A new dialectometric approach applied to the Breton language

Guylaine Brun-Trigaud, Tanguy Solliec, Jean Le Dû

HAL Id: hal-01551178
https://hal.univ-brest.fr/hal-01551178

Submitted on 1 Oct 2020

Chapter 8

A new dialectometric approach applied to the Breton language

Guylaine Brun-Trigaud

CNRS & Université de Nice-Sophia-Antipolis, France

Tanguy Solliec

Université de Paris Descartes, France

Jean Le Dû

Université de Bretagne Occidentale, France

This paper presents a new dialectometric approach applied to the Breton language in the heart of Breton-speaking Brittany. It is based on data from the Nouvel Atlas Linguistique de la Basse-Bretagne Le Dû (2001). We process qualitative data using the Levenshtein algorithm which allows us to accurately measure and take into account the discrepancies or similarities between different pronunciations of a given word. This contribution aims to determine whether linguistic distance is caused by a frequent repetition of the same phenomenon or whether it is the outcome of multiple changes. Our first results suggest new ways of analysing Breton data.

1 Introduction

At a time when the number of languages in use is bound to reduce drastically in the next decades, documenting endangered languages is a major challenge for linguists, as is the urgency to study and analyse their internal variation. Nowadays, Breton is considered to be an endangered language, according to the definition proposed by UNESCO (Moseley 2010). The intergenerational transmission of the language ceased during the 60's. Most of its speakers are now over 70 years of age and, at present, the language is disappearing quickly. We propose to present a dialectometric approach applied to the Breton language based on Le Dû's (2001)

Guylaine Brun-Trigaud, Tanguy Solliec & Jean Le Dû

Nouvel Atlas Linguistique de la Basse-Bretagne (henceforth NALBB) (Le Dû 2001) so as to re-examine the geolinguistic landscape of Lower Brittany.

Little dialectometric work has been done on the Breton language (German 1984, German 1991, Costaouec 2012). Our approach is based on a new methodology: we have applied the principles of edit distance (i.e., the Levenshtein Distance, henceforth LD), which allows us to measure the linguistic distance between two strings of characters (phonetic transcriptions). This distance is defined as the minimal number of characters that need to be deleted, inserted or replaced, in order to transform one string of characters to another. We have introduced modifications in processing the data.

Our work aims to test this new approach and to observe its advantages and disadvantages when applied to Breton. It was tested for the first time on dialects in the Occitan area. This research constitutes the very first step of a PhD thesis by Solliec (2014), whose goal is to process the data available in the NALBB in order to sketch the geolinguistic configuration of the area based on a dialectometric approach. This first study focuses on the phonetics of Breton because the data contained in the NALBB are richer in phonetics than in the lexicon.

2 Preliminary research

2.1 A first test on Occitan dialects

Dialectometry was initiated by Séguy (1971; 1973), and Guiter (1973). Its aim is to quantify linguistic variation with the help of mathematical tools. This approach allows us not only to account for the differences or similarities between the dialects studied, but also to display them with the help of maps and tables. Nowadays, many research groups work with various methods. One of them, the Groningen School, led by Nerbonne and Heeringa (Nerbonne & Heeringa 2001; 2010, Heeringa 2004), uses the Levenshtein algorithm. This methodology allows for the accurate measurements of the number of operations (replacement, deletion or insertion) necessary to transform one string of characters to another and displays the results as a number or a percentage of similarity.

As this method is intended to provide an overall characterization, it accounts most of the time for dialectal distance from a quantitative point of view. However, it does not describe the results qualitatively in a direct way. In others words, it does not indicate the exact proportion of each difference involved in dissimilarity. Does the dissimilarity result from one sole phoneme correspondence (replacement) which occurs frequently while other kinds of variations remain very

Table 1: An example of a quantitative measurement between 2 locations of enquiry for the word 'a day (long)' in the Occitan language.

Clermont (pt 22)	ʒ	u	r	'	n	a	ð	ɔ
Fauillet (pt 20)	ʒ	u	ʀ	'	n	a	d	œ
			1				1 1	3

3 differences = 63% of similarity / 37% of dissimilarity

minor or, on the contrary, does the dissimilarity result from an important array of different correspondences which are more or less of equal importance? In the first case, mutual comprehension between the speakers is not really affected, while in the second case, it is likely to be jeopardized. In order to contrast the two kinds of variation, Brun-Trigaud decided to modify the algorithm. To accomplish this, she implemented it as a function in Visual Basic for Applications (VBA) in an Excel spreadsheet so as to obtain, not a percentage or a number of characters, but rather a record of the phonemes undergoing the operations and their frequency. She then treated them statistically. She carried out a first experiment (Brun-Trigaud 2014) using data taken from the *Atlas Linguistique et Ethnographique du Languedoc Occidental*, gathered in a region located in the Occitan area in southern France. This language area is distinguished by both a large and relatively homogeneous central zone around the city of Toulouse and by two mixed zones, one to the North in contact with Limousin and the other to the South, bordering the Catalan-speaking area.

Table 2: An example of a qualitative measurement between 2 locations of enquiry for Occitan

segments 22 > 20	concept	answ (22)	answ (20)	LD1	LD2	LD3
22 > 20	day	ʒurˈnaðɔ	ʒuʀˈnadœ	repl. r by ʀ	repl. ð by d	repl. ɔ by œ
22 > 20	he snores	ˈr:ũŋkɔ	ˈʀũŋklœ	repl. r: by ʀ	ins. of l	repl. ɔ by œ
22 > 20	fair	ˈfjɛrɔ	ˈfjɛʀɔ	repl. r: by ʀ	repl. ɔ by œ	

When all the data taken from two different locations were compared, as in Table 2, she obtained: replacing r by ʀ (= 36%) and replacing ɔ by œ (= 22%). Brun-Trigaud demonstrated that, for the area studied, from East to West, divergences were caused by the preponderance of a single phonetic correspondence, appearing very frequently, whereas in the North of the area differences in nu-

merous phonetic variables appear, which affects mutual understanding between speakers of different areas.

In order to validate these initial results, she invited her fellow Breton dialectologists, Solliec and Le Dû, the author of the NALBB, to test the new version of the algorithm on Breton data.

3 Earlier dialectometric works on Breton

Studies by Falc'hun (1981) based on the *Atlas Linguistique de la Basse-Bretagne* (Le Roux 1924–1963, henceforth ALBB) have deepened the geolinguistic understanding of the Breton language. They constitute an important reference on the issue.
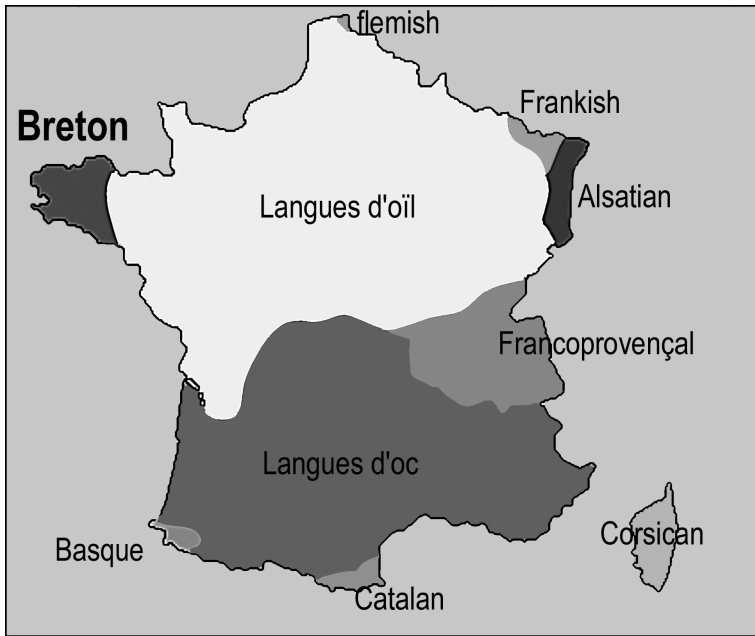

Figure 1: Situation of the Breton language.

Until now very little quantitative work has been done on the Breton language. German (1984; 1991) is the very first to have applied a dialectometric approach to Breton data. A cluster analysis using the Lerman algorithm allowed him to classify the dialect areas according to their linguistic similarities and differences with the Breton of Saint-Yvi he described in his PhD thesis (1984). More recently

Costaouec (2012) correlated the internal dialect borders inside a small area of the Breton speaking region with other aspects of social, cultural or economical factors around the village of La Forêt-Fouesnant, whose speech he described in his thesis (Costaouec 1998). These two studies are based on Le Roux's ALBB, where the data were collected between 1913 and 1920 for a range of 87 sites of enquiry published in 600 maps. Le Dû began working on his NALBB (2001) in 1968 (187 sites and 601 maps). He himself transcribed all the data in order to avoid transcription biases. Since no recent study has tackled the Breton language as a whole, as we mentioned before, further works by Solliec will aim to fill this gap. In order to accomplish this, he will take into account the NALBB data. This article constitutes the first step in this process.

4 The area investigated

We have deliberately chosen to investigate a restricted and quite homogeneous area for the following reasons: on the one hand, we do not want to confront this modified version of the LD algorithm, from a technical point of view, with an area where linguistic variation is very intense, as in the case of the *Vannetais* dialect (South-East of Lower-Brittany). Moreover, as was shown by Falc'hun (1981), innovations have been spreading across this region for centuries. It would then be interesting to compare them with sites from more conservative peripheral areas. Over the past few years, Solliec has been carrying out a project to describe and document local varieties in this area and is therefore aware of the local variation phenomena.

On the methodological side, we have decided to use an inter-site approach: to do so, each location of the area is linked to its closest geographical neighbours in order to establish a comparison we called the "segment". Doing so allows the researcher to detect spatial continuities and discontinuities (breaches) on the phonetic level (Goebl 2012: 137). In this view we selected 23 investigation sites for 165 phonetic maps (i.e. using only one lexeme and its phonetic variations for each concept). We had not beforehand determined which variables to observe in our corpus: our objective was to consider the data as a whole and to observe which phenomena produce linguistic distance and at which frequencies.

It led us to divide the area into a net of 53 "segments", each of which connect a site in the atlas with its neighbours on all sides. This brings us to a total of 8745 comparisons between strings of characters.

Our aim has been on the one hand to gather quantitative indicators and, on the other hand, to identify qualitatively the most common correspondences involved
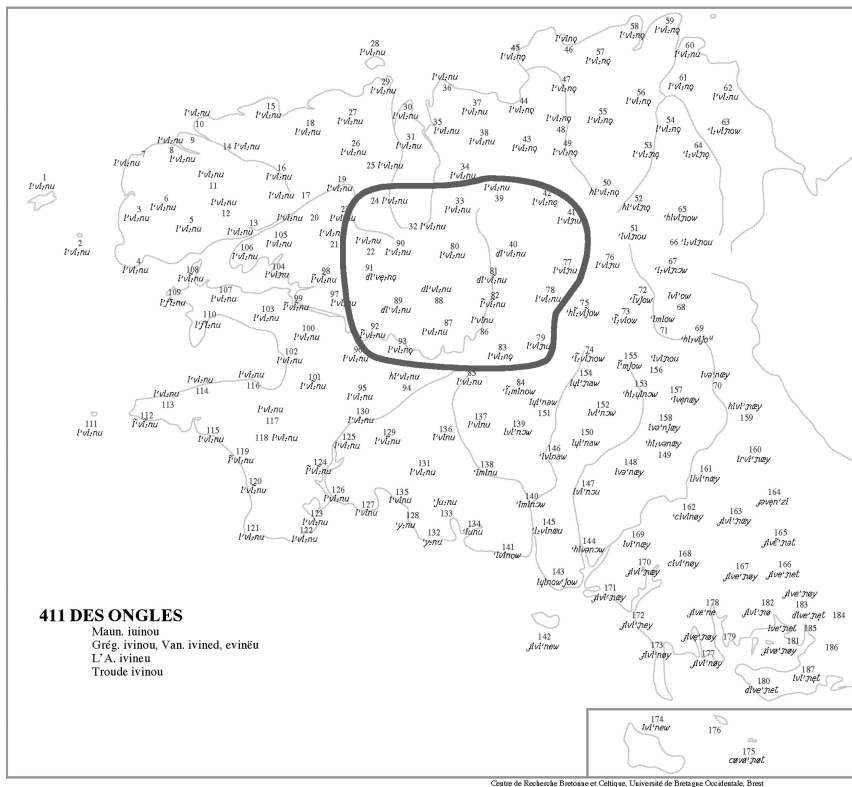
Figure 2: NALBB Map 411 'Fingernails'. The area investigated is delimited in black.

in phonetic variation. Even though the latter are well known (Falc'hun 1981; Jackson 1967), we do not know how frequently they occur.

5 The problems LD meets with Breton

In confronting LD with Breton language data, we noticed difficulties: Table 3 shows an excerpt of the data where the algorithm repeatedly failed to analyze the data, while at the same time the FuzzyMatch function did not work as we can see in Table 3. This last function returns a percentage reflecting the probability of correspondences (and differences) between two words.[1]

---

[1] See https://x443.wordpress.com/2012/06/25/fuzzymatch-in-vbavbscript/

Table 3: Examples of the difficulties encountered by LD for the concept 'that one (masc.)'

segments	answ1	answ2	test Exact	FuzzyMatch	Levenshtein1	Levenshtein2	Levenshtein3	Levenshtein4
40 > 41	hẽːs	hẽːs	TRUE	Not a match				
32 > 80	ˈẹnes	ˈẹnnəs	FALSE	50% (2)	Rep. n by nn	Rep. e by ə		
32 > 33	ˈẹnes	ˈhẹnnəs	FALSE	25% (3)	#VALUE! (Ins. of h)	#VALUE! (Rep. n by nn)	#VALUE! (Rep. e by ə)	
33 > 39	ˈhẹnnəs	hẽː	FALSE	**Not a match**	Del. of s	Del. of ə	Del. of nn	Rep. ẹ by ẽː

Although we were careful to use a unique character for each phonetic segment and its diacritics (including diphthongs and geminates), it appears that our implementation of the Levenshtein algorithm experiences difficulties when treating languages like Breton for which variation in the number of syllables is frequent (e.g., site 80 Berrien [ˈparọs] vs. site 81 Poullaouen [paːʁs] 'parish' (NALBB map 7)).

Nevertheless, after checking and adjusting the function, we ended by incorporating most observations. On the one hand, the raw results obtained from the FuzzyMatch function produces the following distribution (see map in Figure 3):

The map shows several areas involving significant degrees of similarity, particularly in the North around sites 24 on the one hand and 39 on the other, and in the South around site 88. It is clear that the geographic distance between the surveyed localities often correlates with phonetic similarity; for example, there is high phonetic similarity between nearby sites 87 and 88 and low similarity between distant sites 83 and 93. However, there are some exceptions, for instance, between sites 22 and 24 or, vice versa, between 91 and 89. We will later see why they occurred.

On the other hand, the results returned by the new function of the algorithm bring us to the following conclusions: first of all, amongst the 11,949 non-identical phonetic correspondences, "alternations" or "differences" according to our terminology, we had to manage an important dispersion of the distinctive correspondences (nearly 450). They were generated by the narrow phonetic transcription. With the consent of Le Dû, the author of the atlas, we grouped some alternations in order to have to deal with fewer details. This limited the number of the different kinds of alternations to 200. The most frequent ones are listed in Table 4.

These alternations are distributed as follows on the map (Figure 4): for each segment (i.e. a pair of enquiry sites) we calculated the most frequent alternation proportional to its frequency (identical correspondences are excluded).
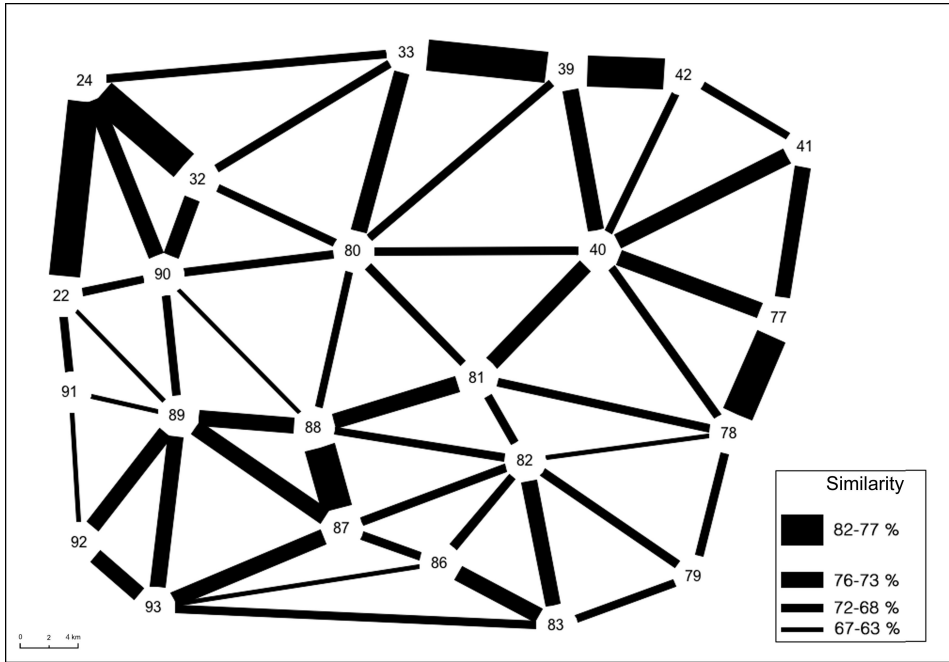
Figure 3: FuzzyMatch results for the area on a schematic map

Table 4: List of the most frequent alternations

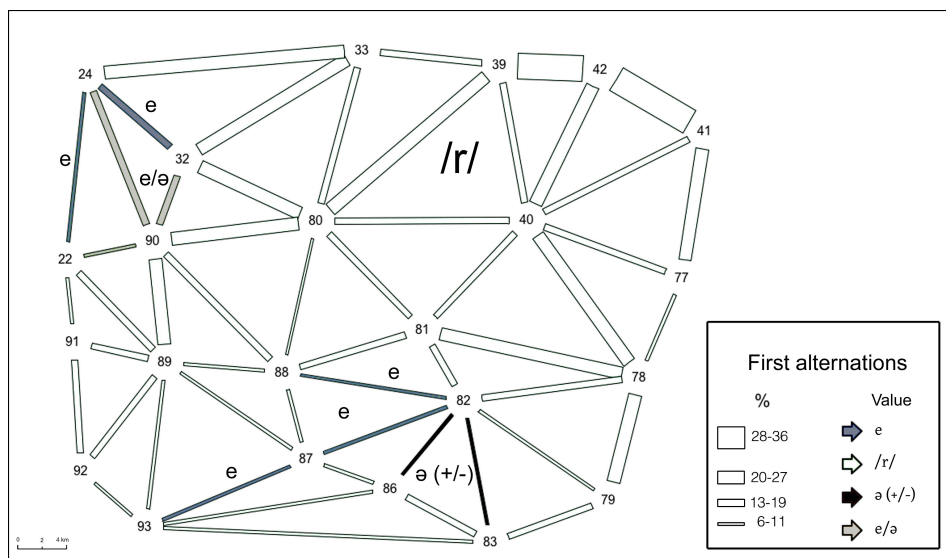|				
1	r/ʁ/ʀ	14.70%		
2	e (ẹ/e/eː)	6%	25%	
3	a/ə	4.5%		
4	r (+/-)	4.3%		
5	e/ə	4%		
6	ə (+/-)	3.5%		50%
7	e/ɛ	3%		
8	o (ọ/o/oː)	3%	25%	
9	e/i	2%		
10	n (n/nn/ṇ)	2%		
11	æ/a	1.7%		
12	i (i/iː)	1.6%		

Figure 4: Distribution of the most frequent alternations across the area

Apart from the [r] variations, the proportions of changes are relatively small, but they are quite similar to the results found in the Occitan region, with one notable difference: there were far more changes in the consonants Brun-Trigaud (2014: 135). In addition, as noted by Le Dû, [r] variations are probably idiolectal, so that we consequently decided to concentrate on other more relevant changes with his agreement.

The nature of the differences is more varied in the second most frequent alternation (map in Figure 5).

Taking into account the fact that the proportions of changes are relatively low – 3 to 13% at most — the map in Figure 5 shows that the south-eastern area is marked by the presence or absence of schwa (dark colour), the central area by alternations between [a] and [ə] (white colour), and, finally, that the north-west is characterized by an alternation between [e] and [ə] (dark grey).

6 Analysing the first results

The values the algorithm returned provided a considerable amount of data. They had to be explored carefully. The following analyses are instances of what can be done when scrutinizing geolinguistic data from a statistical perspective when
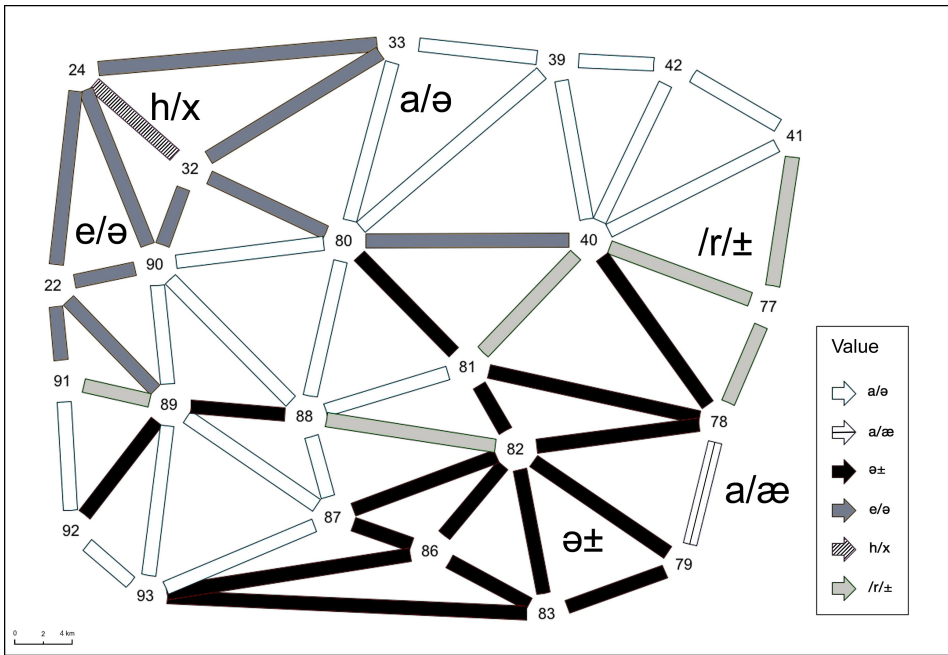
Figure 5: Distribution of the second most frequent alternation across the area

associated with a qualitative approach. First, we will observe how one specific alternation occurs across the area, involving [a] and [ə]. Secondly, we will study the results for one locality, Collorec, site 88 of the NALBB.

6.1 Examining only one kind of difference across the area

The most important kind of change from a numerical point of view, apart from the variation of the rhotics, is the one which affects the vowels [a] and [ə]. In our corpus, this alternation ([a]/[ə])[2] constitutes 9% of the non-identical phonetic correspondences we have gathered and appears in 54 of the 165 maps we selected.

The realizations [a] or [ə] occur in the following phonetic contexts. Each one can interplay with another:

Eighteen of the 165 maps we used included a definite article and 11 had infinitive word forms. These are instances of the alternations under study (see examples a. and e. in Table 5).

---

[2] The alternation is indicated by the brackets ( ) in order to distinguish it from the specific sounds [a] and [ə]

Table 5: Phonetic contexts for the alternations ([a]/[ə])

Type of context	Map number & concept	Examples
a. Definite article's initial vowel	263 'the stable'	(81) Poullaouen [ə hʁow] vs. (80) Berrien [a hʁow]
b. Final unstressed syllables	464 'clothes'	(93) Lennon [ˈdiʎət] vs. (89) Lannédern [ˈdiʎat]
c. Cluster of phones /-uwar/ or /-uarn/	196 'blackberries'	(87) Landeleau [ˈmuːwəʁ] vs. (93) Lennon [ˈmuwaʁ]
	170 'iron'	(32) Plounéour-Ménez [ˈuːaʁn] vs. (33) Plougonven [ˈuːrən]
d. Epenthesis vowel	301 '(a) scythe' (literary form: [falx])	(78) Locarn [ˈvalax] vs. (82) Plounévézel [ˈvʰaləx]
e. Infinitive mark	50 'to count'	(90) Botmeur [ˈkɔntə] vs. (32) Plounéour-Ménez [ˈkɔnta]

This strong presence may result from our selection of maps but, on the other hand, these alternations are very frequent in this language. On a more general level, ([a]/[ə]) alternation is interesting because it offers a perspective on the degree of centralization of vowels by Breton speakers. Breton final-syllable vowels tend to be centralized when unstressed, especially in the central dialect area (Wmffre 1998: 8–10). The occasional appearance of a clear [a] in post-stress context goes against the general economy of the language (Martinet 1955).

The ([a]/[ə]) alternation occurs regularly but with only a slight number of occurrences across the area. The data for this alternation can be classified according to their frequency of appearance in each location under enquiry. Three zones can be spotted according to the spatial distribution of the difference under examination (see Table 6 and Figure 6).

This kind of difference occurs the most intensely in the central area in dark grey. In the two other areas located on the fringes, the frequency is less important. At present, it is still difficult to draw a conclusion because we do not understand how this small area is connected to the rest of the Breton-speaking region, continuously or discontinuously.

Finally, a closer look at these results led us to identify a particular realization of the kind that occurs recurrently. Some words may end with an [a] and a final consonant (a+Cons.) in post-stressed contexts when a more common realization with a schwa and a final consonant (ə+Cons.) would have been expected as in the

Table 6: Values for the alternation ([a]/[ə])

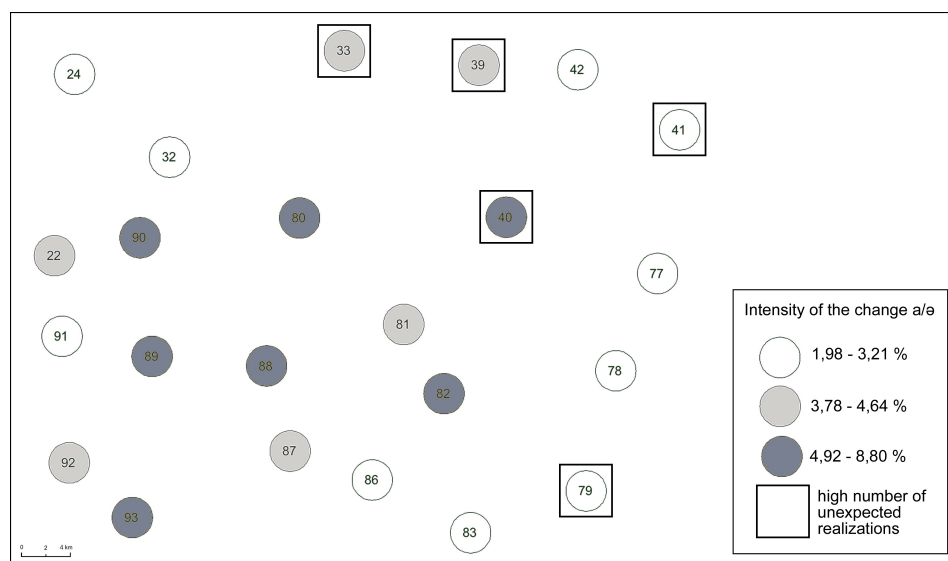| N° of the location	Name of the location	Number of occurrences for the alternation ([a]/[ə])	Proportion amongst these category of alternation (%)	Proportion of the [ə] words involved in the alternation (%)
080	Berrien	93	8.80	65.60
090	Botmeur	80	7.57	46.25
040	Plourac'h	79	7.48	65.80
089	Lannédern	78	7.38	66.60
093	Lennon	66	6.25	13.60
088	Collorec	60	5.68	55.00
082	Plounévézel	52	4.92	61.50
039	Guerlesquin	49	4.64	20.40
033	Plougonven	48	4.54	37.50
022	Saint-Cadou	42	3.97	38.00
087	Landeleau	41	3.88	58.50
081	Poullaouen	40	3.78	67.50
092	Pleyben	40	3.78	67.50
042	Loguivy-Plougras	35	3.31	51.42
091	Saint-Rivoal	34	3.21	52.90
032	Plounéour-Menez	33.00	3.12	27.27
041	Plougonver	32	3.03	34.37
024	Guimiliau	31	2.93	12.90
078	Locarn	30	2.84	40.00
086	Cléden-Poher	26	2.46	8.70
083	Motreff	24	2.27	62.50
077	Saint-Servais	22	2.08	77.27
079	Paule	21	1.98	19.04
		1056	99.9	

Figure 6: Intensity of the alternation ([a]/[ə]) across the area & locations with a high number of specific realizations

following example: NALBB map 109 'a day (long)' site (33) Plougonven [ˈdẹːvas] vs. site. (80) Berrien [ˈdẹːvəs].

This specific realization operates in a quite clear context especially in nouns and in the past participle forms of the verbs. This alternation cannot be explained by etymology.

The realization of [a] for [ə] in final unstressed syllables occurs only in 45 cases in this category that is, 4.26% of the total of the alternations ([a]/[ə]) and only 15 different lexemes in our corpus were affected by this specific realization. But it occurs in one little area and this specific realization is distributed as Table 7 shows.

This appears to be simply anecdotal on the surface but this specific realization regularly occurs in our corpus. Interestingly, the phenomenon is concentrated in the same micro-area around Plougonven (site 33). This means that this specific and not so common realization of a final unstressed (a+Cons.) could be salient from a sociolinguistic perspective as a sign of local identity and explain why an additional unneeded articulatory effort after the stressed syllable is produced for some words. Nevertheless, only a deeper scale study on these issues could help us to better understand the centralization of vowels in Breton.

Table 7: Distribution of the different occurrences of (a+Cons.) according to the
location investigated

Number of the point of enquiry	Location	Number of comparisons displaying the realization (a+Cons.)	Number of lexemes affected by the realization (a+Cons.)
33	Plougonven	15	9
40	Plourac'h	10	2
41	Plougonver	4	2
79	Paule	3	1
88	Collorec	5	2

6.2 The differences around one locality

Another approach is a close examination of the data for one location in order to determine the phonetic similarity between neighbouring sites. We decided to do this for Collorec (site 88), which is worth analysing because of its centrality and on account of the large number of results gathered for this place: 1354 alternations. They account for nearly 11% of all the non-identical correspondences we observed in our corpus. This number is far higher than the average of changes for each locality (519) of the area. The following figures display the linguistic distance between Collorec and its closest neighbouring locations from a phonetic point of view.

Table 8 shows how the differences observed between Collorec and its neighbouring sites are represented spatially. For each segment the number of alternations we noticed corresponds to the average observed for all the other segments across the area (225 on average).

Figure 7 shows that the level of intensity with respect to the differences varies according to the site-pair involved: three locations (80, 82, 90) are phonetically more distant from Collorec than the others. The distance we have noticed for sites 80 and 90 could be explained by a geographic factor: a bog named *Yeun Ellez* spreads out between these sites and Collorec. These findings about phonetic traits match Solliec's own research in the area. This marshy bog, which is an obstacle to travel, seems to promote linguistic differences more than the hills

Table 8: Number of differences to neighbouring sites around Collorec (site 88)

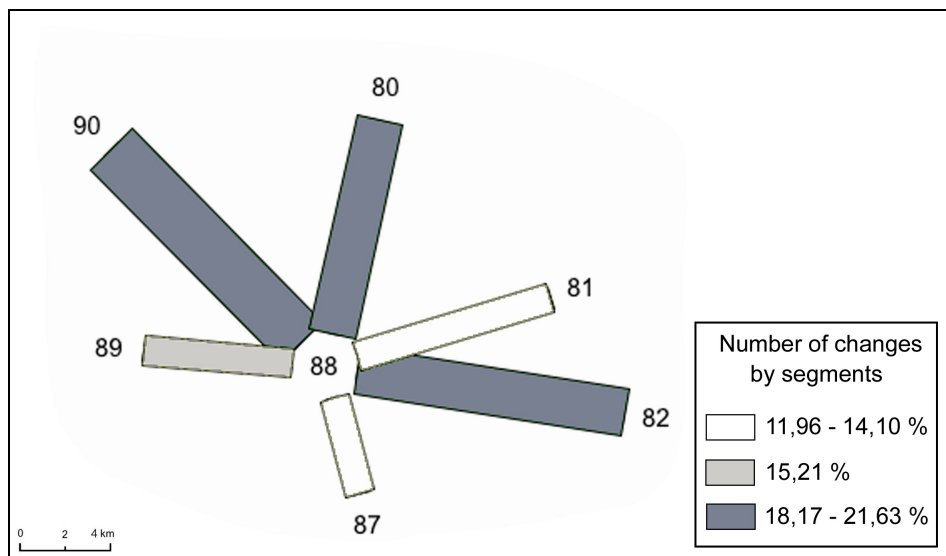| Site Comparison	Number of differences	Percentage compared to the number of differences
88>90	293	21.63
82>88	256	18.90
80>88	246	18.17
88>89	206	15.21
81>88	191	14.10
87>88	162	11.96
	1354	99.97

Figure 7: Number of differences to neighbouring sites around Collorec (point 88)

(named *Monts d'Arrée*), on the North-Western edge of the area, which are easy to get over (see Figure 3: FuzzyMatch map).

Moreover, through the statistical use of the data returned by the LD algorithm, the features of the linguistic variation in Collorec can be examined. The alternations observed for Collorec are detailed in the following figures.

Partition of the alternations for Collorec

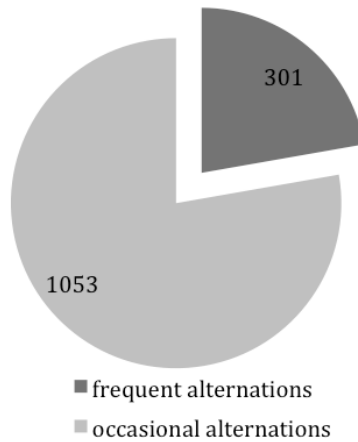■ frequent alternations
■ occasional alternations

Figure 8: Nature of the alternations between Collorec and its surrounding neighbours

The diagrams 8 and 9 show how the non-identical phonetic correspondences we found between Collorec and its surrounding locations are distributed. As these figures show, two types of alternations can be distinguished. The first one groups together the alternations that only show up once or twice in the results. In the second category are gathered the alternations, which can be clearly individualized (Figure 9) and which happen on a quite frequent basis.

These alternations are the same as those occurring across the whole area and with the same frequency, such as the different kinds of rhotics or the ([a]/[ə]) correspondence. Nevertheless, one must keep in mind that the number of occurrences for each alternation is low. For instance, 20 occurrences out of 1354 only account for 1.5% of them. We did not find any alternation unique to Collorec in spite of the large amount of data we have at our disposal for this specific location,

Nature of frequent alternations

R. ʁ by ʀ :
52

Ins. of ə :
27

Ins. of w :
21

R. ʁ by r :
27

R. a by ə :
25

R. ę by e :
41

R. ʀ by ʁ :
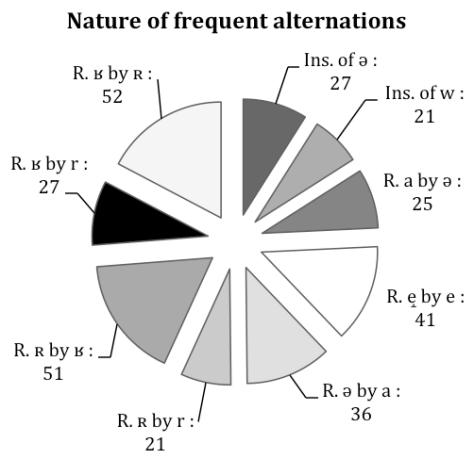51

R. ə by a :
36

R. ʀ by r :
21

Figure 9: Nature of the alternations between Collorec and its surrounding neighbours

whose features thus share in the general tendencies of the area, which is quite uniform in phonetic terms.

These two different approaches, the first about a specific alternation and the second dealing with the phonetic features of one location are the type of investigation opportunities this new usage of the LD algorithm offers for geolinguistic research.

7 Conclusion

To conclude, we want to stress that the sample of data used for this investigation was small because we deliberately restricted the size of the area under investigation as well as the amount of data. This paper reflects the first step of a more ambitious research to present a dialectometric analysis covering the whole Breton-speaking area. Our specific use of the LD algorithm allows us to treat qualitative data statistically. Linguistic variation can be described from a quantitative viewpoint while taking into account the specific features at stake so as to present a more comprehensive view of the data under analysis.

Furthermore, this new method allows us to divide the results obtained into 3 main categories: vowels, consonants and rhotics. These categories will have to be analysed individually while keeping in mind they can interact. A further step to this research will be to elaborate a dialectometric analysis on a larger

scale, based on the data of the NALBB, being conducted by Solliec as part of his PhD at the University of Brest in France. For this purpose, he will compare different dialectrometric techniques. The Breton-speaking area is interesting as it constitutes a real linguistic laboratory thanks to its important inner variations and to its distinctiveness from the neighbouring Romance varieties.

Acknowledgement

References

Brun-Trigaud, Guylaine. 2014. Un usage particulier de l'algorithme de Damerau-Levenshtein dans le domaine occitan. In F. Diémoz & D. Aquino-Weber (eds.), *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*, 127–147. Geneva: Droz.

Costaouec, Denis. 1998. *Le breton parlé à La Forêt-Fouesnant (Finistère). Pratique actuelle, problèmes de phonologie et de syntaxe*. Paris: Université Paris 5 PhD Thesis.

Costaouec, Denis. 2012. Linguistic geography of Breton and sociocultural motivations. *STUF - Sprachtypologie und Universalienforschung* 65(1). 47–64.

Falc'hun, François. 1981. *Perspectives nouvelles sur l'histoire de la langue bretonne*. Paris: UGE.

German, Gary. 1984. *Une étude linguistique sur le breton de Saint-Yvi*. Brest: Université de Bretagne Occidentale PhD Thesis.

German, Gary. 1991. Une méthode dialectométrique (assistée par ordinateur) pour l'analyse des atlas linguistiques. *La Bretagne linguistique* 7. 177–213.

Goebl, Hans. 2012. Introduction aux problèmes et méthodes de l'« école dialectométrique de Salzbourg » (avec des exemples gallo-, italo- et ibéroromans). In X. A. Álvarez Pérez, E. Carrilho & C. Magro (eds.), *Proceedings of the international symposium on limits and areas in dialectology (LimiAr). Lisbon 2011*, 117–166. Lisbon: Centro de Linguística da Universidade de Lisboa.

Guiter, Henri. 1973. Atlas et frontières linguistiques. In G. Straka & P. Gardette (eds.), *Les dialectes romans de la France à la lumière des atlas régionaux*, 61–109. Paris: Éditions du CNRS.

Heeringa, Wilbert. 2004. *Measuring dialect pronunciation differences using Levenshtein distance.* Groningen: University of Groningen PhD Thesis.

Jackson, Kenneth Hurlstone. 1967. *A historical phonology of Breton.* Dublin: Dublin Institute for advanced studies.

Le Dû, Jean. 2001. *Nouvel atlas linguistique de la Basse-Bretagne.* Brest: Centre de recherche bretonne et celtique, Université de Bretagne occidentale.

Le Roux, Pierre. 1924–1963. *Atlas linguistique de la Basse-Bretagne.* 77 sites of investigation. 600 maps. Rennes; Paris: Plihon et Hommay.

Martinet, André. 1955. *Économie des changements phonétiques. Traité de phonologie diachronique.* Bern: Éditions A. Francke S. A.

Moseley, Chris (ed.). 2010. *Atlas of the world's languages in danger.* 3rd edn. Paris: UNESCO Publishing.

Nerbonne, John & Wilbert Heeringa. 2001. Computational comparison and classification of dialects. *Dialectologia et Geolinguistica* 2001(9). 69–83.

Nerbonne, John & Wilbert Heeringa. 2010. Measuring dialect differences. In Jürgen Erich Schmidt & Peter Auer (eds.), *Language and space: Theories and methods* (Handbooks of Linguistics and Communication Science), 550–567. Berlin: Mouton de Gruyter.

Solliec, Tanguy. 2014. *Dialectométrie de la Basse-Bretagne : Enjeux, méthodologie et applications. Quantifier et qualifier la variation linguistique, vision symbolique de l'espace et écho du dynamisme synchronique en breton.* Brest: Université de Bretagne Occidentale PhD Thesis.

Séguy, Jean. 1971. La relation entre la distance spatiale et la distance lexicale. *Revue de Linguistique Romane* 35(138). 335–357.

Séguy, Jean. 1973. La dialectométrie dans l'Atlas linguistique de la Gascogne. *Revue de Linguistique Romane* 37. 1–24.

Wmffre, Iwan. 1998. *Central Breton* (Languages of the World/Materials 152). Munich: Lincom.