Geolinguistic complexity in Berber. Structural and algorithmic perspectives

Mena B. Lafkioui

Geolinguistic complexity in Berber.
Structural and algorithmic perspectives

Mena B. Lafkioui

Abstract

The present article discusses how the combination of structural (qualitative) and algorithmic (quantitative) perspectives provides valuable insights into geolinguistic patterning and variability, and hence testifies to the importance of the integrating approach in addressing geolinguistic complexity. In doing so, it shows how language is constantly modulated in the form of innovations that emerge in structurally layered and causal formations, dictated by a subtle interplay between system-based and system-external properties. A case that accounts for this kind of geolinguistic complexity is provided by this data-driven study on Berber (Afro-asiatic), which shows how certain phonological and morphological innovation processes triggered by the vocalisation of the liquids /r/, /ṛ/, /rr/ and /ṛṛ/ in Rif Berber (North, Northeast, and Northwest Morocco) create language variation and change. Furthermore, the Berber data examined demonstrate the significant role of certain system-internal factors, such as economy and code conformity, in the diffusion of new phonetic, phonological, and morphological items. In order to better understand the intricacy of the various vocalisation phenomena addressed in the study, the results of the qualitative analysis (synchrony and diachrony) are also contrasted with the algorithmic results ensuing from computing geolinguistic distances by means of the Levenshtein distance calculating method with phone strings tokenised in pair-wise alignments (pondered variables).

Keywords

1 Introduction

This data-driven study examines from both a structural (qualitative) and algorithmic (quantitative) perspective how certain phonological and morphological innovation processes triggered by the vocalisation of the liquids /r/, /ṛ/, /rr/, and /ṛṛ/ in Rif Berber (aka Tarifit; Northern Berber/Afro-asiatic; North, Northeast, and Northwest Morocco) create language variation and change and how these innovations are geolinguistically patterned. From a structural perspective, the study

addresses the synchronic variability and structuring of the vocalisation of these liquids as well as the relating diachronic transformations. This structural part of the research builds further on the results discussed in Lafkioui (2006b, 2011) and focuses on recent developments. These developments are considered in contrast with both longitudinal findings covering more than two decades of research on the field and the findings included in the first grammatical descriptions of the Rif area (Biarnay 1917; Renisio 1932). The algorithmic perspective, on the other hand, deals with the computing of the geolinguistic distances of the vocalisation of the liquids concerned by means of the Levenshtein distance calculating method, with phone strings tokenised in pair-wise alignments.

The combination of structural and algorithmic approaches is not only necessary so as to better understand the intricacy of the linguistic phenomena addressed but also provides valuable insights into the relating geolinguistic patterns and their changeability, and hence testifies to the importance of the "holistic" point of view in addressing geolinguistic complexity (Léonard et al. 2014; O'Sullivan 2004). Moreover the data and analyses will demonstrate that 1) linguistic and social factors interact in the diffusion of language forms and that 2) language change is gradual and non-linear on both an extra-linguistic level (geographical and social variation) and a linguistic level (Rif Berber geolinguistic continuum; cf. Lafkioui 2007, 2008, 2018b).

The data examined come from the Atlas linguistique des variétés berbères du Rif (Lafkioui 2007), which also provides the basic map with its 141 georeferenced points belonging to 32 groups of the geolinguistic area. The data also stem from numerous linguistic, sociolinguistic, and ethnographic fieldwork investigations in the Rif area, which started in 1992 and of which the last one was in autumn 2017.

From a typological perspective, the Berber languages, including Rif Berber, are predominantly synthetic (inflection, derivation and compounding) and inflecting. Common features are their VSO base word order, their pro-drop feature, their preposition-noun sequence, their possessive suffixes and their mixed morphological plural formation (affixation and/or apophony). Aside from noun-verb oppositions, all other word class distinctions are unclear in Berber. The Berber languages also provide accounts of the marked-nominative.

Section 2 of the article will examine vocalisation in Rif Berber from a structural perspective, whereas section 3 will focus on the algorithmic investigation of the relating data. The conclusion will bring up the main findings of the study while highlighting the interrelatedness of the two perspectives discussed.

2 Structural geolinguistic complexity

Due to the vocalisation of the liquids /r/, /ṛ/, /rr/, and /ṛṛ/, Central Rif Berber has developed a set of vowels that are absent in most Berber languages, which merely distinguish the following basic vowels: the closed front vowel /i/ [ɪ] , the closed back vowel /u/ [ʊ], and the open central vowel /a/ [æ] or [ɛ], as well as the mid central vowel [ə], which is usually not regarded as a phoneme but as an epenthetic realisation feature (except in a small number of Berber languages, such as Berber of Jerba in Tunisia). The subsequent sections will account for these vocalisation developments.

2.1 Vocalisation of the tap /r/ and the trill /ṛ/ (VOC)

Vocalisation – henceforth VOC – is a diachronic process which in the Berber varieties of the Central Rif area (cf. Fig. 1: 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 a significant quantitative value due to compensatory lengthening, as is shown in Table 1.

Table 1. 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 [ɑː]
iṛ	ị̄	monophthong [eˤː]
uṛ	ụ̄	monophthong [oˤː]
aṛ	ạ̄	monophthong [ʌˤː]

The vocalised variants displayed in Table 1 are thus the outcome of diachronic phonetic processes, which have reached different stages according to the variety in question. These stages, which may overlap, include the following phones depending on the preceding vowel:

- */ir/   >   [ɪɾ], [ɪ̞], [ɛæˤ], [ɛæː], [ɪː]
- */ur/   >   [ʊɾ], [ʊ̞], [ɔɑˤ], [ɔɑː], [ʊː]
- */ar/   >   [ɛɾ] or [æɾ], [ɛ̞] or [æ̞], [ɛˤˤ] or [æˤˤ], [ɛː] or [æː]
- */ər/   >   [əɾ], [ɛɾ] or [æɾ], [ɛ̞] or [æ̞], [ɛˤˤ] or [æˤˤ], [ɛː] or [æː]
- */iṛ/   >   [erˤ], [e̞ˤ], [eˤˤ], [eˤː]
- */uṛ/   >   [orˤ], [o̞ˤ], [oˤˤ], [oˤː]
- */aṛ/   >   [ʌrˤ], [ʌ̞ˤ], [ʌˤˤ], [ʌˤː]

The long vowels /ī/, /ū/, /ā/ and /ą̄/ belong to the phonological system of Rif Ber-
ber (Lafkioui 2006b, 2007: 29–37, 2011). Note that /ṛ/-VOC is a marginal phe-
nomenon compared to /r/-VOC and that the monophtongs [ɪː] and [ʊː] are con-
fined to the varieties of the Ayt Weryaɣel (cf. dark blue area of Fig.12). 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.

VOC is bound to a general formal restrictive rule, that is, it is allowed in the
syllable coda only, as is demonstrated in examples 1 to 4:

(1)   *īḏən*   ⇐   *irḏən*   (= ir + ḏən)   'wheat'
(2)   *asəkkū*   ⇐   *asəkkur* (= a + sə + kkur)   'partridge (male)'
(3)   *anẓā*   ⇐   *anẓar*   (= an + ẓar)   'rain'
(4)   *aɣāḍa*   ⇐   *aɣərḍa*   (= a + ɣər + ḍa)   'mouse, rat'
(5)   *ṭā̄wa*   ⇐   *ṭarwa*   (= ṭaṛ + wa)   'offspring'

Furthermore, VOC is part of the overall phonetic process of weakening of con-
sonants, 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;
Lafkioui 2006a, 2007: 38–58) and the vocalisation of semi-consonants (Lafkioui
2007: 27–28). Phonetic economy is probably the functional triggering factor be-
hind these transformations.

Given that the VOC of the trill /ṛ/, like in (5), 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 the functional factor of economy has
triggered either processes or that /ṛ/-VOC is a formal development created by
analogy with /r/-VOC.

From a pan-Berber comparative perspective, VOC is an idiosyncrasy, a spec-
ificity of Central Rif Berber, as is shown in Lafkioui (2011) and on the distribu-
tion map in Fig. 1, which pertains to /r/-VOC of the sequence /ər/ of the item
asərdun 'mule'. On this map, only the blue areas vocalise: in the light blue areas,
VOC as well as retention of the tap are attested, whereas in the dark blue areas,

complete VOC is primarily attested. In the green-blue areas, on the other hand, the features [ər] and [ɛɾ] are retrieved. 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], [ħ], and [ʕ].
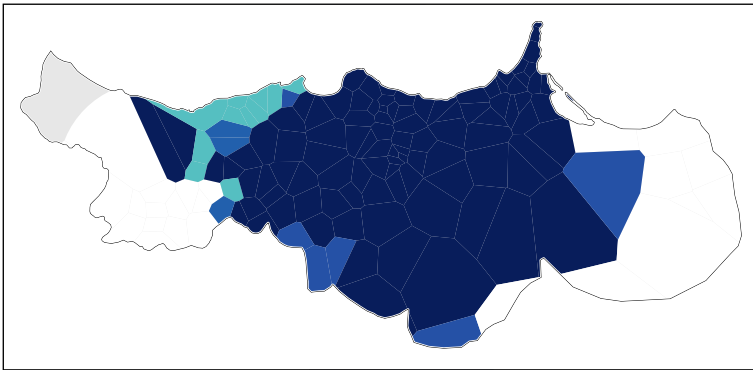

Fig. 1. Distribution map of /r/-VOC of the item *asərdun* 'mule' in the Rif Berber continuum.

Moreover, VOC is characterised by a gradual diffusion in terms of both linguistic and extralinguistic features. Different overlapping variants representing the different phases of vocalisation are gradually diffused over the Rif geographic area according to the lexical item, the language variety, and the speaker in question. A case in point showing at glance the gradual diffusion of VOC is given in Figs. 2 to 6, which concern the same /r/-VOC phenomenon as in Fig. 1 (sequence /ər/ of the item *asərdun* 'mule').
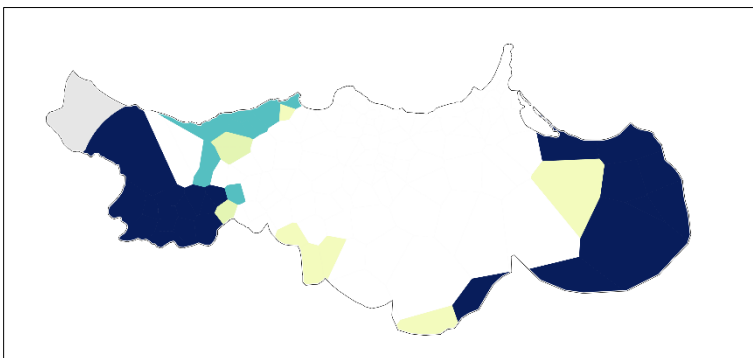

Fig. 2. Distribution map of feature [ər] of the item asərdun 'mule'.
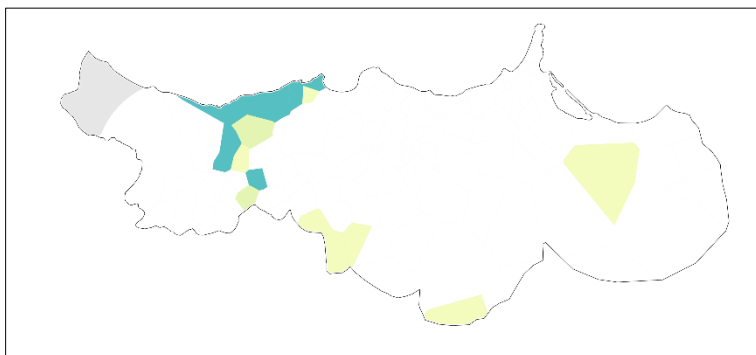
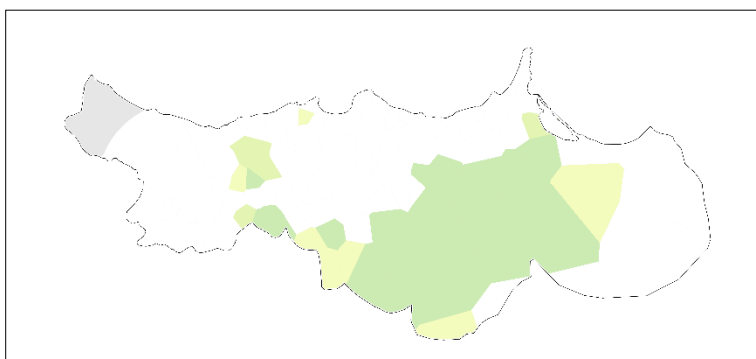Fig. 3. Distribution map of feature [ɛɾ] of the item *asərdun* 'mule'.

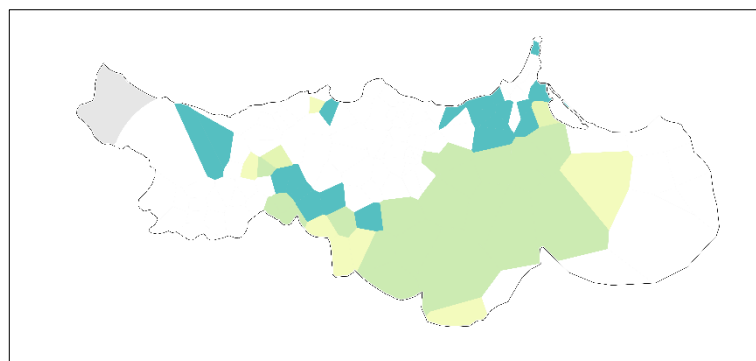Fig. 4. Distribution map of feature [ɛɽ] of the item *asərdun* 'mule'.

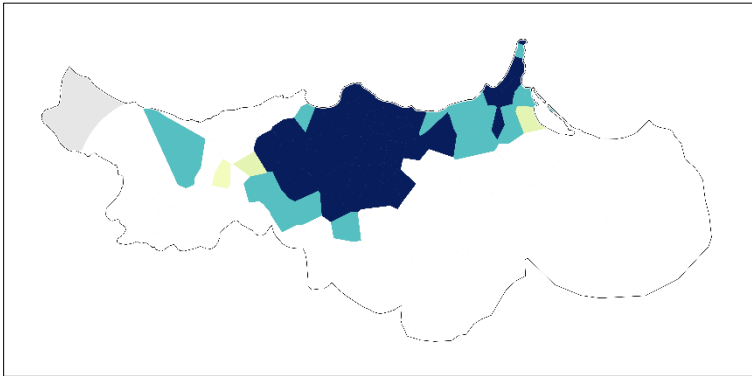Fig. 5. Distribution map of feature [ɛˤr] of the item *asərdun* 'mule'.

Fig. 6. Distribution map of feature [ɛː] of the item *asərdun* 'mule'

Compared to the distribution displayed in Fig. 1, the following distribution map in Fig. 7 accounts for the fact that the degree and type of /r/-VOC are connected to the lexical item in which it occurs and its syllabic and phonetic composition.

For instance, the same sequence /ər/ as in *asərdun* 'mule' (Fig. 1 to 6) has a different distribution when it is part of the item *aɣərḍa* 'mouse, rat' (Fig. 7). The most significant difference concerns the eastern part of the Rif Berber continuum, where the variety of Zayyu does not vocalise for the item *aɣərḍa* 'mouse, rat' (Fig. 7), whereas for the item *asərdun* 'mule' it does, but at an initial stage (Fig. 1). The most western varieties (i.e groups 1-9 of Fig. 8), on the other hand, do not vocalise at all, even though they contain the feature [ɛɾ] along with [əɾ] for the item *aɣərḍa* 'mouse, rat' (green-blue area). In fact, the feature [ɛɾ] is produced here under the influence of the preceding velar [ɣ], which is in accordance with the Berber predilection to lower the articulation of the mid central vowel when it is in contact with consonants articulated at the back of the throat (e.g. pharyngeals, uvulars, velars).
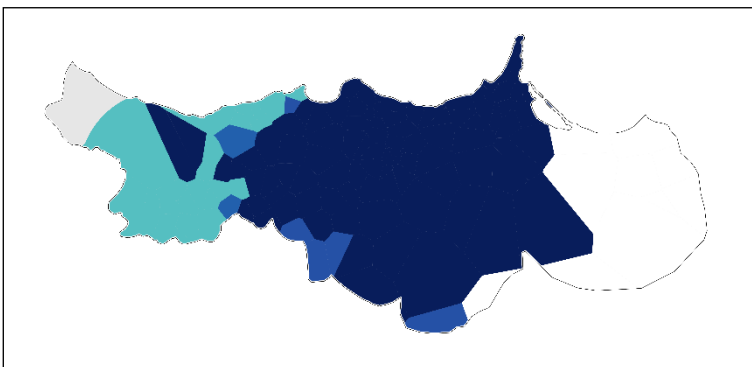

Fig. 7. Distribution map of /r/-VOC and /ə/-lowering of the item *aɣərḍa* 'mouse, rat'.

Consequently, if the items are not carefully analysed and selected at the qualitative level before proceeding to the quantitative analysis, the algorithmic classifications may indicate erroneous geolinguistic correlations.
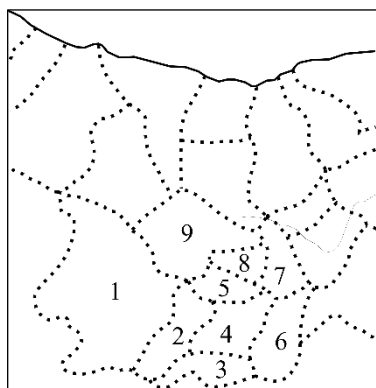

Fig. 8: Map of the Western Rif Berber-speaking groups without VOC and with /ə/-lowering of the item *aɣərḍa* 'mouse, rat'.

1	Ktama
2	Taɣzut
3	Ayt Bušibet
4	Ayt Ḥmed
5	Ayt Bunsar
6	Ayt Bšir
7	Zerqet
8	Ayt Ḥennus
9	Ayt Seddat

Moreover, there is a phonetic gradual and complementary link between the tap /r/ and its vocalised variants: the etymological /r/ can serve as a continuous glide sound when a long vowel precedes another vowel in a quickly articulated utterance, like in 6. In addition, in many vocalising varieties, the semi-vowel /y/ has taken over the glide function of the etymological tap /r/, like in 7, which confirms the phonological status of the long vowels in Rif Berber. The latter case is a formally triggered innovation, wherein the semi-vowel /y/ has been reinterpreted as a glide in the phonetic context [v̄ + v] by analogy with the glide function of /y/ in the context [v + v], as in 8.

(6) *amžā i sɣiɣ* + high cadence ⇒ *amžar i sɣiɣ* 'the sickle that I bought'
(7) *amžā i sɣiɣ* + high cadence ⇒ *amžā y sɣiɣ* 'the sickle that I bought'
(8) *ṭamža i sɣiɣ* + high cadence ⇒ *ṭamža y sɣiɣ* 'the flute that I bought'

2.2 Extended vocalisation (EXTVOC)

Apart from standard vocalisation (VOC), which is a widespread phenomenon in Central Rif Berber (cf. Fig. 1), two "extended" vocalisation types (EXTVOC) are also attested in the area:

1) VOC of the liquids /r/ and /ṛ/ in onset position, i.e. EXTVOC1
2) VOC of the transformed lateral approximant /l/ (/r ⇐ l/ or /r̂/), i.e. EXTVOC2

Since vocalisation is very productive and significantly extending over the Rif geolinguistic area, the term EXTVOC that is used here matches the phenomena examined better than the formerly used term of "alternative vocalisation" (Lafkioui 2006b, 2007: 37, 2011).

2.2.1 *Vocalisation in onset position (EXTVOC1)*

EXTVOC1 occurs in both prevocalic and intervocalic positions, as is shown in examples 9 to 18, and hence has a serious impact on the phonetic, phonological and morphological structure of the lexemes concerned (Lafkioui 2006b, 2007: 37, 2011). This phenomenon is mainly attested in the Ayt Weryaɣel varieties (cf. dark blue area of Fig. 12).

– Prevocalic position:

(9)   *aḇīḏ* [æβɪːð] ⇐ *aḇriḏ* (aḇ + riḏ), [æβrɪˈð] 'way'

(10)   *aġūm* [aɣʊːm] ⇐ *aġrum* (aġ + rum), [aɣrʊˈm] 'bread'

(11)   *amqqān* [æmqqaːn] ⇐ *amqqran* (am + qqran) [æmqqraˈn̩] 'big'

– Intervocalic position:

(12)   *īy* [ɪːj], [ɛæːj] or *yīy* [jɪːj], [jɛæːj̩] ⇐ *iri* (i + ri), [ɪrɪ] 'neck'

(13)   *āy* [aːj] or *yāy* [jaːj̩] ⇐ *ari* (a + ri), [arɪ] 'to write'

(14)   *ḏiġūy* [ðɪɣʊːj] ⇐ *ḏiġuri* (ḏi + ġu + ri), [ðɪɣʊrɪ] 'study'

– /r/-VOC in intervocalic/onset position:

(15)   *āyṯi* [æːjθɪ] ⇐ *ariṯi* (a + ri + ṯi), [ærɪθɪ] 'dough'

– /ṛ/-VOC in intervocalic/onset position:

(16)   *dṵ̄(w)* [doˤː(w)] ⇐ *duṛu* (du + ṛu), [dorˤo] 'duro', 'coin'

(17)   *ā̰wmi* [ʌˤːwmɪ] ⇐ *aṛumi* (a + ṛu + mi), [arˤʊmɪ] 'christian'

An interesting recent finding is that the long vowel /ṵ̄/ in *dṵ̄* [doˤː] tends to be significantly shortened in the variety of Imzuren (city), sometimes even into *dṵ* [doˤ].

Cases of double vocalisation, combining EXTVOC1 and /r/-VOC, do also occur in this region. A case in point is given in (18):

(18)   *aʕū* [æʕʊː] or [æʕɔaː] ⇐ *aʕrur* (aʕ + rur) [æʕrʊr] 'back'

EXTVOC1 provides thus for syllabic restructuring of the lexical item, with a reduction of the number of syllables, in particular when occurring in intervocalic position (as in 12 to 17). In the latter case, EXTVOC1 produces an extension of the first vowel and a disyllabification of the second vowel, which is a formal adaptation strategy of two possible innovation types: 1) a functionally triggered innovation in which the economy principle is driven to extremes or 2) a formally triggered innovation in which analogy is made between EXTVOC1 and VOC.

Another interesting finding is that even proper names may be affected by EXTVOC1 lately, which indicates that this phenomenon is very productive in the

area. An example is presented in (19), which forms a perfect distinctive opposition with (20):

(19) *fāyda* [fɛːjdɛ] ⇐ *farida* (fa + ri + da), [fɛɾɪdɛ]

(20) *fayda* [fɛjdɛ] 'utility' (loanword from Arabic **fāʔida*)

In addition, EXTVOC1 is bound to the formal restrictive sub-rule which excludes its instantiation in an absoluter Anlaut position before a full vowel, like in example 21. But my recent investigations in the area indicate a shift to disregarding this rule in certain Ayt Weryaɣel varieties, especially in the rural areas surrounding Imzuren (cf. area indicated by a star in Fig. 12), like in example (22).

(21) **āža* ⇐ *raža* [ɾaʒa] 'to wait'

(22) *āža* [aːʒa] ⇐ *raža* [ɾaʒa] 'to wait'

> ➢ EXTVOC1a [*absoluter Anlaut] ⇒ EXTVOC1b [absoluter Anlaut]

As a result, EXTVOC1 has been recently developed into two subtypes, EXTVOC1a, which does not allow for VOC in absoluter Anlaut position before a full vowel and EXTVOC1b, which does to a variable degree.

EXTVOC1a also includes VOC in absoluter Anlaut position before the mid central vowel [ə], as in (23), or before a consonant, as in (24); both phenomena are regularly attested in loanwords from Arabic.

(23) *ạbbi* [ʌˤːbbɪ] ⇐ *ṛəbbi* (ṛə + bbi), [rˤəbbɪ] 'God'

(24) *āḇəḥ* [æːβəħ] ⇐ *rḇəḥ* (monosyllabic), [rβəħ] 'to win'

In example (23), the sequence /ṛə/ undergoes a permutation under the influence of the sonority rule before vocalisation may occur: /cṛe/ ⇒ /ceṛ/ ⇒ /ạ/. In example (24), on the other hand, /ə/ is inserted before /r/ so as to undergo EXTVOC1a.

Furthermore, EXTVOC1b is a typologically interesting phenomenon, as it occurs in absoluter Anlaut position (before a full vowel), which is considered to be a strong consonant position and therefore less prone to vocalisation (Escure 1977; Foley 1977), and it concerns the VOC of a tap corresponding to the first radical of the base.

2.2.2 *Vocalisation of the mutated liquid /l/ (EXTVOC2)*

EXTVOC1 is in co-variation with another diachronic phenomenon which regularly occurs in the Central Rif area, i.e. the phonetic mutation of the lateral approximant /l/ into (Lafkioui 2007: 69–71, 2011):

a) /r ⇐ l/ = 1) a voiced tap [r] with an ultra light friction or 2) a voiced trill [r]

b) /r̂/   = fricative [ɹ]

The difference between the two taps is difficult to perceive as such, but becomes more apparent when vowels precede them: /v/ + /r/ ⇐ l/ ≠ /v/ + /r/; only the original tap /r/ has a lowering effect on the vowels that it follows:

(25)   *ari* ⇐ *ali* 'to go up': [ɛɾɪ] ≠
(26)   *ari* ⇐ *ari* 'to write': [aɾɪ]

These consonant mutations are in direct correlation with the *phonetic restriction* that excludes the vocalisation of all instantiations derived from /l/. The varieties of the Ayt Weryaɣel and certain adjacent varieties, however, form an exception to this formal restrictive rule, as the mutated lateral can indeed be vocalised here, following a course of evolution analogous to the original tap /r/, with long diphthongs as maximum transformations. This innovation makes up EXTVOC2, which is illustrated in (27).

Examples of EXTVOC2; Ayt Weryaɣel (West-Central Rif):

(27a) /irs ([ɪrs], [ɪrs]) ⇐ ils/ + /r/-VOC ⇒ [ɛæːs], [jɛæːs], and [ɪjɛæːs] 'tongue'
(27b) /irs ([ɪrs], [ɪrs]) ⇐ ils/ + /r/-retention ⇒ [ɪrs], [ɪrs] 'tongue'

Examples of blocking rule of EXTVOC2, which stem from Iqelʕiyen (East-Central Rif), where both /r/ and /r̂/ are retrieved:

(27c) /irs ⇐ ils/ + /r/-retention ⇒ irs ([ɪrs], [ɪrs] 'tongue')
(27d) /ir̂s ⇐ ils/ + /r̂/-retention ⇒ ir̂s ([ɪɭs] 'tongue')

Recent developments indicate that EXTVOC2 also takes place in intervocalic position, as in (28), which is the vocalised variant of (25) (*ali* 'to go up') and which forms a distinctive opposition with the vocalised variant of (26) (*ari* 'to write'), as is shown in (29):

(28)   *āy* [ɛːj] or *yāy* [jɛːj] ⇐ *ari* [ɛɾɪ] 'to go up' (< *ali*)
(29)   *āy* [aːj] or *yāy* [jaːj] ⇐ *ari* [aɾɪ] 'to write' (< *ari*)

The prolific diffusion of EXTVOC2 in this area is reflected in the wide range of cases that it affects, which lately also include proper names, which are usually not concerned by this phenomenon or by any consonant mutation of the lateral /l/. A well attested case is displayed in (30), which testifies to intervocalic vocalisation as well.

(30)   *māyka* [mɛːjkɛ] ⇐ *malika* [mɛlɪkɛ] (*[mɛrɪkɛ])

2.3 Vocalisation of the geminate trill /rr/

VOC of the geminate trill /rr/ does not necessarily go in parallel with that of the simple tap /r/ (Lafkioui 2006b, 2007: 34), as is demonstrated in (31), which lists the different attested developments:

(31)   *ərr* 'give back'

> ➢ Retention of the geminate trill *ərr* ([ɛrr])
> ➢ Qualitative timbre change *arr* ([ɛrr], [ærr])
> ➢ Qualitative and quantitative timbre change, *ārr* ([ɛːrr], [æːrr])
> ➢ Compensatory lengthening + simplification of articulatory tension, *ār* ([ɛːɾ], [æːɾ])

Note that the tap [ɾ] changes its articulation mode and becomes the trill [rr] when geminated. It is also worth mentioning that VOC of the trill in Rif Berber is a remarkable and typologically uncommon phenomenon because it allows to apply a phonetic rule (i.e. vocalisation) to a part of the geminate (first part), and consequently invalidates the phonetic generalisations (Integrity and Inalterability) which claim the opposite (such as in Kenstowicz & Pyle 1973; Guerssel 1977).

Moreover, the geminate trill /rr/ has an intricate relationship with important morphological declination and derivation patterns. For instance, the opposition [simple or vocalised tap /r/] versus [long or reduced trill /rr/] is used for TAM marking (cf. section 3) and for derivation (e.g. noun verbs and plural formation; Lafkioui 2011).

3 Vocalisation-induced morphological changes

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 (2011). In this section, I will briefly present some of these remarkable morphological innovations with respect to the verb system, so as to give a general idea of the degree of linguistic complexity attested in Rif Berber. The Rif Berber verb system, like all other Berber verbal systems, is of a Root-Pattern type, structured around a complex aspectual hierarchical configuration, which is indicated by both consonant length and intraradical vowel alternation.

Example:

(32) *y-swa* [jəswɛ] 'he has drank' (3MSG-PFV) = PM *y-* (3MSG) + base *-swa* [swɛ]
      (root *sw* action 'to drink' + pattern /--a/)

The first major morphological innovation triggered by VOC concerns the extension of the verbal paradigms of the type /ccv/, /vcc/ and /cvc/ at the expense of the predominant root /ccc/, like in (33) to (35):

(33) ccc + vocalisation ⇒ ccv; *ḥzr* [xzər] ⇒ *ḥzā* [xzaː] 'to look'
(34) ccc + vocalisation ⇒ vcc; *rẓm* [rzˤəm] ⇒ *āẓm* [aːzˤəm] 'to open'
(35) ccc + vocalisation ⇒ cvc; *ḵrf* [çrəf] ⇒ *ḵāf* [çɛːf] 'to tie'

The second major morphological innovation instigated by VOC pertains to the reorganisation of the paradigmatic structure of the verb, including the Berber fundamental aspectual opposition of the perfective versus the imperfective for the positive aspects and that of the perfective versus the negative perfective for the negative aspects.

As regards the opposition PFV-IPFV, a diversification of the morphological markers for /crc/-verbs is attested all over the vocalising Rif area. This morphological variation results from the different solutions invented so as to counterbalance the eroded aspect markers, as we can see in Table 2:

Table 2. Perfective ~ Imperfective Markers - *ḵrf* 'to tie'

PFV	IPFV
/r/ - *ḵrf*	/rr/ - *ḵrrf*
/r/-VOC - *ḵāf*	/ār/ - *ḵārf* /ārr/ - *ḵārrf* /(t)t-/ - *tḵāf* /(t)t-/+ /ār/ - *tḵārf* /(t)t-/+ /ārr/ - *tḵārrf*

So, new variants are engendered as compensatory adjustments for the eroded morphological configurations of the PFV and the IPFV by means of the following procedures: 1) the creation of new morphemes resulting from the VOC process (/r/-VOC for PFV; /ār/ and /ārr/ for IPFV); 2) prefixing the existing IPFV marker /t-/ or /tt-/ to the base of the PFV without any changes; 3) a combination of both developments (/t-/ or /tt-/ + /ār/ and /ārr/ for IPFV). The structural adequacy of these specific and central morphophonemic oppositions has been a significant factor in their general diffusion in the vocalising Rif areas, as is demonstrated in Lafkioui (2011).

Regarding the opposition PFV-NPFV, verbs with a vocalised tap /r/ as second or third consonant display for the NPFV the base /cv̄c/ or /ccv̄/ with /v̄/ as the same long vowel as in the PFV, like in *ḵāf* 'to tie' (35), for instance, which stands for both PFV and NPFV, while the expected respective form *ḵīf* 'to tie' is not attested. The NPFV of /ccc/-verbs in Berber generally has /ccic/ as its base, e.g. *ḵrif* 'to tie'. Consequently, the bases /cv̄c/ and /ccv̄/ of the PFV have been formally re-analysed as bases of the NPFV, with /v̄/ as a long invariable vowel. Their widespread diffusion, corresponding to the whole dark blue area on the map in Fig. 1, correlates with the systematisation of the new variants with a morphologically unmarked base for the NPFV. The unmarked nature of the variants has probably facilitated their diffusion. However, this diffusion is not wholesale; it is rather gradual and relates to the combinatorial opposition conditions shown in Table 3, which is based on the verb *mžr* 'harvest':

Table 3. Combinatorial oppositions for PFV-NPFV for the verb *mžr* 'harvest'

mžā (PFV)	*ur* (NEG) + *mžā* (NPFV, -mark)	optional frequent opposition in varieties with complete VOC (focal area)
mžā (PFV)	*ur* (NEG) + *mžir* (NPFV, +mark)	optional frequent opposition in all VOC areas
mžr (PFV)	*ur* (NEG) + *mžā* (NPFV, -mark)	optional frequent opposition in VOC areas
mžr (PFV)	*ur* (NEG) + *mžir* (NPFV, +mark)	obligatory opposition in areas without VOC

4 Algorithmic geolinguistic complexity

This part of the study builds upon the methods and results obtained from the algorithmic analyses and classifications discussed in Lafkioui (2008, 2009, Forthcoming), which provide evidence for the validity of the Levenshtein distance calculating method, especially when the phone strings are tokenised in pairwise alignments. Among the many techniques to visualise aggregate distances, the Multi Dimensional Scaling technique – MDS henceforth – was proven to be the best suited technique 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. In addition, it is one of the most stable techniques, compared to classical clustering, for instance (Lafkioui 2008, 2018b; Nerbonne et al. 2011).

In this study, these techniques will be further used and accommodated to the Berber data. They are based on Kleiweg's free software tools (http://www.let. rug.nl/kleiweg/ L04/) as well as on the more recent web application GABMAP

(Nerbonne et al. 2011), in addition to the data conversion programmes developed for this purpose and for which I am grateful to Bart Cocquyt for his assistance. The adaptive conversions also include a systematic conversion to UTF-8 for the geolinguistic data (including the data of the Atlas) and to KML for the geographic data (http://www.opengeospatial.org/stand-ards/kml/), which was a time and energy consuming task.

As for the digital data that are compared and classified in this study, they consist of 31 phonetic items regarding vocalisation. The data sample amounts to 16278 tokens and to 26 unique tokens, which are plotted on a selection of 141 georeferenced points. The data and map stem from a vast geolinguistic corpus built by means of specific methodological procedures concerning data gathering, their systematisation, and their archiving (Lafkioui 2007, 2015).

The tokenized and pair wise aligned phonetic data used for this research meet the criteria for good quality data, as is indicated by the two relating measures: Cronbach's α, which has a value of 0.98 here, taking into account that the closer to 1 the better the score, with a minimum of 0.7. As for the local incoherence measure, the data has a value of 2.57 (the closer to 0 the better the score), while, for instance, in the study of Nerbonne and Kleiweg (2007), which aims at providing a yardstick for dialectology research, values range from 1.75 to 2.05.

The Classical MDS map (with r= 0.98) displayed in Fig. 9 aggregates the phonetic differences concerning all vocalisation types in Rif Berber – including /r/-VOC, /r/-EXTVOC, /ṛ/-VOC, and /rr/-VOC – as is measured between the various geolinguistic sites considered. It shows 3 main aggregates, corresponding to the 3 main colour themes. Apart from the white zones at the edges of the continuum, all zones vocalise at some extent. The light shaded areas vocalise to a lesser degree and still have the unvocalised tap and trill, whereas the darker areas vocalise extensively; the dark blue ones also provide evidence of complete vocalised variants. The royal blue area – the Ayt Weryaɣel and its adjacent varieties – even account for a systematic use of the extended vocalisation processes.

As one can also distinguish on this map, the diffusion of the vocalisation phenomena is wavelike. This means that there is a *focal area* – the dark blue area – where the phonetic change originated and is nowadays regular. There are also *transition areas* – light shades of blue, yellow and green – where the change is less regular and which still have the unvocalised tap and trill. And there are also *relic areas* – white coloured areas – in which vocalisation is totally absent. This matches the 'diffusion model' of Chambers and Trudgill (1998: chap. 11), which is not that surprising since the Rif is a geographic area with a long settlement history.
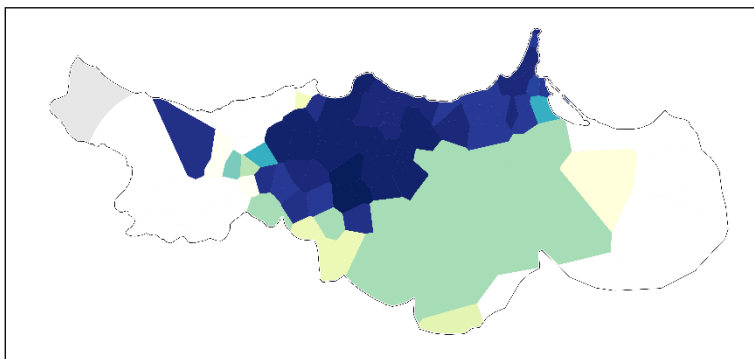
Fig. 9. Classical MDS map of all vocalisation types in Rif Berber.

The algorithmic classification projected onto the map in Fig. 9 also confirms Laf-
kioui's previous qualitative results (2011), which indicate that varieties with the
highest phonetic diversity for the instantiation of the tap /r/ are mostly located in
contact areas, which correspond to the light green and blue shaded varieties on
the map. These varieties are generally part of border zones (e.g. Ayt Ḥdifa of the
Ayt Weryaɣel, West-Central Rif; nr. 42 in Fig. 10), transit areas (e.g. Ayt ʕazza
of Targist, West Rif; nr. 19 in Fig. 10), and socio-economic centres (Imezzužen
of Ayt Naḍur-Iqelʕiyyen, East-Central Rif; nr. 131 in Fig. 10). The difference
map of /r/-VOC of the sequence /ər/ shown in Fig. 10 corroborates this distribu-
tion. Note that the darker the colour of the links, the higher the phonetic differ-
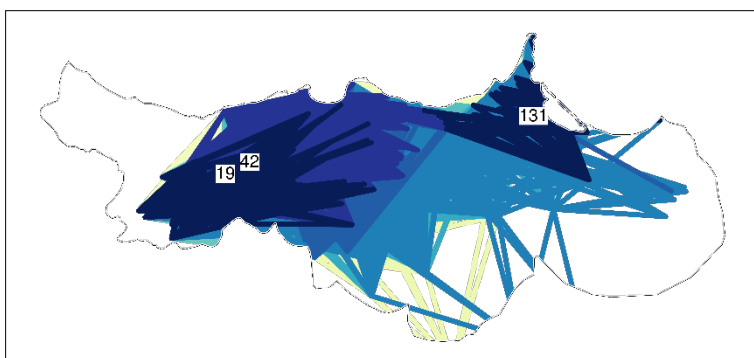ence attested.

Fig 10. Difference map of /r/-VOC of the sequence /ər/ in Rif Berber.

The map in Fig. 10 also provides an example of the numerous results obtained from the systematic data mining analysis completed by means of two quantitative measures of GABMAP, that is, representativeness and distinctiveness. This type of quantitative examination of the data is important, as it allows to objectively identifying which items primarily determine the formation of the different aggregates. Its outcomes support once more Lafkioui's qualitative findings (2007, 2011) and indicate the following features as the primary aggregate discriminators:

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

Given that vocalisation occurs in one large region, with the Central Rif 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 (Cf. dark blue area in Fig. 1). In addition, the great size of the diffusion area also points to social acceptance.

The obtained aggregates represented in Fig. 9 are very stable, as is substantiated by the probabilistic clustering analysis, 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 dendrogram displayed in Fig. 11 represents a clustering with a noise level of 0.2 added.

The most stable clusters in the dendrogram correspond to the areas where vocalisation is commonly attested, which proves that the innovated vowels were most likely engendered in a particular location from which several variants were distributed. The location in question is probably the region wherein vocalisation has been extensively developed and diffused, that is, the Ayt Weryaɣel region (cf. dark blue area in Fig. 12), as is also demonstrated in the structural part of the article (with the various EXTVOC phenomena) as well as in Lafkioui (2011). The reference point map in Fig. 12, which has Imzuren as its reference point, testifies to this development.

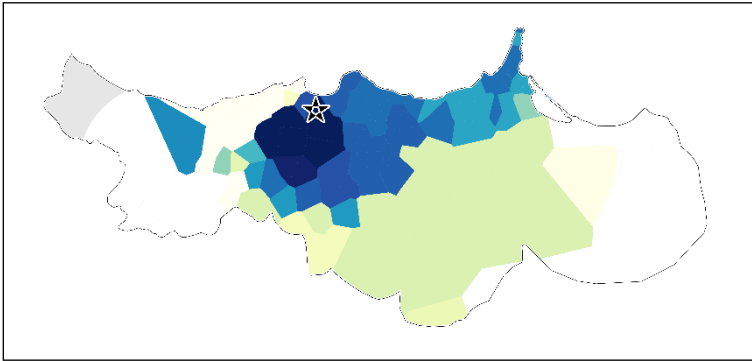Fig. 11: Probabilistic dendrogram of all vocalisation types in Rif Berber.

Fig. 12: Reference point map of all vocalisation types in Rif Berber – Imzuren.

Furthermore, the Ayt Weryaɣel region is not only one 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 socio-cultural yardstick until today. This region even plays a central role in current political and ideological debates concerning the Berber claim for linguistic and cultural recognition. Its capital city, Alhoceima, is also the socioeconomic centre of a large part of the Rif area, mainly corresponding to its central-western and western part.

Interestingly, EXTVOC2 phenomena, which concern the vocalisation of the transformed lateral, are not systematically attested in the city of Alhoceima, at least not among its "native" citizens, but rather in the rural areas surrounding it. The first systematic occurrences of EXTVOC2 that I have observed go back to 20 years ago and were chiefly attested in the rural areas adjoining the small socio-economic centres of Ayt Buʕiyyaš and Imzuren, both belonging to the Ayt Weryaɣel group. While the village of Ayt Buʕiyyaš has not made any significant socioeconomic progress since that time, the village of Imzuren, on the other hand, has developed into an important town, which nowadays functions as a directing centre in this rural area. Its socioeconomic success is due to different factors that interact, among which the following key ones:

1) Its recent increased economic activity (trade and transport) thanks to the substantial financial investments from Rif Berbers abroad and to the recently improved transport infrastructure, whose construction started after the serious earthquake in 2004.

2) Its central geographical location in the area, which facilitates communicative exchanges, including trade and socio-cultural dissemination. It should be

mentioned that local schools form one of the main contact and diffusion centres in the area. Indeed, Imzuren has an important boarding school (primary and secondary levels), which receives pupils from all over the Alhoceima province, and hence plays a pivotal role in the rapid social extension of its language features.

My various linguistic and sociolinguistic longitudinal surveys in this area indicate that many of the EXTVOC phenomena have been created among children and teenagers from this region, who until recently were reprimanded by adults for their non-normative language use, especially regarding EXTVOC2. Meanwhile, several generations of speakers, even the grandparents, of most varieties of the Ayt Weryaɣel and several adjacent varieties have adopted these language practices, though not always for the same lexemes and to the same degree. Once more, gradual diffusion is at work here.

Daily contact is thus important in the spread of vocalisation phenomena, which is diffused through networks of neighbouring sites in the Rif area, as is displayed on the difference map of adjacent sites in Fig. 13, representing all vocalisation types attested in Rif Berber.
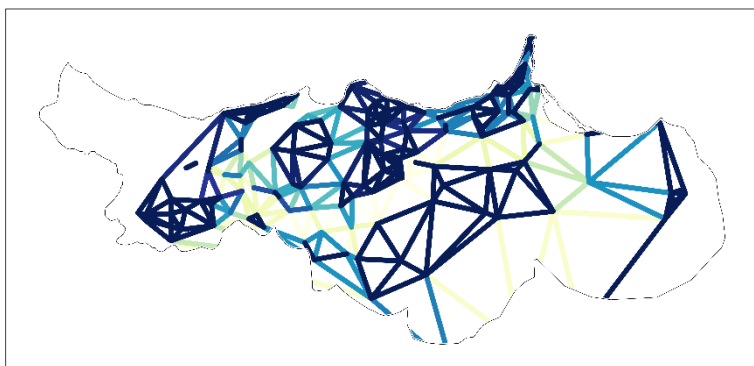

Fig. 13: Difference map of adjacent sites of all vocalisation types in Rif Berber.

As shown on this map, several networks are formed and spread over the entire Rif Berber continuum. The darker the colour of the lines, the more alike are the linguistic sites they connect and hence the stronger the linguistic networks they constitute and the more frequent and intense the contacts between the Berber speakers involved.

Surely, 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, the functional factors of economy and code conformity also contribute to the transmission of the vocalised variants (cf. section

2). These linguistic factors may explain the absolute preference in the focal area for complete vocalised variants, since these frequent forms cannot be directly associated with specific social functions and patterns, such as status attribution.

In fact, most of the EXTVOC phenomena, 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. Rif Berber-speakers who use the EXTVOC variants in their verbal interactions are well aware of the language representations and attitudes regarding these variants. 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.

Furthermore, it is remarkable to observe that there is, since very recently, a generational distinction with respect to the language attitudes towards EXTVOC 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.

In addition, as demonstrated in the structural part of the article (cf. section 2), EXTVOC1 and EXVOC2 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 (onset-vocalisation) in particular has a remarkable success in this region. Code conformity is a probable functional explanation for this trend, along with the system-external factors previously discussed here.

However, in some recent cases of EXTVOC1, formal factors also play an important role in their development and diffusion. 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

parents. Accordingly, it is reasonable to infer that the recent EXTVOC innovations are system-based. In other words, their creation and diffusion is formally motivated (extension by analogy), just as with the older alternative vocalisations *aḇīḏ* and *amqqān* (functional and/or formal trigger).

5 Conclusion

This study on vocalisation in Rif Berber has demonstrated from both a qualitative and quantitative perspective that language is continuously transformed by means of innovations that are engendered in structurally layered and causal formations, directed and diffused by a subtle interaction between system-based properties (i.e. linguistic factors) and speaker-based properties (i.e. extralinguistic factors). Moreover, the vocalisation phenomena discussed here account for the complex and gradual nature of language change and diffusion, on a linguistic level, i.e. phonetic, phonological, morphological and lexical variation pertaining to vocalisation, as well as on an extra-linguistic level, i.e. geographical and social variation. 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.

Abbreviations

3MS	=	third masculine singular
IPFV	=	imperfective
NEG	=	negation, negator
NPFV	=	negative perfective
PFV	=	perfective
PM	=	personal marker
VOC	=	vocalisation
EXTVOC	=	extended vocalisation

References

Biarnay, Samuel. 1917. *Étude sur les dialectes berbères du Rif : lexique, textes et notes de phonétique* (Publications de la Faculté des lettres d'Alger 54). Paris: E. Leroux.

Boudot-Lamotte, Antoine. 1964. "Notes éthnographiques et linguistiques sur le parler Berbère de Timimoun". *Journal Asiatique* 252: 487–558.

Chambers, Jack K. & Peter Trudgill. 1998. *Dialectology*. 2nd ed. (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.

Escure, Geneviéve. 1977. "Hierarchies and phonological weakening". *Lingua* 43: 55–64.

Foley, James. 1977. *Foundations of theoretical phonology*. Cambridge: Cambridge University Press.

Guerssel, Mohammed. 1977. "Constraints on phonological rules." *Linguistic analysis* 3: 267–305.

Kenstowicz, Michael & 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 & Charles Kisseberth (Janua Linguarum). Berlin/New York: Mouton de Gruyter.

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*, eds. Dymitr Ibriszimow, Rainer Vossen & Harry Stroomer, 175–184 (Berber Studies 14). Köln: Rüdiger Köppe.

Lafkioui, Mena B. 2007. *Atlas linguistique des variétés berbères du Rif*. Köln: Rüdiger Köppe.

Lafkioui, Mena B. 2008. "Dialectometry analysis of Berber lexis." *Folia Orientalia* 44: 71–88.

Lafkioui, Mena B. 2009. "Analyses dialectométriques du lexique Berbère du Rif." In *Studien zur Berberologie/Etudes Berbères*, 133–150, eds. Dymitr Ibriszimow, Rainer Vossen & Harry Stroomer (Berber Studies 4). Köln: Rüdiger Köppe..

Lafkioui, Mena B. 2011. "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. 2015. "Méthodologie de recherche en géolinguistique." *Revue Corpus* 14: 139–164.

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 ge-olinguistic study." In *Hulde-Album voor Jacques Van Keymeulen.*, eds. Timothy Colleman, Johan De Caluwe, Veronique De Tier, Anne-Sophie Ghyselen, Liesbet Triest, Roxanne Vandenberghe & Ulrike Vogl. Gent: Skribis.

Léonard, Jean Léo, Gilles Polian, Els Heinsalu & Marco Patriarca. 2014. "Variación dialectal de la morfología tseltal (maya occidental) en los ámbitos morfológico, fonológico y léxico: Un enfoque holístico del diasistema." In *Patterns in Mesoamerican Morphology*, 280–303, eds. J. L. Léonard & A. Kihm. Paris: Michel Houdiard.

Nerbonne, John, Rinke Colen, Charlotte S. Gooskens, Therese Leinonen & Peter Kleiweg. 2011. "Gabmap – a web application for dialectology." *Dialectologia* 2: 65–89.

Nerbonne, John & Peter Kleiweg. 2007. "Toward a dialectological yardstick." *Journal of Quantitative Linguistics* 14: 148–166.

Nerbonne, John, Peter Kleiweg, Wilbert Heeringa & Franz 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*, 647–654, eds. Christine Preisach, Lars Schmidt-Thieme, Hans Burkhardt & Reinhold Decker. Berlin: Springer.

O'Sullivan, David. 2004. "Complexity science and human geography." *Transactions of the Institute of British Geography* 29: 282–295.

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.

Samuels, Michael L. 1972. *Linguistic evolution, with special reference to English.* Cambridge: Cambridge University Press.

Mena B. Lafkioui · Université Sorbonne Paris Cité/LLACAN-UMR8135 · FRANCE
Mena.Lafkioui@cnrs.fr