

Etudes de géographie linguistique berbère du Rif: variation géolinguistique et classification dialectométrique*

par

Mena LAFKIOUI**

MOTS-CLES. — Géographie linguistique; *Tarifit*; Dialectologie berbère; Grammaire comparative berbère; Dialectométrie.

RESUME. — Notre article se veut une présentation de deux études de géographie linguistique berbère reliées: l'examen dialectologique du phénomène de la vocalisation des liquides /t/ et /t'/ dans les variétés berbères du Rif (Maroc du nord) et l'étude de classification dialectométrique de ces variétés moyennant diverses techniques automatisées. Cette contribution contient plusieurs cartes géolinguistiques résultant des études respectives.

TREFWOORDEN. — Linguïstische geografie; *Tarifit*; Berber dialectologie; Berber comparative grammatica; Dialectometrie.

SAMENVATTING. — *Taalgeografische studies van het Rif-Berber: geolinguïstische variatie en dialectometrische classificatie*. — Dit artikel is een presentatie van twee gerelateerde studies over Berber taalgeografie: een dialectologiestudie van het vocalisatiefenomeen van de liquida /t/ en /t'/ in de Berbervariëteiten van het Rifgebied (Noord-Marokko) en een studie over de dialectometrische classificatie van deze taalvariëteiten door middel van verschillende computergestuurde technieken. Deze bijdrage bevat meerdere geolinguïstische kaarten die het resultaat zijn van de respectievelijke studies.

KEYWORDS. — Linguistic Geography; *Tarifit*; Berber Dialectology; Berber Comparative Grammar; Dialectometry.

SUMMARY. — *Linguistic Geography Studies on Rif-Berber: Geolinguistic Variation and Dialectometric Classification*. — This paper presents two related studies of Berber linguistic geography: the dialectological study of the vocalization of the liquids /t/ and /t'/ in the Berber varieties of the Rif area (North Morocco) and the dialectometrical classification study of these varieties based on different computational techniques. This contribution contains several geolinguistic maps which are the result of both studies.

* Communication présentée à la séance de la Classe des Sciences morales et politiques tenue le 21 novembre 2006. Décision de publication prise le 17 octobre 2008. Texte définitif reçu le 21 octobre 2008.

** Universiteit Gent, Afrikaanse Talen en Culturen, Rozier 44, B-9000 Gent (Belgique); Università di Milano-Bicocca, Dipartimento di Scienze Umane per la Formazione «Riccardo Massa», Piazza dell'Ateneo Nuovo 1, Milan (Italie).

1. Introduction

Cet article, réalisé dans le cadre d'une recherche postdoctorale du FWO (*Fonds voor Wetenschappelijk Onderzoek-Vlaanderen*), présentera deux études de géolinguistique berbère reliées. La première est un examen dialectologique du phénomène de la vocalisation des liquides /r/ et /ʀ/ en tarifit (point 2). Elle permet d'exposer la variation synchronique accusée des formes vocalisées, et ce, en rapport avec leur diffusion géographique particulière sur le terrain rifain, ainsi que de dégager les processus diachroniques responsables de ces renouvellements vocaliques. La seconde étude portera sur la classification dialectométrique du lexique rifain suivant des méthodes assistées par ordinateur (point 3). Cette contribution a pour objectif principal de montrer comment la recherche sur la variation linguistique peut servir à la fois la dialectologie — en l'occurrence intégrée dans une perspective de géographie linguistique — et la dialectométrie.

Toutes les analyses exposées dans cet article sont fondées sur des données provenant de l'«Atlas linguistique des variétés berbères du Rif» (LAFKIOUI 2007), dorénavant ARL, qui offre une étude géolinguistique des variétés berbères du Maroc du nord. Ces variétés linguistiques — que l'on regroupe souvent sous la dénomination de «tarifit» ou de «rifain» — appartiennent aux langues berbères nord et font donc partie du phylum afro-asiatique. Le tarifit est géographiquement délimité à l'ouest par les variétés de la tribu de Ktama (aussi appelées variétés des Senhajas), au sud par la koinè de Gersif (dernier point géographique rifain avant le corridor de Taza), et à l'est par les variétés de la tribu des Iznasen qui sont en contact direct avec des variétés arabophones vers la frontière algéro-marocaine (cf. annexe I). Outre les processus phonologiques complexes qui influent sur la structure morphologique de la langue, le tarifit est surtout connu par sa variation dialectale considérable.

2. Diversité vocalique des variétés berbères du Rif

Parmi les faits linguistiques principaux distinguant le tarifit des autres langues berbères se range son vocalisme qui montre une complexité remarquable issue historiquement de la vocalisation des liquides /r/ et /ʀ/. Ce processus phonétique a résulté en une série de transformations diachroniques à répercussion phonologique importante sur le système vocalique du tarifit, car il l'a augmenté par des voyelles longues, absentes dans les autres langues berbères. La plupart des langues berbères ne connaissent que les voyelles de base /i/, /u/ et /a/ [1]*.

* Les chiffres entre crochets [] renvoient aux notes, pp. 455-456.

2.1. VOCALISATION DES LIQUIDES NON TENDUES /r/ ET /ʀ/

La battue /r/ ([r]) et la vibrante pharyngalisée /ʀ/ ([r̥]) font l’objet du processus de vocalisation dans de nombreuses variétés berbères appartenant principalement au Rif central (cf. annexe III). Selon la variété en question, la vocalisation de ces liquides est arrivée à différents stades d’évolution (cf. annexe 2; LAFKIOUI 2006a, 2007). En fonction de la voyelle qui les précède, elles sont réalisées comme suit au stade final (tab. 1):

Tableau 1
Voyelles longues issues de la vocalisation des /r/ et /ʀ/

Forme diachronique	Forme synchronique	Forme phonétique
ir	ī	diphtongue [ɛæ:] monophtongue [ɪ:]
ur	ū	diphtongue [ɔɑ:] monophtongue [u:]
ar/er	ā	monophtongue [ɛ:] monophtongue [æ:]
ir̥	ī̥	monophtongue [e̞:]
ur̥	ū̥	monophtongue [o̞:]
ar̥	ḁ̄	monophtongue [ʌ̞:]

Ainsi, ce phénomène est à la base d’une extension du système vocalique par des timbres qui se distinguent des voyelles de base par une qualité vocalique modifiée — un abaissement compensatoire net pour /ī/ et /ū/ (*lowering*) — et une valeur quantitative considérable à la suite d’un allongement compensatoire (*compensatory lengthening*) [2].

En raison de leur réalisation phonétique nettement différente, leur identification en tant qu’unités distinctives, le rendement fonctionnel considérable des oppositions distinctives qu’elles forment et leur fréquence d’emploi élevée, les voyelles longues /ī/, /ū/, /ā/ et /ḁ̄/ ont été intégrées dans le système phonologique rifain (LAFKIOUI 1999, 2000, 2002, 2006a, 2007). Autrement dit, le système vocalique rifain distingue au total sept phonèmes dont trois voyelles de base, trois voyelles longues et une voyelle longue et pharyngalisée (tab. 2).

Tableau 2
Système vocalique du tarifit (Rif central)

Voyelles de base	Voyelles dérivées
i	ī
u	ū
a	ā
	ḁ̄

La vocalisation du /r/ est toutefois conditionnée par le fait qu'elle n'est permise, en principe, qu'en position de coda de la syllabe. Exemples des Ayt Temsaman (Rif central):

- (01) *at bir* (aṭ + bir) [æθβɪˈr] + vocalisation ⇒ *at bī* [æθβɛæː] (pigeon)
 (02) *urṭu* (ur + tu) [vrθu] + vocalisation ⇒ *ūrṭu* [ʊɑːθu] (figuier)
 (03) *aryaz* (ar + yaz) [æryæˈz] + vocalisation ⇒ *āyaz* [æːyæˈz] (homme)
 (04) *tarwa* (taɾ + wa) [θaɾˈwɛ] + vocalisation ⇒ *tāwa* [θaɾˈwɛ] (progéniture)

Cette règle restrictive oriente donc le processus de vocalisation. Celui-ci fait partie du processus linguistique général d'«affaiblissement articulatoire» caractérisant les variétés centrales (BIARNAY 1917, RENISIO 1932) [3]. Il serait probablement déclenché par le paramètre fonctionnel d'économie phonétique. La distribution fonctionnelle des nouvelles variantes dans le système vocalique rifain — dans lequel le paramètre de désambiguïsation (transparence) joue un rôle central — est en lien direct avec le paramètre formel de distinction phonologique.

Cependant, il arrive chez les Ayt Weryaḡel (cf. annexe III: zone délimitée sur la carte par la ligne en gras) que la vocalisation ait lieu en position d'attaque (vocalisation alternative), non sans conséquences importantes sur la structure phonétique, phonologique et même morphologique des lexèmes en question (LAFKIOUI 2006a, 2007). Exemples pré vocalique (05) et intervocalique (06) des Ayt Weryaḡel:

- (05) *aḡūm* [aɣuːm] ⇐ *aḡrum* [aɣruːm] (pain)
 (06) *īy* [iːj] of [ɛæːj] ⇐ *iri* [ɪɪɪ] (cou)

Il importe de signaler que la vocalisation en position intervocalique (06) a comme résultat l'allongement de la première voyelle et la désyllabification de la seconde. En outre, on constate que la vocalisation ne peut être effectuée devant une voyelle en position d'Anlaut absolu:

- (07) **āža* ⇐ *raža* (attendre, Aoriste-S)

La vocalisation en position d'attaque implique une restructuration syllabique du lexème, ce qui peut être associé à une réduction du nombre de syllabes. Ceci est une stratégie adaptative formelle aux deux types d'innovation possibles:

- Une innovation fonctionnelle dans laquelle le principe d'économie est poussé au maximum;
- Une innovation formelle où il est question d'analogie entre la vocalisation en position d'attaque et celle en position de coda.

Bien que ces transformations altèrent sérieusement la structure lexicale de base du tarif et qu'elles puissent empêcher l'intercompréhension, elles connaissent une adoption sociale dans les variétés des Ayt Weryaḡel (LAFKIOUI 2007, p. 37).

La vocalisation en position d'attaque est en covariation avec un autre phénomène diachronique, à savoir les mutations phonétiques de l'approximante latérale /l/ en /r ⇐ l/ ou /ʁ/ (Lafkioui 2007, pp. 69-71) [4]:

- /r ⇐ l/= battue rhotique à friction ultralégère [r] dont l'articulation, au plan perceptif, est très proche ou parfois même similaire au /r/ = vibrante [r]
- /ʁ/= fricative [ɹ]

Ces mutations consonantiques sont principalement attestées dans le Rif central [5]. Elles sont en corrélation avec la restriction phonétique qui exclut, en principe, la vocalisation du /r/ issu du /l/. Mais les variétés des Ayt Werayghel font toutefois exception à cela, car elles permettent la vocalisation de /r ⇐ l/ selon un parcours évolutif analogue à celui du /r/ ayant des diphtongues longues comme transformations finales (LAFKIOUI 2006a; 2007, p. 37). En voici un exemple des Ayt Weryaġel (cf. annexe III: zone délimitée sur la carte par la ligne en gras):

- (08a) /irs ([ɪrs], [ɪrs]) ⇐ ils/ + vocalisation /r/ ⇒ [ɛæ:s], [jɛæ:s] et [ɪjɛæ:s] (langue)
- (08b) /irs ([ɪrs], [ɪrs]) ⇐ ils/ + conservation /r/ ⇒ [ɪrs], [ɪrs] (langue)

2.2. DIFFUSION GEOGRAPHIQUE DE LA VOCALISATION DES LIQUIDES /r/ ET /ʁ/

Vu la variation et la complexité considérables que laisse voir la diffusion géographique de la vocalisation des liquides /r/ et /ʁ/ dans le Rif, nous nous limiterons dans cette section à une présentation générale. Des cartes géolinguistiques de LAFKIOUI (2007, pp. 29-33), il ressort que seules les variétés aux extrémités du Rif (Rif occidental et oriental) et la majorité des variétés des Ibeqquyen ne connaissent pas ce phénomène; le /r/ y reste intact (cf. annexe III). Partout ailleurs dans le Rif, les différentes phases du processus de vocalisation sont représentées par des formes diverses, souvent de manière chevauchée (diffusion graduelle). Le stade le plus avancé de la vocalisation est repéré dans des variétés où les liquides ont complètement disparu («noyau central»): Ayt Weryaġel (plupart des variétés), Ayt Temsaman, Ayt Tuzin, Tafersit, Ayt Wlišek, Igzenayen du nord, Ayt Seïd de l'ouest et Iqelēiyen du nord-ouest (cf. annexe III: zone en gris foncé). À l'exception des Ayt Weryaġel, toutes les variétés de ce groupe ont en stade final pour le phonème /i/ la diphtongue [ɛæ:], pour /ū/ la diphtongue [ɔɑ:], pour /ā/ les monophthongues allophones [ɛ:] ou [æ:] et pour /ā/ la monophthongue [ɪ:]. Chez les Ayt Weryaġel, par contre, il existe une alternance libre entre les formes [ɛæ:] et [ɪ:] pour /i/ et entre [ɔɑ:] et [u:] pour /ū/ (cf. annexe III: zone délimitée sur la carte par la ligne en gras).

Les variétés les plus concernées par cette multiplicité de réalisations phonétiques des /r/ et /ʁ/ appartiennent le plus souvent à des zones de contact: des zones de frontière telles les Ayt Buġay de la tribu des Ayt Mezduy (frontière entre Rif occidental et Rif central); des lieux de passage comme Izemmuren des Ibeqquyen (Rif central) et Zayyu des Wlad Settut (Rif oriental); des centres socio-économiques comme Targist (Rif occidental) et Imezzužen d'Ayt Naḍur-Iqelēiyen (Rif oriental). La variation linguistique fort élevée dans ces zones géographiques est essentiellement due à l'augmentation et l'intensification des contacts constituant une source idéale pour le renouvellement linguistique.

Etant donné que le phénomène de vocalisation apparaît dans une seule grande région, le Rif central, et qu'elle s'étend de plus en plus vers l'ouest et l'est — comparé aux données des premières descriptions du tarifit (BIARNAY 1917, RENISIO 1932) —, les nouvelles voyelles seraient, selon toute vraisemblance, formées dans une zone bien précise à partir de laquelle les diverses variantes auraient été diffusées. La taille considérable de la zone de propagation indique une adoption sociale établie. Il va de soi que le succès de la diffusion de cette innovation est déterminé par des paramètres extralinguistiques tels que le contact linguistique. Mais les paramètres fonctionnels d'«économie» et de «conformité de code» (paramètres linguistiques) contribuent toutefois aussi à la diffusion des variantes vocalisées. Ils expliqueraient, par exemple, la préférence de formes complètement vocalisées dans la zone «noyau central» qui ne sont pas en corrélation directe avec des fonctions et structures sociales spécifiques telles que l'attribution de statut entre autres.

3. Classification dialectométrique du lexique berbère du Rif [6]

La dialectométrie est une méthode quantitative pour calculer les distances linguistiques entre des variétés linguistiques. Des différentes méthodes utilisées, nous préconisons celles assistées par ordinateur, car elles permettent de manier des corpus de données étendus avec une certaine commodité, tout en garantissant la précision et la cohérence des analyses, par le fait que:

- Les distances et les fréquences sont mesurées de manière automatique;
- Les données sont classifiées numériquement;
- La cartographie peut être assistée par ordinateur;
- Des analyses statistiques et non statistiques peuvent être effectuées et visualisées automatiquement.

Les analyses dialectométriques, que nous présenterons dans cet article, ont été effectuées avec le logiciel informatique libre de Kleiweg (RuG/L04) [7].

Afin d'accomplir une analyse dialectométrique visualisée, toutes les étapes de la procédure résumée ci-dessous (tab. 3) sont indispensables (LAFKIOUI 2009):

Tableau 3
Procédure générale de l'analyse dialectométrique automatisée

Etape 1	Atlas linguistique = source des données géoréférencées
Etape 2	Matrice des données
Etape 3	Matrice des distances
Etape 4	Analyses
Etape 5	Visualisation

3.1. L'ATLAS LINGUISTIQUE DU RIF COMME SOURCE DES DONNEES

Les matériaux lexicaux comparés et classifiés dans cette étude, provenant de l'ALR (LAFKIOUI 2007), constituent un corpus numérique de soixante-deux lexèmes portant sur le corps humain (cartes 295 à 315), les liens de parenté (cartes 316 à 321), les animaux (cartes 322 à 327), les couleurs (cartes 328 et 329), les numéraux (cartes 330 à 332) et un sous-ensemble de noms et verbes divers (cartes 333 à 356). Parmi ces lexèmes, onze ne disposent que d'une seule variante par variété; tous les cinquante et un autres lexèmes exposent tous une co-occurrence de multiples variantes pour chaque lexème.

En raison de la réalisation automatisée de l'ALR, les données qui en sont extraites sont déjà en format numérique, ce qui a évité un gros travail de numérisation. Cependant, une conversion adaptative au logiciel RuG/L04 (Kleiweg) a été nécessaire. L'ALR offre, en outre, une carte géographique numérique précise de la région du Rif (fig. 1). Cette carte est essentielle à la visualisation des analyses, exception faite du dendrogramme.

Cent quarante et un points géoréférencés — appartenant à trente-deux tribus rifaines — ont été sélectionnés d'un ensemble de quatre cent cinquante-deux localités du Rif selon leur degré de variation linguistique (LAFKIOUI 2007) [8].

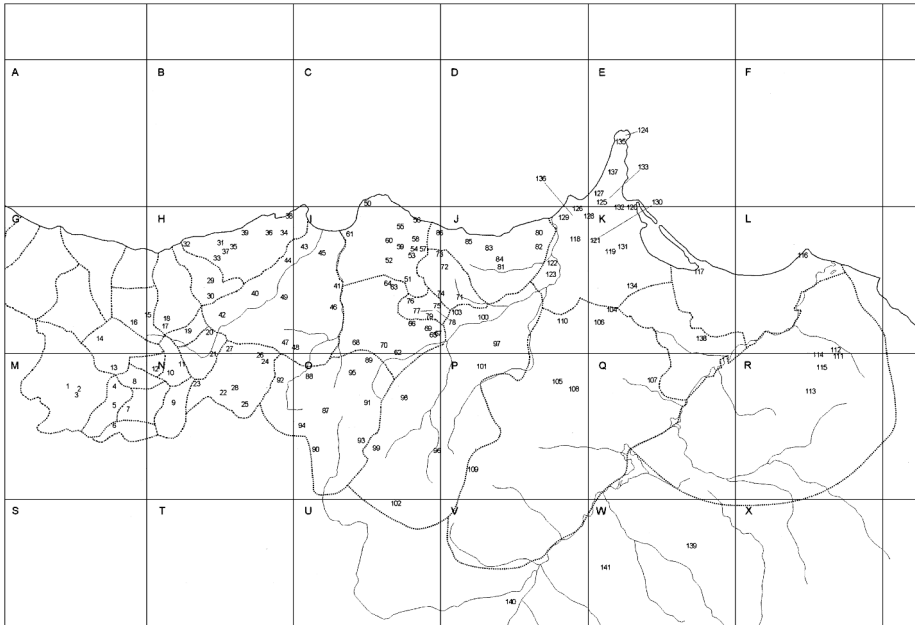


Fig. 1. — Carte des points d'enquête géoréférencés du Rif (LAFKIOUI 2007).

3.2. MATRICE DES DONNEES LEXICALES BERBERES DU RIF

La matrice des données est composée des matériaux lexicaux numériques extraits de l'ALR (LAFKIOUI 2007) et convertis suivant le format du logiciel RuG/L04 (Kleiweg). En voici un petit échantillon en format numérique de l'ALR (format de Mapinfo Professional; tab. 4) et en format texte du logiciel RuG/L04 (tab. 5):

Tableau 4
Extrait des données en format numérique de l'ALR

Sector	Tribe	Full Name ND	LF461	LF462	LF463	LF464	LF465	LF466	LF467	LF468	LF469	LF470
1	Ktama	Asammer	31	51	52	32	54	51	32	33	53	52
1	Ktama	Lmexzen	31	51	52	32	54	51	32	33	53	52
1	Ktama	Ssahel	31	51	52	32	54	51	32	33	53	52
2	Taghzut	Lqel'a	31	54	52	32	54	51	34	33	54	52
2	Taghzut	Ssaqya	31	54	52	32	54	51	34	33	54	52
3	Ayt Bucibet	Tarya	31	54	52	32	54	51	32	33	54	52
4	Ayt Hmed	Mazuz	31	51	52	32	54	51	32	33	54	52
5	Ayt Bunsar	Luta	31	32	52	32	54	51	32	33	51	52
6	Ayt Bcir	Tizirt	31	51	52	32	52	51	33	31	51	52
7	Zerqet	Aghennuy	31	32	52	32	52	51	33	31	51	52
7	Zerqet	Wersan	31	32	52	32	52	51	33	31	51	52
8	Ayt Xennus	A'raben	31	32	52	32	54	51	32	33	53	52
9	Ayt Seddat	Azila	31	32	52	32	54	51	32	33	53	52
9	Ayt Seddat	Tamadda	31	32	52	32	54	51	32	33	53	52
A	Ayt Gmil	Azru n tili	53	53	53	32	51	54	31	52	53	53
A	Ayt Gmil	Tizi	53	53	53	32	51	54	31	52	53	53
B	Ayt Bufrah	Igzennayen	53	32	52	32	51	13	31	52	53	52
B	Ayt Bufrah	Iharunen	53	32	52	32	51	13	31	52	53	52
C	Targist	Ayt' Azza	54	52	53	53	51	12	53	52	53	53
D	Ayt Mezduy	Bni Budjay	53	53	53	32	51	54	31	52	53	53
D	Ayt Mezduy	Bu'di	53	32	52	32	51	51	31	52	53	52

3.3. MATRICE DES DISTANCES POUR LE LEXIQUE BERBERE DU RIF

Dans cette section, nous mettons en contraste les trois techniques de comparaison numérique les plus usitées: la mesure de distance binaire (algorithme de Hamming), la mesure de distance *Gewichteter Identitätswert* (identité pondérée), et la mesure de distance Levenshtein. Nous les appliquerons au lexique berbère du Rif afin de tester leur validité et d'en sélectionner la plus appropriée au berbère. Chaque mesure de distance permet d'obtenir des valeurs numériques précises issues de la comparaison linguistique entre les variétés du Rif. Ces valeurs

composent les matrices de distance (matrices symétriques N x N; N = somme des variétés) dont la configuration diverge selon l’algorithme adopté.

Tableau 5

Extrait des données en format texte du logiciel RuG/L04

: Asammer	: A'raben	: Asammer
- aqnin	- aqnin	- aqenni
: Lmexzen	: Azila	: Ayt Hmed
- aqnin	- aqnin	- aqenni
: Ssahel	: Tamadda	: Sidi Bucetta
- aqnin	- aqnin	- aqenni
: Lqel'a	: Azru n tili	: Tazrut
- aqnin	- aqenni	- aqenni
: Ssaqya	: Tizi	: Ufis
- aqnin	- aqenni	- aqenni
: Tarya	: Igzennayen	: Wad Mahkim
- aqnin	- aqenni	- aqenni
: Mazuz	: Iharunen	: L'ars
- aqnin	- aqenni	- aqenni
: Luta	: Ayt 'Azza	: Tufist-Imuruten
- aqnin	- aqenniy	- aqenni
: Tizirt	: Bni Budjay	
- aqnin	- aqenni	
: Aghennuy	: Bu'di	
- aqnin	- aqenni	
: Wersan	: Aghir Hmed	
- aqnin	- aqenni	

3.3.1. La mesure de distance binaire

La mesure binaire (Bin) permet de classifier les unités lexicales comme étant identiques ou non identiques. Le rapport de comparaison est donc de 0-1; 0 = ressemblance et 1 = différence. Voici un extrait de la matrice de distance binaire pour le lexème «talon» (ALR, carte 312) (tab. 6):

Tableau 6

Extrait de la matrice de distance binaire pour le lexème «talon»

	Tizirt	Aghennuy	Wersan	A'raben	Azila	Tamadda	Azru n tilli	Tizi
Wersan	0	0	0	0	0	0	1	1
A'raben	0	0	0	0	0	0	1	1
Azila	0	0	0	0	0	0	1	1
Tamadda	0	0	0	0	0	0	1	1
Azru n tilli	1	1	1	1	1	1	0	0
Tizi	1	1	1	1	1	1	0	0
Igzennayen	1	1	1	1	1	1	0	0
Iharunen	1	1	1	1	1	1	0	0
Ayt'Azza	1	1	1	1	1	1	1	1

3.3.2. La mesure de distance Gewichteter Identitätswert

La mesure *Gewichteter Identitätswert* (GIW) diverge de la mesure binaire par le fait que la fréquence de la variante lexicale joue également un rôle dans la comparaison : les variantes à fréquence basse pèsent plus lourd que les variantes à fréquence élevée. Les valeurs de distance obtenues par cette technique sont comprises entre 0 et 1, soit $\{0 \leq d \leq 1\}$. En voici un exemple provenant de la matrice de distance du lexème «talon» (tab. 7):

Tableau 7

Extrait de la matrice de distance GIW pour le lexème «talon»

	Tizirt	Aghennuy	Wersan	A'raben	Azila	Tamadda	Azru n tilli	Tizi
Wersan	0.00501792	0.00501792	0	0.00501792	0.00501792	0.00501792	1	1
A'raben	0.00501792	0.00501792	0.00501792	0	0.00501792	0.00501792	1	1
Azila	0.00501792	0.00501792	0.00501792	0.00501792	0	0.00501792	1	1
Tamadda	0.00501792	0.00501792	0.00501792	0.00501792	0.00501792	0	1	1
Azru n tilli	1	1	1	1	1	1	0	0.215054
Tizi	1	1	1	1	1	1	0.215054	0
Igzennayen	1	1	1	1	1	1	0.215054	0.215054
Iharunen	1	1	1	1	1	1	0.215054	0.215054
Ayt' Azza	1	1	1	1	1	1	1	1

3.3.3. La mesure de distance Levenshtein

Les valeurs de distance qui dérivent de la comparaison fondée sur la mesure Levenshtein — algorithme qui tient compte de l'ordre séquentiel des unités phoniques dont les lexèmes sont composés — varient entre 0 et 1, ($\{0 \leq d \leq 1\}$), comme le montre le tableau 8.

Ces valeurs sont le résultat de la sélection du calcul le moins coûteux pour transformer une unité lexicale — en tant que chaîne d'unités phoniques — en une autre. Le tableau 9 expose les coûts les plus bas des opérations permettant de modifier les chaînes de caractères de *awrez* (talon) en *inerz* (talon).

Le coût le moins élevé des opérations modifiant *awrez* en *inerz* est de 3, de sorte que la distance entre ces deux lexèmes est de 3/5 (5 étant le total des caractères); autrement dit, la distance de Levenshtein est de 60 %. Ces calculs sont basés sur des opérations qui coûtent 0.5 pour une insertion ou une suppression et 1 pour une substitution. Ainsi, par exemple, le tableau 10.

Tableau 8

Extrait de la matrice de distance Lv pour le lexème «talon»

	Tizirt	Aghennuy	Wersan	A'raben	Azila	Tamadda	Azru n tilli	Tizi
Wersan	0	0	0	0.	0.	0.	0.6	0.6
A'raben	0	0	0.	0	0.	0.	0.6	0.6
Azila	0	0	0.	0.	0	0.	0.6	0.6
Tamadda	0	0	0.	0.	0.	0	0.6	0.6
Azru n tilli	0.6	0.6	0.6	0.6	0.6	0.6	0	0
Tizi	0.6	0.6	0.6	0.6	0.6	0.6	0	0
Igzennayen	0.6	0.6	0.6	0.6	0.6	0.6	0	0
Iharunen	0.6	0.6	0.6	0.6	0.6	0.6	0	0
Ayt'Azza	0.555556	0.555556	0.555556	0.555556	0.555556	0.555556	0.111111	0.111111

Tableau 9

Coûts des opérations permettant de modifier *awrez* en *inerz* (talon)

		a	w	r	e	z
	0	0.5	1	1.5	2	2.5
i	0.5	1	1.5	2	2.5	3
n	1	1.5	2	2.5	3	3.5
e	1.5	2	2.5	3	2.5	3
r	2	2.5	3	2.5	3	3.5
z	2.5	3	3.5	3	3.5	3

Tableau 10

Exemple du calcul de distance Lv pour modifier *awrez* en *inerz* (talon)

Tamadda	a	w	r	e		z	
Tizi	i	n		e	r	z	
Distance Lv	1	1	0,5	0	0,5	0	3/5 * 100 = 60 %

3.4. ANALYSES DIALECTOMETRIQUES NUMERIQUES DU LEXIQUE BERBERE DU RIF

A partir des matrices de distance, des analyses de comparaison numériques du lexique berbère peuvent être réalisées par le biais de deux types de technique: *Cluster Analysis* (analyse de regroupement) et *Multidimensional Scaling* (analyse de graduation multidimensionnelle). La technique de *Cluster Analysis* (CA) consiste à regrouper les données par réduction de la matrice de distance moyennant des algorithmes variés. A la suite de Kleiweg (RuG/L04), nous avons adopté l'algorithme de Ward (variance minimum) qui est généralement considéré comme un des algorithmes les plus adéquats. *Multidimensional Scaling* (MDS), par contre, est:

[...] a technique that, using a table of differences, tries to position a set of elements into some space, such that the relative distance in that space between all elements corresponds as close as possible to those in the table of differences (Kleiweg, RuG/L04) [9].

3.5. VISUALISATION DES ANALYSES DIALECTOMETRIQUES DU LEXIQUE BERBERE DU RIF

Pour être visualisée, la classification par regroupement (CA) a nécessairement recours à un dendrogramme qui est une sorte d'arborescence complexe, généralement en couleur, dont les branches représentent les variétés. Ce dendrogramme peut être associé avec une carte géographique numérique, ce qui donne comme résultat une carte géolinguistique qui montre la répartition des variétés selon les différences linguistiques et les critères de classification retenus. Les analyses par graduation multidimensionnelle (MDS), en revanche, offrent directement des cartes où la variation relative est représentée de façon graduelle par des nuances de couleurs différentes.

3.5.1. *Visualisation et interprétation des analyses CA*

La structure hiérarchique du dendrogramme et la répartition des variétés sur la carte du Rif qui y est associée varient sensiblement selon l'algorithme de distance (Bin, GIW ou Lv) appliqué. Ainsi, l'on aperçoit pour la mesure binaire (fig. 2) une structure composée de sept groupes principaux, regroupés dans deux sous-groupes: le sous-groupe mineur contenant les groupes 6 et 7 et le sous-groupe majeur contenant les groupes 1 à 5; la distance entre ces deux sous-groupes étant de 16.17. Cette valeur de distance relativement élevée indique une frontière linguistique nette après le groupe 7 qui est délimité à droite par les variétés des Ayt Weryagel et des Ayt 'Ammart. Le sous-groupe majeur connaît une subdivision assez équilibrée ($d = 9.34$) entre, d'une part, les groupes 4 et 5 (variété de Targist incluse) et, d'autre part, les groupes 1 à 3 qui, à leur tour, ont également subi une subdivision. La seconde frontière linguistique importante coïncide donc avec les variétés limitrophes des groupes 4 (Igzennayen) et 5 (Ayt Seïd et Ayt Tuzin).

La classification fondée sur l'algorithme GIW diverge considérablement de celle fondée sur l'algorithme Bin, car elle aboutit à un ensemble de cinq regroupements (fig. 3), dont le regroupement 1 englobe les sous-groupes 1, 2 et 4 de la classification Bin (fig. 2). Cependant, la frontière linguistique principale détectée par le biais de GIW — frontière (est) tracée après les variétés du groupe 5 — est identique à celle que dégage le dendrogramme Bin, bien que la distance entre les deux sous-ensembles principaux soit moins élevée pour GIW ($d_{GIW} = 10.87$) que pour Bin ($d_{Bin} = 16.17$), en raison de l'intégration de la fréquence dans la comparaison.

La classification issue des analyses dialectométriques fondées sur la mesure de distance Lv résulte dans une configuration asymétrique de sept groupes répartis en deux sous-ensembles distants l'un de l'autre de 8.08 (fig. 4). Ce dendro-

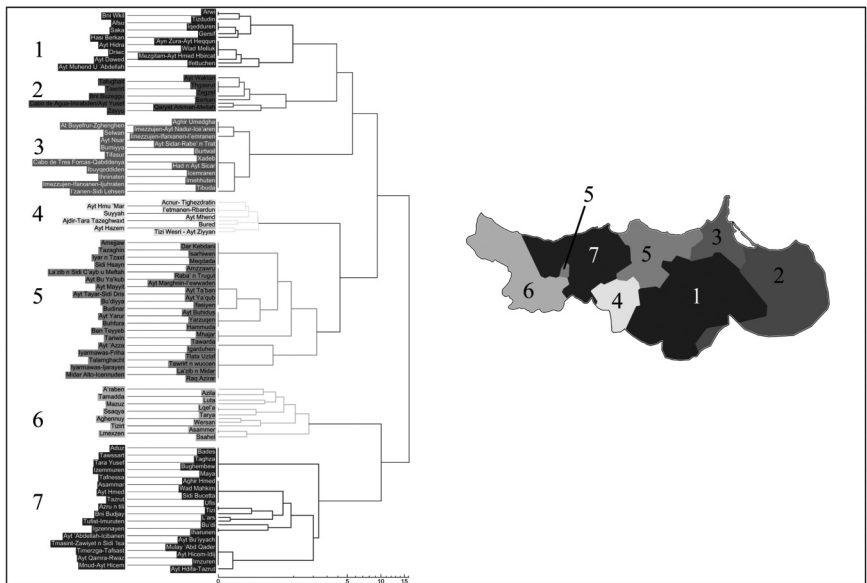


Fig. 2. — Dendrogramme vs Carte CA - Bin – Tout le lexique.

gramme partage la même délimitation linguistique prépondérante (entre groupes 6 et 3-4) avec les autres dendrogrammes. Ce constat est corroboré par les cartes CA_{LV} présentées dans la figure 5, dont celle à deux regroupements indique claire-

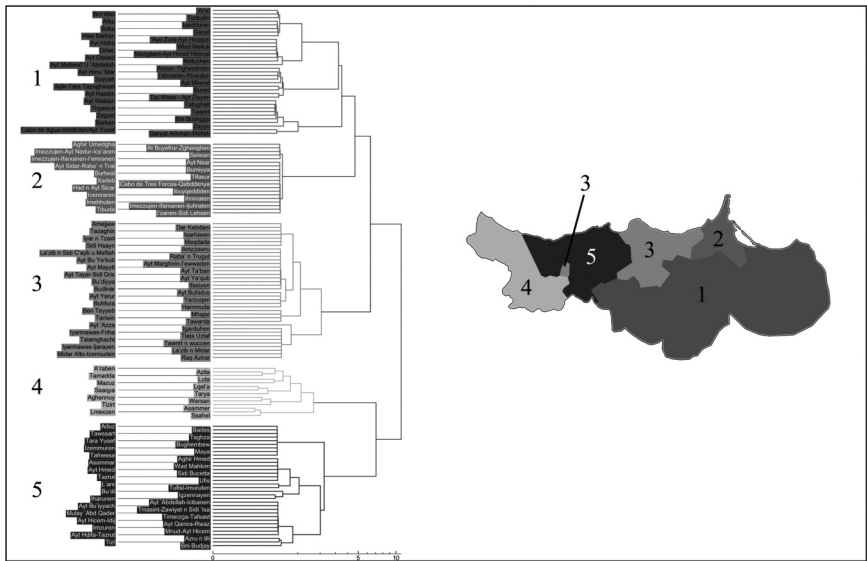


Fig. 3. — Dendrogramme vs Carte CA - GIW – Tout le lexique.

ment la frontière linguistique la plus distinctive. Il importe de remarquer que la carte CA_{Lv} (fig. 4) affiche une répartition des variétés analogue à celle de la carte CA_{Bin}, bien que la composition de leur dendrogramme respectif soit divergente.

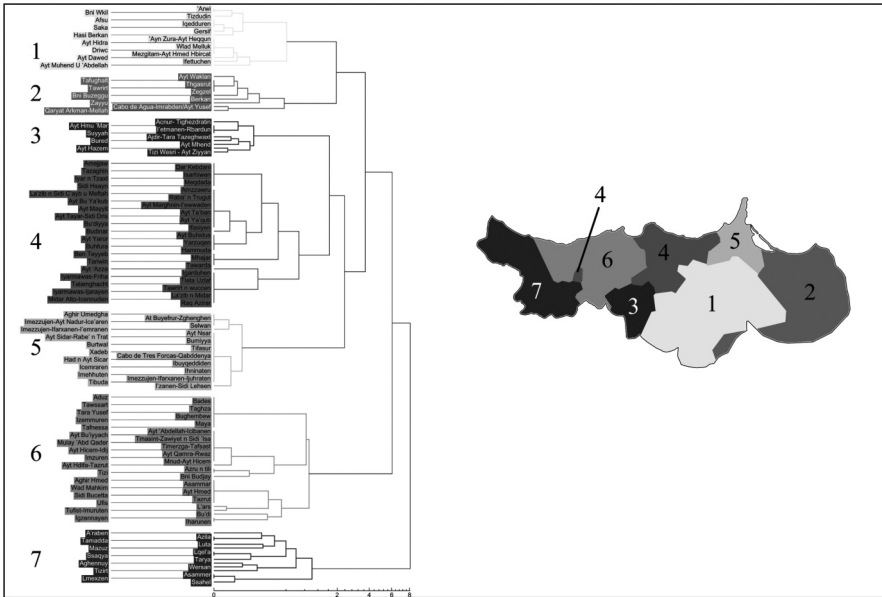


Fig. 4. — Dendrogramme vs Carte CA - Lv – Tout le lexique.

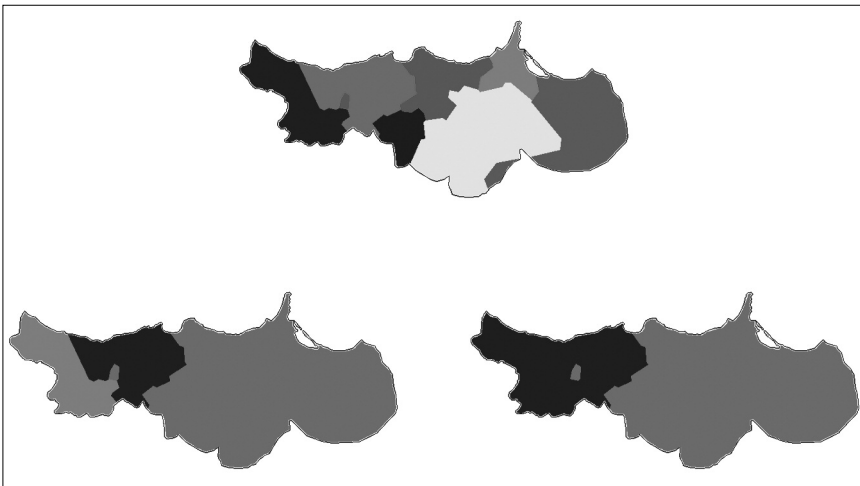


Fig. 5. — Cartes CA_{Lv} – 7 groupes vs 3 groupes vs 2 groupes – Tout le lexique.

3.5.2. Visualisation et interprétation des analyses MDS

La technique MDS offre l'avantage majeur de garantir une objectivité et une précision lors de la phase d'analyse des matériaux, parce qu'elle en exclut tout paramétrage externe. Ainsi, par exemple, on n'y peut pas modifier le nombre de regroupements. C'est le système d'analyse qui le fournit automatiquement. Chaque variété y a sa propre couleur. Ce sont les contrastes de couleurs qui servent à l'interprétation des données linguistiques comparées: une continuité de couleur indique une corrélation parfaite entre les lexèmes, alors qu'une mosaïque de couleurs dévoile une corrélation faible entre eux.

L'aire du Rif connaît, par le biais de MDS, une répartition en sept grandes zones, quelle que soit la mesure de distance appliquée (fig. 6). La répartition des variétés sur les cartes MDS est quasi similaire pour Bin et GIW; seules quelques différences mineures de nuances de certaines couleurs ont été observées. La carte MDS_{Lv} ressemble fortement aux deux autres; la seule distinction significative constatée est l'apparition d'une petite subdivision à l'intérieur du groupe des variétés occidentales.

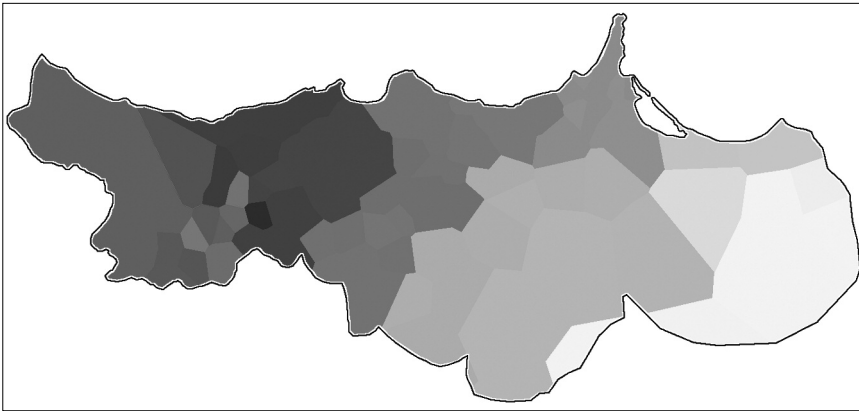


Fig. 6. — Carte MDS_{Lv} – Tout le lexique.

3.6. RESULTATS COMPARATIFS

Du fait de son exactitude, le procédé MDS est le plus approprié pour l'analyse dialectométrique du lexique berbère. Il forme, de ce fait, un cadre de comparaison adéquat pour vérifier les résultats obtenus des autres méthodes dialectométriques. Parmi les classifications par regroupement (CA), les classifications CA_{Bin} et CA_{Lv} rejoignent le mieux la distribution affichée par les cartes MDS (sept groupes). La classification CA_{Lv} montre, en outre, une précision supérieure du fait qu'elle prend en considération la variation phonique des occurrences autant que leur agencement dans les lexèmes. Cependant, toute analyse fondée sur la mesure de distance Lv (CA aussi bien que MDS) ignore la hiérarchie qui existe entre les

unités phoniques (unités phonétiques = unités phonémiques), à moins qu'on leur accorde des poids divers moyennant un paramétrage spécifique, ce qui revient à construire un système phonologique à l'intérieur du logiciel, impliquant un travail laborieux et trop coûteux par rapport aux profits.

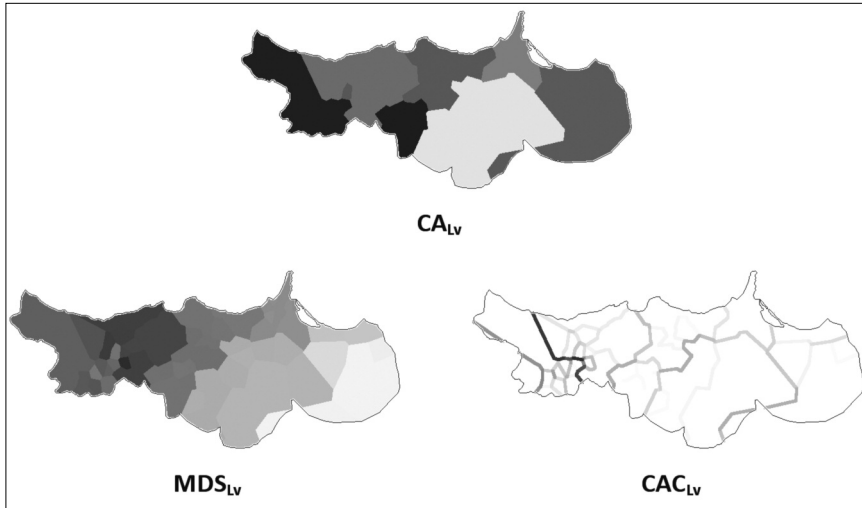


Fig. 7. — Cartes CA_{Lv} vs MDS_{Lv} vs CAC_{Lv}

La classification par regroupement (CA) offre l'avantage d'indiquer précisément les frontières linguistiques significatives. Les cartes CAC (*Composite cluster map* = carte de regroupement composite; figs 7 et 8) les marquent par des lignes à couleur foncée. Comparée aux délimitations distinctives dégagées par les dendrogrammes et les cartes CA correspondantes des figures 2 à 5, la frontière linguistique principale de la carte CAC de la figure 7 est tracée encore davantage vers l'ouest. Il importe, cependant, de signaler que les cartes CAC ne nous paraissent pas les plus aptes à visualiser la classification du lexique berbère du Rif, en raison de la difficulté d'interprétation des données, due à leur représentation assez chaotique (fig. 8) [10].

4. Conclusion générale

Les analyses exposées dans les sections 2 et 3 montrent comment la variation rifaine peut être examinée suivant deux approches géolinguistiques différentes mais néanmoins connectées, à savoir la dialectologie (approche qualitative) et la dialectométrie (approche quantitative). Pour la dialectologie, nous avons présenté le cas de la vocalisation des liquides /r/ et /ʀ/ en tarifit. Ce processus diachronique a engendré une extension du vocalisme rifain par le biais de diverses innovations

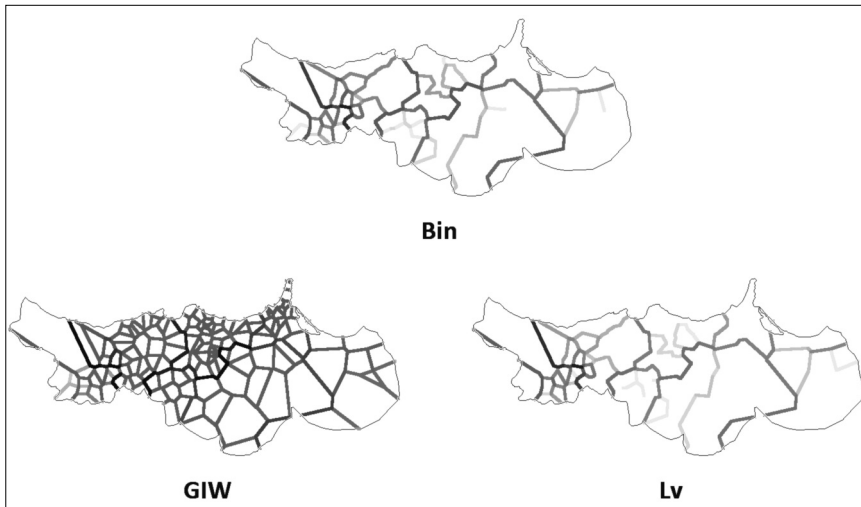


Fig. 8. — Cartes CAC – Bin vs GIV vs Lv – Tout le lexique.

phonétiques et phonologiques qui, au plan synchronique, se manifestent sous forme de nombreuses variantes vocaliques. Celles-ci connaissent une répartition inégale sur l'aire du Rif qui est déterminée par des paramètres extralinguistiques (sociaux et démographiques) aussi bien que par des paramètres linguistiques (formels et fonctionnels). Pour la dialectométrie, c'est la variation lexicale qui a servi de point de départ aux analyses de classification qui ont abouti à des résultats convergeant vers des frontières linguistiques distinctives.

NOTES

- [1] La majorité des langues berbères disposent de ces trois voyelles de base: la voyelle antérieure (étirée) /i/ d'aperture minimum étant réalisée [i]: [izi] (mouche); la voyelle postérieure /u/ (arrondie) d'aperture minimum étant réalisée [u]: [ul] (cœur); la voyelle médiane /a/ d'aperture maximum étant réalisée [æ] ou [ɛ]: [æmæ'n] ou [ɛmɛ'n] (eau). Outre ces vocoïdes de base, les langues berbères connaissent également la voyelle centrale [ə] qui est caractérisée par, à la fois, la vibration des cordes vocales et l'absence d'occlusion ou de friction. A l'encontre des voyelles fondamentales, le [ə] n'est généralement pas considéré comme un phonème dans les études berbères mais comme un «lubrifiant» évitant l'entassement des consonnes.
- [2] A l'exception de TANGI (1991), DELL & TANGI (1993), LOUALI & PUECH (1997) et LOUALI-RAYNAL (2000, 2002), la plupart des berbérissants les identifient comme des voyelles «longues» (BIARNAY 1917; RENISIO 1932; CHAMI 1979; CADI 1987; CHTATOU 1982, 1994; HAMDAOUI 1985; ALLATI 1986; EL AISSATI 1989; LAFKIOUI 1999, 2000, 2002, 2006a, 2007). Cependant, la longueur de ces voyelles peut varier selon les régions et les locuteurs en question.

- [3] La spirantisation des occlusives (LAFKIOUI 2006b, 2007) et la vocalisation des semi-voyelles /w/ (⇒ /u/) en /y/ (⇒ /i/) (LAFKIOUI 2007) en font aussi partie.
- [4] La différence phonétique entre la battue issue du /l/ (/r ⇐ l/ avec légère friction) et la battue d'origine (/r/ sans friction) est toutefois très difficile à détecter. La friction de la battue /r ⇐ l/ est tellement faible qu'elle ne permet pas une distinction au plan perceptif. Cependant, leur divergence articulatoire devient plus nette une fois qu'elles sont précédées par des voyelles: /V/ + /r ⇐ l/ ≠ /V/ + /r/. Car seule la battue /r/ a un effet d'abaissement sur les voyelles qu'elle suit (LAFKIOUI 2007, p. 69). Exemples: ari ⇐ ali (monter): [æɾɪ] ou [ɛɾɪ] ≠ ari ⇐ ari (écrire): [aɾɪ].
- [5] D'autres mutations consonantiques ont été repérées dans le Rif occidental: /l/ ⇒ /ʒ/ ([ʒ]), /y/ ([j]). Ce phénomène est, en revanche, absent dans les variétés orientales (LAFKIOUI 2007, pp. 69-71).
- [6] Cette section reprend en partie les analyses de LAFKIOUI (2009).
- [7] [Http://odur.let.rug.nl/~kleiweg/L04](http://odur.let.rug.nl/~kleiweg/L04).
- [8] Les points d'enquête ont été sélectionnés suivant le principe d'équidistance divisant le terrain d'enquête en plusieurs mailles dont chacune d'elles a été affectée d'un point qui pouvait correspondre sur le terrain à une localité. Plus la variation était grande, plus les mailles ont été réduites. Les quatre cent cinquante-deux localités retenues pour cette recherche ont été, en majeure partie, choisies de façon qu'elles puissent, a priori, indiquer des frontières linguistiques, ce qui découlait principalement de la connaissance empirique et scientifique de l'enquêteur des différentes variétés parlées au Rif.
- [9] [...] une technique qui, au moyen d'un tableau de différences, tente de positionner un ensemble d'éléments dans l'espace, de telle manière que les distances relatives entre les éléments de cet espace correspondent autant que possible à celles du tableau des différences.
- [10] Kleiweg (RuG/L04) propose certaines alternatives pour l'algorithme Ward qui semble être à la base de ce désordre visuel par lequel les cartes CA_{GIW} sont les plus touchées.

BIBLIOGRAPHIE

- ALLATI, A. 1986. Phonétique et phonologie d'un parler amazigh du Nord-Est marocain (le parler des Aït Saïd). — Aix-en-Provence, Université de Provence, Thèse de 3^e cycle.
- BIARNAY, S. 1917. Etude sur les dialectes berbères du Rif (Ibeqqoyen, Ait Ouriaghel, Ait Touzin, Tamsaman, Ikebdanen, Ait Itteft). — Paris, Leroux.
- CADI, K. 1987. Système verbal rifain. Forme et sens. — Paris, SELAF.
- CHAMI, M. 1979. Un parler amazigh du Rif marocain: approche phonologique et morphologique. — Paris, Paris-V, Thèse de 3^e cycle.
- CHATOU, M. 1982. Aspects of the Phonology of a Berber dialect of the Rif. — London, University of London/SOAS, PhD Dissertation.
- CHATOU, M. 1994. La représentation vocalique dans les dialectes berbères du Rif. — *Etudes et Documents berbères*, 11: 177-196.
- DELL, F. & TANGI, O. 1993. On the vocalisation of /r/ in ath-Sidhar Rifain Berber. — *Linguistica communicatio*, 5 (1-2): 211-224.
- EL AISSATI, A. 1989. A Study of the Phonotactics of Asht Touzine Tarifit Dialect. — Rabat, Université Mohamed V, Thèse de 3^e cycle.

- HAMDAOUI, M. 1985. Description phonétique et phonologique d'un parler amazigh du Rif marocain (Province d'Al Hoceima). — Aix-en-Provence, Université de Provence, Thèse de 3^e cycle.
- LAFKIOUI, M. 1999. Propositions pour la notation usuelle à base latine du berbère et application sur le rifain. — *In: L'enseignement/Apprentissage de l'Amazighe: expériences, problématiques et perspectives*, Actes du colloque de l'Université d'été d'Agadir (juillet 1996, Agadir), pp. 66-82.
- LAFKIOUI, M. 2000. Propositions pour la notation usuelle à base latine du rifain. — *Comptes Rendus du G.L.E.C.S.*, **XXXIII**: 189-200.
- LAFKIOUI, M. 2002. Le rifain et son orthographe: entre variation et uniformisation. — *In: CAUBET, D., CHAKER, S. & SIBILLE, J. (Eds), Codification des langues de France*. Paris, L'Harmattan, pp. 355-366.
- LAFKIOUI, M. 2006a. La vocalisation des alvéolaires /r/ et /rr/ dans les variétés berbères du Rif. — *Studien zur Berberologie/Etudes Berbères*, **3**: 175-184.
- LAFKIOUI, M. 2006b. La spirantisation dynamique de la vélaire occlusive simple /k/ dans les variétés berbères du Rif. — *Studi berberi e mediterranei*, **3** (NS): 219-228.
- LAFKIOUI, M. 2007. Atlas linguistique des variétés berbères du Rif. — Köln, Rüdiger Köppe Verlag.
- LAFKIOUI, M. 2009. Analyses dialectométriques du lexique berbère du Rif. — *Studien zur Berberologie/Etudes Berbères*, **4**: 19 pp.
- LOUALI-RAYNAL, N. 2000. Vocalisme berbère et voyelles touarègues. — *In: CHAKER, S. (Ed.), Etudes berbères et chamito-sémitiques. Mélanges offerts à Karl-G. Prasse*. Paris/Louvain, Peeters, pp. 263-276.
- LOUALI-RAYNAL, N. 2002. Les mutations du l et du r pan-berbère. — *In: NAIT-ZERRAD, K. (Ed.), Articles de linguistique berbère. Méorial Werner Vycichl*. Paris, L'Harmattan, pp. 301-334.
- LOUALI, N. & PUECH, G. 1997. Le vocalisme en berbère rifain. — *In: Journées d'Etudes Linguistiques: la voyelle dans tous ses états* (Nantes), pp. 38-43.
- RENISIO, A. 1932. Etude sur les dialectes berbères des Beni Iznassen, du Rif et des Senhaja de Srair. — Publications de l'Institut des Hautes Etudes Marocaines, 22 pp.
- TANGI, O. 1991. Aspects de la phonologie d'un parler berbère du Maroc: Ath-Sidhar (Rif). — Paris, Paris VIII, Thèse de Doctorat.

ANNEXE 2
Vocalisation de la battue simple /r/, /atIR/ (pigeon) (d'après LAFKIOUI 2008, p. 31)



