Unity in Diversity? A Fine-grained Approach to Linguistic Geography of Breton by Means of Dialectometry.[1]

Tanguy Solliec

1. Introduction

Diversity is a very common feature a researcher has to deal with when describing linguistic phenomena. However, its importance is often downplayed in the standard language models promoted by the so-called nation states, yet languages vary. Languages like Breton, with no official recognition and a low level of literacy among its speakers present situations in which linguistic variation abounds. Such languages can be viewed as linguistic laboratories, where internal linguistic diversity interacts intensively with geography. Dialectologists have described and aimed to model such situations for more than a century.

Describing and recognizing diversity in language is mostly done on a macro-level scale by comparing sets from different languages (Donohue et al. 2011; Hammarström 2016; Nettle 1998). On the other hand, the studies of determinants and of the processes of internal differentiation in minority languages are few and far between.

I propose to study local phonetic variation in Breton by means of a quantitative approach called *dialectometry*. Firstly, my main concern is to determine how the Breton-speaking area is structured according to the category of linguistic distance, and to observe how such findings fit the dialectal structure of Breton. Secondly, I will analyse different phenomena involved in linguistic distance across various areas. My objective is then to identify the determinants of linguistic variation in Breton and to gain a better understanding of the language dynamics. Moreover, exploring linguistic variation from a quantitative perspective offers the possibility to weigh the role of dialectal variables in shaping linguistic distance in Breton.

The paper will start with presenting the data and the area under investigation. Following that, I will present the methodology I have used to obtain my results. Then, I will focus on the results of linguistic distance between different locations of this area. Fourthly, I will examine in detail the linguistic facts the linguistic distance is made of, before determining their respective importance. Finally, I will

1. I would like to thank Professor Gary German, Dr Maxim Fomin and an anonymous reviewer for their gracious help with revising my English and for their valuable remarks.

summarise the outcomes of my research, making observations on the nature of linguistic distance in Breton.

2. The area and the data under scrutiny

Observers and speakers often describe Breton as a language in which internal variation is significant. This is regularly counted as the main factor to explain why mutual intelligibility is not always possible for speakers from different regions (Broudic 2004). However, even if this general perception is widespread, it does not give any precise indication as to how close or distant the different varieties are from each other and it is difficult to find out about the facts which account for such distance.

2.1. *Nouvel Atlas Linguistique de la Basse-Bretagne* (NALBB)

In order to take a more comprehensive approach, this study is based on the data provided by the *Nouvel Atlas Linguistique de la Basse-Bretagne* (Le Dû 2001, henceforth abbreviated as NALBB). This work follows the pioneering methodology proposed by Gilliéron in the *Atlas Linguistique de la France* (Edmont & Gilliéron 1902–1910) and includes 600 maps. Each map presents a distinct lexical item. The area covers all of western Brittany with a network of 187 locations, scattered all across the area at regular intervals.

The fieldwork for the NALBB started in 1969 (Le Dû 1972) and was carried out by Le Dû with the assistance of his co-workers. The project came to its end with the publication of the NALBB in 2001. This atlas was intended to be the first part of a much larger study, with the aim to update the *Atlas Linguistique de la Basse-Bretagne* by Le Roux (1924–63) on a much finer scale (see figs. 1 and 2 for the extent of work carried out by Le Roux and Le Dû).

The two volumes of the NALBB are mainly dedicated to phonetic material and contain nearly 110,000 forms. All the data werre transcribed by Le Dû himself. Such a precaution preserved considerable unity in the transcriptions and therefore avoided the confusion of having fieldworkers interpreting the data according to different standards. The notation used by Le Dû renders phonetic details very precisely. For instance, four degrees of aperture are distinguished for the mid vowels (ɛ̝ ɛ e e̞ / ɔ̝ ɔ o o̞) and three different degrees of diphthongization (ej ei ei) are differentiated.

Fig. 1: Map of locations investigated for the ALBB

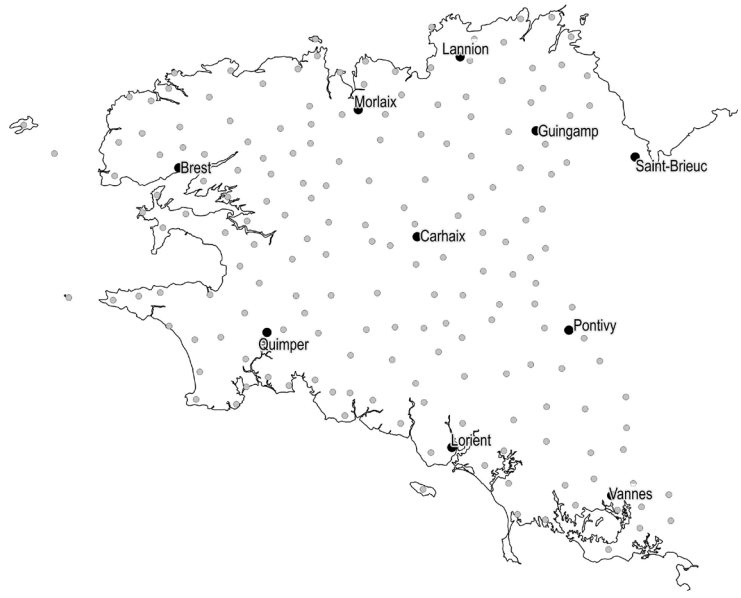

Fig. 2: Map of locations investigated for the NALBB

2.2. The area selected in central Brittany

For this study, I have selected the area in central Brittany that I have previously investigated in (Solliec 2017; see also Brun-Trigaud et al. 2016). The area is situated between different dialect regions but is also not far from Carhaix,[2] a nexus of linguistic innovations for centuries. I propose to analyse it by means of linguistic distance and to observe which phonetic facts contribute to it.

The dataset I have explored for my analysis consists of the data taken from the NALBB at 23 locations (fig. 3).[3] Each point of the atlas was linked to other neighbouring locations in order to form a network. The 53 relations or segments constitute the frame that I have used to compare the different linguistic forms in the corpus.

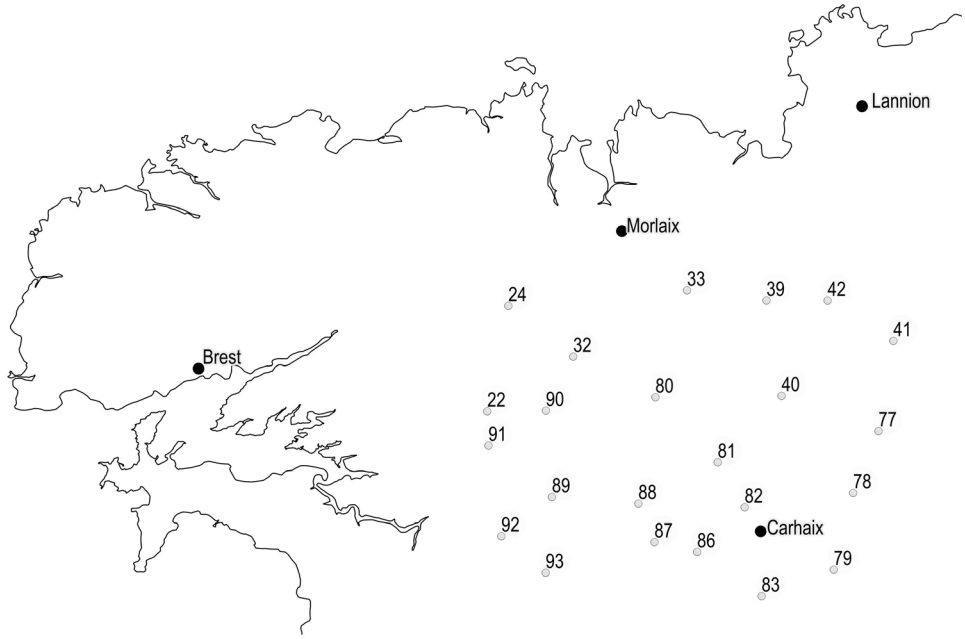

Fig. 3: The locations investigated for this study

2. Carhaix does not lie far from Pounévezel **(Plounevezel), point 82 of the network.**

3. The numbers on the map refer to the numerical code associated with each location in the NALBB. See Appendix 1 for a list of the codes.

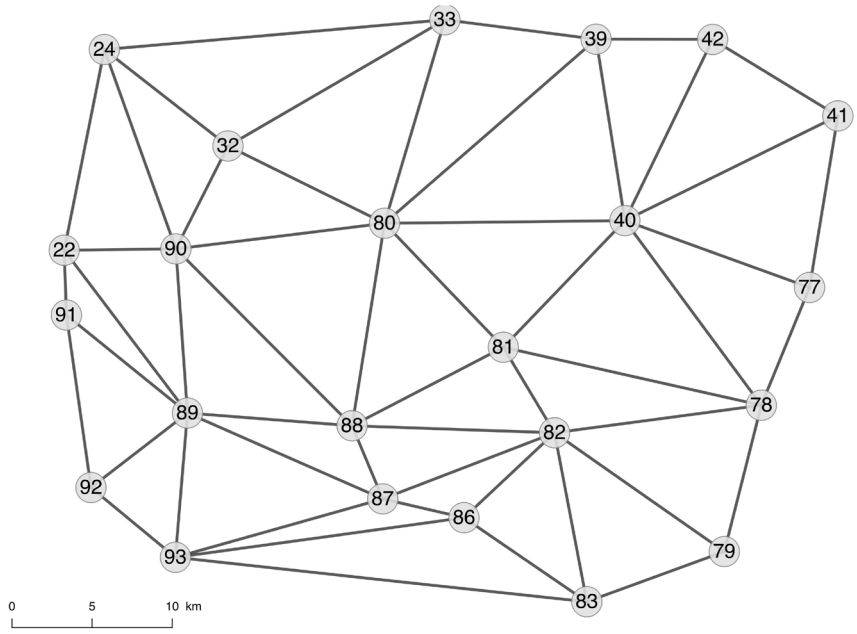

Fig. 4: The corresponding network of locations for the comparisons

The data surveyed is analysed according to 165 maps taken from the NALBB,[4] each map presenting one lexeme. 3,795 phonetic forms for 165 different lexemes were analysed. All in all, this produced 8,745 comparisons in which each element was monitored.

3. Dialectometry as a method to weigh linguistic distances

3.1. The concept of linguistic distance

Linguistic distance can be described as the totality of differences which exist between two varieties, be they in related but different languages or in varieties of the same language. Paul Heggarty defines it as "the degree of similarity/difference between any two language varieties" (Heggarty 2000: 531). It is generally expressed as a percentage of resemblance.

4. See Appendix 2 for a detailed inventory.

Jean Séguy, the leader of the *Atlas linguistique de Gascogne* project (1954–73), initiated the calculation of similarity rates between different dialectal varieties of Gascon. He laid the foundations for the study of their spread across space (1971, 1973), but unfortunately, Séguy passed away shortly afterwards. One of his followers, Hans Goebl (1982, 2005) developed the methodology of dialectometry further by implementing classificatory techniques used in biosciences, computerizing his workflow and developing the statistical and cartographical aspects of the methodology. In the last two decades, John Nerbonne and Wilbert Heeringa, amongst others, have developed a specific use of the Levenshtein algorithm applied to dialectology (of which see below) (Nerbonne et al. 1999; Heeringa & Nerbonne 2010; Wieling & Nerbonne 2015). Western Brittany is not *terra incognita* for dialectometry. Previous work had already been carried out by German (1984, 1987, 1993) and Costaouec (1998, 2012).

3.2. The Levenshtein algorithm

The Levenshtein algorithm (Levenshtein 1966) was introduced into the field of dialectometry by Kessler (1995) in order to test the distribution of Irish dialects statistically. Nerbonne and Heeringa made substantial contributions by expanding its use and confirming its validity (Heeringa 2004; Nerbonne & Heeringa 2010; Wieling & Nerbonne 2015). In a few words, the Levenshtein algorithm compares two chains of characters and calculates a similarity rate between them. To do so, the algorithm aligns two strings of characters. It then calculates the numbers of operations needed to transform the first form into the second by using simple operations such as 'replace one character by another', 'delete one character' and 'insert one character'. The tool then calculates a similarity rate between the two forms by counting the number of operations needed to go from one string to another.

3.3 A specific use of the Levenshtein algorithm

In this study, I have focused more specifically on the measurement of phonetic distance between two forms of the same word. Since I intended to identify the components of the linguistic distance, I have used a version of the Levenshtein algorithm customized by Guylaine Brun-Trigaud (2014), which has already been applied to the Breton language (Brun-Trigaud et al. 2016; Solliec 2017). The version of the Levenshtein algorithm used here, not only operates alignments of phonetic pronunciations and counts the differences between the chains, it also returns the nature of the transformations provided by the algorithm. My hypothesis is that it reflects the nature of the diatopic changes in language variation when moving from one location of the NALBB to another one. Collecting them over a lot of comparisons would therefore permit us to identify and to value the main phonetic changes across the Breton-speaking area.

If this version of Levenshtein distance is applied to two forms in map 163 NALBB (the sun *an heol*), one would obtain the following calculation:

(1) Saint-Servais (77) [ən ˈɛwəl] ~ Locarn (78) [ən̥ ˈhɛo̥l]

									Change in the number of syllables
(77) St-Servais	ə	n	ˈ		ɛ	w	ə	l	
Nature of the operations		Replacement _n by _n̥		Insertion of h_		Suppression of _w_	Replacement of ə by o̥		
(78) Locarn	ə	n̥	ˈ	h	ɛ		o̥	l	
Number of differences	0	1	0	1	0	1	1	0	0
	Four differences								

Fig. 5: Levenshtein comparison of two pronunciations

Four differences for this case accounts for 42.8% of similarity between the two forms. The difference rate is obtained by dividing the number of differences obtained by the number of characters of the longest string of characters and then, multiplying it by 100. In order to calculate a similarity rate between these two pronunciations, one needs to subtract the previous rate to 100. For this example, the operation can be summed up as follows:

$$100 - ((4/7)^{*}100) = 42.8$$

For each segment (i.e. a pair of locations) such as the segment (77)-(78), I collected all the similarity rates for 165 lexemes altogether before calculating an average rate of linguistic similarity which is 79.9% in that case. In addition, so that not to confuse the operations returned by the algorithm with the notion of 'linguistic change', I have called them as 'modifications'. I have also collected all the different modifications accompanied by their opposites.[5] I call these groupings 'alternations'.

5. When one segment is examined, it actually could be read in two ways. Firstly, when the segment is presented as the (77)-(78) sequence, the modification of the segment is the 'replacement of [_n] by [_ n̥]'. Secondly, when the segment is examined from an opposite perspective, i.e. as the (78)-(77)

Each alternation has a number of occurrences, which correspond to the number of modifications made and then counted by the algorithm. In this study, the possible direction of linguistic change has not been investigated but I have looked for specific sounds involved in linguistic distance.

4. Distribution of the similarity rate across the area

Once all the operations of comparisons were completed, I obtained a sum of 12,066 occurrences of modifications gathered into 761 distinct alternations (see table 1). The average rate of similarity for the area is 74.84%, which seems to be quite high and gives an impression of relative linguistic uniformity. However, it is also interesting to observe how the similarity rate is distributed spatially for each of the 53 segments under consideration.

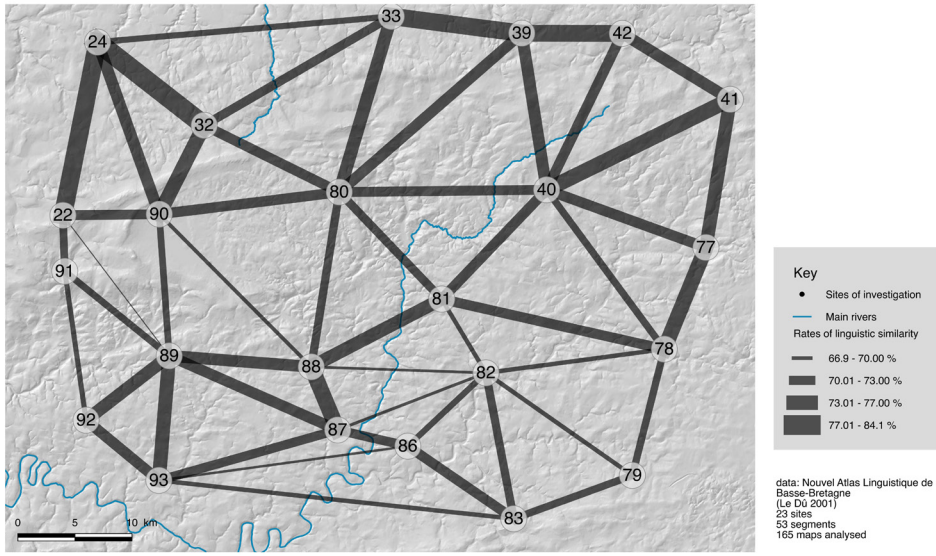

Fig. 6: The distribution of the similarity rate across the area

sequence, the modification is the 'replacement of [_n̥] by [_n]'. In order to avoid such confusion, both operations were merged into one grouping: 'alternation of [_n]/[_ n̥]'.

In figure 6, the different segments have been classified into four categories depending on their importance (ranging between 67% and 84% of phonetic similarity). At the first glance, there is a general impression of convergence between the data examined. More precisely, three different areas where the similarity rates are over 77% can be distinguished. They show a close linguistic affinity despite the geographical distance and constitute, each in their own way, centres of linguistic convergence.

By contrast, specific areas where the rate of similarity is below 70% can also be identified. For Guiter (1973: 79), one of the first dialectometricians, this level corresponds to the difference between two dialect areas for the Romance languages. Mapping these different areas as in figure 7 shows a clear distinction between the spaces of convergence (segments in dark grey) and those of divergence (areas in light grey).

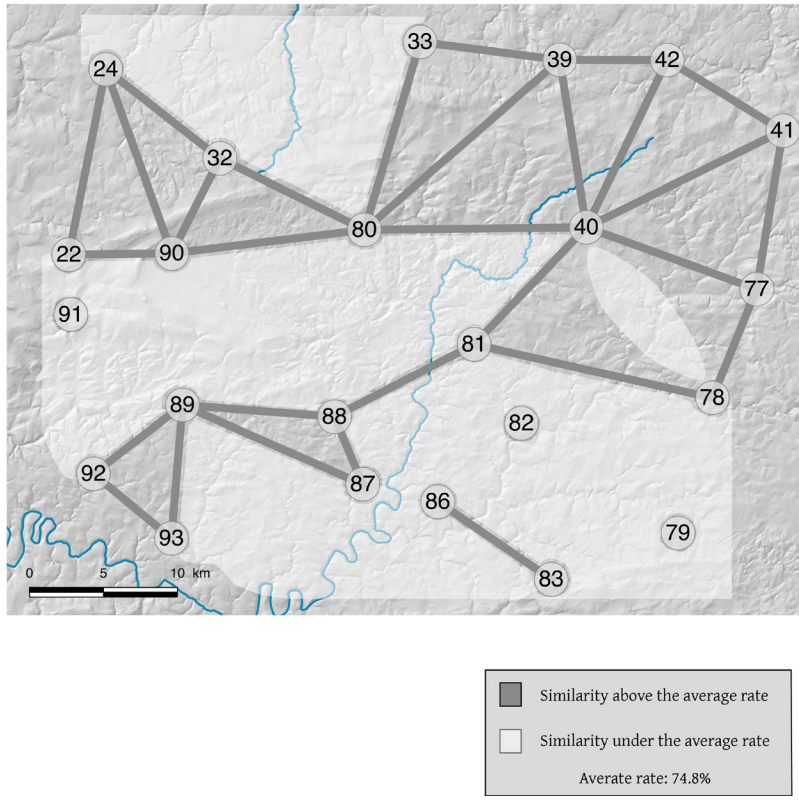

Fig. 7: Areas of convergence and divergence

Three micro-areas are reported in light grey. They have been identified according to a statistical criterion: they are made up of the pairs of locations whose similarity rate is under 70%, the lowest category in the sample. It is possible to break them down into three groups:

- segments (24)-(33) and (32)-(33), in the upper part of the figure
- the area drawn by the lines made of locations (22), (90) and (80) on the one side and the line made by the sites (92), (89), (88) and (81) on the other side, in the centre of the map
- the south-eastern part of the area, whose similarity rate is the lowest rates of the sample; more specifically, site (82), Plounévezel, appears to differ strongly from its neighbours on a phonetic basis (Brun-Trigaud et al. 2016); segment (40)-(78) can also be associated with this large area.

Broadly speaking, the first two areas can be correlated quite well with the local geography. In the first case, it matches with the valley and the river Queffleuth. The second area corresponds to a marshy area named Yeun Ellez.

The explanation for the last area is more difficult to formulate since in some way it challenges the model proposed by Falc'hun (1963, 1981) in his description of the dialectal structure of Breton. He convincingly showed that Carhaix has been a linguistic centre in which linguistic innovations originated for a long time. This city was the place where a kind of central Breton emerged (Wmffre 1998; Favereau 1992: v) which was quite easily understood by most Bretons. I will later discuss a few possible explanations to clarify this paradox, which will assist us in identifying a divergent area in the space of convergence.

Of course, one major bias in this analysis could be the method designed and employed for this study. However, the same kind of distribution appears when using another tool and a different way of weighting the similarity, namely the Gabmap web application (Nerbonne et al. 2011; Snoek 2014), as can be seen in figure 8. These matching results validate the specific methodology designed for this study and the findings confirm the previous observations from figures 6 and 7.

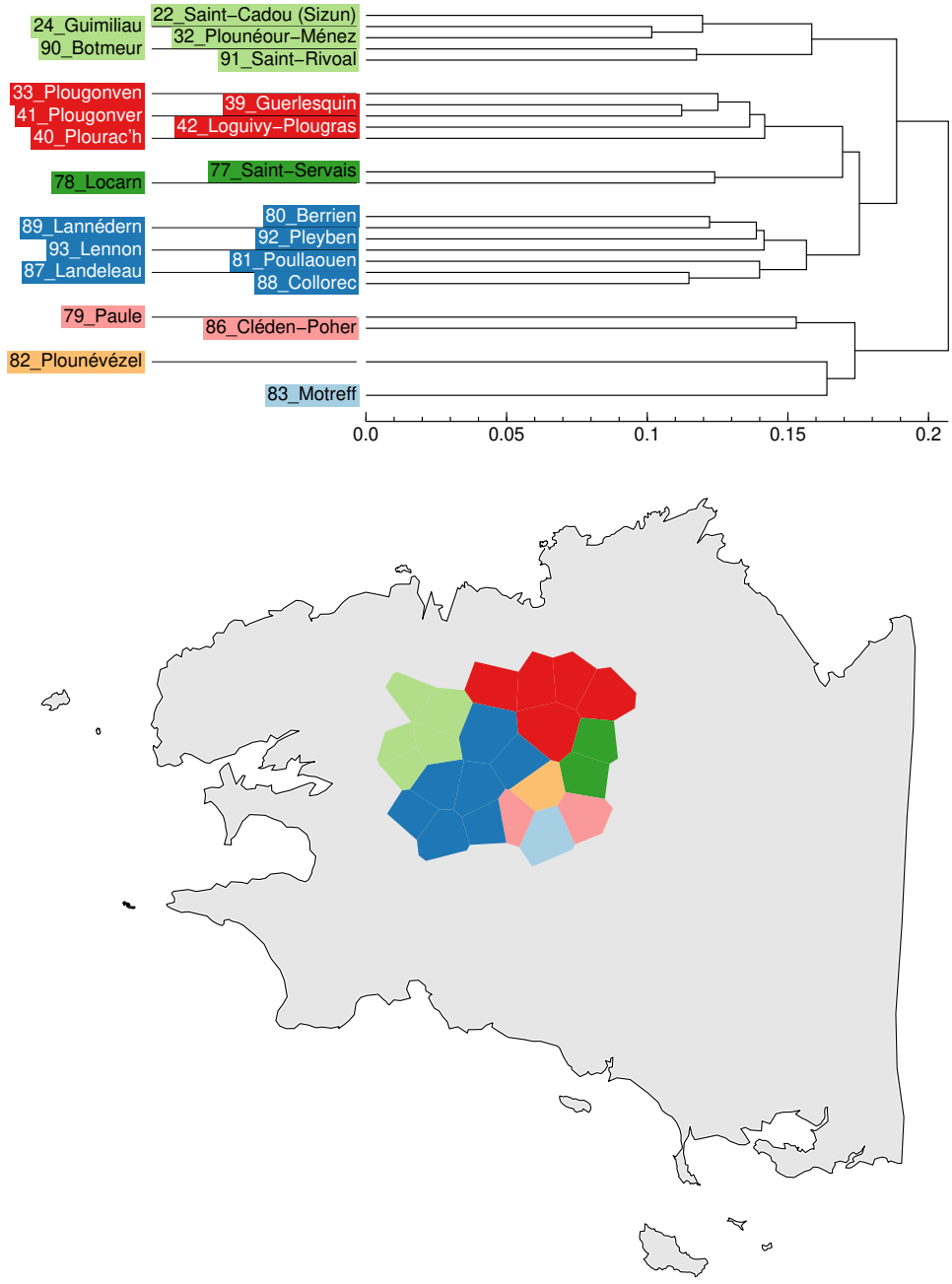

Fig. 8: Clusters of the locations according to the Group Average Clustering method (seven clusters) and their geographical distribution

The degree of difference between the data of the different locations investigated is shown on the cluster distribution in figure 8.[6] The most similar locations are all grouped together into small clusters. This confirms the initial observations:

- the gap between sites (32) and (24), on the one hand, and (33), on the other hand
- the divergence between locations (90), (91) and those in the dark blue cluster (locations (80), (89), (92), (93), (81), (87), (88))
- the phonetic divergence between the locations (79), (86), (83), (78) and (79), on the one hand and (82) and (83), on the other
- the central position of Berrien, NALBB point (80), at the crossroads of different clusters (cf. Plonéis 1983).

Although the level of discrepancy is quite high in these zones, the distribution of the similarity rate across the area shows even so good connection between the different locations since no segment displays a similarity rate under 68%. Such values are not that surprising when one takes into account the fact that, on the one hand, a linguistic continuum is under study here and, on the other hand, the data analysed is provided by only one level of the language: its phonetics. It reduces therefore the number of variables (morphology or vocabulary, amongst others) that may have increased the linguistic distance.

An overall approach to linguistic similarity tends to result in hiding the particular phenomena occurring in the area. Thanks to the specific use of the Levenshtein algorithm made in this study, it is possible to investigate the data and evaluating its relative importance in producing linguistic distance.

5. Elements of variation by segments

The degree of linguistic similarity varies across the area, as one would expect from a vernacular language. The different values have been divided into four distinct categories. However, the features this variation is composed of and how these phenomena are distributed across the area has not been identified yet. Therefore, the components of phonetic variation involved for each segment (i.e. the pair of locations whose data is compared) must be specified in order to observe whether it is made up of the same characteristics or whether some of them are prevalent in some places more than in others.

6. This must be distinguished from a phylogenetic tree, which indicates genetic relationships. Here the clustering is based only on statistical comparisons.

	Numbers of occurrences	Proportion	Number of alternations	Proportion
Total	12066	100%	761	100%
Vowels	6921	57.35%	488	64.12%
Rhotic consonants	2429	20.13%	66	8.67%
Non-rhotic consonants	2303	19.21%	133	17.47%
Morphology	443	3.8%	90	11.82%

Table 1: Repartitioning of the results into different categories

When analysing all the modifications returned by the algorithm during the calculation of the similarity rates, I have identified patterns that structure the phonetic distance across the area. Their distribution can be characterized quantitatively and geographically.

I have previously observed (Solliec 2017) that the statistical distribution followed the Pareto principle. It means that the majority of the modifications gathered for this study are distributed between a few alternations only. On the other hand, most of the alternations have barely a few occurrences. The most important alternations from a numerical perspective have been selected for each segment in order to determine the main facts involved in the phonetic distance across the area. I previously detailed (*ibid.*) an argument in favour of dividing the modifications into three distinct categories: the vowels, the non-rhotic consonants and the rhotic consonants. Each one exhibits a specific phonological and statistical behaviour as well.

One more observation must be made. The different lines established in the following figures (figure 9 to figure 15) are not to be read and understood as classical isoglosses. In reality, they distinguish the area where the labelled alternation happens frequently. It does not mean that the labelled phenomenon does not occurs on the other side of this line; it could take place but at a lower frequency.

5.1. The vowels

Focusing on the vowels, the overwhelming importance of the data involving schwa is striking. As table 2 shows, one is dealing mainly with two alternations:

- the realization or not of [ə] (alternation [ə]+)
- an alternation involving the strength of the realization: [ə] realized as [ᵊ] on some occasions (alternation [ə]/[ᵊ]).

The following alternations can also be noticed:

- that between [a] and [ə] (alternation [a]/[ə]; see Brun-Trigaud et al. 2016: 144–8)
- the switch between [e] and [ə] (alternation [e]/[ə])
- that between [e] and a more closed equivalent, [e̝] (alternation [e]/[e̝])

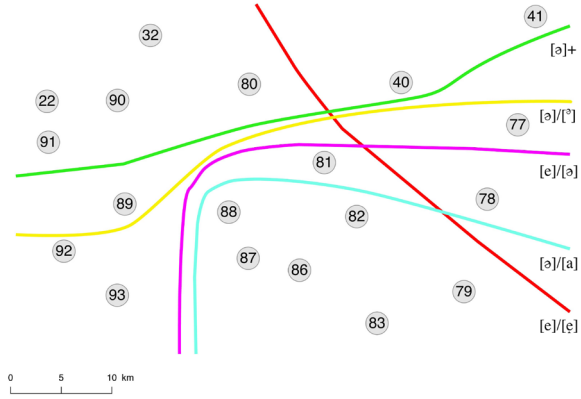

Fig. 9: Main vocalic alternations

Rank	Vocalic alternations	Number of occurrences	Proportion in the category
1	[a]/[ə]	386	5.57 %
2	[e]/[e̝]	347	5.01 %
3	[ᵊ]/[ə]	290	4.19 %
4	[ə]+	263	3.8 %
5	[e]/[ə]	220	3.17 %

Table 2: Numbers for the main vocalic alternations

The importance of the alternations which entail the schwa highlights the importance of the centralization of vowels in central Breton in a post-tonic context (Wmffre 1998). Many different possibilities are gathered here. This phenomenon results from a strong tonic accent, which leads post-tonic vowels to lose their quality or even to be elided (alternation [ə]+).

The different lines reflecting the distribution of the vocalic alternations correspond to the same geographical distinction between a north-western area and a central zone that Falc'hun (1981: 251–66, 505–10) noticed for the diffusion of some lexical innovations. Their distribution confirms his findings on a phonetic level (*ibid*., pp. 213–32) with the exception of the alternation [e]/[e̬]. Moreover, the presence of successive lines of that kind in a small area suggests a transition area between two core dialectal zones.

5.2. Rhotics

Rhotics are a phonological category which is hard to describe in phonetic terms. Nevertheless it has been shown that they tend to share similar phonological behaviour (Lindau 1985; Lagofeld & Maddieson 1996; Wiese 2011) and they are also liable to alternate easily with each other (Scobbie 2006). In the area selected for this study, only three different occurrences of the sounds [r], [ʁ] and [ʀ] have been encountered, even if other variants are present in the NALBB.

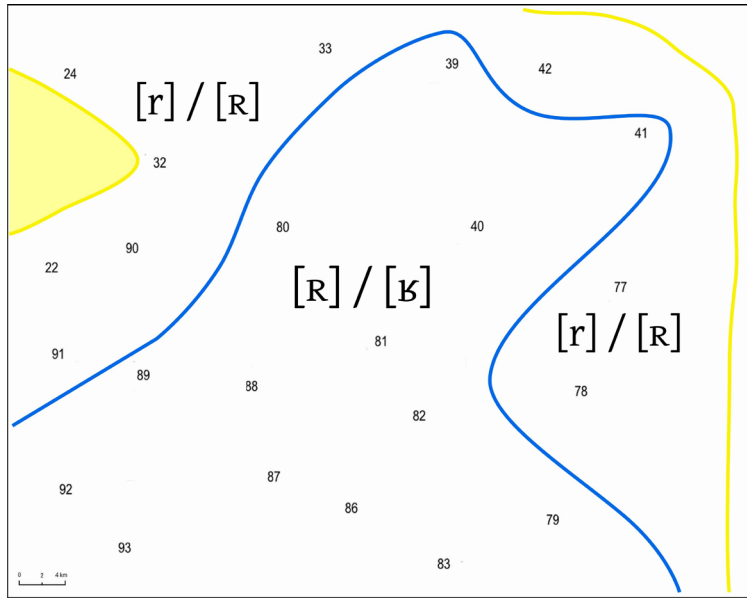

Fig. 10: Main rhotic alternations

Rank	Rhotic alternations	Number of occurrences	Proportion in the category
1	[_ʀ]/[_ʁ]	347	14.28 %
2	[_ʀ_]/[_r_]	299	12.3 %
3	[_ʀ_]/[_ʁ_]	293	12.06 %
4	[_ʁ]/_r]	283	11.65 %
5	[_r]/[_ʀ]	196	8.07 %

Table 3: Numbers for the main rhotic alternations

The main alternation involves [ʁ] and [ʀ] in two different positions (final and intervocalic) in the central part of the area under study. As figure 10 shows, the alternation of the sounds [r] and [ʀ] prevails at each edge of this central zone. This suggests that a change is currently taking place there, whereby a uvular trill is being replaced by a fricative. The older way to pronounce the sound is favoured in peripheral areas such as north-western Léon.

This specific distribution of the sounds across the area then reflects the different historical steps of the realization of the phoneme /r/. The linguistic geography suggests then that /r/ was first pronounced [r], then [ʀ] and finally [ʁ]. Otherwise, somewhat surprisingly, location (82), which is isolated when considering the linguistic distance in an aggregate view (cf. section 4), does not differ from the neighbouring locations at the level of rhotic sounds.

This specific distribution also illustrates the role played by Carhaix in the diffusion of innovations across the centuries, as shown by Falc'hun (1981). This uvular realization could also be an echo of a more general trend initiated from French (Trudgill 1974) and promoting the pronunciation of /r/ as a uvular fricative in the north-west of Europe. Thus, Central Breton still seems to be leading the way to changes in the language. This view is also supported by Wmffre (1998: 6).

5.3. The non-rhotic consonants

For this category of sounds, no clear pattern of the results distribution can be discerned at the first glance. Moreover, the non-rhotic consonants are the most stable category in the sample and more generally in human languages (Campbell 1999; Wälchli 2010). Thus, they can be studied as isoglosses when investigating diatopic variation. They do not vary too much and, when they do, it is meaningful for dialectology (see the case of the intervocalic [_z_] in section 5.3.2). However, no clear general pattern can be distinguished at first, as table 4 indicates.

Rank	Consonantal alternations	Number of occurrences	Proportion in the category
1	[_n]/[_ṇ]	162	7.03 %
2	[h_]/[x_]	152	6.6 %
3	[h_]+	93	4.03 %
4	[_w_]+	76	3.03 %
5	[_ʒ_]/[_z_]	75	3.25 %

Table 4: Numbers for the main consonantal alternations

Nevertheless, when observing the nature of the alternations gathered in this category, the most striking one is the alternation of the sounds [h] and [x] in initial or final position. The latter is found more specifically around Plounévézel (location (82) of the NALBB). This important variability across the area may lead us to reassess the distinction between [x] and [h] as distinct phonemes in this area and consider them as allophones of a phoneme /h/. Both sounds seem to belong to the same phonological space (Moulton 1962). It seems difficult to establish a clear opposition between [x] and [h] in minimal pairs. Of course, only a more precise study on the subject could determine the in-depth relationship between the sounds [h] and [x] across the area. The pronunciation of [x] instead of [h] is mocked by some speakers (Favereau 1997),[7] which can only indicate a local realization as a free variant. However, German (1984) mentions a clear phonemic opposition between both sounds for Saint-Yvi in southern Cornouaille.

Nevertheless, for the category of the consonants, it is not enough to focus on the nature of the alternations of sounds. Their position in words should also be taken into into account so as to gain a clearer picture of the linguistic facts.

5.3.1. Initial position

Concerning the initial position, the area appears to be divided along a thick diagonal line, along which an initial [h] is realized before a vowel as figure 11 shows. Alongside this trend, a fricative alternation [x]/[h] occurs also frequently all over the investigated area.

7. Cf. the alliteration associated with the people of Spézet (Finistère, south to locations 86 and 87 of the NALBB) : *Ar c'hi hag ar c'hazh o c'hrognal 'barzh ar c'hogn* 'The dog and the cat are purring in the fireplace'. (Favereau 1997 : 159)

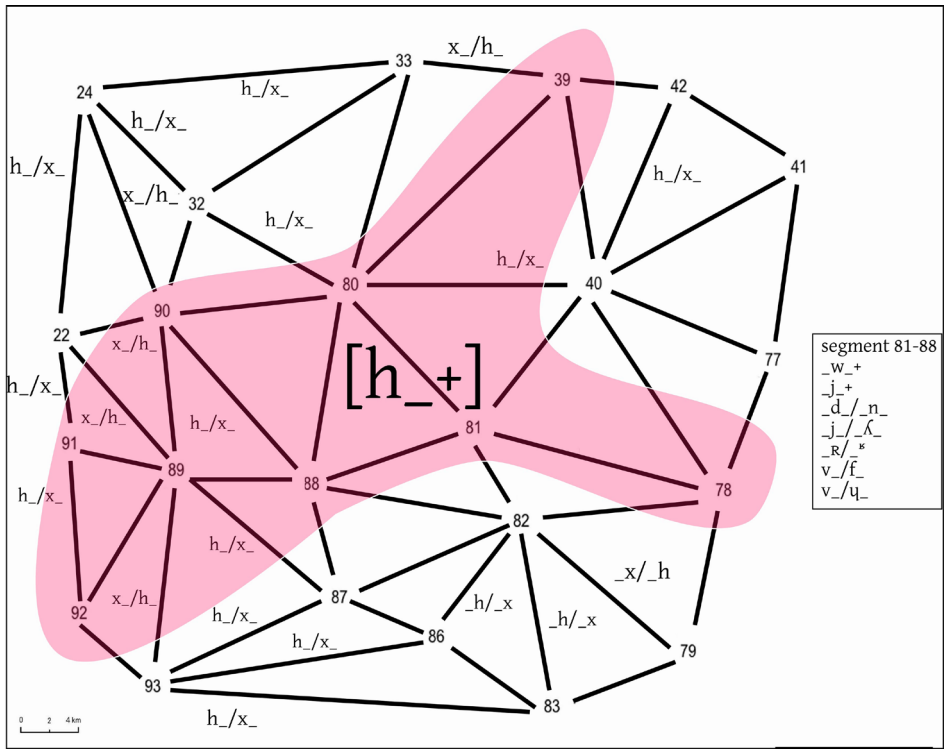

Fig. 11: **Main fricative alternations**

This phenomenon is well known in Breton dialectology and was chosen by German (1993) as a diachronic variable in his study. This initial [h] is the reflex of a corresponding sound in Old Breton, which had been lost very early in the north-western part of Lower Brittany (example (2)). Moreover, some words presenting a similar context can also exhibit an initial [h], which had developed epenthetically in those cases as in example (3):

(2) NALBB map 287 to sow *hadañ* (etymological [h])
Saint-Cadou (22) [ˈaːda] ~ Paule (79) [ˈhaːd̥o̥]

(3) (3) NALBB map 335 foal *eubeul* (epenthetic [h])
Saint-Cadou (22) [ˈøːbøl] ~ Paule (79) [ˈhøːbøl]

The NALBB confirms the previous observations of the phenomenon for the sample area, which is deeply rooted in the diachronic development of Breton as presented by Jackson (1967: 557–60) and Falc'hun (1981: 354–60).

After investigating more closely the main alternations gathered for each segment, it can be noted that a group of them appears to be relevant. It is made up of the following elements: [j_]/[ɟ_], [g_]/ [ɟ_] and [k_]/[c_] and they have been labelled 'palatalisation'. They are in fact instances of a trend to realize the corresponding sounds in the palatal area of the mouth. Palatalisation is a common feature in Breton and it is particularly widespread in some varieties as in the *Vannetais* dialect and its surroundings (Falc'hun 1981: 325–334; Le Pipec 2015; Plourin 2005: 20). In this study, the situation is quite restricted as the following figure 12 displays it.

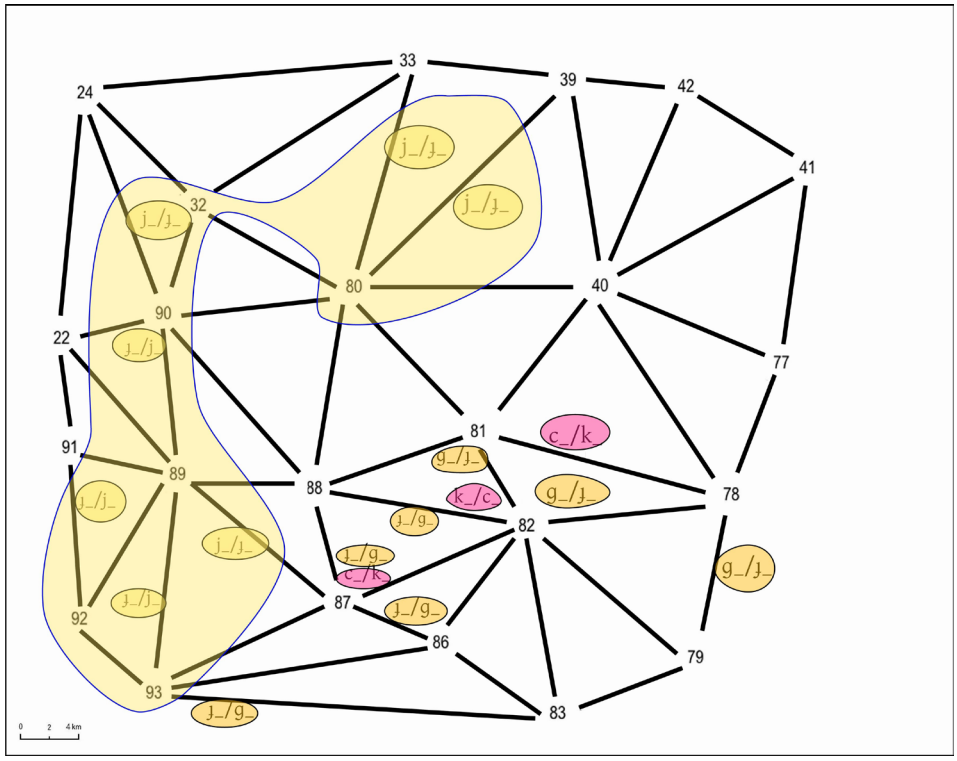

Fig. 12: Palatalisation phenomena in initial position

In the area studied, a specific trend towards the palatalisation of [j] into [ɟ] occurs in a few locations. This feature in fact constitutes a kind of local isogloss following a slim strip around the Yeun Ellez area.

The initial position in words therefore constitutes a place for the possible realization of a [h_]. Of course, a more detailed view of the data would establish to what extent a correlation between the palatalisation and the presence of fronted or palatal vowels is relevant.

5.3.2. Intervocalic position

The behaviour of the consonants in intervocalic position is intriguing since different trends happen in this position. In the northern part of the area, the nasal and the rhotic consonants tend to geminate (alternation [_n_]/[_nn_] or [_r_]/[_rr_]). Simultaneously, the sound [_ʎ_] occurs in roughly the same area whereas outside of it, its counterparts are simpler such as [_l_] or [_j_]. This suggests that the word-internal boundary of the syllables in this part of the area tend to be clearly structured.

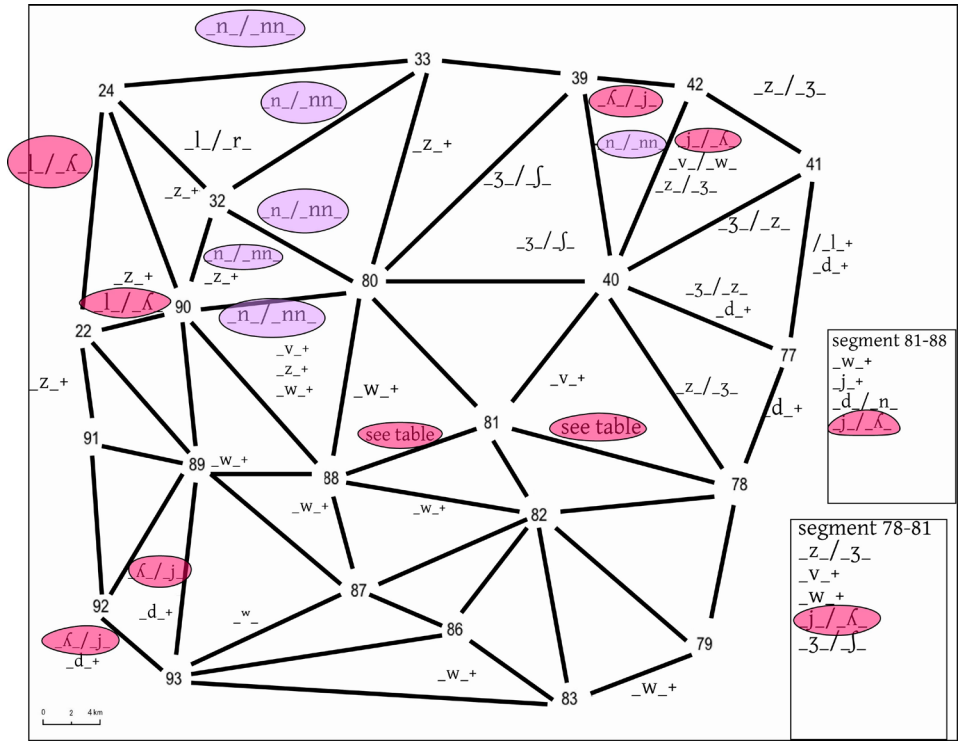

Fig. 13: Gemination in intervocalic position

Next, there is a strong tendency for intervocalic voiced consonants to be elided, such

as [_d_], [_z_], [_w_] or [_v_], as can be seen in the following example and figure 14:

(4) NALBB map 126 a week *ur sizhun* (alternation [_z_]+)
Saint-Cadou (22) [ər ˈziːzyn] ~ Botmeur (90) [ar ziːn]

(5) NALBB map 210 mouse *logodenn* (alternation [_d_]+)
Plourac'h (40) [loˈgɔːdən] ~ Saint-Servais (77) [loˈgoːˑˀn]

(6) NALBB map 141 wind *awel* (alternation [_v_]+)
Lannédern (89) [ˈaeˌl] ~ Botmeur (90) [ˈaːvəl]

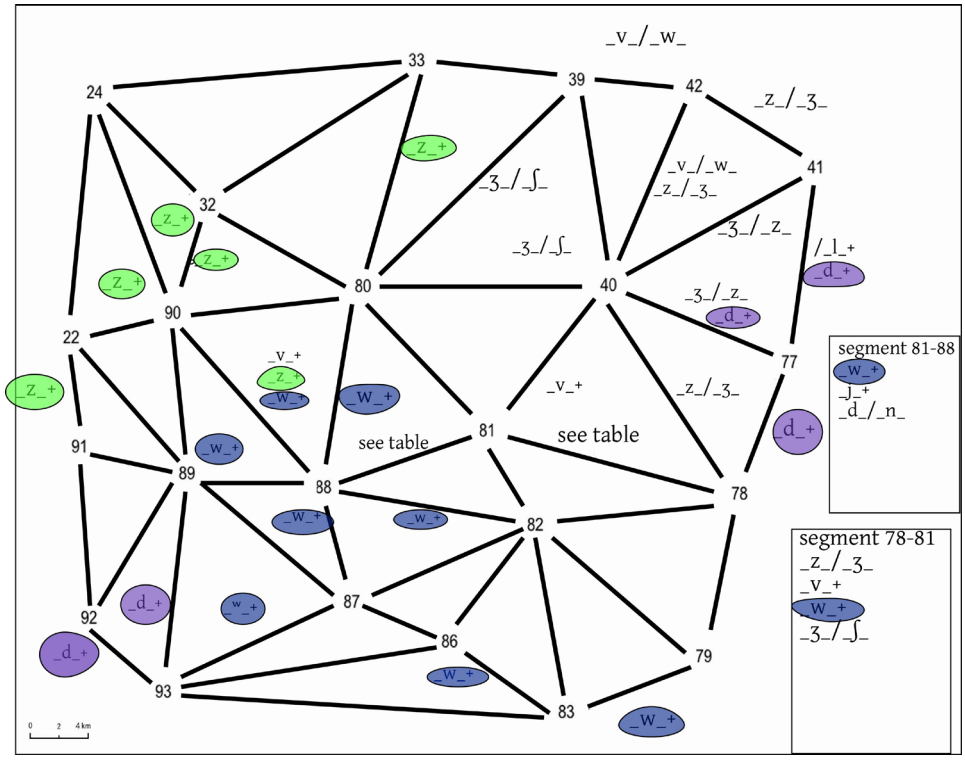

Fig. 14: Voiced intervocalic alternations

These results suggest that voicing as a phonetic feature contributes to the shaping of the phonetic distance across the area. Interestingly, the implication of voicing for the possible realization of an intervocalic consonant to some extent echoes the process of initial

mutations (when involving lenition) or final devoicing but inside the word. This feature tends to make the sounds it is associated with less stable than their unvoiced counterparts. It would explain, then, the tendency of these sounds to vary.

More specifically, the realization of [_z_] does fit in with a well-known dialectal variable: the evolution of the historical interdental fricative [ð] into [z] in Léon and its elision elsewhere across the area.

(7) NALBB 89 late *diwezad*
Saint-Cadou (22) [diˈveːzat] ~ Paule (79) [diˈvɛːɛt]

Figure 14 illustrates well that this isogloss separates the northwest fringe from the central area of Brittany where [z] as a reflex of Old Breton [ð] is practically never realized. The intervocalic position, therefore, is a position where non-rhotic consonants are liable to vary and then to increase the phonetic distance.

5.3.3. Final position

The main results that have been observed for the final position largely concern the syllabic nature of final [_n] (the alternation '_n/_n̥' in my notation).

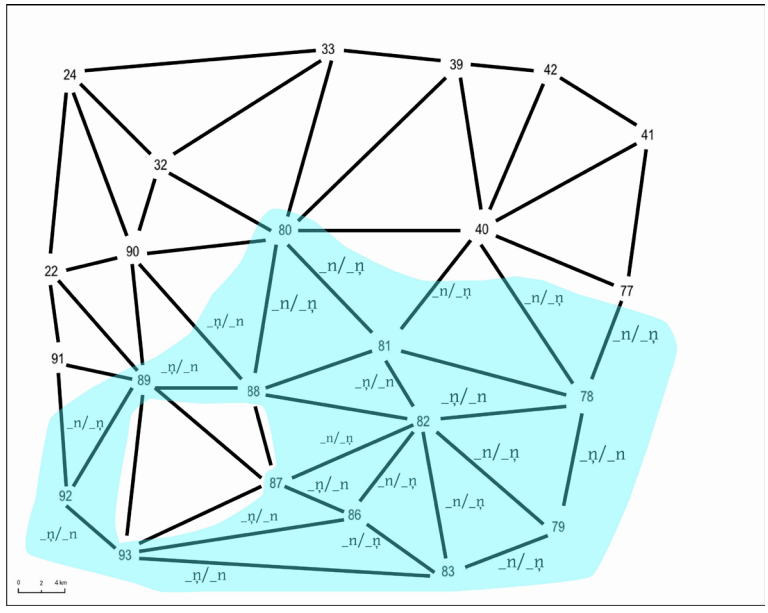

Fig. 15: Syllabification of final [_n]

This specific alternation is correlated to the possible elision of a preceding vowel,

mostly schwa (the alternation [ə]+). The strong intonation on the penultimate syllable in Central Breton (Wmffre 1998) results in the neutralization of the following vowel or even, very frequently, in its elision, in a post-tonic context, and as a consequence, the final [_n̥] is syllabified. German (1984, 2008) also reports this feature for Saint-Yvi. This phenomenon is indeed widespread in the central part of the Breton-speaking area.

In this view, considering the consonantal category of alternations in the light of their respective position is fruitful. The data for the consonantal category display interesting features:

- Palatalisation is frequent and can give rise to local isoglosses. However, the extent of this phenomenon is somewhat limited.
- [h] and [x] interchange very often and in different positions. This trend suggests that they share the same phonological space and each sound seems to operate as a favourite realization of the /h/ phoneme in each investigated location in central Brittany.
- The sonority of the consonants appears to be a factor of variation for a consonant in intervocalic position. This category of sounds in that specific position is affected by a significant amount of variability in the area under study.
- Some facts of a different nature that belong to different categories may be correlated. On the one hand, the possible realization of an initial [h] and, on the other hand, the gemination of [_n_] in internal position can be brought together. Each of these two variants occurs in the neighbouring areas: the first in the centre of the sample area, the second at the north-western edge of it.

6. Discussion

Different points have to be discussed in order to fully appreciate the results obtained. First, to what extent they are meaningful from a dialectological perspective? Secondly, what can be said of the components of the linguistic distance and their distribution across space? Thirdly, how do linguistic distance and geography interact? And finally, the distinction between linguistic distance and linguistic variation has to be drawn more precisely.

6.1. A paradox to resolve or an approach to redefine?

First, my findings contradict the classical model of the dialectal tripartition of Lower Brittany established by Falc'hun (1963, 1981). The central part of the area forms a zone going from the north-east to the south-west (roughly from Lannion in the Côtes d'Armor to Pont-l'Abbé, south-west of Quimper, in the Finistère) with Carhaix as its centre. All across this area, linguistic innovations have spread over the centuries. On each side of this area, there are Léonais Breton to the north-west

and Vannetais Breton to the south-east, which corresponds more or less to the medieval Léon and Vannetais bishoprics and their adjacent areas. All of them have retained archaic linguistic features, unlike central Brittany.

This model is still appropriate, as far as it goes. Other researchers share the same opinion (Favereau 1992; Wmffre 1998; Hewitt 2010: 293). In this view, I had expected the area around Carhaix to display an important level of linguistic similarity. Nevertheless, as figure 8 shows, the locations in this area diverge greatly from each other since they are divided into six different clusters. Location 82, Plounévezel, which is the closest neighbour of Carhaix in the NALBB data (Wmffre 1998), should therefore have displayed a phonetic behaviour more in line with the common features of the central area. However, interestingly, the opposite can be observed. Plounévezel greatly diverges from its neighbours at a phonetic level.

So, why do the findings of this study not match Falc'hun's model? As far as I can get it, I focused only on phonetics, whereas Falc'hun had a broader scope: he incorporated morphology and vocabulary into his analysis. Furthermore, in this analysis, I have taken vowels into account. This could have contributed to providing a different perspective to the results I obtained. Indeed, I have previously observed that the sounds belonging to that category vary in Breton considerably (Solliec 2017). They constitute the major part of the modifications I have gathered (57.35 % of all occurrences). Previously, German (1993) and Costaouec (2012) based their work on weighting the modifications which only affect the consonants. Falc'hun devoted only 4 maps to the vocalic phenomena (1981: 148, 149, 159–60). Therefore, the consideration of the vowel category contributes to an increase in the level of the phonetic distance in the sample surveyed and offers a different picture of the area.

One more thing to be considered is that Falc'hun worked from the data of the ALBB, with only 77 locations investigated (as opposed to 187 for the NALBB). I have therefore increased the probability to record more phenomena of differentiation and to get higher levels of linguistic difference, as a consequence. Moreover, the high rate of linguistic similarity I have set for distinguishing between convergence and divergence (74.84% average similarity rate across the area) could have contributed to this new picture. Henri Guiter, one of the founding fathers of dialectometry, considered that such a level of similarity was to be found between subdialects (based on his analysis of data from the Romance languages (1973: 79)).

However, this new perspective originates more probably in the major difference between Falc'hun's work and the approach of this study. I am primarily investigating linguistic similarity across Breton-speaking Brittany whereas Falc'hun mainly analysed how linguistic innovations had diffused along specific areas. Since I deal with different objectives, it is no surprise that the results are structured differently.

The specific use of an aggregate approach to linguistic similarity in this work also explains why the results do not entirely fit in with Falc'hun's model. Nerbonne (2009: 193–4) speaks about analysing dialectal data according to the methodology based on the aggregation of variables:

> Single-feature studies risk being overwhelmed by noise, i.e., missing data, exceptions, and conflicting tendencies, which are common in this and most areas of linguistics. We aggregate in order to obtain a more reliable signal… [W]e claim that aggregate analyses provide a level at which very general laws concerning linguistic variation might be formulated. This section was quite programmatic, but dialectology is in sore need of more general theoretical work, and aggregating analyses are promising.

In this sense, the results obtained do not contradict Falc'hun's observations but they result from an investigation at a different level of analysis.

Last but not least, I have analysed the data of the NALBB without isolating specific variables with regard to their respective importance for the dialectal structure of the Breton. Therefore, this approach has led to a cumulative analysis of all kinds of dialectal and more local variation. Different kinds of data are subsumed into numbers thanks to an aggregate analysis. However the specific use of the Levenshtein distance I used (see section 3.3) offers the opportunity to scrutinise the nature of the linguistic facts involved in linguistic distance.

6.2. Understanding better how the phonetic components of the linguistic distance are connected.

I noted in section 5 that different phonetic facts involved in shaping the linguistic distance can be correlated. Nevertheless, these links do not appear clearly even if there is a clear phonetic tie between them. For example, the elision of schwa (alternation [ə]+) occurs very often in association with the syllabification of a final [_n] (alternation [_n]/[_n̩]) as in the following example:

(8) NALBB map 149 thunder *kurun*
Saint-Rivoal (91) [ˈgy:rən] ~ Lannédern (89) [ˈgy:ʁn̩]

The same occurs quite frequently also with other words such as *chadenn* 'chain', *kistinenn* 'chesnut', *logodenn* 'mouse'. In all, the following results were gathered (see table 5 below):

Nature of the alternations	[ə]+	[_n]/[_n̥]
Number of occurrences	263	162
Proportion in %	2.18	1.34
Rank	8th	14th

Table 5: Comparison of the frequency of the alternations [ə]+ and [_n]/[_n̥]

Although the two alternations do not behave similarly statistically, it does not mean that they are not linked. There could also be a loose tie between them and one observation would be that the elision or the realization of schwa does not always depend on the syllabification of a final [_n]. However, only a real statistical analysis could support the validity of this view and, in particular, correlate the realization of the second alternation to the first one.

Establishing statistical correlations will be helpful in order to identify a greater system of rules underlying the phonetic phenomena involved in linguistic distance across the area, as Uriel Weinreich (1954: 354) suggested, rather than focusing on a precise repartition. It will exclude, therefore, the simple possibility of coincidence. For example, I have noticed an opposite distribution of two alternations: on the one hand, the possible realization of an initial [h_] (alternation [h_+]) before a vowel and, on the other hand, the gemination of an intervocalic [_n_] (alternation [_n_]/[_nn_]), as can be seen in figures 11 and 13. For now, I have not identified a link between them apart from a correlation. Perhaps the explanation lies in a constraint on the structure of the syllable since the insertion of an epenthetic [h] provides an initial consonant at the beginning of the syllable. In the area where the tonic accent is very strong, it could be a way to clearly demarcate the beginning of a syllable. On the other hand, the possible gemination of intervocalic consonants could help to differentiate distinctly between the different syllables of a word when the tonic accent is not as strong as it is in the centre of Lower-Brittany.

Another fruitful approach to explore would be to change the level of analysis of the phonetic facts and focus on more specific sound features. This would then offer the possibility of gathering different alternations under a more general label such as [+/- palatal] or [+/- open].[8] Such results would allow us to gain a better understanding of the dynamics of the language and how the phonemic system varies geographically. In this data, the first distinction that has been studied was the alternation

8. The notation in the NALBB distinguishes four distinct degrees of aperture for the mid vowels (Le Dû 2001), cf. section 2.1.

between [e]/[ę] that cuts the area into two parts, as figure 9 shows. The aggregation of the other vocalic alternations, which display the same feature could confirm this distribution or present another pattern.

Such an investigation leads one to scrutinize the dynamics of the language (Martinet 2005; Wmffre 2013) perceived through variation phenomena. This kind of approach also allows us to observe how the results of diachronic evolution can interact with the synchronic facts of local variation. For instance, palatalisation, which can affect initial [g_], as well as other consonants (section 5.3.1, figure 12), may be mentioned here. The specific position at the beginning of a word before a vowel is interesting because a full range of phenomena occurs in this position such as the possible realization of an initial [h_] before a vowel (cf. examples (2) and (3) and figure 11). In the case of the initial [g], the diachronic background also interacts with synchronic phenomena since, in a few words, this specific sound was added epenthetically in the Old Breton period (Jackson 1967: 427–440).

(9) NALBB map 119 Friday *Gwener* from Latin *veneris*
Landeleau (87) [ˈgwẽ̯ːnəʁ] ~ Cléden-Poher (86) [ˈʝɥœ̃nᵊʁ]

This indicates here a convergence between a diachronic and a synchronic analysis of this specific pattern of variation. Explaining it convincingly will help us gain a better understanding of the phonetic parameters involved in the variability of Breton. The realization of extra consonants in the initial position of words in the sample area seems to be an answer to a constraint of building words upon a CV_ syllabic structure.

These few examples illustrate more specifically that the content of linguistic distance is not an amount of uncorrelated facts. However, it seems difficult to assess more than they are the result of a simple coincidence. A thorough statistical analysis may be useful to eliminate the factor of coincidence and to show the correlation between those facts. Nevertheless, focusing on the content of linguistic distance is also a way to observe which phonetic facts are involved in the general dynamics of Breton.

6.3. The relationship between geography and linguistic distance

The moving relationships between the different facts aggregated in linguistic distance leads us to another question, which is the influence of geography on linguistic distance. Geography in geolinguistics is very often reduced to a factor of Euclidean distance to the detriment of other dimensions, as summarized by Nerbonne and Kleiweg (2007) as the 'Fundamental Dialectology Principle'. This states that "geographically proximate varieties tend to be more similar than distant ones" (Kleiweg 2007: 154). Focusing on the lexicon, Séguy (1971) showed that

the relationship of linguistic distance to the geographical distance is logarithmic. However, Szmrecsanyi (2012) and Nerbonne (2013) concluded recently that Euclidean distance only partly explains linguistic distance.

I explained in section 3 that areas of linguistic divergence frequently correlate with geographic physical structures (such as valleys or basins occupied by swamps), and as Nerbonne puts it: "geography is indeed structured more complexly than simple distances, and also […] geography influences linguistic variation deeply" (Nerbonne 2013: 222). It would be interesting to draw a clearer picture of the distribution of all the patterns I have observed previously in order to gain a better understanding of it and, in this way, determine which geographic factors they are based on. In his work, Falc'hun (1963, 1981) highlighted the importance of the historical road networks in the propagation of linguistic innovations and also the role played by small cities as centres for their propagation: "Immediately, I felt as if the words were spreading along the roads, especially from Carhaix towards Tréguier" (1981: 18).[9]

Trudgill (1974), who was working on the phonetic change in the realization of the vowel /æ/ in southern Norway, suggested that a trend to a more open realization of this vowel was linked to neighbouring cities, whose influence had been increasing due to the development of cars and a good road network. He concluded then by the following: "We see illustrated very clearly, for example, the jumping of the innovation from one central place to another, and the subsequent operation of the neighbourhood effect …" (*ibid*., p. 232). This work shows the correlation between the diffusion of linguistic innovations, the development of roads and the opportunity for people to circulate.

Falc'hun (just like Trudgill) focuses on the diffusion of linguistic innovations but linguistic distance cannot be reduced to the spread of innovations alone. It does not entirely explain how geography and diatopic variation interact across a given area. Moreover, the linguistic distance cannot be reduced to only a few variables, whose intensity varies across space. The geographical dimension in dialectometry is indeed hard to consider in all of its aspects since it can also interact with most human activities. An aggregate analysis of the repartition of the dialectal data by German (1984, 1987, 1993) and Costaouec (2012) led them to identify correlations at different levels between linguistic variation and sociocultural facts such as the distribution of traditional dress, specific local dances and social networks for collaboration during seasonal work or the distribution of archaeological remains (German 1993: 185–186; Costaouec 2012: 7).

9. Original French: "Aussitôt j'eus comme l'impression de voir les mots courir par les routes, spécialement de Carhaix vers le Tréguier".

6.4. Disentangling linguistic distance from dialectal variation

Such investigation of linguistic distance gave an opportunity to reconsider a part of the Breton language area from this point of view. I have moved away from the accepted model of the dialectal division in Breton. Moreover, the dialectometrical approach has helped to identify phonetic variables across the area without eclipsing the whole context. It, therefore, offers a good opportunity to measure the relative importance of dialectal variables. For example, the reflexes associated with the Old Breton final interdental fricative [_ð] are good indicators to define whether data belong to the NW dialectal area or not (Falc'hun 1981: 198–9; Jackson 1967: 652–4). In the sample, they contribute 0.29% of the modifications with only 35 occurrences (this corresponds to the alternations [_s+] or [_z+]).[10] This is not amongst the most important in numerical terms but the alternation [_s+] is at the 79th position in a series of 761 when ranked decreasingly according to the their respective number of occurrences.

Since a language is not only an addition of phonetic innovations, it is interesting to be able to determine what a dialect area is made up of and how it fits in with the traditional dialectological variables. For instance, I reported the importance of the centralization of the vowels linked to a post-tonic context (see section 5.1, figure 9). Apart from the work by Wmffre (1998), this feature has rarely been reported by dialectologists (German 1984; Plourin 2005: 23, 28–9). Traditional dialectal variables tend to play little role in shaping the linguistic distance. On the contrary, local phenomena as observed in section 5 seem to have a bigger incidence.

These remarks lead to a more general interrogation. Since the variability of the language is not limited to dialectal variables, what is the status and the nature of linguistic distance and how it interacts with other variation phenomena? Linguistic distance offers an opportunity to study data in the aggregate. This is why the notion of linguistic distance is to be distinguished clearly from dialectal variation since they cannot be used indistinctively. However, the specific use made of the Levenshtein distance in this study has made one observe different phenomena which contribute to shaping the dynamics of language. On the other hand, dialectal variables help to elaborate internal classification of languages. This is why so much attention has been paid to the selection of variables in traditional dialectological research (Stankiewicz 1957: 46 for the idea and Möhlig 1974 for an illustration). Each approach offers a different view on the linguistic facts. Linguistic distance does not always match the dialectal variation and vice versa. The dialogue of both perspectives helps then to gain a more precise picture of the internal structure of

10. [_z]+ reflects the cases when the last consonant was not devoiced before a pause. Derived words for which this final [z] would be in intervocalic position have not been taken into account, for example *gwez 'trees' but gwezenn 'a tree'*.

the language under study. This is why it would be fruitful to confront the dialect areas in to the convergence areas (i.e. an area where the rate of linguistic similarity is high) and to look for possible correspondences.

7. Conclusion

I started my investigation by asking myself how the linguistic distance, more precisely, the phonetic distance, was distributed across the sample area and what it was made of. In line with Nerbonne (2009), I proceeded with an aggregate analysis of a sample of phonetic data from a linguistic atlas of Breton, the NALBB. Subsuming different variables into an overall rate of linguistic similarity has allowed to gain a close view of the distribution of linguistic distance across the studied area. Furthermore, the specific use of the Levenshtein algorithm to evaluate linguistic distance has provided the opportunity to examine its components, to weight the features under investigation and, therefore, to evaluate their impact on linguistic distance. Consequently, this study has developed along the two levels of analysis:

- the aggregate level
- the content of linguistic distance

The results displayed at the aggregate level match geographic features, especially the local relief (section 4). This situation suggests that Euclidean distance cannot explain the distribution of linguistic distance in its entirety. It is therefore necessary to reassess the impact of geography in all its aspects on the shaping of linguistic distance. This view converges with those of Nerbonne (2013) and Szmrecsanyi (2012) that geography considered as Euclidean distance is not enough to explain linguistic variation, even if it clearly contributes to it.

In looking for an explanation for the apparent paradox between these results and the classic model of Breton language dialectal structure established by Falc'hun (1963, 1981), the notion of linguistic distance has been refined for Breton. Linguistic similarity consists then not only of shared innovations like dialectal variables but also of a full range of other phenomena. This explains why the distribution of the linguistic similarity observed sometimes contrasts Falc'hun's conclusions. Thus, Plounévezel, the location 82 of the NALBB differs phonetically from its neighbours although it is situated in the middle of a linguistically convergent area. Linguistic distance offers then an alternative view for considering the distribution of linguistic data in space. It complements the classical dialectological perspective on language differentiation without being intended to replace it. In this view, investigating linguistic distance is a heuristic tool to observe diatopic structures invisible at the first glance. In this work, one can notice that the observations for each level

do not really converge. This is why it would be interesting to investigate further on that specific matter.

The second point I have focused on is to determine as precisely as possible the content of the linguistic distance (section 5). I have noticed some intriguing patterns, having demonstrated that linguistic distance covers a range of different phenomena whose distribution and intensity across the area vary a lot. The Levenshtein distance has supplied me with percentages of linguistic distance alongside with the nature of its constituents. Some of the phonetic phenomena that create the linguistic distance have been identified. However, they are distributed unevenly across the investigated area. Therefore, this central zone of Breton-speaking Brittany provides a linguistic illustration to the phrase that I have included in the title of my contribution — 'unity in diversity'.

The content of linguistic distance varies and an aggregate view on it does not display all its variety. As in all quantitative approaches, the facts are subsumed into numbers and lose their individuality, which is the strength as well as the weakness of this approach. This method, based on the result and on the exploitation of the operations processed by the Levenshtein distance, offers the possibility to deal with these two aspects of analysis. It has been noticed that the dialectal variables contribute clearly to linguistic distance but they constitute only a minor part. Other linguistic facts whose distribution is more restrained shape such linguistic distance even more importantly. More generally, the phonetic facts identified reflect some trends in the dynamics of the language.

This research is still in progress and, for that reason, the results will be refined and completed in due course. Extending the scope of my analysis to the Breton-speaking area of Brittany in its entirety in the near future, I will see to which extent the 'unity in diversity' pattern can be extended to the territory occupied by Breton. Moreover, this will also present an opportunity to deal with the following points in depth:

- Can the distribution of linguistic distance be correlated or not with the socio-historical structures?
- What happens when dialectal and non-dialectal variables are weighted separately?

At that point, a possibility will arise to examine how each one of them contributes to the overall linguistic distance and their respective frequency (Chambers 1982).

However, I have determined potential trends to analyse, and I have also gained a glimpse of the parameters involved in phonetic distance in Breton. I expect my results to display how geography contributes to shaping the linguistic distance alongside with language dynamics. The specific use of the Levenshtein distance has

given the possibility to explore the distribution of linguistic distance as well as its content. It offers a way to associate a few descriptive features with a quantitative approach. One step further will be to examine how the material supplied by this approach can be used to have a view on the phonological structure of the language.

The dialectometrical approach I have used, due to the association of two different perspectives on linguistic data, opens a new prospect for the analysis of the diatopic variation in the centre of Breton-speaking Brittany. Preliminary observations underline the heuristic value of linguistic distance for appreciating linguistic differentiation in space. Extending this approach to the whole Breton-speaking area of Brittany could open new ways of understanding it is structured.

Université de Bretagne Occidentale, Brest

References

Broudic, F., 2004, 'La perception de la variation diatopique à travers le temps', *La Bretagne linguistique* 13, 31–52.

Brun-Trigaud, G., 2014, 'Un usage particulier de l'algorithme de Damerau-Levenshtein dans le domaine occitan', in: Diémoz, F., & Aquino-Weber, D., 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, Genève, 127–47.

Brun-Trigaud, G., **Solliec** T., & **Le Dû** J., 2016, 'A New Dialectometric Approach Applied to the Breton Language', in: Côté, M.-H., Knooihuizen, R., & Nerbonne, J., (eds.), *The Future of Dialects*, Berlin, 135–54.

Campbell, L., 1999, *Historical Linguistics. An Introduction*. Cambridge, MA: The MIT Press.

Chambers, J. K. 1982. *Geolinguistics of a Variable Rule*. Discussion Papers in Geolinguistics, 5. Stafford, North Staffordshire.

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

Costaouec, D., 2012, 'Linguistic Geography of Breton and Sociocultural Motivations', *STUF — Sprachtypologie und Universalienforschung* 65(1), 47–64.

Donohue, M., **Wichmann**, S., & **Albu** M., 2011, 'Typology, Areality, and Diffusion', *Oceanic Linguistics* 47(1), 223–32.

Edmont, E., & **Gilliéron** G., 1902–1910, *Atlas Linguistique de la France*, Paris, Champion, 8 vols., 639 sites of investigation, 1920 maps.

Falc'hun, F., 1963, *Histoire de la langue bretonne d'après la géographie linguistique*, Paris.

Falc'hun, F., 1981, *Perspectives nouvelles sur l'histoire de la langue bretonne*, Paris (3rd edition).

Favereau, F., 1992, *Geriadur ar brezhoneg a-vremañ* [Dictionary of Contemporary Breton], Morlaix.

Favereau, F., 1997, *Grammaire du breton contemporain*, Morlaix.
German, G., 1984, *Une étude linguistique sur le breton de Saint-Yvi*, PhD thesis, Brest.
German, G., 1987, 'Une description du déterminant défini *an* à St-Yvi en Cornouaille', *La Bretagne Linguistique* 3, 157–68.
German, G., 1993, 'Une méthode dialectométrique (assistée par ordinateur) pour l'analyse des atlas linguistiques', *La Bretagne linguistique* 7, 177–213.
German, G., 2008, Language Shift, Diglossia and Dialectal Variation in Western Brittany: The case of Southern Cornouaille', in: Tristam, H. L. C., ed., *The Celtic Languages in Contact. Papers from the Workshop within the Framework of the XIII International Congress of Celtic Studies*, Bonn, 26–27 July 2007, Postdam, 146–92.
Goebl, H. 1982. *Dialektometrie. Prinzipien und Methoden des Einsatzes der Numerischen Taxonomie im Bereich der Dialektgeographie*, Denkschriften der Österreichischen Akademie der Wissenschaften, Vienna.
Goebl, H., 2005, 'La dialectométrie corrélative. Un nouvel outil pour l'étude de l'aménagement dialectal de l'espace par l'homme', *Revue de linguistique romane* 69, 321–67.
Guiter, H., 1973, 'Atlas et frontières linguistiques', in: Straka, G., & Garde, P., eds., *Les dialectes romans de la France à la lumière des atlas régionaux*, Paris, 61–109.
Hammarström H., 2016, 'Linguistic Diversity and Language Evolution', *Journal of Language Evolution* 1(1), 19–29.
Heeringa, W., 2004, *Measuring Dialect Pronunciation Differences Using Levenshtein Distance*, PhD thesis, Groningen.
Heeringa, W., & **Nerbonne**, J., 2010, 'Dialect Areas and Dialect Continua', *Language Variation and Change* 13(3), 375–400.
Heggarty, P., 2000, 'Quantifying Change over Time in Phonetics', in: Renfrew, C., McMahon, A., & Trask, L., eds., *Time Depth in Historical Linguistics*, Cambridge, 531–62.
Hewitt, S., 2010, 'Mood in Breton', in: Rothstein, B., & Thieroff, R., eds., *Mood in the Languages of Europe*, Amsterdam, Philadelphia, 292–308.
Jackson, K. H., 1967, *A Historical Phonology of Breton*. Dublin.
Kessler, B., 1995, 'Computational Dialectology in Irish Gaelic', in: *Proceedings of the Seventh Conference of the European Chapter of the Association for Computational Linguistics (EACL, Dublin)*, San-Francisco, 60–67.
Lagofeld, P., & **Maddieson** I., 1996, *The Sounds of the World's Languages*, Oxford.
Levenshtein, V. I., 1966, 'Binary Codes Capable of Correcting Deletions, Insertions and Reversals', *Soviet Physics Doklady* 10, 707.
Le Dû, J., 1972, 'Le Nouvel Atlas Linguistique de Basse-Bretagne', *Etudes Celtiques* XIII(1), 332–45
Le Dû, J., 2001, *Nouvel atlas linguistique de la Basse-Bretagne*, Brest, 2 vols., 187 sites of investigations, 600 maps.
Le Dû, J., & **Brun-Trigaud**, G., 2003, 'La bipartition de la langue bretonne : quelques

question sur la zone vannetaise', in: Caprini, R., ed., *Parole Romanze. Scritti per Michel Contini*, Alessandria, 245–60.
Le Pipec, E., 2015, 'La palatalisation vannetaise', *La Bretagne Linguistique* 19, 297–327.
Le Roux, P., 1924–1963, *Atlas linguistique de la Basse-Bretagne*, Rennes, Paris, 6 vols., 77 sites of investigation. 600 maps.
Lindau, M., 1985, 'The Story of /r/', in: Fromkin, V.A., ed., *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*, Orlando, FA, 157–68.
Martinet, A, 2005, Économie *des changements phonétiques. Traité de phonologie diachronique*. Paris (2nd edition).
Möhlig, W. J. G., 1974, *Die Stellung des Bergdialekte im Osten des Mt. Kenya. Ein Betrag zur Sprachgliedering im Bantu*, Berlin.
Moulton, W. G., 1962, 'Dialect Geography and the Concept of Phonological Space', *WORD* 18 (1–3), 23–32.
Nerbonne, J., 2009, 'Data Driven Dialectology', *Language and Linguistics Compass* 3(1), 175–98.
Nerbonne, J., 2013, 'How Much Does Geography Influence Language Variation?', in: Auer, P., Hilpert, M., Stukenbrock A., & Szmrecsanyi, B., eds., *Space in Language and Linguistics: Geographical, Interactional, and Cognitive Perspectives*, Berlin, New York, 220–36.
Nerbonne, J., **Heeringa**, W. & **Kleiweg**, P., 1999, 'Edit distance and dialect proximity', in: Sankoff, D., & Kruskal, J. (eds.), *Time Warps, String Edits, and Macromolecules. The Theory and Practice of Sequence Comparison*, Stanford (2nd edition), v–xv.
Nerbonne, J. & **Heeringa** W., 2010, 'Measuring dialect differences', in: Schmidt J. E. & Auer P., eds., *Language and space: theories and methods*, Berlin, New-York, 550–567.
Nerbonne, J. & **Kleiweg**, P., 2007, 'Toward a Dialectological Yardstick', *Journal of Quantitative Linguistics* 14(2), 148–66.
Nerbonne, J., **Colen**, R., **Gooskens**, C., **Kleiweg**, P., & **Leinonen** T, 2011, 'Gabmap — a Web Application for Dialectology', *Dialectologia Special Issue* II, 65–89.
Nettle, D., 1998, 'Explaining Global Patterns of Language Diversity', *Journal of Anthropological Archeology* 17, 354–74.
Plonéis, J.-M., 1983, *Au carrefour des dialectes Bretons — Le parler de Berrien. Essai de description phonématique et morphologique*, Paris.
Plourin, J.-Y., 2005, *Tammoù Gwaskin. Au coeur du breton légitime*, Brest
Scobbie, J. M., 2006, '(R) as a Variable', in: Brown, K., ed., *Encyclopedia of Language & Linguistics*, Oxford (2nd ed.), 337–44.
Séguy, J., 1971, 'La relation entre la distance spatiale et la distance lexicale', *Revue de Linguistique Romane* 35, 335–357.
Séguy, J., 1973, 'La dialectométrie dans l'Atlas linguistique de la Gascogne', *Revue de Linguistique Romane* 37, 1–24.

Séguy, J. et al., 1954–1973, *Atlas linguistique et ethnographique de la Gascogne*, Toulouse, Paris, 6 vols.
Snoek, C., 2014, 'Review Of Gabmap. Doing Dialect Analysis on the Web', *Language Documentation & Conservation* 8, 192–208.
Solliec, T., 2017, 'Explorer la variation en breton grâce à la dialectométrie, la Basse-Bretagne considérée par la distance linguistique', *La Bretagne Linguistique* 21, 29–61.
Stankiewicz, E., 1957, 'On Discretenes and Continuity in Structural Dialectology', *WORD* 13(1), 44–59.
Szmrecsanyi, B., 2012, 'Geography is Overrated', in: Hansen, S., Schwarz, C., Stoeckle, P., & Streck, T., eds, *Dialectological and Folk Dialectological Concepts of Space,* Berlin, 215–31.
Trudgill, P., 1974, 'Linguistic Change and Diffusion: Description and Explanation in Sociolinguistic Dialect Geography', *Language in Society* (3), 215–46.
Wälchli, B., 2010, 'The Consonant Template in Synchrony and Diachrony', *Baltic Linguistics* 1, 215–244.
Weinreich, U., 1954, 'Is a Structural Dialectology Possible?', *WORD* 10(2–3), 388–400.
Wiese, R., 2011, 'The Representation of Rhotics', in: van Oostendorp, M., Ewen C., Hume E., & K. Rice, eds., *The Blackwell Companion to Phonology*, Malden, MA, Oxford, 711–729.
Wieling, M., & **Nerbonne**, J., 2015, 'Advances in Dialectometry', *Annual Review of Linguistics* 1, 243–64.
Wmffre, I., 1998, *Central Breton*, Munich.
Wmffre, I., 2013, *Dynamic Linguistics, Labov, Martinet, Jakobson and Other Precursors of the Dynamic Approach to Language Description*, Oxford, Bern, Bruxelles, Frankfurt am Main, New York, Wien: Peter Lang.

Appendix 1. List of locations from the NALBB for this study (with their corresponding number code)

22 Saint-Cadou (Sizun)	42 Loguivy-Plougras	86 Cléden-Poher
24 Guimiliau	77 Saint-Servais	87 Landeleau
32 Plounéour-Ménez	78 Locarn	88 Collorec
33 Plougonven	79 Paule	89 Lannédern
39 Guerlesquin	80 Berrien	90 Botmeur
40 Plourac'h	81 Poullaouen	91 Saint-Rivoal
41 Plougonver	82 Plounévézel	92 Pleyben
	83 Motreff	93 Lennon

Appendix 2. List of the maps of the NALBB which provided the data

7 paroisse (*parish*)
9 le bourg (*the village center*)
21 bas (adj.) (*low*)
25 celui-ci (*this one; masc.*)
27 celui-là (*that one; masc.*)
43 petit (*small*)
44 long (*long*)
46 lourd (*heavy*)
47 léger (*light*)
50 compter (*to count*)
51 un (*one*)
54 trois masculin (*three; masc.*)
56 quatre masculin (*four; masc.*)
57 quatre féminin (*four; fem.*)
58 cinq (*five*)
59 six (*six*)
60 sept (*seven*)
61 huit (*eight*)
62 neuf (le chiffre) (*nine*)
73 vingt (*twenty*)
77 maintenant (*now*)
80 souvent (*often*)
89 tard (*late*)
95 hier (*yesterday*)
99 midi (*noon*)
104 nuit (*night*)
106 jour (*day*)
108 tous les jours (*everyday*)
109 journée *((a) day long*)
110 des journées (*days*)
111 le lundi (*the Monday*)
112 lundi (adv.) (*Monday*)
114 mardi (adv.) (*Tuesday*)
115 mercredi (nom.) (*Wednesday*)
117 jeudi (nom.) (*Thursday*)
119 vendredi (nom.) (*Friday*)
121 le samedi (*Saturday*)
123 le dimanche (*Sunday*)
126 une semaine (*a week*)
130 un mois (*a month*)
131 des mois (*months*)
138 au revoir (*goodbye*)
139 temps (*weather and time*)
141 du vent (*wind*)
144 de la pluie (*rain*)
146 de la neige (*snow*)
149 tonnerre (*thunder*)
158 sécheresse (*drought*)
161 les cieux (*heavens*)
162 paradis (*paradise*)
163 le soleil (*the sun*)
165 de la terre (soil)
166 de l'eau (*water*)
168 de la cendre (*ashes*)
170 du fer (*iron*)
171 un arbre (*a tree*)
176 de l'herbe (*grass*)
184 pomme (*apple*)
185 des pommes (*apples*)
186 (une) châtaigne (*chestnut*)
187 des châtaignes (*chestnuts*)
196 des mûres (*blackberries*)
199 des loups (*wolves*)
200 renard (*fox*)
202 voleur (*thief*)
203 des voleurs (*thieves*)
209 des rats (*rats*)
210 souris (animal) (*mouse*)
211 des souris (*mice*)
223 un corbeau (*a crow*)
224 des corbeaux (*crows*)
226 des tourterelles (*turtledoves*)
230 des crapauds (*toads*)
231 grenouille (*frog*)
241 (une) mouche

((*a) fly*)
242 des mouches (*flies*)
243 (une) guêpe ((*a) wasp*)
245 (une) abeille ((*a) bee*)
248 un chien (*a dog*)
249 des chiens (*dogs*)
252 un chat (*a cat*)
253 des chats (*cats*)
254 matou (*tomcat*)
256 (une) chatte ((*a) female cat*)
261 puits (singulier) (*well sg.*)
262 des puits (*wells*)
263 l'étable (*the cowshed*)
266 du fumier (*manure*)
274 fermier (*farmer*)
276 domestique (de ferme) ((*farm*) *servant*)
277 des domestiques (de ferme) ((*farm*) *servants*)
281 journalier (tâcheron) (*day labourer*)
288 sarcler (*to weed*)
290 charger (*to load*)
293 traire (*to milk*)
294 trait (elle a) (*milked (she has*))
295 une pelle (*a shovel*)
300 (une) faucille ((*a*) *sickle*)
301 (une) faux ((*a*) *scythe*)
306 (une) chaîne ((*a*) *chain*)
307 des chaînes (*chains*)
308 (une) paire de tenailles ((*a) pair of pincers*)
312 une charrette (*a cart*)
317 sarcleuse (*weeder*)
322 du grain (*grain*)
323 de la semence (*seed*)
325 du trèfle (*clover*)
327 des chevaux (espèce) (*horses*)
335 poulain (*foal*)
337 pouliche (*filly*)
338 des pouliches (*fillies*)
341 âne (*donkey*)
344 des cochons (*pigs*)
345 du bétail (*cattle*)
349 des vaches (pl. régulier) (*cow*; *regular pl.*)
352 une vache (*a cow*)
354 les vaches (*the cows*)
355 veau (*calf*)
357 génisse (*heifer*)
359 du cuir (*leather*)
360 paître (*to graze*)
361 bélier (*ram*)
364 des moutons (*sheeps*)
371 bouc (*billy goat*)
375 une poule (*a hen*)
376 poules, volaille (*hens*, *poultry*)
386 pondre (*to lay eggs*)
388 couver (*to brood*)
390 des cheveux (*hair*)
392 barbe (*beard*)
393 œil (*eye*)
394 des yeux (paire) (*eyes, a pair of*)
396 oreille (*ear*)
404 main (*hand*)
405 des mains (paire) (*hands, a pair of*)
406 pied (*foot*)
407 des pieds (*feet*)
408 doigt (*finger*)
409 des doigts (*fingers*)
410 ongle (*fingernail*)
411 des ongles (*fingernails*)
418 vivant (*alive*)
428 transpirer (*to sweat*)
429 transpiré (*sweated*)
434 souffler (*to blow*)
444 (l') odeur ((*the) odour*)
452 penser (*to think*)
456 vous (*you -plural-*)
464 des vêtements (*clothes*)

471 nœud (*knot*)
472 des nœuds (*knots*)
475 de la laine (*wool*)
477 boire (*to drink*)
484 dîner (repas du soir) (*dinner, evening meal*)
492 des verres (*glasses*)
493 plein (*full*)
504 du vin (*wine*)
507 du pain (*bread*)
511 du beurre (*butter*)
513 des œufs (*eggs*)
519 pêcher (*to fish*)
521 pêcheur (*fisherman*)
536 (la) cheminée ((*the*) *fireplace*)
540 descendre (*to go down*)
548 (la) fenêtre ((*a*) *window*)