisant intervenir des catégories de différente nature.

Notre approche nous a permis de reconsidérer l'objet que nous décrivons en l'inscrivant dans un continuum pluridimensionnel. Nous avons délaissé les facteurs géographiques abordés auparavant pour nous concentrer sur la lecture et la compréhension des faits recueillis. Cette étude se prolongera par la prise en compte d'un nombre plus important de données et de points d'enquête afin de recouvrir l'ensemble de la Basse-Bretagne.