Rif Berber: From Senhaja to Iznasen
A qualitative and quantitative approach to classification

Mena B. Lafkioui

Abstract

By combining qualitative (synchronic and diachronic) and quantitative (algorithmic) approaches, this study examines the nature, structure, and dynamics of the linguistic variation attested in Berber of the Rif area (North, Northwest, and Northeast Morocco). Based on a cross-level corpus of data obtained from the Atlas linguistique des varieties berbères du Rif (Lafkioui 2007) and from numerous linguistic, sociolinguistic, and ethnographic fieldwork investigations in the area since 1992, this study shows that these Berber varieties form a language continuum with the following five stable core aggregates, which cut across administrative and political borders: Western Rif Berber, West-Central Rif Berber, Central Rif Berber, East-Central Rif Berber, and Eastern Rif Berber. Furthermore, data mining studies made it possible to objectively identify the principal aggregate discriminators of the Rif Berber continuum, which are dealt with in the study. A special focus in the article is put on the interplay between system-internal and system-external parameters for the selection, diffusion, and transformation of variants in Rif Berber.

Keywords

Berber, cross-level classification, composite diffusion, language continuum

1 Introduction

The present data-driven study demonstrates from both a qualitative and quantitative perspective that the Berber varieties of the Rif area (North, Northwest, and Northeast Morocco, Figures 1 and 2) – including the varieties of the Senhaja (westernmost group) and of the Iznasen (easternmost group) – form a language continuum with a number of stable core aggregates, obtained through algorithmic classifications and verified by means of structural (synchronic and diachronic) classifications. The evidence supporting these claims is consistent with the data and qualitative analysis and classifications provided in the Atlas linguistique des varieties berbères du Rif (Lafkioui 2007; freely downloadable from https://atlasrif.wordpress.com/), the ALR henceforth, as well as with the quantitative classifications presented in Lafkioui (2008a; 2018b; 2020). Compared to these latter classifications, two new major outcomes are presented in this study. The first one is the emergence of a new core aggregate, that is East-Central Rif Berber, thanks to the enhancement of the corpus by cross-level data, which allowed to accomplish comprehensive algorithmic classifications, imperative to improve and deepen the related qualitative explanations. The notion of "cross-level" refers here to the involvement of different linguistic levels, which are the phonetic, phonological, morphological, syntactic, and lexical levels. This new outcome brings

the number of stable core aggregates of the Rif Berber continuum to five, which correspond to the following geolinguistic subdivisions: Western Rif Berber (WRB), West-Central Rif Berber (WCRB), Central Rif Berber (CRB), East-Central Rif Berber (ECRB), and Eastern Rif Berber (ERB) (Figure 3). The second major outcome is that data mining studies on the cross-level corpus made it possible to objectively identify the principal aggregate discriminators of the Rif Berber continuum, which will be examined here from a qualitative perspective as well. It is the validation of these discriminators from both a quantitative and qualitative perspective that determined the phenomena selected for further examination in this study.

The five core aggregates of the Rif Berber continuum cut across the traditionally – and often erroneously – used groupings of Senhaja, Rif, Iznasen, and many other smaller groupings, such as Iqelɛiyen, Ibḍalsen, and Igzennayen, which are in fact ethnonyms and hold no classification value of any kind, neither do they correspond to the sociolinguistic landscape of the Rif area, which shows considerable complexity. Even more, language groupings such as those presented in Biarnay (1917) for the Rif area are questionable because of their impressionistic and biased viewpoints, which are directly related to the colonial backdrop in which the studies were accomplished.

This study builds further on the quantitative methods and results obtained from the algorithmic classifications of Rif Berber's lexis discussed in Lafkioui (2008a, 2018b, 2020), which give evidence for the validity of the Levenshtein distance calculating method, also called edit distance, especially when the phone strings are tokenised in pair-wise alignments. Furthermore, among the many techniques to analyse and visualise aggregate distances, Multi-Dimensional Scaling – MDS henceforth – was proven to be the best suited for studying language continua, which is the case of Rif Berber. The MDS technique has also the advantage to visualise the aggregates, as well as the degree of their intra- and inter-linguistic divergence. Moreover, it is one of the most stable techniques, compared to classical clustering, for instance (Nerbonne et al. 2011). I will continue using these techniques here, which draw on Kleiweg's free software tools (See http://www.let.rug.nl/ kleiweg/L04/), as well as on the more recent web application GABMAP (Nerbonne et al. 2011). In addition, the study is also based on numerous data conversion programmes developed for this purpose, and for which I am grateful to Bart Cocquyt for his assistance, as well as for his input in applying the k-means clustering algorithm (Section 2).

Before getting into the details, an introduction to Rif Berber is in order. Rif Berber (aka Tarifit, *Tmaziɣt n Rrif*, or the Rif Amazigh language) belongs to the Northern Berber language type and thus is part of the large Berber language family, which forms a branch of the Afro-Asiatic language phylum (Lafkioui 2017). The area of the Rif stretches from the Strait of Gibraltar in the West of Morocco to the Algerian frontier in the East, and from the Mediterranean Sea in the North

to the corridor of Taza in the South, where Moroccan Arabic is mostly spoken (Figures 1 and 2). There are two regions in the Rif area that are mainly Berber-speaking: the small isolated region of Ghomara (Camps and Vignet-Zunz 1998; Colin 1929; El Hannouche 2010) and the extensive territory where Rif Berber is spoken and which forms a geolinguistic continuum, which is delimited (Figure 3 and Table1):

- In the West, by the varieties of the Ktama group (nr. 1), which belong to WRB and hence also to the so-called Senhaja Berber group. Senhaja Berber includes all varieties of WRB and of westernmost WCRB (nrs. 1 to 13). The term Senhaja Berber is used here when the relating 13 varieties are specifically concerned, otherwise I refer to the aggregates WCRB and WRB, which are more accurate denominations, geolinguistically speaking.
- In the South, by the koinè of Gersif, which is the ultimate geographic point where Rif Berber is spoken before reaching the corridor of Taza (nr. 31).
- In the East, by the varieties of Iznasen, which have spread to the regions of Arabic- speaking varieties towards the Moroccan-Algerian border (nr. 26).

The Ghomara Berber varieties, on the other hand, are not part of this continuum but are separated from it by the Arabic varieties of the Jbala, whose great impact on Ghomara Berber has significantly contributed to their linguistic distinctiveness (El Hannouche 2010, Mourigh 2016; also verified by my own fieldwork in the area; see arrow in Figure 2). Its substantial contact-induced linguistic singularity and its isolated location imply that Ghomara Berber forms a kind of distinct geolect within the larger Moroccan Berber continuum. The latter is part of the super-continuum covering entire North Africa, including the Sahara and the North and Northwest Sahel. Indeed, the whole Berber linguistic branch is one vast continuum containing various subcontinua, which progressively blend into each other regardless of administrative and political borders. Smaller and isolated geolects are scattered here and there over this super-continuum (Lafkioui 2018d).

Fig. 1. The Rif area
(© OpenStreetMap contributors)

Fig. 2. The Rif Berber continuum
(© OpenStreetMap contributors)

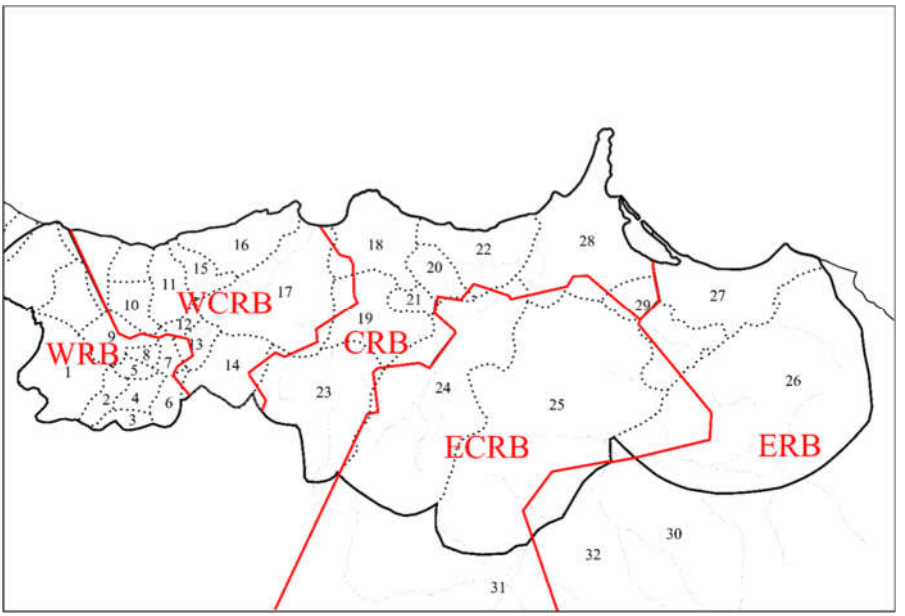

Fig. 3. Aggregates of the Rif Berber continuum and their respective Berber-speaking groups

WRB	WCRB	CRB	ECRB	ERB
1 Ktama	10 Ayt Gmil	18 Ayt Temsaman	24 Ibḍalsen	26 Iznasen
2 Taɣzut	11 Ayt Bufraḥ	19 Ayt Tuzin	25 Ayt Buyeḥya	27 Ikebdanen
3 Ayt Bušibet	12 Targist	20 Ayt Wlišek	31 Gersif	29 Wlad Settut
4 Ayt Ḥmed	13 Ayt Mezduy	21 Tafersit		30 Ayt Buzeggu
5 Ayt Bunsar	14 Ayt Ɛammart	22 Ayt Sɛid		32 Tawrirt
6 Ayt Bšir	15 Ayt Iṭṭeft	23 Igzennayen		
7 Zerqet	16 Ibeqquyen	28 Iqelɛiyen		
8 Ayt Ḥennus	17 Ayt Weryaɣel			
9 Ayt Seddat				

Table1: Aggregates of the Rif Berber continuum and their respective Berber-speaking groups

In what follows, Section 2 will present the map, data, and aggregate discriminators on which this study is based. Section 3 will investigate the cross-level classifications of Rif Berber. Section 4, on the other hand, will examine the aggregate discriminators selected on a phonetic and phonological level (vocalisation and spirantisation), whereas Section 5 will deal with the morphological and syntactic level (pronoun) and Section 6 with the lexical level (time). Section 7 will discuss

the complex makeup of the Rif Berber continuum and the importance of combining quantitative and qualitative perspectives for a better understanding of language variation and change. The article will end with a conclusion in Section 8.

2 Map, data, and aggregate discriminators

The data examined in this study mainly come from the ALR (Lafkioui 2007), of which the basic map with its 141 georeferenced points, belonging to 32 Rif Berber-speaking groups (Figure 3), is extracted and presented in Figure 4. These points are a selection of the 452 points that are examined and chosen by their degree of linguistic variation and comparativeness in the ALR. Initially, the survey points were selected on the basis of the principle of equidistance, which divides the inquiry field into several grids to which were assigned points that could match with localities on the field. The greater the variation was, the more the grids were reduced. All data investigated here stem from a vast geolinguistic corpus built by means of specific methodological procedures concerning data gathering, their systematisation, and their archiving (Lafkioui 2007, 2015). They were obtained by means of numerous linguistic, sociolinguistic, and ethnographic fieldwork investigations in the Rif area, which started in 1992, the last one being in autumn 2018.

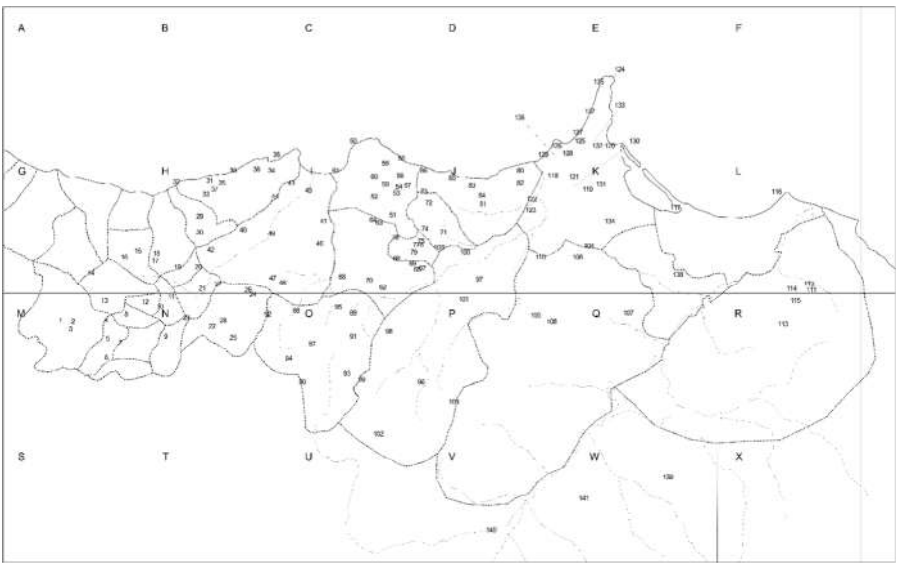

Fig. 4. Map of the selected georeferenced points of the Rif area
(Lafkioui 2007: 15)

In this study, the selected digital cross-level data are compared and classified according to the specific linguistic level to which they belong (i.e., phonetics, phonology, morphology, syntax, and lexicon), as well as to certain configurations

that combine the different levels in structural layers, such as, for instance, the overall cross-level configuration, which combines all levels.

The phonetic and phonological corpus is composed of 229 items and a selection of 141713 tokens, which correspond to the primary characteristics of Rif Berber's phonetic and phonological system, including the following phenomena: the vocalisation of the liquids *r*, *ṛ*, *rr*, and *ṛṛ* and the related extensions of the vowel system; phenomena pertaining to spirantisation and palatalisation, such as the synchronic spirantisation of the bilabial *b*, the diachronic spirantisation of the velars *k*, *g*, *kk*, and *gg*, the synchronic and diachronic spirantisation of the interdental *ṯ*, and the spirantisation of the pharyngeal *ɣ*; consonant mutations regarding the liquids *l*, *ll*, and the sequence *lṯ*; gemination; the vowel system, including vowel timbre, the initial vowel and its particular treatments, the central vowel schwa and its relating syllabic configurations; velarization of the uvulars *q* and *qq*; various assimilation phenomena; hiatus treatment; among others (see Lafkioui 2007: 17-95 for the relating qualitative analysis).

Regarding the morphological and syntactic corpus, it covers a wide range of phenomena pertaining to the nominal system (e.g., gender and plural formation, noun state), the pronominal system (e.g., independent and clitic pronouns), the verbal system (e.g., verb formation, PNG marking, standard and labile verbs, derivation, verb conjugation and valency, verbal nouns), word order, the negation system, and numerous invariable morphemes (e.g., demonstratives, prepositions, preverbs, adverbs, copula, ordinals, conjunctions, subordinators, negation and interrogation markers); see Lafkioui (2007: 97-241) for the relating qualitative analysis. The number of items examined is 195, corresponding to 398930 tokens.

As for the lexical corpus, it comprises 195 items regarding the human body, kinship, animals, colours, numbers, along with a subset of various nouns and verbs. This lexical selection is an augmentation of the corpus examined in Lafkioui (2018b) by 26 items and amounts to 371737 tokens; see Lafkioui 2007: 243-279 for the relating qualitative analysis.

Algorithmic classifications based on the ALR were possible only after an adaptive conversion of its data to the formats used by the RuG/L04 software and by GAPMAP, which also consisted of a laborious systematic conversion to UTF-8 for the geolinguistic data and to KML (http://www.opengeospatial.org/standards/kml/) for the geographic data.

The tokenized and pair wise aligned data used for this research is of excellent quality, as is shown by the following two relating measures of Cronbach's α and of local incoherence:

a) For the phonetic and phonological data, Cronbach's α has a value of 0.98 here, while the local incoherence measure has a value of 0.89.

b) For the morphological and syntactic data, Cronbach's α has a value of 0.99 here, while the local incoherence measure has a value of 0.92.

c) For the lexical data, Cronbach's α has a value of 0.99 here, while the local incoherence's value is 0.90.

Note that the closer to 1 the better the score (with a minimum of 0.7) for Cronbach's α. As for the local incoherence measure, the optimal score is 0, but values ranging from 1.75 to 2.05 correspond to what may be regarded as a yardstick for dialectology (Nerbonne and Kleiweg 2007). This implies that the local incoherence values of the present studies are far better than the average values used.

In order to adequately interpret the colour shades representing linguistic variation and the respective aggregate formations for MDS-classification – one of the most accurate and stable techniques for quantitative linguistic classification (Lafkioui 2008a, 2018b; Nerbonne et al. 2011) – the three-dimensional GABMAP colour cube in Figure 5 is very useful and works as follows. Based on the MDS-projection of the 141 dimensions (relating to the georeferenced points, Figure 4) to 3 dimensions per variety, each variety takes a specific position in the cube with a relating colour. Linguistically comparable varieties are sited next to each other in the cube and so take similar colours (e.g., shades of red), whereas dissimilar varieties have positions further apart in the cube and therefore take distinct colours (e.g., blue compared to yellow).

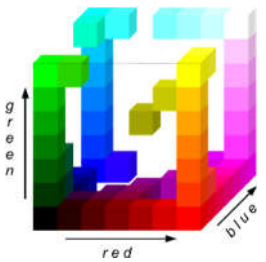

Fig. 5. Colour cube representing three-dimensional space for MDS (Leinonen 2010)

In addition to the numerous algorithmic classifications that are carried out on the large, cross-level, and representative corpus, this study also involves data mining tasks, which allow to objectively identify which features determine the emergence of the different stable aggregates of the Rif Berber continuum. This means that the study examines systematically which phonetic, phonological, morphological, syntactic, and lexical items are accountable for the major geolinguistic differences attested in the Rif Berber area. For this purpose, two different techniques are used. The first one is provided by GABMAP and consists of the quantitative measures of representativeness and of distinctiveness (Nerbonne et al. 2011). The second one – used here in order to verify the GABMAP technique – is based on the k-means clustering algorithm, which "searches for a pre-determined number of clusters within an unlabelled multidimensional dataset" (online access on GitHub: https://jakevdp.github.io/ PythonDataScienceHandbook/;

MacQueen 1967). The k-means approach adopted in this study consists of the following three steps: First, the k-means clustering algorithm is applied to the difference matrix of the cross-level dataset. This overall clustering serves as the comparative baseline. Second, the same k-means clustering algorithm is applied to the difference matrix of each individual feature, which results in a set of individual clustering classifications. Third, the resulting cluster discriminator score for a given feature is the sum of the number of sites with the same clustering as the baseline clustering; the maximum score being the number of sites, which is 141, while the minimum score is 1.
Subsequently, the outcomes of these two data mining studies are compared with the qualitative classifications and results from the ALR (Lafkioui 2007) for validation, which lead to the following phenomena as being the primary aggregate discriminators of the Rif Berber continuum (in order of prevalence according to the highest k-means score):
1. Lexicon: time expressions (score 121);
2. Phonetics-phonology: vocalisation of both the simple rhotic *r* and the geminate trill *rr* (score 113), and spirantisation and palatalisation of the velars *k* and *g* and their geminate counterparts, and spirantisation of the interdental *ṯ* (score 107);
3. Morphosyntax: pronoun (score 108);
These specific phenomena will be investigated from a quantitative and qualitative perspective in the following sections and will be ordered according to the linguistic level to which they belong.

3 Cross-level algorithmic classifications of Rif Berber

The algorithmic classifications presented in this section are the outcomes of numerous cross-level examinations of data concerning phonetics, phonology, morphology, syntax, and lexicon (see Section 2 for more details). These outcomes support once more the continuum makeup of the Rif Berber geolinguistic area (Lafkioui 2007, 2008a, 2018b), to which testifies the MDS-map displayed in Figure 6, which aggregates the linguistic differences quantified.
As for the internal structure of this continuum, there is a significant difference compared to previous classifications, which were based on lexical material only (2008a, 2018b), in that a new aggregate is revealed, which is plotted in fuchsia pink on the map in Figure 6. This new aggregate, which I coin East-Central Rif Berber (ECRB), mainly contains the varieties of Ayt Buyeḫya (nr. 25, Figure 3) and of Ibḍalsen (nr. 24), and also stands in a somewhat looser connection with the southern varieties of the Igzennayen (nr. 23), as is indicated by the colour shade of this area. Although ECRB forms an aggregate on its own, it correlates well with ERB, as shown by the colour continuity. In other words, compared to the lexical classifications, the overall cross-level classifications make the varieties of ECRB stand more out while still matching with the ERB varieties. The

phenomena responsible for the emergence of ECRB as a separate aggregate mainly pertain to the phonetic and phonological level, as will be shown in Section 4. Hence, the Rif Berber continuum is made up by the following five core aggregates (Figure 6 and relating aggregate partition in Figure 3): WRB (dark green), WCRB (light green and blue/bluish), CRB (orange, yellow-orange), ECRB (fuchsia pink), and ERB (cherry red). Furthermore, there is much internal variation attested in the WCRB aggregate, which is reflected in the mosaic pattern of its colour shades, signifying a relatively lower correlation between the items compared. This is especially the case of the small light brown aggregate of Targist (nr. 12), whose variety of the Ayt Ɛazza is particular in that its speakers form a small community of ancient immigrants, who originally stem from Iqelɛiyen (nr. 28) and thus from a region situated much more to the east of the continuum. This explains the light brown shade of this isolated area, as it contains linguistic features from both WCRB (light green) and CRB (orange).

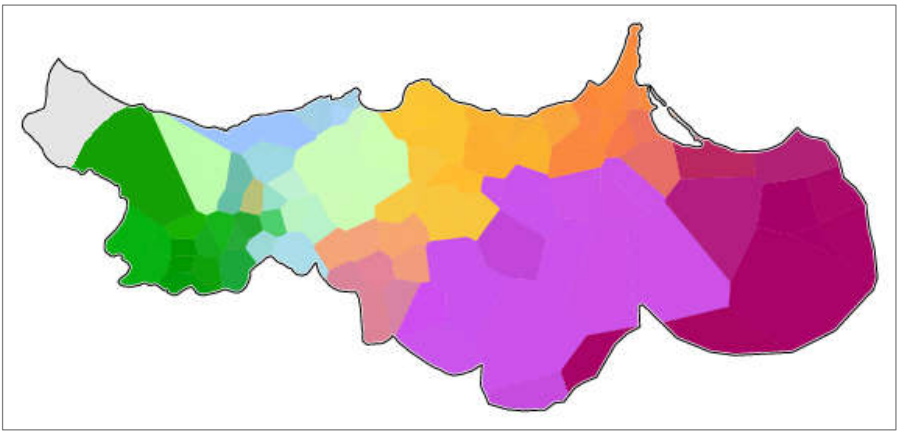

Fig. 6. Cross-level classical MDS Map of Rif Berber

The aggregates in Figure 6 are very stable, as they have been validated by means of a corresponding MDS scatter plot, a stable technique that GABMAP offers for this purpose and which usually stands for more than 80% of the variation in the data. The distances as measured by the plot show a high correlation with the distances given in the linguistic distance table, with a value of r=0.99.
Another technique that confirms that these obtained aggregates are highly stable and so accounts for an adequate aggregate display of the data is the probabilistic clustering technique, which basically consists of constantly adding quantities of noise while clustering and maintaining the cophenetic distance of the sites compared (Nerbonne et al. 2008). Even after 0.8 of noise added – while the default extra noise is 0.2 – the aggregates remain stable.

The stability of these major cross-level aggregates is also verified by other algorithmic classification techniques, as is shown for instance in Figure 7, which presents the results of a clustering classification based on the following weighted average algorithm (GABMAP):

$$d_{k[ij]} = (\frac{1}{2} \times d_{ki}) + (\frac{1}{2} \times d_{kj})$$

In doing so, the Berber data corroborate that this algorithm has the advantage of delivering consistent and representative clusters, as it allocates equal weight to the clusters that merge, despite the unequal number of sites that make up each cluster. Note that these clusters are also validated by means of the GABMAP cluster validation technique, which draws on MDS and its two dimension plots.

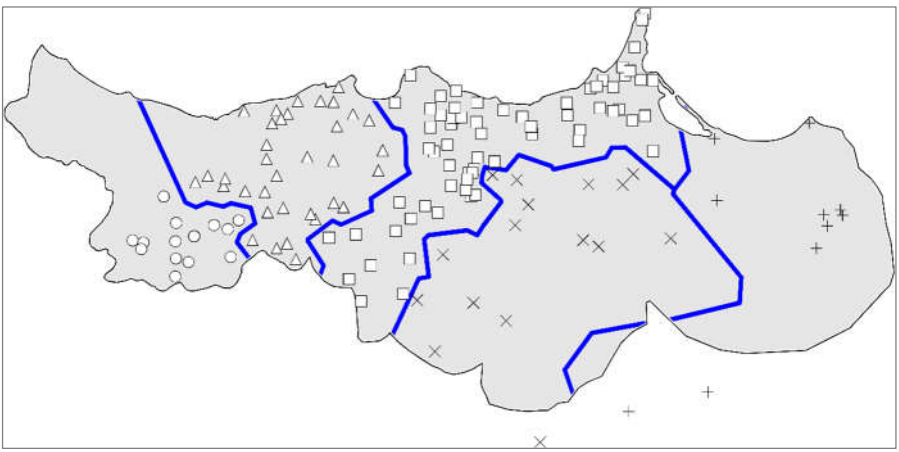

Fig. 7. Cross-level weighted average cluster map of Rif Berber

The following Figure 8 shows at glance that Rif Berber has the shape of a continuum, even when considered from different reference points. For instance, when the variety of Zegzel from the Iznasen (nr. 26), located in the east end of the Rif area, is taken as a reference point (see star on the map), the continuum outline of Rif Berber remains the same. In addition, the classification of Figure 8 points to the varieties of the Ayt Weryaγel (light green, WCRB, nr. 17) as forming an important transitional boundary and the varieties of Ktama (white, westernmost end of WRB, aka Senhaja, nr. 1) as being the most distinctive varieties compared to the variety of Zegzel.

It is the outcomes of all these classifications, including those presented in Figures 6 to 8, that are behind the subdivisions of the Rif Berber continuum displayed in Figure 3 and the respective Table 1 above (Section 2).

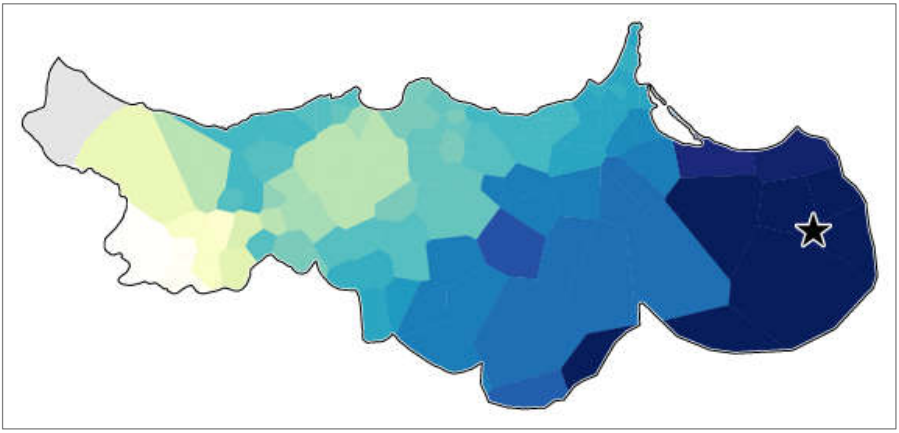

Fig. 8. Cross-level reference point map (Zegzel, Iznasen, nr. 26)

4 Phonetics and phonology

In this section are examined the phonetic and phonological classifications of the Rif Berber continuum (Section 4.1), as well as the two primary phenomena that determine its distinctive aggregates, which are vocalisation (Section 4.2) and spirantisation-palatalisation (Section 4.3), and which result from data mining tasks and qualitative verification. Both phenomena are part of a larger phonetic-phonological process which typifies Northern and Eastern Berber – to which Rif Berber belongs – and which probably traces back to ancient times, namely "weakening".

4.1 Algorithmic classifications

The algorithmic classifications of the phonetic and phonological phenomena of Rif Berber considered in this study corroborate the continuum makeup of this geolinguistic area, for which accounts the MDS-map presented in Figure 9. Different techniques were employed to verify the stability of its aggregates, among which the MDS scatter plot, which has a high correlation value of r=0.99 here. It is the phonetic and phonological data, and in particular those relating to vocalisation and spirantisation-palatalisation, that are largely determinant for the emergence of the ECRB aggregate (light green), which in the morphological, syntactic, and lexical classifications is less apparent as a distinct aggregate (Figures 12 and 16). Moreover, compared to the overall cross-level classification in Figure 6, the phonetic-phonological classification in Figure 9 reveals a closer affiliation between certain varieties of WCRB (i.e. varieties in red) and CRB (i.e. those in fuchsia). It also brings the northern varieties of the Igzennayen (orange-brown, nr. 23, Figure 3) closer to those of WCRB (i.e. varieties in shades of red, brown, and brown-green), while these – and all other varieties of the Igzennayen – are

usually more associated with CRB (see Figures 6, 12, and 16). This is because the northern Igzennayen varieties have certain important distinctive phonetic and phonological features, such as those pertaining to the vocalisation of the liquid *r* (Section 4.2), which they share with WCRB, and especially with the varieties of Ayt Ɛammart (nr. 14).

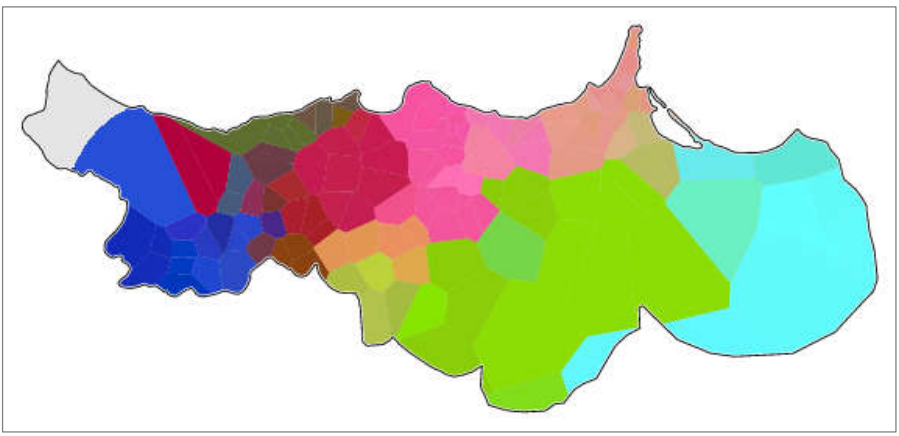

Fig. 9. Classical MDS Map of Rif Berber phonetics and phonology

4.2 Vocalisation and the relating extensions of the vowel system

The vocalisation processes that are examined here are responsible for the unique makeup of the Rif Berber vowel system. As a matter of fact, the varieties belonging to WCRB (most of them), CRB, and ECRB have developed through the vocalisation – henceforth VOC – of the liquids *r*, *ṛ*, *rr*, and *ṛṛ* a set of vowels that are lacking elsewhere in the Berber language family. Most Berber languages only distinguish the basic vowels *i* [ɪ] (the closed front vowel), *u* [ʊ] (the closed back vowel), and *a* [æ] or [ɛ] (the open central vowel), together with the mid central vowel [ə] (schwa), which is generally used as an epenthetic realisation feature, except in a small number of Berber languages, such as Berber of Jerba in Tunisia. There are some Southern Berber languages (Tuareg and Zenaga Berber, basically), were the vowel system would also comprise certain extensions, although not through the same diachronic processes (e.g., Prasse 1972 for Tuareg Berber). From a pan-Berber comparative perspective, VOC is an idiosyncrasy, a specificity of CRB, WCRB, and ECRB, as is shown in Lafkioui (2006b, 2007: 29-37, 2011a, 2018c) and on the classification map in Figure 10, which aggregates all phenomena pertaining to vocalisation of the liquids in question. On this map, only the pale yellow areas do not vocalise, whereas complete VOC is primarily attested in the green, green-blue, and blue areas. Both VOC and a retention of the

liquids are attested in the remaining areas, which are contact areas, including border zones (e.g., Ayt Ḥdifa of the Ayt Weryaɣel, nr. 17, Figure 3), transit areas (e.g., Ayt Ɛazza of Targist, nr. 12), and socio-economic centres (e.g., Imezzužen of Ayt Naḍur-Iqelɛiyyen, nr. 28). It should be mentioned that the liquid *r* undergoes certain transformations – but no VOC – in Timimoun, in the Gourara region of the Algerian Sahara (Boudot-Lamotte 1964), which mainly are: 1) at the end of a lexeme, *r* disappears without leaving any trace and, 2) in preconsonantal position, it transforms into [h], [ɦ], and [ʕ].

Given that VOC occurs in one large region, with the Central Rif area as focal area, and that it extends a considerable distance to the West and the East, further than what is indicated by data from the first Rif Berber descriptions (Biarnay 1917, Renisio 1932), the new vocalic variants are probably engendered in a particular location from which several variants are distributed (green and green-blue area in Figure 10), and more precisely in the Ayt Weryaɣel region (grass green area; nr. 17), with its various extended vocalisation phenomena, which are dealt with below. This region is not only one of the most innovating regions of the Rif area, language-wise (Lafkioui 2017, 2018a), but has also played an important role in its political and cultural history, and still functions as a sociocultural yardstick until today. In addition, the great size of the diffusion area of VOC also points to social acceptance.

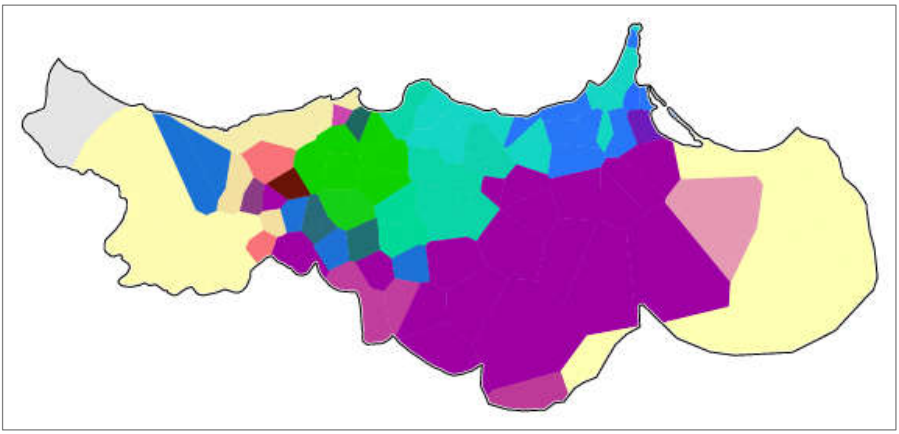

Fig. 10. Classical MDS map of all vocalisation types in Rif Berber

From a quantitative perspective, the systematic data mining analysis accurately identifies the following features as the primary aggregate discriminators of Rif Berber, which supports the qualitative findings presented in what follows and in Lafkioui (2007: 29-37, 2011a, 2018c):

- *r*-VOC for sequence *ur* ([ɔɑː] and [ʊː] as ultimate vocalised variants for *ū*) and *ir* ([ɛæː] and [ɪː] as ultimate vocalised variants for *ī*)
- EXTVOC1 and EXTVOC2

Vocalisation of the tap *r* and the trill *ṛ* is a diachronic process which in the Berber varieties of WCRB and CRB (Figure 10, green, green-blue, and blue zones) has caused an extension of the vowel system through the creation of units that are distinct from the basic vowels by: 1) a changed quality, that is, an apparent compensatory lowering for **ir* and **ur*, and 2) a significant quantitative value due to compensatory lengthening, as is shown in Table 2. Note that *ṛ*-VOC is a marginal phenomenon compared to *r*-VOC and that the monophtongs [ɪː] and [ʊː] are confined to the varieties of the Ayt Weryaɣel (grass green area in Figure 10, nr. 17 in Figure 3). The latter long monophtongs may vary in quantitative value, depending on the speaker in question; slightly shortened variants do occur, especially in the final position of the lexeme.

Table 2. Long vowels resulting from *r*-VOC and *ṛ*-VOC

Diachronic Form	Synchronic Form	Phonetic Form
ir	*ī*	diphthong [ɛæː] monophthong [ɪː]
ur	*ū*	diphthong [ɔɑː] monophthong [ʊː]
ar/ər	*ā*	monophthong [ɛː], [æː] or [aː]
iṛ	*ī̤*	monophthong [eˤː]
uṛ	*ṳ̄*	monophthong [oˤː]
aṛ	*ā̤*	monophthong [ʌˤː]

The vocalised variants displayed in Table 2 are thus the outcome of diachronic phonetic processes, which have reached different stages according to the variety in question. The long vowels *ī*, *ū*, *ā*, and *ā̤* belong to the phonological system of Rif Berber, as displayed in Table 3.

Table 3. Vocalic phonemes of Rif Berber

Basic vowels	Derived vowels
i	*ī*
u	*ū*
a	*ā*
	ā̤

As a general rule, VOC only occurs in the syllable coda, for which account examples 1 to 4:

(1)	*t̲asīt̲*	[θɛsɛæːθ]	<	*t̲asirt̲*	(= *t̲a* + *sirt̲*)	[θɛsɪˈrθ] ‘mill’
(2)	*t̲ammūt̲*	[θɛmmɔɑːθ]	<	*t̲ammurt̲*	(= *t̲am* + *murt̲*)	[θɛmmʊˈrθ] ‘land’
(3)	*ad̲rā*	[ɛðrɛː]	<	*ad̲rar*	(= *ad̲* + *rar*)	[ɛðrɛˈr] ‘mountain’
(4)	*t̲ā̤wa*	[θʌˤːwɛ]	<	*t̲aṛwa*	(= *t̲aṛ* + *wa*)	[θʌrˤwɛ] ‘offspring’

However, two types of "extended" vocalisation (EXTVOC) – with their various subtypes – have been developed in the Ayt Weryaɣel area and its neighbouring varieties (see grass green area in Figure 10; nr. 17), which is exemplified in (5) to (8):
- VOC of the liquids *r* and *ṛ* in onset position, i.e. EXTVOC1 (ex. 5 to 9)
- VOC of the mutated lateral approximant *l* (*r* < *l* or *r̂*), i.e. EXTVOC2 (ex. 10 and 11)

(5) *aɣūm* [aɣʊːm] < *aɣrum* (*aɣ* + *rum*), [aɣrʊˈm] 'bread'
(6) *dūṛ(w)* [doˤːw)] < *duṛu* (*du* + *ṛu*), [dorˤo] 'duro', 'coin'

EXTVOC1 transformations are very productive and pervasive in the area, since they affect even proper names (7), loanwords (8), and rhotics or taps in an absoluter Anlaut position before a full vowel (9), which is considered to be a strong consonant position and therefore less prone to vocalisation (Escure 1977; Foley 1977). EXTVOC1 is therefore a typologically remarkable phenomenon, even more when it concerns a rhotic or tap corresponding to the first radical of the stem, as in (8) and (9).

(7) *fāyda* [fɛːjdɛ] < *farida* (*fa* + *ri* + *da*), [fɛɾɪdɛ]
(8) *āḇəḥ* [æːβəħ] < *rḇəḥ* (monosyllabic), [ɾβəħ] 'to win'
(9) *āža* [aːʒa] < *raža* (*ra* + *ža*) [ɾaʒa] 'to wait'

EXTVOC2, on the other hand, allows the mutated lateral *l* (*r* < *l* or *r̂*) to be vocalised, which is also exceptional. Usually all instantiations derived from *l*, even the voiced tap [ɾ] with an ultra-light friction, the voiced trill [r], and the fricative [ɹ], do block VOC in Rif Berber. The following are examples in point; while (10a) is a case of EXTVOC2, (10b) is a retention case, which is regular practice outside the Ayt Weryaɣel area.

(10a) *irs* ([ɪrs], [ɪrs]) (< **ils*) + r-VOC > [ɛæːs], [jɛæːs], and [ɪjɛæːs] 'tongue'
(10b) *irs* ([ɪrs], [ɪrs]) (< **ils*) + r-retention > [ɪrs], [ɪrs] 'tongue'
(11) *malika* [mɛlɪkɛ] (*[mɛɾɪkɛ]) > *māyka* [mɛːjkɛ]

EXTVOC2 is subject to a strong diffusion in the area of the Ayt Weryaɣel and its adjacent varieties, which is echoed in the diversity and number of cases concerned, recently also involving proper names (as in 11), which are usually not affected by this phenomenon or by any consonant mutation of the lateral *l*. The frequently used case of (11) also accounts for intervocalic vocalisation.
As for VOC of the geminate trill *rr*, various developments are attested, which do not necessarily match with those of the simple tap *r*, as in (12). The tap [ɾ] changes its articulation mode and becomes the trill [rr] when geminated. Furthermore, *rr*-VOC in Rif Berber is a typologically remarkable phenomenon in that it

allows to apply a phonetic rule (i.e., VOC) to a part of the geminate (first part) and, therefore, invalidates the phonetic generalisations (Integrity and Inalterability) which claim the opposite (such as in Kenstowicz and Pyle 1973; Guerssel 1977).

(12) *kərrəz* 'ploughing'
> Retention of the geminate trill *kə**rr**əz* [ərr]
> Qualitative timbre change *ka**rr**əz* [ɛrr], [ærr]
> Qualitative and quantitative timbre change, *kā**rr**əz* [ɛːrr], [æːrr]
> Compensatory lengthening + simplification of articulatory tension, *kā**r**əz* [ɛːɾ], [æːɾ]

Once more, the varieties of the Ayt Weryaɣel (nr. 17) – and not those of the Ayt Temsaman (nr. 18) as claimed in Louali and Puech (1998) – account for the most developed diachronic vocalisation stage, and in this case that of the trill *rr*, since they only have the complete VOC stage, that is, *ār* ([ɛːɾ], [æːɾ]), involving both a compensatory lengthening and a simplification of the articulatory tension.
It is worth highlighting that the opposition [simple or vocalised tap *r*] versus [long or reduced trill *rr*] plays an important morphophonological role in Rif Berber's declination and derivation systems; e.g., it is used for TAM marking, verbal noun derivation, and plural formation. Moreover, VOC and EXTVOC have caused various significant morphological transformations in Rif Berber's linguistic typology, such as the restructuring of morphological patterns of its verbal and nominal system, which are discussed in detail in Lafkioui (2011a, 2018c).
From a diachronic perspective, all VOC developments are part of the global phonetic process of weakening of consonants, which is an essential feature of Rif Berber phonetics and phonology (Biarnay 1917; Renisio 1932) and for which ample evidence can be found; such as, for instance, the spirantisation of plosives (> fricatives > approximants > zero; see Section 4.3) and the vocalisation of semi-consonants (Lafkioui 2007: 27–28). Phonetic economy is probably the functional – and hence system-based – triggering factor behind these weakening transformations.
Given that the VOC of the trill *ṛ* shows a homologous course of transformation with the tap *r* – i.e. progressive change of quality and quantity of the preceding vowel – it is probable that economy has triggered either processes or that *ṛ*-VOC is a formal development created by analogy with *r*-VOC. The same goes for the VOC of the geminates *rr* and *ṛṛ*: the triggering factor is either functional (i.e. economy, code conformity) and/or formal (i.e. analogy).
In the same line of thought, the EXTVOC phenomena are language processes wherein the functional trigger of economy exerts such a strong pressure that the functional properties of "transparency" and "intelligibility" – of crucial importance for the form-function balance of language structure – are at risk. From a formal perspective, these innovations have a great impact and are sometimes

even pernicious because of their significant eroding effect on the basic syllable and lexical structure of Rif Berber. Nevertheless, these side effects do not deter speakers of the Ayt Weryaɣel from adopting them. EXTVOC1 in particular has a remarkable success in this region. Code conformity is a probable functional explanation for this trend, along with certain system-external factors, especially those engaging sociocultural and economical contact, such as schooling, transportation, and commercial activities (Lafkioui 2011a, 2018c). These system-external (speaker-based) factors strongly determine the success of the diffusion of the EXTVOC innovations in the phonetic and phonological system of Rif Berber. However, speaker-based factors do not explain the absolute preference in the focal area for complete vocalised variants, since the EXTVOC variants cannot be directly associated with specific social functions and patterns, such as status attribution. In fact, most of these variants, particularly when they are relatively recent innovations, are regarded as rural and are even stigmatised outside their habitat, especially in more urban areas, like in the city of Alhoceima (directing centre). Rif Berber-speakers who use the EXTVOC variants in their verbal interactions are well aware of the language representations and attitudes regarding them. They even adjust their language use to the more normative and common practices of VOC by restoring the etymological tap *r* as much as possible – sometimes with hypercorrections – when outside their local environment or comfort zone. Youngsters are particularly sensitive to the stereotypes concerning the EXTVOC variants and, lately, try to avoid them or to use the re-established tap and trill, even in the peri-urban city of Imzuren, where these variants are common practice. Since recently, a generational distinction with respect to the language attitudes towards EXTVOC is attested in Imzuren, especially among the schooled speakers. While youngsters seek to steer their vocalisation practices away from the rural tendencies and toward the variety spoken in Alhoceima, their parents and most second generation speakers either do not pay any attention to these specific variants or, on the contrary, they highlight them so as to use them as local and ethnic identity markers, in particular in the context of the Berber claim or local political matters.

Linguistic-internal factors rather explain the success of the adoption of EXTVOC innovations. Apart from the functional factors of economy and code conformity, certain formal factors also have had a hand in the transmission of the vocalised variants. This is for instance the case of certain recent EXTVOC1 innovations. For example, the new form *aɣūm* (**aɣrum* ‘bread’), initially only encountered among children and adolescents, is an extension by analogy with *aḇīḏ* (**aḇriḏ* ‘way’) and *amqqān* (**amqqran* ‘big’). The latter are long-established vocalisation cases: they occur among older generations, including speakers with nearly a century of language experience who claim always to have known this language practice as such, even among their grandparents. Accordingly, it is reasonable to infer that the recent EXTVOC innovations are formally motivated (extension by

analogy), probably just as with the older alternative vocalisations *aḇīḏ* and *amqqān* (functional and/or formal trigger).
Both hierarchical diffusion, from more to less speaker groups, and contiguous diffusion, from geographically adjacent speech communities, shape the geolinguistic patterns of vocalisation in the Rif Berber continuum. And in the specific case of EXTVOC2 (i.e. vocalisation of the approximant lateral), Rif Berber, and in particular the varieties of the rural area of Imzuren, which represent the focal diffusion area, provide counterevidence to Samuels' assumption (1972: 93) that innovations usually originate in large cities than in their surrounding areas and that these cities function as directing and diffusion centres. Consequently, Rif Berber also accounts for contra-hierarchical diffusion (Wikle & Bailey 1997).

4.3 Spirantisation and palatalisation

The second most significant phonetic-phonological determiner of the main aggregates composing the Rif Berber continuum concerns the phenomena of spirantisation and palatalisation of the velars *k*, *kk*, *g*, and *gg* and the spirantisation of the interdental *ṯ*.
Spirantisation corresponds to the phonetic process of weakening of plosives (*begadkefat*, Vycichl 1975), which in Rif Berber follows this trajectory or some of its stages: plosives > fricatives > approximants > zero (Lafkioui 2006a, 2007: 38–58). While "synchronic spirantisation" only involves a weakening of the articulatory tension, "diachronic spirantisation" engages both a weakening of the articulatory tension and a change of the place of articulation. Both these spirantisation types concern simple consonants generally. Moreover, there exists a hierarchy in Berber spirantisation, which goes as follows: velars > dentals > labials. This means that spirantisation at lower levels (labials and dentals) necessarily implies that of the corresponding higher levels; in other words, when labials are spirantised, dentals and velars are, too, and when dentals are spirantised, velars are as well. Phonetic economy (Louali 1999) and avoidance of homonymical conflicts (Armas 2019) are proposed as possible explanations to why Berber follows this hierarchy; the latter proposition is more plausible but is subject to further empirical confirmation.
Just like in the other spirantising Berber languages, which generally belong to Northern and Eastern Berber (including, e.g., Middle Atlas Berber, Kabyle Berber, and Jerba Berber), synchronic spirantisation is of limited relevance in Rif Berber. It is not consistent and it does not affect all simple plosives or it affects them to varying degrees, depending on the variety and the lexeme in question (Lafkioui 2007: 38-40). The only plosive whose spirantisation has a certain regular geolinguistic realisation and diffusion pattern is the bilabial *b*. Except the varieties of Ayt Iznasen (nr 26, Figure 3) and some irregular cases, all varieties

of Rif Berber spirantise this plosive to a variable degree in both Berber lexemes and loanwords.
Despite the leading trend of spirantising, plosives do exist in Rif Berber, although in most cases they are conditioned by the phonetic context. Indeed, the liquids *l*, *m*, and *n* often prevent spirantisation of dentals. The following are examples in point:

(13) *ulṯma* [ʊltmɛ] 'sister'
(14) *ṯammənṯ* [θɛmmənt] 'honey'
(15) *anḏu* [ɛndʊ] 'plate'

Simple unconditioned plosives are scarce in Rif Berber and generally derive from underlying long consonants, which have been simplified by contact with contiguous vowels. In addition, there are plosives derived from homomorphic long consonants – corresponding to radicals – as in (16) and those derived from heteromorphic long consonants, which result from the assimilation of a dental final radical and the feminine or diminutive marker *ṯ*, as in (17).

(16) *ṯaḇrat* < *ṯaḇratt* 'letter'
(17) *ṯaḇrit* < *ṯaḇritt* < *aḇriḏ* + *ṯ* _ *ṯ* (feminine marker) 'alley'

Consequently, it is synchronic spirantisation and a simplification of the articulatory tension that are at the basis of the tripartite opposition [long plosive vs simple plosive vs simple fricative], which is generally attested in spirantising Berber languages (see e.g., Saib 1974 for Middle Atlas Berber, Morocco). Examples from Ayt Temsaman (nr. 18, CRB) are displayed in (18) and (19).

(18) *tt* vs *t* vs *ṯ*
nətta 'him' vs *yə-sɣi=t* 'he bought it' (3MS-buy.PFV=FS)
vs *yə-sɣi=ṯ* 'he bought it (3MS-buy.PFV=MS)

(19) *dd* vs *d* vs *ḏ*
yə-ddā 'he lives' (3MS-buy.PFV) vs *d* (VENT) vs *ḏ* 'with'

As for diachronic spirantisation, the transformations concerned have reached different stages according to the varieties of Rif Berber (Lafkioui 2006a). Regarding the simple plosive velars *k* and *g*, the various phonetic realisations produced by diachronic spirantisation are often combined with secondary palatal realisations. The result is a range of polymorphic variants which, within the same variety, can be in free co-occurrence or can be conditioned by phonetic and morphological constraints. For example, instances of the type *ksi* 'take' (type 8 of the 10 possible types attested in Rif Berber; Lafkioui 2007: 41–48) have gotten to the final stage of spirantisation with a complete vocalisation of the velar *k* into *i*, as is shown in

(20). Complete spirantisation generally relates to the absence of a vowel in the immediate vicinity of the velar and is facilitated by the presence of the vowel *i* in word final position, which triggers vowel harmony (metaphony). In addition, the *ksi* type of spirantisation is of particular interest, in that it gives evidence of the morphological interplay between the plosive *k* (and variants) and its spirantised correlates, as in (21); certain varieties with a spirantised *ḵ* for the perfective correlate with a plosive *k* for the imperfective (e.g., *ḵsi* vs *kəssi* 'take').

(20) *ksi* > *ḵsi, šsi, šṯi, yšṯi, ysi, isi* 'take' (AOR)
(21) *ḵsi, šsi, šṯi, yšṯi, ysi, isi* 'take' (spirantised *ḵ* in PFV)
vs *kəssi, ḵəssi* 'take' (non-spirantised *k*/spirantised *ḵ* in IPFV)

The presence of the palatal *y* in these spirantisation transformations has probably to do with the following transformation rule 1 (22), which is generally observed in the so-called "Zenet" Berber languages; Zenet stands here for a group of Northern and Eastern Berber languages of North Africa which share a number of linguistic features:

(22) k + $C_{+dental/alveolar,+voiceless}$ > yC (rule 1)

However, rule 1 does not apply to all Rif Berber varieties or it affects them to varying degrees. The articulation realisations containing both the palatal *y* and a fricative – typical of many varieties of CRB, ECRB, and ERB – would have been the subject of an assimilation of sound and articulation point following rule 2 in (23).

(23) y + $C_{+fricative,+voiceless}$ > k̲C > (> šC) (rule2)

Rule 2, which is historically posterior to rule 1, would have thus reversed the course of the articulatory changes of the palatal *k* and would have therefore neutralised the effects produced by rule 1. The regularity of rule 2 is so that it also relates to old borrowings, such as the case in (24), borrowed from Arabic *zayt* 'oil'.

(24) *zzəyṯ* > *zzəyḵṯ, zzəḵṯ, zzəyḵšṯ, zzəḵšṯ, zzəyšṯ, zzəšṯ* 'oil'

Rules 1 and 2 would also explain the different hybrid realizations of spirantisation that are used in Rif Berber to indicate certain common morphological oppositions, like in (21).

Among the five possible types attested for the spirantisation of the velar *g* in Rif Berber (Lafkioui 2007: 49–56), type 3, represented by the form *ṯigəẓẓal* 'kidneys', is the type wherein *g* has completely disappeared, as is displayed in (25)

and exemplified in (26). Note that the remaining *i* in (26) may also be the outcome of an assimilation of the lexical *i* and the *i* resulting from *g*-spirantisation.

(25) *g - ḡ - yḡ - y (ey/iy) - i - ø*
(26) *ti**g**əẓẓal* > *ti**i**ẓẓal* 'kidneys'

There are also a number of exceptional cases pertaining to the spirantisation of the velar *g*, for which I refer to the ALR (Lafkioui, 2007: 49–56).

Another interesting process of diachronic spirantisation is that of the interdental fricative *t* [θ], generally instantiated through weakening until its complete disappearance (27), as is exemplified in (28) and (29) from Ktama (nr. 1, WRB).

(27) *t* [θ] > *h* [h] > *ø*
(28) *tə-nγa* > *hə-nγa* > *nγa* (3FS-kill.PVF) 'she killed'
(29) *t-aḥəčiw-t* > *h-aḥəčiw-t* > *aḥəčiw-t* (FS-girl-FRS) 'girl'

The fricative *t* generally corresponds to or is part of a morpheme, like, e.g., personal markers (28), pronouns (Section 5.3), and the feminine marker (29). It frequently occurs in absoluter Anlaut position (28, 29), which is regarded as an unfavourable position for spirantisation (Ferguson 1978). *t*-spirantisation is extensively attested in WRB and WCRB and to a lesser degree in the other aggregates of the Rif Berber continuum. As a matter of fact, in those varieties where it is of regular usage, it has affected certain morphological paradigms, by adding innovated variants functioning as allomorphs, such as for instance PNG marking (Lafkioui 2008b) and the pronominal paradigms (Section 5.3). This type of diachronic spirantisation does also occur outside Rif Berber, usually in the Zenet type of languages, like for instance in Shawi Berber (Aures, Algeria), where it is quite common (Basset 1961, Lafkioui & Merolla 2002: 16-17 and texts). Furthermore, in certain WCRB varieties, the fricative *t* [θ] undergoes voicing into *d* [ð] (30, 31) instead, which in contact with certain consonants, like nasals for instance, undergo assimilation into *d* [d] (31b). The following are examples in point.

(30) *t* [θ] > *d* [ð]
(31a) *tammənt* > *dammənt* [ðɛmmənt] 'honey'
(31b) *tammənt* > *dammənd* [ðɛmmənd] 'honey'

This kind of voicing is very consistently used with morphemes (e.g., feminine marker *t* — *t* in 31) in the varieties of the Ayt Weryaγel (nr. 17) and their neighbours, and is even applied to loanwords, such as in (33) from Arabic *ḥātəm* 'ring'; its realization goes hand in hand with maintaining the voiceless interdental *t* within the lexeme, whereas the varieties east of the Ayt Weryaγel generally have the voiced *d* as a correlate. In other words, outside the Ayt Weryaγel, *d*-voicing

occurs within the lexeme, often in post-vocalic position; both positions being favourable to voicing. Hence, the following correlations are attested in Rif Berber:
- *ḏ* in initial position goes with *ṯ* in the middle position (32a, 33a)
- *ṯ* in initial position goes with *ḏ* in the middle position (32b, 33b)

(32a) *ṯaṯbirṯ* > *ḏaṯbirṯ* ‘pigeon’
(32b) *ṯaṯbirṯ* > *ṯaḏbirṯ*

(33a) *ṯḥāṯəmṯ* > *ḏḥāṯənṯ* / *ḏḥāṯənḏ* ‘ring’
(33b) *ṯḥāṯəmṯ* > *ṯḥāḏəmṯ* / *ṯḥāḏənṯ*

Note that *ḏ*-voicing in word-initial position may be reproduced in final position, like in (33a). Contact with the varieties of Djebel Habib as an explanation for this kind of voicing, as suggested in Louali (1999), is unlikely since all Rif Berber varieties that distinguish this phenomenon are not in close and regular contact with the Djebel Habib area. Moreover, the varieties of WRB (i.e. Senhaja Berber), which are closer to Djebel Habib but still far away and with no particular cultural and socioeconomic connection, do not distinguish this phenomenon at all. Rather, it is more likely that *ḏ*-voicing has been motivated by system-internal factors, that is, by weakening – once more – as one step further in the spirantisation process, which in this case modifies the voiceless *ṯ* into the voiced *ḏ*. This sound change in Rif Berber perfectly reflects the last stages of the well-known Proto-Indo-European spirantisation case of the labialized voiceless velar plosive **kw*, the predecessor of English preaspired [hw], which would have gone through the following stages: **kw* > Germanic, Old English, Scots [xw] > Middle English, modern Scots [hw] > modern English [w] (Chambers 2013: 307-308; Schreier 2005). While a weakening from plosive to fricative to approximant generally concerns the velars discussed above, a weakening from voiceless to voiced merely pertains to the dental fricative *ṯ*.
Concerning palatalisation of the velars *k*, *kk*, *g*, and *gg*, it is mainly the well-known Zenet correlations (Kossmann 1995; Lafkioui 2006a, 2007: 59–68) that are attested among the majority of the Rif Berber varieties. In what follows, (34a) concern the *k*-*kk* correlations and (34b) the *g*-*gg* correlations.

(34a)	*k* (non-Zenet)	~	*š* (Zenet)	*kk* (non-Zenet)	~	*č* (> *šš*) (Zenet)
	kəm, *ḵəm*	~	*šəm* ‘you’	*nəkk*	~	*nəč*, *nəšš* ‘me’
(34b)	*g* (non-Zenet)	~	*ž* (Zenet)	*gg* (non-Zenet)	~	*ǧ* (> *žž*) (Zenet)
	gar	~	*žar* ‘between’	*aggag*	~	*aǧǧaǧ*, *ažžaž*

When it comes to velars in pre-vocalic or final position, WRB and westernmost WCRB (Senhaja) follow the non-Zenet and hence non-palatalising scenario, depending on the velar and lexeme in question; for instance, optional spirantisation for *k* may occur in this area, as in *ḵəm* 'you' in Ktama Berber (nr. 1).
As regards the origin of the spirantisation and palatalisation of the velars in Berber (i.e., Northern and Eastern Berber), both phenomena clearly pertain to the phonetic process of weakening (Biarnay 1917; Lafkioui 2006a, 2007: 36–68; Louali 1999; Renisio 1932). It is unclear, however, whether this general weakening process has been motivated by linguistic-internal or linguistic-external parameters, or by a combination of both, but this is out of the scope of this study. Yet some recent findings, based on toponymical material, point to contact as the main parameter (Armas 2019). This latter study suggests a system-external origin to spirantisation that traces back to at least Phoenician times (at least 2nd century BC) and thus way before the contact of Berber with Latin, as proposed in Kossmann (2013: 179). This ancient contact between Berber and Carthage and subsequently also the Roman Empire would explain the current diffusion of spirantisation along the Mediterranean coastline, from East to West, whereas elsewhere in Berber-speaking North Africa this phenomenon is absent or only sporadically attested. Although further study is needed, the diffusion of spirantisation and palatalisation in Rif Berber corroborates this claim, as is patterned on the corresponding classification map in Figure 11. Indeed, the more we go West, the less spirantisation and palatalisation are attested in Rif Berber, or at least not on a regular basis, which is the case of WRB and westernmost WCRB (Senhaja Berber; brown, black and deep purple coloured area in Figure 11).

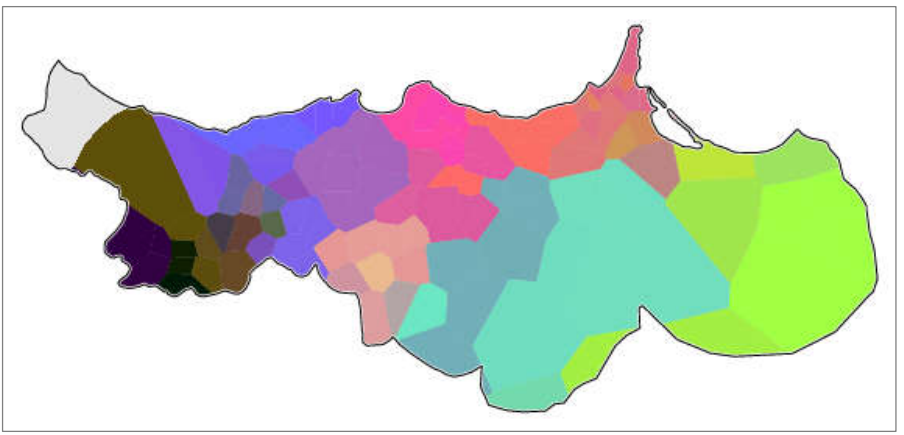

Fig. 11. Classical MDS map of all spirantisation and palatalisation types in Rif Berber

5 Morphology and syntax

This section discusses first the algorithmic classification of the morphological and syntactic data in Section 5.1. Next, it addresses in Section 5.2 the pronominal system and in particular the phenomena that resulted as primary aggregate discriminators from the data mining research.

5.1 Algorithmic classifications

As one can infer from the colour configurations on the MDS classification map in Figure 12, Rif Berber's morphology and syntax is patterned following an East-West axis, just as in the overall cross-level classification (Figure 6), as well as in the phonetic-phonological (Figure 9) and lexical (Figure 16) classifications. The five core aggregates of Rif Berber are very stable (r= 0.99) and well-demarcated here, although there is a close correlation between ECRB (pink) and ERB (fuchsia). Moreover, apart from the exceptional case of Targist (light brown area; nr. 12, Figure 3), WCRB displays a homogeneous colour pattern (light green) and hence a high morphosyntactic correlation between its varieties, whose properties extend further to the West (dark green; western Senhaja) but clearly differ from the varieties to the East (orange and pink).

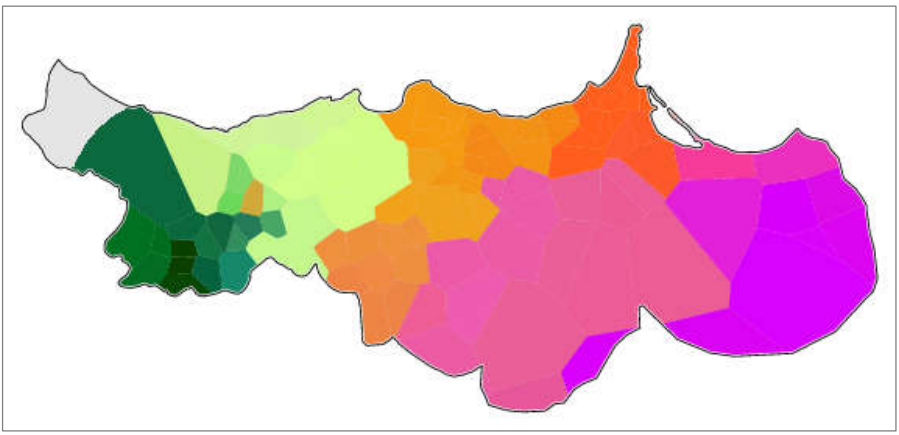

Fig. 12. Classical MDS map of Rif Berber morphology and syntax

5.2 The pronoun

The pronominal system of Rif Berber shows a remarkable morphological variation and complexity pertaining to its combinatorial and distributional properties (morphosyntax) and its geolinguistic diffusion (Lafkioui 2007: 116–163). Rif Berber distinguishes both independent pronouns and pronominal clitics; the latter

are grouped into several series according to their morphosyntactic function, that is, the direct object (DO) and indirect object (IO) clitics, the kinship clitics, the prepositional clitics, the predicator clitics, and the presentational clitics. The algorithmic classification of the Rif Berber pronouns presented in Figure 13 divides the geolinguistic Rif area in four main subdivisions (instead of five of Figure 12), wherein ECRB and ERB are merged into one single aggregate (ochre yellow). The pronominal system is one of the most significant aggregate discriminators of the Rif Berber continuum, with a k-means score of 108 and with the IO clitics at the top of the classification. In the light of these findings, DO and IO clitics – which are interrelated in Rif Berber – will be examined in next section.

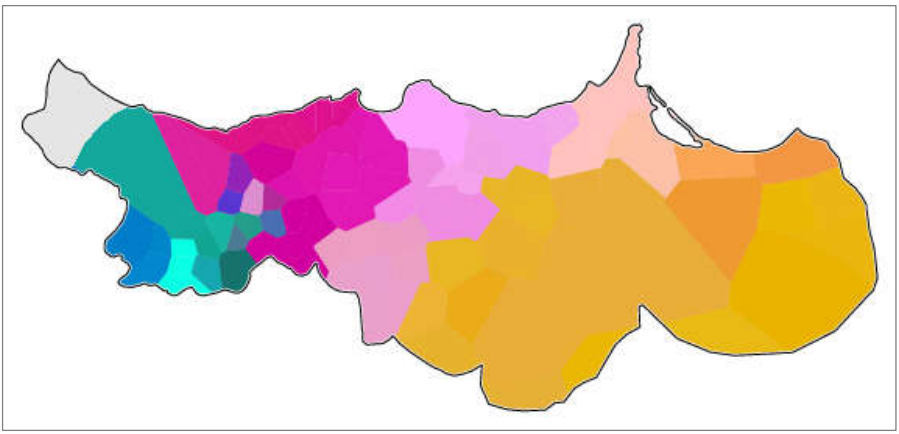

Fig. 13. Classical MDS map of all pronoun types in Rif Berber

5.3 DO and IO pronoun clitics

The DO and IO clitics distinguish two sub-series depending on their position in the verb phrase, that is, the postverbal series and the preverbal series, as is shown in Tables 4 and 5.

Concerning the DO clitics, the postverbal series are grouped into two sub-series based on the morphosyntactic structure of the verb phrase, that is, sub-series I and II, the latter being distinct by an initial *i*, except in certain WRB varieties (Senhaja) where the vowel in question is *a*, with the exception of the 3S and 3P where it remains *i*. The initial vowel *a* is probably an innovation triggered by analogy with the initial *a* of the IO correlates (see Table 4; on the initial vowel of this type of clitics in Berber, see Galand 2010: 116–17).

1) The postverbal sub-series I clitics are used after a verbal form which ends with:

- A full vowel of the verb stem, with the exception of the perfectives of the type c1c2 and c1c1 in most varieties of WCRB, CRB, ECRB, and ERB; e.g.,

ṯ-rəbbu=šəm ‘she carries you on her back’ (3FS-carry.IPFV=2FS) from Igzennayen (nr. 23, Figure 3). In WRB and most of western WCRB (Senhaja), on the other hand, this rule is valid for all types of verbs ending with a vowel; e.g., *i-čča=ṯən* ‘he ate them’ (3MS.eat.PFV=3MP) from Ayt Bšir (nr. 6).
- A consonant which is part of the PNG marker; e.g., *ṭṭf-ən=š* ‘they took you’ (PFV.take-3MP=2MS) from Ayt Temsaman (nr. 18).
2) The post-verbal sub-series II clitics, which are marked by the presence of an initial *i* or *a* (Senhaja), are used after a verbal form ending with:
- The vowel *a* of the perfective of verbs of the type c1c2 and c1c1, which is deleted when clitics are added (mainly in CRB, ECRB, ERB, and most of WCRB); e.g., *i-nɣ=išḵum* (< *i-nɣa=išḵum*) ‘he killed you’ (3MS-kill.PFV=2MP) from Ayt Weryaɣel (nr. 17), which in WRB is usually rendered by *i-nɣ=awən* (< *i-nɣa=awən*) (3MS-kill.PFV=2P) from Ayt Bušibet (nr. 3).
- A consonant which does not represent the PNG marker; e.g. *yə-ḏfr=ikənniw* ‘he followed you’ (3MS-follow.PFV=2MP) from Wlad Settut (nr. 29). But certain ERB varieties follow the subseries I rule when a schwa precedes the final consonant; e.g., *yə-ḏfər=ḵən* ‘he followed you’ (3MS-follow.PFV=2MP) from Iznasen (nr. 26).
So, apart from the exceptions and restrictive rules cited above, C-initial clitics are generally used after a full vowel of the verb stem or a suffixal PNG marker (postverbal sub-series I), whereas V-initial clitics appear after a consonant of the verb stem or perfectives of the type c1c2 and c1c1 ending with a (postverbal subseries II). C-initial clitics may be preceded by the extension *ya*, which is attested for the first person all over the Rif area, e.g. *yay* (1S) and *yanəɣ* (1P) and their local variants, and in particular in WCRB and WRB, where it also occurs in the 2S and 2P in many varieties. In the latter aggregates (especially in Senhaja), the extension undergoes metathesis into *ay* when combined with V-initial clitics following a consonant of the verb stem, as in e.g. *i-ssn=ayam* (3MS.know. PFV=2FS) from Ktama (nr. 1).
As for the variants containing the spirantised *h* < **ṯ* (WRB, western WCRB; see spirantisation Section 4.3), they appear in postvocalic position, like for instance with the following cases of Ktama (nr. 1): the clitic *h* (3MS), which is usually followed by the ventive, as in *i-nɣa=h=d* ‘he killed him’ (3MS-kill.PFV=3MS =VENT); the clitic *hən* (3P) *i-nɣa=hən=d* ‘he killed them’ (3MS-kill.PFV=3MP =VENT) of the subseries I. In case of verbs ending with a consonant, the variant *ihən* (3P) of the subseries II is used, as in *i-ssn=ihən* ‘he knew them’ (3MS-know.PFV=3MP).
Regarding the 3FS, the data divide the Rif area into two, basically: one aggregate contains ERB, ECRB, and a large part of CRB, which freely alternate the clitics *t* and *t(t)əṯ* (and corresponding variants with *i*), and a second aggregate is mainly composed of WCRB and WRB, which use the variant *t* and its correlate *it*, only.

The WRB aggregate contains, however, a small number of varieties (mainly located around the area of Ayt Hmed, nr. 4, and Taghzut, nr. 2), which have the uncommon affricate *ṯ* as clitic for the 3FS, which is also attested in Kabyle Berber (North Algeria). In certain WRB varieties (especially the more western ones; western Senhaja), the presence of the ventive triggers a number of phonetic changes of the clitics, such as for example that of the 3FS *t* into *h* in an intervocalic position (just as with 3MS *t̲* > *h*) and that of *t* into *t̲* in prevocalic position. Consequently, the 3MS and 3FS clitics converge here. They are dissociated by means of the ensuing ventive, which, for instance, changes into *id̲* after the 3MS clitic as opposed to the regular *d* (*əd*) after the 3FS.

Concerning the 1P clitic, its variants may be subject to devoicing, e.g., *anəḫ*, *aḫən*, *ax* compared to *anəɣ*, *aɣən*, *aɣ*, which frequently occurs in CRB. In westernmost WCRB and WRB (Senhaja), on the other hand, the corresponding *ɣ* may undergo vocalisation and result into *ana* and its reduced form *na* (< *anaɣ* < **anəɣ*), which is reinstated when the ventive follows (see Lafkioui 2007: 82-83 for vocalisation and devoicing of *ɣ*). The variants *aɣən* and *aḫən* have been affected by metathesis, which is also attested in other Berber-speaking areas of North Africa (Brugnatelli 1993, Lafkioui & Merolla: 2002).

WRB (western Senhaja) generally does not distinguish gender for the plural in all pronominal series, whereas western WCRB (eastern Senhaja; i.e., Zerqet, Ayt Gmil, Ayt Bufraḥ, Targist, and Ayt Mezduy) does, just as in all other Rif Berber varieties.

As regards the 2P DO clitic, Rif Berber can be split into two, a group with clitics based on *wən* and its numerous variants, spread all over the Rif area, and another one based on *kum* and its variants, mainly attested in WCRB and WRB (see distribution maps in Lafkioui 2007: 131). Note that clitics with *t*-voicing (into *d*) are commonly attested among the Ayt Weryaɣel varieties (nr. 17), not only for the 2FP, e.g., *ikənd*, *ik̲ənd*, and *išk̲ənd*, but also for the 3FP, e.g., *it̲ənd* and *ind*. On the other hand, in more western varieties of WCRB and those of WRB (Senhaja), a spirantisation of *t̲* > *h* > *ø* is regularly retrieved among the 3P clitics, as in *t̲ən*, *hən*, *n*; the latter variant usually occurs in preverbal position.

However, the 3P *n* (and variant *in*), may also stem from another – and probably older (see Brugnatelli 1993 about the loss of *t̲* in Berber) – variant with which it is homophonous and which is frequently used among a number of non-Senhaja varieties of Rif Berber (mainly eastern WCRB, CRB, ECRB), where it is in perfect correlation with the 3FP clitic *nt* (and variant *int*); both may appear here in postverbal and preverbal position, although a slight preference for the preverbal position is recorded. The short variants *n* (3MP) and *nt* (3FP) are employed after verbs ending with a vowel, as in (35a) from Ayt Temsaman (nr. 18). Otherwise, a long variant of the clitic is used, as in (35b), or a variant of type II with an initial *i-*, as in (36).

(35a) *yə-ẓri=n*
3MS-see.PFV=3MP
'He saw them.'

(35b) **ẓri-γ=n > ẓri-γ=ṯən*
see.PFV-1S=3MP
'I saw them.'

(36) *yə-ṭṭf=in*
3MS-take.PFV=3MP
'He took them.'

Concerning the preverbal DO clitics, their fronted position is triggered by certain preverbal markers, such as those pertaining to the irrealis (e.g., *a*, *ad*, *la*), negation (e.g., *u*, *wa*, *war*), and subordination (e.g., *a*, *ay*, *y*). There is only one preverbal series, which does not alter according to the morphological features of the verb (except in westernmost WRB; western Senhaja), but it may alter according to the phonetic context and the geolinguistic variety in question. As such, short and spirantised variants containing *h* are preferred (and sometimes even obligatory) in preverbal position in certain aggregates (especially in WRB). The preverbal series is morphologically distinct from the post-verbal series I for the first person in most of the Rif Berber varieties, for which the clitics begin with the interdental *ḏ* in the majority of the Rif Berber aggregates (except WRB) or with *l* in certain ERB varieties (see Table 4).
In WRB and westernmost WCRB (Senhaja), on the other hand, the 1S preverbal clitic is the same *y* as the postverbal clitic, with the exception of some varieties of westernmost WRB, mainly belonging to Ktama (nr. 1), where *yṯ* is attested, and to Taghzut (nr. 2), where *yḏ* is recorded (see blue and blue-green aggregate in Figure 13); these particular variants generally go with the 3MS and the 3P of the verb, whereas the regular *y* is used with all other persons (2S, 2P, 3FS), except with the 1S. A similar variant is attested in Ghomara Berber (El Hannouche 2010: 116). It is also in Senhaja Berber that certain varieties (e.g., Ktama, nr. 1) have *hən* as the preverbal correlate of postverbal variants such as *na*. One of the most regular 1P preverbal clitics in Senhaja Berber is *ġən*, which results from a metathesis of the pan-Berber *nəġ*. Metathesis affecting preverbal clitics is also common elsewhere in Rif Berber, especially in WCRB, CRB, and ECRB, where not only variants such as *ḏayən* and *ḏaxən* are frequently used, but also their short correlates *ḏay* and *ḏax*. The less common metathesis of the 3MP *ṯən* into *nt* is also attested in Rif Berber, especially in CRB, where it appears in preverbal position mainly, just as in the other Berber languages where it occurs (Brugnatelli 1993). Note also that the 3MS preverbal clitic is *t* instead of the more common *ṯ* (or spirantised *h*) in certain WRB varieties (e.g. Ayt Hmed, nr. 4) and in parallel the 3P is *tən* instead of *ṯən* or *hən*.

Some general observations concerning all DO series are the following: relatively less variation is attested for the 3S and 3P, a phenomenon that Rif Berber shares with the majority of Berber languages. Moreover, variants containing the spirant *š* – which is a typical Zenet feature – are mainly retrieved in ERB, ECRB, and CRB, to which are added, as one goes towards the West, variants containing the velars *k* and *ḵ* (CRB and WCRB), which end up completely supplanting the former in the varieties of WRB (western Senhaja). It is also worth mentioning that the 2S and 2P of the postverbal I and preverbal series have free variants which are similar to their independent correlates; e.g. *kəm*, *ḵəm*, *šəm* for the 2FS. The 2S and 3P also have short variants. In CRB, ECRB, and WCRB, these short variants tend to appear with perfectives of the c1c2 and c1c1 type for the 3S and 1P, as in the following examples from Ayt Temsaman (nr. 18): *yə-nγa=m* compared to *yə-nγ=išəm* 'he killed you (FS)' (3MS.kill.PFV=2FS), *yə-nγ=iš* compared to *yə-nγ=išək* 'he killed you' (3MS.kill.PFV=2MS), *yə-nγ=in* compared to *yə-nγ=iṯən* (3MS.kill.PFV=3MP) 'he killed them', *yə-nγ=int* compared to *yə-nγ=iṯənt* 'he killed them (FP)' (3MS.kill.PFV=3FP). In WRB and westernmost WCRB (Senhaja), both short and long variants do also occur but generally in free alternation (although a preference for short forms is recorded) and for the 2S; e.g., *m* instead of *ḵəm* for the 2FS (Ayt Bunsar, nr. 5).

As for the IO clitics, they consist of a postverbal and a preverbal series. The latter series has a prefixed *ḏ* (in most Rif Berber varieties) or *l* (in certain ERB varieties), except in WRB and western WCRB (Senhaja), where the IO series is largely similar to the DO series, except for the 3S and 3P, which correspond to the respective pan-Berber variants *as* and *asən*, but do not distinguish gender, though. Elsewhere in Rif Berber, gender is mainly marked by a *t* or its voiced correspondent *d* (Ayt Weryaγel, nr. 17). Apart from voicing, the IO clitics, may also be subject to diachronic spirantisation and to vocalisation (see Table 5). Just like with the DO clitics, the variants of the 2MP are subdivided into two sets; a set containing the bilabial *w* (e.g. *awen*), scattered all over the Rif area, and a set having the velar *k* or *ḵ* (e.g. *aḵum*), mostly attested in WCRB and WRB. Concerning the 2FP, the Ayt Weryaγel varieties (nr 17) provide interesting findings, as they do not only account for a voicing of the final dentals (e.g. *akənd*, *aḵənd*) but also for the irregular variant *aḵumt*, which is the feminine corelate of *aḵum* (see Figure 14 for a geolinguistic distribution of the postverbal 2FP). When in contact with a vowel, variants of the post-verbal series alternate freely with allomorphs preceded by *y*, as in (37) from Ayt Weryaγel (nr. 17).

(37) *ini=asən* > *in=asən* / *ini=yasən*
tell.AOR.IMP.S=3MP
'Tell them!'

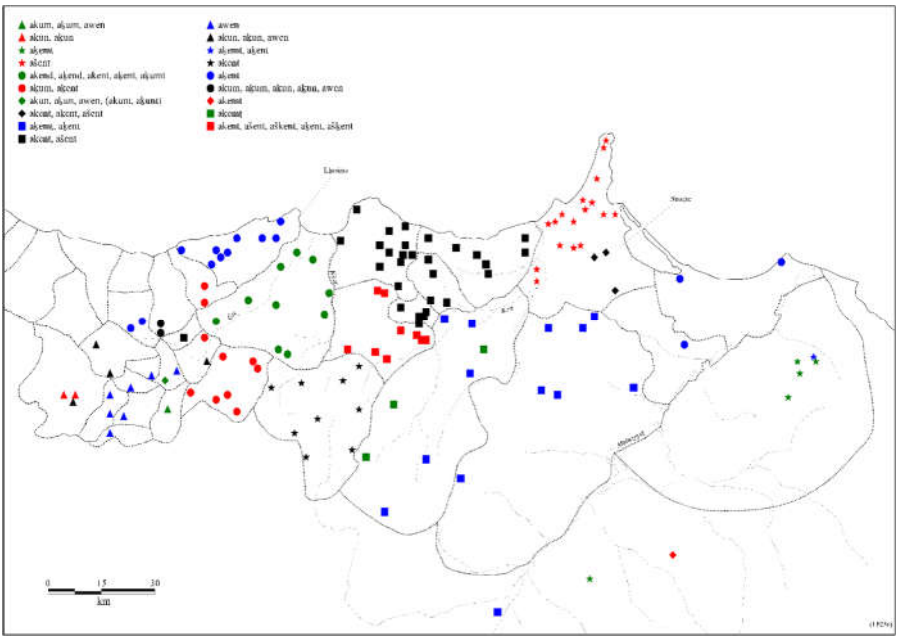

Fig. 14. Postverbal IO 2FP clitic in Rif Berber (ALR, Lafkioui 2007: 131)

Preverbal clitics, whether they are DO or IO, are employed in certain morpho-syntactic contexts that cause fronting, such as the irrealis, negation, subordination, and interrogation. Pronominal and ventive clitics precede the verbal head (but follow the marker), usually without changing their respective order, namely [IO clitic + DO clitic + ventive clitic]. Clitic fronting is a common phenomenon in Rif Berber and in Berber in general. The following examples from Ikebdanen (nr. 27) account for it. Example (38a) is a regular case of fronting, whereas (38b) is a case of partial fronting, which is also attested in Rif Berber, especially in ERB, WCRB, and WRB, and generally when the ventive is involved. Clitic doubling is also attested in these aggregates of Rif Berber, to which (38c) testifies.

(38a) *ur* ***ḏay***=*t*=*id*=*yə-sɣi* *ša.*
NEG 1S=3FS=VENT=3MS-buy.PFV.NEG NEG
'He did not buy me that.'

(38b) *ur* ***d***=*yə-sɣi*=***t*** *ša.*
NEG VENT=3MS-buy.PFV.NEG=3FS NEG
'He did not buy it.'

(38c) *ur* ***d***=*yə-sɣi*=*t*=***id*** *ša.*
NEG VENT=3MS-buy.PFV.NEG.=3FS=VENT NEG
'He did not buy it.'

Although partial fronting and clitic doubling is attested elsewhere in Berber, such as in Shawi Berber (Lafkioui & Merolla 2002), it is not a widespread phenomenon. A more common phenomenon involving clitics in Berber is the double expression of the IO referents, generally based on the verb 'to say', such as *yə-nna=as* 'he said to him' in *yə-nna=as i mmi=s* 'He said to him to his son > He said to his son' (3MS-say.PFV=3MS to son=3MS). Some of these double dative expressions, such as e.g. *yə-nn=as* (3MS-tell.PFV=3MS) and *yə-nn=aš* (3MS-tell.PFV=2MS), are grammaticalised and often employed to manage the dynamics of the discourse (see Souag 2015 on dative agreement in Berber).

From the data mining research, the IO clitics came out as the most important pronouns in shaping the aggregate formations of the Rif Berber continuum on a morphosyntactic level, which is plotted on the MDS classification map in Figure 15. In addition, the postverbal IO 2FP clitic was identified as the primary aggregate discriminator, corresponding to a k-means score of 108; the overall geolinguistic distribution of this clitic in the Rif area is displayed in Figure 14, which is extracted from the ALR (Lafkioui 2007: 131).

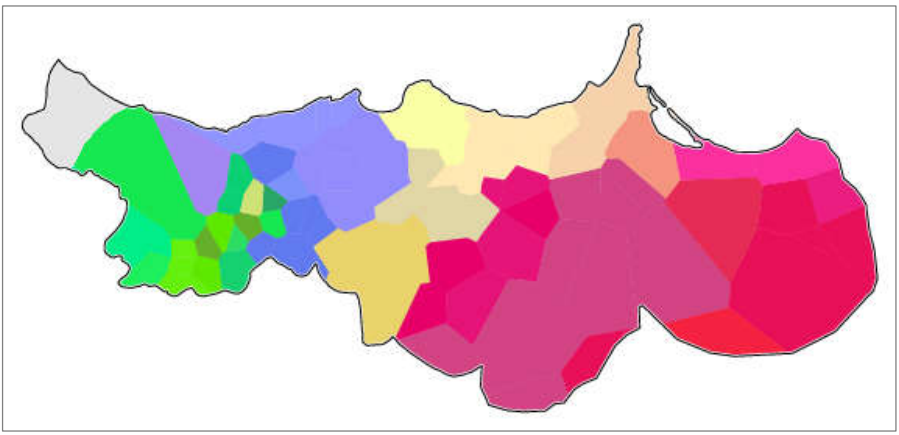

Fig. 15. Classical MDS map of IO clitics in Rif Berber

PNG	DO		
	Postverbal I	Postverbal II	Preverbal
1S	*ay, ayi, əyyi, iyyi, y, yi*	*ay, ayi, əyyi, iyyi, y, yi*	***d*ay, *d*əyy, *d*iyy, *d*y, *d*yi, *d*i, *l*i, *y, yt, yd***
2MS	*k, k, š, šək, šək, kəž, kəžž, kəž, ak kəžž, kəğ, kəğ*	***i*k, *i*k, *i*š, *i*šək, *a*k**	*k, k, š, šək, šək, kəž, kəžž, kəž, kəžž, kəğ, kəğ*
2FS	*kəm, kəm, šəm, m, akəm, akəm, am*	***i*kəm, *i*kəm, *i*šəm, *a*kəm, *a*kəm, *a*m**	*kəm, kəm, šəm, m*
3MS	*t, h*	***i*t, *h***	*t, h, t*
3FS	*t, ttət, tət, t*	***i*t, *i*ttət, *i*tət, *i*t**	*t, ttət, tət, h*
1P	*anəɣ, anaɣ, aɣ, aɣən, ana, na, anəx, axən, ax*	*anəɣ, anaɣ, aɣ, aɣən, ana, na, anəx, axən, ax*	***d*anəɣ, *d*anaɣ *d*aɣ, *d*aɣən, *d*anəx, *d*axən, *d*ax, *d*ana, *l*anəɣ, aɣən, ɣən, hən**
2MP	*wəm,wən, awən kən, kən, kum, kum, kun, kun, šwəm, šwən, škum, kənniw, kənniw*	***i*wəm, *i*wən, *a*wən, *i*kən, *i*kən, *i*kum, *i*kum, *i*kun, *i*kun, *a*kən, *a*kən, *a*kum, *a*kum, *a*kun, *a*kun, *i*šwəm, *i*šwən, *i*škum, *i*kənniw, *i*kənniw**	*wəm,wən, kən, kən, kum, kum, kun, kun, šwəm, šwən, škum, kənniw, kənniw*
2FP	*kənt, kənt, kənd kənd, kəmt, kəmt, kəmt, škənt, škənt, škənd, šənt, kunt, kunt, wənt, kənnint, kənnind, kənnint, kənnimt, kənnind, kənniw, wən, awən, wəm, kən, kən, kun, kun, kum, kum, škum, šwəm, šwən*	***i*kənt, *i*kənt, *i*kənd, *i*kənd, *i*kəmt, *i*kəmt, *i*kəmt, *i*škənt, *i*škənt, *i*škənd, *i*šənt, *i*kunt, *i*kunt, *i*wənt, *i*kənnint, šwəm, *i*kənnind, šwən, *i*kənnint, *i*kun, *i*kənnimt, *i*kun, *i*kənnind, *i*kum, *i*kənniw, *i*wən, *i*wəm, *i*kən, *i*kən, *i*kum, *i*škum, *a*kən, *a*wən, *a*kən, *a*kum, *a*kum, *a*kun, *a*kun**	*kənt, kənt, kənd kənd, kəmt, kəmt, kəmt, škənt, škənt, škənd, šənt, kunt, kunt, wənt, kənnint, kənnind, kənnint, kənnimt, kənnind, kənniw, wən, wəm, kən, kən, kun, kun, kum, kum, škum, šwəm, šwən*
3MP	*tən, hən, n*	***i*tən, *i*hən, *i*n**	*tən, hən, n*
3FP	*tənt, hənt, nt, tənd, nd, tən hən, n*	***i*tənt, *i*hənt, *i*nt, *i*tənd, *i*nd, *i*tən, *i*hən, *i*n, tənt, hənt, nt**	*tənt, hənt, nt, tən, hən, n*

Table 4. Predominant DO clitic pronouns in Rif Berber

PNG	IO	
	Postverbal	Preverbal
1S	*ay, ayi, əyyi, iyyi*	*ḏay, lay, ḏayi, ḏəyyi, ḏiyyi , y, yṯ, yḏ*
2MS	*ak, aḵ, aš, ay*	*ḏak, ḏaḵ, ḏaš, laḵ ḏay, ak, aḵ*
2FS	*akəm, aḵəm, am*	*ḏakəm, ḏaḵəm, ḏam, lam, akəm, aḵəm, am*
3MS	*as*	*ḏas, las, as*
3FS	*as*	*ḏas, las, as*
1P	*anəɣ, anaɣ, aɣ, ana, anəx, ax*	*ḏanəɣ, ḏanaɣ ḏaɣ, ḏaɣən, ḏanəx, ḏaxən, ḏax, ḏana, lanəɣ, aɣən, ɣən, ahən, hən*
2MP	*awəm,awən, akum, aḵum, akun, aḵun*	*ḏawəm, ḏawən, ḏakum, ḏaḵum, ḏakun, ḏaḵun, lawən, awəm, awən, akum, aḵum, akun, aḵun*
2FP	*akənt, akənd, aḵənt, aḵənd, aḵəmt, aḵəmṯ, aḵumt, akunt, aḵunt, ašənt, aškənt, ašḵənt, ašənt, (a)wən, (a)kum, (a)ḵum, (a)kun, (a)ḵun*	*ḏakənt, ḏakənd, ḏaḵənt, ḏaḵənd, ḏaḵəmt, ḏaḵəmṯ, ḏaḵumt, ḏakunt, ḏaḵunt, ḏašənt, ḏaškənt, ḏašḵənt, ḏašənt, ḏawən, ḏakum, ḏaḵum, ḏakun, ḏaḵun, laḵəmt, laḵəmṯ* *akənt, aḵənt, aḵəmt, aḵəmṯ, aḵumt, akunt, aḵunt, (a)wən, (a)kum, (a)ḵum, (a)kun, (a)ḵun*
3MP	*asən, sən*	*ḏasən, lasən,asən*
3FP	*asənt, asənd, (a)sən*	*ḏasənt, ḏasənd, ḏasən, lasənt, (a)sən, (a)sənt*

Table 5. Predominant IO clitic pronouns in Rif Berber

6 Lexicon

The MDS classification projected onto Figure 16 not only confirms that the Rif Berber varieties form a language continuum, lexically speaking, but also provides results analogous to those presented in Lafkioui (2008a, 2018b), regardless of an increase of the data examined from 169 to 195 items (20 nouns and 6 verbs). Therefore, the following four core aggregates are very stable: ERB (fuchsia and pink), CRB (orange and orangey), WCRB (light green and green-blue), and WRB (dark green). Their stability is also validated by the MDS scatter plot, which shows a high correlation with the distances given in the linguistic distance table, with a value of r=0.99. Other techniques that were used to verify the stability of these aggregates, such as the probabilistic clustering technique, also confirm these classification outcomes.

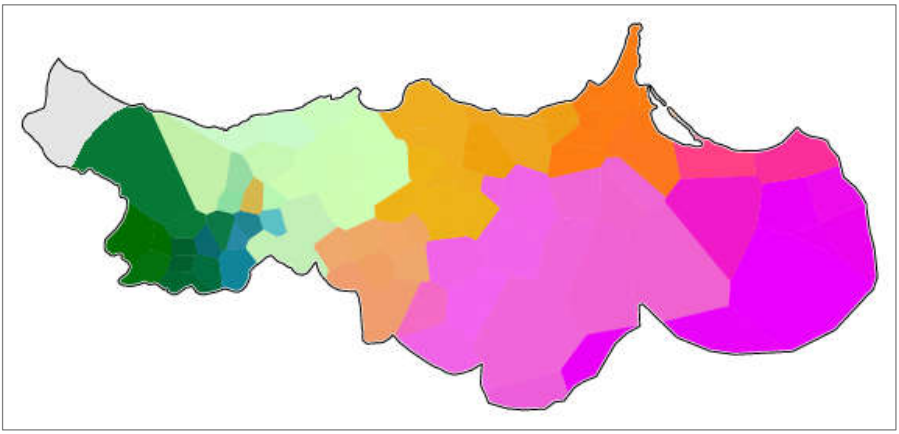

Fig. 16. Classical MDS Map of Rif Berber lexis

Note that ERB (fuchsia and pink) includes here ECRB (pink), which in the other classification plots forms a separate aggregate (see Figures 6, 9, and 12). Moreover, in this lexical classification, WCRB (light green and green-blue) exhibits a high correlation with WRB (dark green), which is conveyed through the colour continuity of the aggregates.

The data mining study of the augmented lexical corpus corroborates Lafkioui's findings (2008a, 2018b) and indicates lexemes belonging to semantic fields of time and space and of the human body as more significant in discriminating the aggregates. Indeed, the most important cross-level aggregate discriminator of the Rif Berber continuum corresponds to the lexical time expression "tomorrow", having the highest k-means score of 121. The variants of this time expression are displayed in Figure 18, which is extracted from the ALR (Lafkioui 2007: 214). The classification presented in Figure 17 below substantiates the importance of time lexemes in Rif Berber's geolinguistic classification.

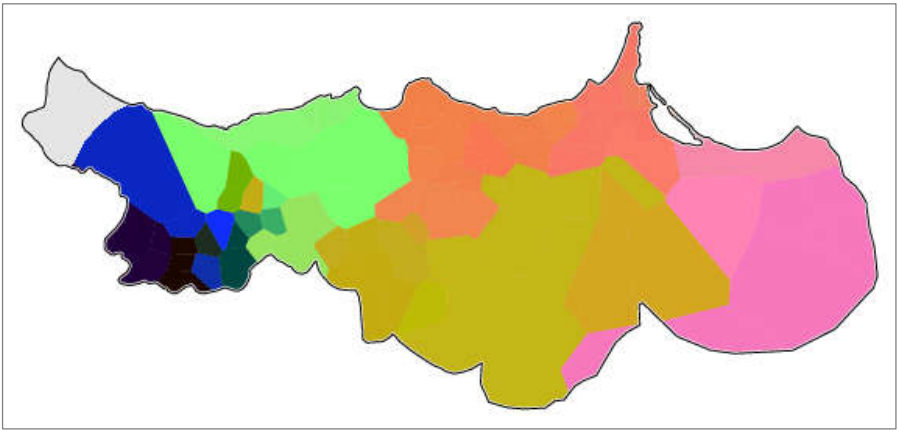

Fig. 17. Classical MDS Map of time lexemes of Rif Berber

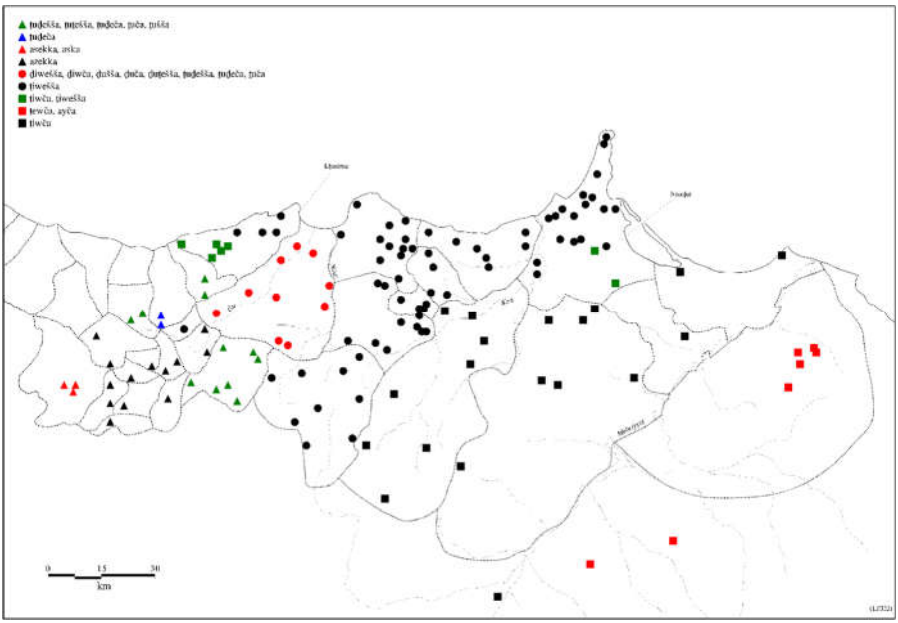

Fig. 18. Lexeme “tomorrow” in Rif Berber (Lafkioui 2007: 214)

7 Discussion

In the light of the data and outcomes presented in the previous sections of this article, Rif Berber accounts for a composite diffusion model involving a hierarchical (gravity model, Trudgill 1974), contra-hierarchical (Wikle & Bailey 1997), and contagion (wave model) diffusion of its linguistic variants. These three diffusion models shape together the geolinguistic patterns of the phonetic, phonological, morphological, syntactic, and lexical phenomena of the Rif Berber

language continuum. As such, a phenomenon like VOC, for instance (Section 4.2), provides abundant evidence for the interaction of these three models in the diffusion of the variants: 1) a hierarchical diffusion from more to less speaker groups and with the city of Alhoceima (WCRB aggregate) as the directing centre; 2) a contra-hierarchical diffusion of especially the innovated EXTVOC variants, steered by the rural area of Imzuren (Ayt Weryaɣel, nr. 17, Figure 3). Accordingly, this diffusion pattern gives counterevidence to Samuels' assumption (1972: 93) that innovations usually emerge in large cities; 3) a contagion diffusion of the VOC and EXTVOC variants in geographically adjacent speech communities, corroborated by the algorithmic classification in Figure 10. This classification clearly indicates the great size of the diffusion area of the VOC transformations – it covers the entire Rif area but the eastern and western periphery – with the Central Rif area (and more precisely the Ayt Weryaɣel) as the focal diffusion area. Daily contact through networks of neighbouring sites is thus important in the diffusion of variants in the Rif area. In addition, certain new and wider linguistic networks have emerged in the area in parallel with the development of certain socio-economic networks, due to increased economic activity (trade and transport) and thanks to the substantial financial investments from Rif Berbers abroad and to the improved local and national transport infrastructure. So, the stronger the socio-economic networks, the more the language practices of Berber speakers are subject to change.

Furthermore, the five core aggregates composing the Rif Berber continuum (Figures 3, 6) cut across administrative and political borders and even across the traditionally used groupings like Senhaja, Rif, and Iznasen (Biarnay, 1917; Renisio 1932). These local ethnonyms do not correspond to the complex sociolinguistic landscape of the Rif area; e.g. the WCRB aggregate (Figure 3) encompasses both varieties of so-called Senhaja Berber (nrs. 10-13) and of what is traditionally considered to be "Rif" Berber (nrs. 14-17). Even if in comparative and historical linguistics these ethnonyms are frequently used as terms referring to groups sharing certain linguistic features, their use is unfortunate when dissociated from their local geolinguistic context and even more when they are used for historical or genealogical subclassifications of the Berber language family, which are generally based on synchronic data, anyway (as in e.g. Kossmann 2020, Naït-Zerrad 2001). As a matter of fact, the Rif Berber area, as well as the other areas of North Africa, have been the subject of many migrations since remote times, which makes it very difficult – as things stand – to accurately identify which ethnic group spoke which kind of variety in which time period. To illustrate this with the case of the Rif area, for example, some well-known migrations are those from Iqelɛiyen (nr. 28, ERB) to Targist (nr. 12, WCRB, Senhaja), from Ibḍalsen (24, ECRB) to Ayt Weryaɣel (nr. 17, WCRB), and from Ayt Weryaɣel to Ktama (nr. 1, WRB); the case of Targist (nr. 12), for instance, is perfectly reflected in the

linguistic data and classifications presented in the article (e.g., small light brown aggregate in WCRB, Figure 6). The continuum nature of the Berber language family – a tightly knit and coherent bloc – makes its subclassification very tricky. Therefore, historical and genealogical classifications of the Berber super-continuum would greatly benefit from combining qualitative linguistic variational and historical approaches with quantitative approaches, for which some promising methods are now available in historical dialectology (e.g., Blaxter 2019). It goes without saying that fieldwork and relating descriptive studies of all Berber-speaking areas of North Africa, as well as comprehensive philological and corpus linguistic studies of the numerous manuscripts that exist in and on Berber are a prerequisite for the success of this venture.

8 Conclusion

Viewed from a qualitative and quantitative perspective, this data-based study has shown that the varieties of Rif Berber form a language continuum according to a composite diffusion model (hierarchical, contra-hierarchical, and contagion diffusion); the following are its five core aggregates: Western Rif Berber, West-Central Rif Berber, Central Rif Berber, East-Central Rif Berber, and Eastern Rif Berber. Moreover, the examination of the aggregate discriminators that resulted from the data mining studies on the cross-level corpus (lexical time expressions, vocalisation, spirantisation and palatalisation, the pronominal system) has demonstrated the complex and gradual nature of language change and diffusion in the Rif area and the importance of combining structural and algorithmic approaches in explaining them.

Abbreviations

1	= first person
2	= second person
3	= third person
AOR	= aorist
AS	= annexed state
DO	= direct object
F	= feminine
FRS	= free state
G	= gender
IMP	= imperative
IO	= indirect object
IPFV	= imperfective
IRR	= irrealis
M	= masculine
NEG	= negation, negator
N	= number

PFV = perfective
PL = plural
P = person
S = subject
SG = singular
TAM = tense-aspect-mood
VENT = ventive
VOC = vocalisation
EXTVOC = extended vocalisation

References

Armas, Jonay A. 2019. Notas sobre la espiratización de */t/ bereber en los guanchismos. *Revista de Filología* 39: 13–70.

Basset, André. 1961. *Textes berbères de l'Aurès (Parler des Aït Frah)*. Paris: Adrien Maisonneuve.

Biarnay, Samuel. 1917. *Étude sur les dialectes berbères du Rif : lexique, textes et notes de phonétique*. Paris: E. Leroux.

Blaxter, Tam. 2019. Ther varom mid j hia. Tracing linguistic diffusion in the history of Norwegian using kernel density estimation. *Dialectologia et Geolinguistica* 27: 5–34.

Boudot-Lamotte, Antoine. 1964. Notes Ethnographiques et Linguistiques Sur Le Parler Berbère de Timimoun. *Journal Asiatique* 252: 487–558.

Brugnatelli, Vermondo. 1993. Quelques particularités des pronoms en berbère du Nord. In *A la croisée des études libyco-berbères. Mélanges offerts à Paulette Galand-Pernet et Lionel Galand*, 229–245, eds. Jeanine Drouin & Arlette Roth. Paris: Geuthner.

Camps Gabriel & Vignet-Zunz Jaques, 1998. Ghomâra. *Encyclopédie berbère* 20: 3110–3119.

Chambers, Jack K. 2013. Patterns of Variation including Change. In *The Handbook of Language Variation and Change*, 297–323, eds. Jack K. Chambers & Natalie Schilling. West-Sussex: Wiley-Blackwell.

Colin G.S., 1929. Le parler berbère des Ghmara. *Hespéris*: 173–208.

El Hannouche, Jamal. 2010. *Arabic influence in Ghomara Berber*. MA thesis, Leiden University.

Escure, Geneviéve. 1977. Hierarchies and Phonological Weakening. *Lingua* 43: 55–64.

Ferguson, Charles. 1978. Phonological processes. In *Universals of Human language*, Volume 2, Phonology, 403–443, eds. Joseph H. Greenberg, Charles A. Ferguson & Edith A. Moravcsik. Stanford: Stanford University Press.

Foley, James. 1977. *Foundations of Theoretical Phonology*. Cambridge: Cambridge University Press.

Galand, Lionel. 2010. *Regards sur le berbère*. Milan: Centro Studi Camito-Semitici.

Guerssel, Mohammed. 1977. Constraints on Phonological Rules. *Linguistic Analysis* 3: 267–305.
Kenstowicz, Michael, and Charles Pyle. 1973. On the Phonological Integraty of Geminate Clusters. In *Proceedings of the Conference On Phonology*. [1971. Urbana, Ill., U.S.A.], 27–43, eds. Michael Kenstowicz and Charles Kisseberth. Berlin/New York: Mouton de Gruyter.
Kossmann, Maarten. 1995. La spirantisation dans les parlers zénètes : aperçu historique. In *Langues du Maroc : aspects linguistiques dans un contexteminoritaire*, 11–19, ed. Petra Bos. Tilburg: Tilburg University Press.
Kossmann, Maarten. 2013. *The Arabic influence on Northern Berber*. Leiden: Brill.
Kossmann, Maarten. 2020. Berber. In *The Oxford Handbook of African Languages*, 281–289, eds. Rainer Vossen & Gerrit J. Dimmendaal. Oxford: OUP.
Lafkioui, Mena B. 2006a. La spirantisation dynamique de la vélaire occlusive simple /k/ dans les variétés berbères du Rif. *Studi Magrebini* 3: 219–28.
Lafkioui, Mena B. 2006b. La vocalisation des alvéolaires /r/ et /rr/ dans les variétés berbères du Rif. In *Etudes berbères: Le nom, le pronom et autres articles*, 175–84, eds. Dymitr Ibriszimow, Rainer Vossen & Harry Stroomer. Köln: Rüdiger Köppe Verlag.
Lafkioui, Mena B. 2007. *Atlas linguistique des variétés berbères du Rif*. Köln: Rüdiger Köppe Verlag.
Lafkioui, Mena B. 2008a. Dialectometry Analysis of Berber Lexis. *Folia Orientalia* 44: 71–88.
Lafkioui Mena B. 2008b. Variation géolinguistique berbère du Rif : le cas des indices personnels verbaux. In *The Amazigh Language at Home and at School. Perspectives on Oral Discourse Structure and Academic Language Skills*, 51–62, ed. A. El Aissati. Köln: Rüdiger Köppe Verlag.
Lafkioui, Mena B. 2011a. How System-Internal Linguistic Factors Indicate Language Change and Diffusion. A Geolinguistic Analysis of Berber Data. *Dialectologia et Geolinguistica* 19: 62–80.
Lafkioui, Mena B. 2011b. *Etudes de la variation et de la structuration linguistiques et sociolinguistiques en berbère du Rif*. Köln: Rüdiger Köppe Verlag.
Lafkioui, Mena B. 2015. Méthodologie de recherche en géolinguistique. *Revue Corpus* 14: 139–64.
Lafkioui, Mena B. 2017. Rif: La Langue (Rifain/Tarifit). *Encyclopédie Berbère* 41: 6916–6956.
Lafkioui, Mena B. 2018a. The Imperfective in Berber: Evidence of Innovated Forms. In *Afroasiatic. Data and Perspectives*, 85–103, ed. Mauro Tosco. Amsterdam: John Benjamins Publishing Company.
Lafkioui, Mena B. 2018b. The Rif Berber Language Continuum: An Algorithmic Geolinguistic Study. In *Woorden om te bewaren. Hulde-Album Voor Jacques Van Keymeulen*, 95-105, eds. Timothy Colleman, Johan De Caluwe, Veronique De Tier, Anne-Sophie Ghyselen, Liesbet Triest, Roxanne Vandenberghe & Ulrike Vogl. Gent: Skribis.
Lafkioui, Mena B. 2018c. Geolinguistic complexity in Berber. Structural and algorithmic perspectives, *Dialectologia et Geolinguistica*, 26: 33–56.

Lafkioui Mena B. 2018d. Berber languages and linguistics. *Oxford Bibliographies*. DOI: 10.1093/OBO/9780199772810-0219.
Lafkioui Mena B. 2020. Dialectology and linguistic geography. In *The Oxford Handbook of African Languages*, 104–124, eds. Rainer Vossen & Gerrit J. Dimmendaal, Oxford: OUP.
Lafkioui Mena B. & Daniela Merolla, 2002. *Contes berbères Chaouis de l'Aurès*. Köln: Rüdiger Köppe Verlag.
Leinonen, Therese. 2010. *An Acoustic Analysis of Vowel Pronunciation in Swedish Dialects*. PhD. Rijksuniversiteit Groningen.
Louali, Naïma. 1999. La spirantisation en berbère. In *Afroasiatica Tergestina*, eds. Marcello Lamberti & Livia Tonelli, 271–298. Padua: Unipress.
Louali Naima & Puech Gilbert. 1998. La partition de l'espace vocalique en berbère rifain. In *Actes des 22èmes Journées d'Études sur la Parole*, 83–85.
MacQueen, James. B. 1967. Some Methods for classification and Analysis of Multivariate Observations. In *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability* 1, 281–297. Berkeley: University of California Press.
Mourigh, Khalid. 2016. *A Grammar of Ghomara Berber (North-West Morocco)*. Köln: Rüdiger Köppe Verlag.
Naït-Zerrad, Kamal. 2001 Esquisse d'une classification linguistique des parlers berbères. *Al-Andalus-Magreb* 8–9: 389–412.
Nerbonne, John, and Peter Kleiweg. 2007. Toward a Dialectological Yardstick. *Journal of Quantitative Linguistics* 14: 148–66.
Nerbonne, John, Peter Kleiweg, Wilbert Heeringa & Manni. 2008. Projecting Dialect Differences to Geography: Bootstrap Clustering vs. Noisy Clustering. In *Data Analysis, Machine Learning, and Applications. Proc. of the 31st Annual Meeting of the German Classification Society*, 647–54, eds. Christine Preisach, Lars Schmidt-Thieme, Hans Burkhardt & Reinhold Decker. Berlin: Springer.
Nerbonne, John, Rinke Colen, Charlotte S. Gooskens, Therese Leinonen & Peter Kleiweg. 2011. Gabmap – A Web Application for Dialectology. *Dialectologia* 2: 65–89.
Prasse Karl-Gottfried, 1972. *Manuel de grammaire touarègue (tahaggart)*. Copenhague: Akademisk Forlag.
Renisio, Amédé. 1932. *Étude sur les dialectes berbères des Beni Iznassen, du Rif et des Senhaja de Sraïr. Grammaire, textes et lexique*. Paris: E. Leroux.
Saib, Jilali. 1974. Gemination and spirantization in Berber: diachrony and synchrony. *Studies in African Linguistics* 5 : 1–26.
Samuels, Michael L. 1972. *Linguistic Evolution, with Special Reference to English*. Cambridge: Cambridge university press.
Schreier, Daniel. 2005. On the loss of preaspiration in Early Middle English. *Transactions of the Philological Society* 103 : 99–112.
Souag, Lameen. 2015. The Development of Dative Agreement in Berber: Beyond Nominal Hierarchies. *Transactions of the Philological Society*, 113 (2): 213–248.

Trudgill, Peter. 1974. Linguistic change and diffusion. Description and explanation in sociolinguistic dialect geography. *Language in society* 3: 215 Akademisk Forlag 246.
Vycichl, Werner. 1975. Begadkefat im Berberischen. In *Hamito-Semitica, Proceedings of a Colloquium held by the Historical Section of the Linguistics Association (GB) at the SOAS*, University of London, March 18–19 1970, 315–317, eds. J. Bynon & Th. Bynon. The Hague: Mouton.
Wikle, Thomas & Guy Bailey. 1997. The spatial diffusion of linguistic features in Oklahoma. In *Proceedings of the Oklahoma Academy of Science* 77: 1–15.