



**HAL**  
open science

## Méthodologie de recherche en géolinguistique

Mena B. Lafkioui

► **To cite this version:**

Mena B. Lafkioui. Méthodologie de recherche en géolinguistique. *Corpus*, 2015, 14, pp.139-164.  
hal-01486739

**HAL Id: hal-01486739**

**<https://hal.science/hal-01486739>**

Submitted on 6 Nov 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

---

## Méthodologie de recherche en géolinguistique

Mena Lafkioui

---

**Electronic version**

URL: <http://corpus.revues.org/2658>  
ISSN: 1765-3126

**Publisher**

Bases ; corpus et langage - UMR 6039

**Printed version**

Date of publication: 1 January 2015  
Number of pages: 139-164  
ISBN: 1 638-9808  
ISSN: 1638-9808

**Electronic reference**

Mena Lafkioui, « Méthodologie de recherche en géolinguistique », *Corpus* [Online], 14 | 2015, Online since 29 August 2017, connection on 29 August 2017. URL : <http://corpus.revues.org/2658>

---

This text was automatically generated on 29 August 2017.

© Tous droits réservés

---

# Méthodologie de recherche en géolinguistique

Mena Lafkioui

---

## 1. Introduction

- 1 Cet article<sup>1</sup> se veut une présentation des méthodes principales que nécessitent les études en géolinguistique, un vaste domaine de recherche consacré à l'analyse de la variation, du changement et de la diffusion linguistiques. La raison principale de mon intérêt pour la recherche géolinguistique (recherches variationnelles incluses) est que de nombreuses aires linguistiques – dont celles du berbère – restent encore pas ou peu étudiées et que celles qui ont été examinées sont, pour des raisons méthodologiques, pratiques ou autres, très limitées géographiquement – souvent une seule variété ou groupe de variétés homologues sont pris comme point de référence pour toute l'analyse et la généralisation qui en est déduite – et très réduites socialement – généralement, les informateurs auxquels il est fait appel sont peu nombreux et ne sont pas sélectionnés en fonction de leurs diverses réalités sociales.
- 2 Dans cette contribution, une attention particulière sera accordée à la dialectologie synchronique et diachronique en berbère. La dialectologie synchronique s'interroge sur la variation linguistique (phonologique, grammaticale et lexicale) en rapport avec sa distribution géographique, tandis que la dialectologie diachronique (ou historique) décrit et explique par le biais des reconstructions sociohistoriques les processus de changement et de diffusion linguistiques qui ont mené à cette variation. Ces deux approches sont reliées de sorte qu'il est difficile de faire une distinction nette entre elles. En effet, en dialectologie, toute diversité géographique renvoie nécessairement à une diversité temporelle (Dalbera 2013 ; de Saussure 1973 [1916] ; Lafkioui 2007). Par ailleurs, la variation sociale et la variation individuelle contribuent aussi de façon directe à cette diversité géolinguistique car, comme l'a bien noté Martinet (1967 : 157), « ce n'est pas la distance elle-même qui produit la différenciation mais le relâchement des contacts ». La

variation sociale couvre plusieurs réalités dont principalement celle de l'habitat (ville/campagne, pays d'origine/pays d'immigration), du sexe, de l'âge, de la classe sociale et du niveau et du type d'instruction. La variation individuelle recoupe en partie la variation sociale, mais intègre aussi l'aspect idéo-historique et psychologique de l'individu (Lafkioui 2011).

- 3 Quant à la dialectométrie – suite pertinente à l'examen de dialectologie –, elle se préoccupe de la classification quantitative de la variation linguistique, ce qui permet de mettre au jour les faisceaux d'isoglosses structurées divisant la région étudiée. Ce genre d'étude offre en outre des résultats utiles à la compréhension de nombreux phénomènes de contact et d'évolution linguistiques, tels le phénomène d'adoption, d'acquisition et de diffusion. Les méthodes dialectométriques les plus usuelles (traditionnelles, perceptuelles et computationnelles) et leurs applications et adaptations au berbère ont été présentées dans Lafkioui (2008, 2009).
- 4 La géolinguistique englobe aussi le domaine d'étude appelé la linguistique aréale (*areal linguistics* ou *linguistic geography*), qui traite des transformations linguistiques issues de contacts stables au niveau inter-générationnel dans des aires géographiques adjacentes (voir p. ex. Clements & Rialland 2008 pour une macro-classification aréale de l'Afrique fondée sur des traits phonologiques). Ses problématiques et ses méthodes ont été exposées en détail pour les langues africaines dans Lafkioui (à paraître).
- 5 Dans ce qui suit, je présenterai d'abord les outils méthodologiques d'enquête de terrain indispensables à la constitution d'un corpus de matériaux oraux et visuels qui permettent des analyses géolinguistiques, dont également celles de type comparatif. Ensuite, je montrerai quelles sont les méthodes cartographiques les plus appropriées pour présenter et représenter la variation et la diffusion géolinguistiques<sup>2</sup>.

## 2. Construction du corpus « empirique concentrique »

- 6 Un corpus empirique de référence doit son opérationnalité et sa pérennité essentiellement au plan d'outils théoriques et méthodologiques définis et à la qualité et quantité des matériaux collectés et systématisés. Car à la différence d'une base de données – correspondant à un rassemblement ou un réservoir de matériaux –, un corpus suppose un projet scientifique sous-tendant. Autrement dit, avant d'entamer des recherches de terrain, il importe de bien déterminer 1) l'objet d'étude pour lequel on souhaite recueillir des données, 2) les approches préconisées, 3) le terrain géographique, démographique et social sur lequel on souhaite effectuer ses enquêtes, et 4) la façon dont on envisage faire la collecte des matériaux. Ces paramètres non seulement garantissent le succès du projet mais déterminent également la nature des résultats obtenus car la manière dont on constitue un corpus a des répercussions définitives sur la manière dont on analyse les données. Ainsi, par exemple, pour l'analyse de l'énoncé non verbal, il convient de tenir davantage compte des aspects suprasegmentaux (p. ex. intonation) et extralinguistiques (contexte d'énonciation) en raison de son fort ancrage dans la situation immédiate (Lafkioui 1999, 2001, 2006a, 2011 : 23-92). Bien qu'il soit nécessaire de prédéfinir le format de base du corpus de référence à réaliser, cela n'implique cependant pas que ce format exclue des changements au niveau des outils méthodologiques ou théoriques présélectionnés. Il arrive que les phénomènes attestés sur le terrain obligent l'enquêteur à modifier l'approche adoptée ou même parfois à abandonner certains aspects de l'objet d'étude choisi. La constitution d'un corpus empirique est donc un

*processus concentrique continu* qui permet tout ajout ou changement ultérieur contribuant à la compréhension des formes et des contenus des langues étudiées (voir Figure 1).

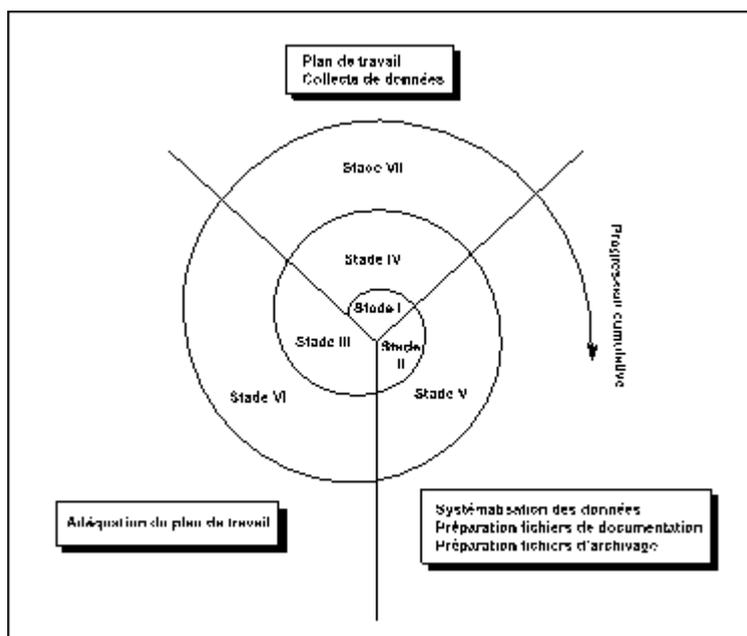


Figure 1. Le corpus empirique concentrique

- 7 Voici les détails concernant les stades du processus continu qui constituent le corpus empirique concentrique :

Tableau 1. Stades du corpus empirique concentrique

<b>Stade I</b>	Plan de travail (p. ex. approches, outils, protocoles)
	Collecte des données 1= collecte préliminaire
<b>Stade II</b>	Systématisation des données = transcription, annotation, traitement et classification
	Mise en place de la partie documentation (données et métadonnées)
	Mise en place de la partie archivage (en formats standards)
<b>Stade III</b>	Adéquation du plan de travail
<b>Stade IV</b>	= Stade I mais de niveau plus avancé = Mise à jour 1 = collecte des données 2 (détaillée)
<b>Stade V</b>	= Stade II mais de niveau plus avancé
<b>Stade VI</b>	= Stade III mais de niveau plus avancé

<b>Stade VII</b>	= Stades I et IV mais de niveau encore plus avancé = Mise à jour 2 = collecte des données 3 (davantage détaillée)
...	...

- 8 Le corpus empirique comporte aussi, en dehors des données linguistiques recueillies, les « métadonnées » pertinentes (voir Stade II de Tableau 1), comme par exemple les profils sociaux et démographiques des informateurs (ou consultants) ainsi que leur répertoire linguistique<sup>3</sup>. Ces métadonnées, conservées de manière systématique, permettent non seulement de mieux gérer le corpus – surtout lorsqu’il est dynamique et donc susceptible d’être mis à jour – mais aussi de mieux comprendre les phénomènes étudiés, dont notamment ceux conditionnés par la situation immédiate. Et lorsque ce corpus est digitalisé et systématisé, les métadonnées contribuent également à la recherche et l’exploitation systématiques des données collectées moyennant des robots confectionnés à ces fins (comme p. ex. Search Indexer) ; ces fonctions automatiques sont indispensables à la géolinguistique computationnelle.
- 9 Parmi les nombreux éléments que la partie métadonnées peut contenir (p. ex. localisation géographique, contexte interactif, date d’enregistrement, protocole, historique des changements succincts, outils d’élicitation, genres de texte, types d’enquête, commentaires extralinguistiques), elle comprend nécessairement une section consacrée aux types de consentement obtenus pour l’exploitation et la diffusion des données collectées. Car sans autorisation des informateurs – éventuellement après anonymisation des documents –, les matériaux ne peuvent être communiqués en partie ou intégralement au public.
- 10 Comme je montrerai plus loin dans cette contribution, le corpus empirique concentrique se donne aussi comme objectif d’être représentatif par rapport au projet défini, et ce à plusieurs niveaux. De ce fait, le corpus géolinguistique tient sa validité à sa représentativité au niveau scientifique (données et approches), géographique et démographique (variation et localisation horizontales), et social (variation verticale). La diversification des paramètres linguistiques et extralinguistiques est une condition *sine qua non* pour la constitution d’un corpus géolinguistique représentatif ; elle permet en outre que les phénomènes moins récurrents y soient aussi représentés.

Tableau 2. Plans de représentativité du corpus géolinguistique

<b>Plan scientifique</b>	Qualité et quantité des données
	Adéquation des approches
<b>Plan géographique et démographique</b>	Variation et localisation horizontales
<b>Plan social</b>	Variation verticale (p. ex. classe sociale et niveau d’instruction)

- 11 Dans les sections suivantes, je mettrai au jour comment un corpus empirique de type géolinguistique est construit (à l’égard des stades de Figure 1) par le biais des matériaux

recueillis durant mes nombreuses enquêtes de terrain en Afrique du Nord. Les méthodes et données présentées dans cet article proviennent principalement de mes travaux géolinguistiques réalisés dans le cadre de l'*Atlas linguistique des variétés berbères du Rif* (ALR, Lafkioui 2007).

- 12 Cet Atlas a comblé un vide dans la recherche en géolinguistique berbère car depuis la dernière publication d'André Basset, « Articles de dialectologie berbère », en 1959, aucune étude systématique n'avait été réalisée dans ce domaine. Il est, en outre, novateur parce que la région du Rif (Maroc du Nord) n'avait jamais fait l'objet d'une étude de géolinguistique et parce que les enquêtes de terrain menées dans le cadre de cette recherche sont de types variés (dirigé, semi-dirigé et non dirigé ; voir § 2.1. ici-même). Par ailleurs, ces enquêtes ont montré la nécessité d'inclure, à part égale, tous les domaines linguistiques (phonétique, phonologie, morphologie, syntaxe, lexique) intervenant dans la construction d'une étude géolinguistique comparative. Ce choix théorique et méthodologique – qu'on retrouve aussi dans le *Nouvel Atlas linguistique de Basse-Bretagne* (NALBB, Le Dû 2001), par exemple<sup>4</sup> – n'avait jamais été adopté dans les études de géolinguistique berbère généralement limitées au domaine du lexique. Enfin, il importe de signaler que l'ALR comporte également un nombre considérable de commentaires explicatifs d'ordre synchronique et diachronique.

## 2.1 Terrain d'enquête : le Rif

- 13 Le Rif est la région septentrionale du Maroc délimitée par la mer Méditerranée au nord, l'océan Atlantique à l'ouest, l'Algérie à l'est et le Moyen Atlas au sud. Il contient principalement deux aires majoritairement berbérophones : le petit isolat géolinguistique de Ghomara (Camps & Vignet-Zunz 1998 ; Colin 1929) et le territoire étendu bordé à l'ouest par Ktama, à l'est par les Iznasen (vers la frontière algérienne), et au sud par Gersif, dernier point géographique rifain avant le corridor de Taza (voir Figure 2). C'est le « rifain » qui est pratiqué sur ce dernier territoire, dénomination représentant un « continuum géolinguistique » de variétés berbères. Ce continuum empêche toute classification linguistique précise en raison de la variation graduelle et hybride des formes et fonctions dont il est composé (Lafkioui 2007, 2008, 2009). J'ai cependant constaté que plus on se dirige vers l'ouest, plus on détecte des traits appartenant au groupe « senhaja » (également dégagés au Maroc central), alors que vers l'est ce sont les traits « zénètes » qui prédominent, traits qu'on repère aussi ailleurs en Afrique du Nord comme dans l'Aurès en Algérie, à Djerba en Tunisie et à Zouara en Libye (p. ex. la palatalisation des vélaires et l'absence conditionnée de la voyelle préfixale). Il importe toutefois de remarquer que les notions ethno-linguistiques de « senhaja » et de « zénète » posent problème car non seulement elles ont tendance à disparaître dans les pratiques langagières des locuteurs berbérophones mais elles peuvent être aussi fort critiquées, ce qui est notamment le cas de « senhaja ». Dans cette contribution, c'est de ce continuum géolinguistique dont il est question et c'est à ce phénomène dans sa globalité et sa complexité que renvoie la notion de « rifain » (ou sa variante néologique berbère « tarifit »<sup>5</sup>) que j'utilise<sup>6</sup>.

## 2.2 Types d'enquête de terrain

- 14 Au vu de la variation linguistique – particulièrement marquée – du tarifit et de la complexité de la problématique, l'élaboration d'un ensemble structuré de méthodes adaptées au terrain géolinguistique du Rif était la démarche la plus appropriée pour obtenir un corpus géolinguistique représentatif. Pour ce faire, trois méthodes de collecte des données (notamment sonores) ont été employées (Lafkioui 2007) : enregistrements dirigés, enregistrements semi-dirigés et enregistrements non dirigés.
- 15 *Enregistrements dirigés.* Ce genre d'enregistrement fait appel à des questionnaires élaborés en fonction de la nature linguistique des variétés et de l'objet d'étude. Par « questionnaire », j'entends tout outil d'élicitation de données linguistiques, servant à la description et à la comparaison des variétés linguistiques. Il inclut des outils plus connus dans le milieu de la linguistique de terrain, tels que le questionnaire écrit, dont font partie par exemple l'élicitation par traduction (vers la variété cible et inversement) et l'élicitation moyennant des questions analytiques, mais il englobe aussi ce qu'on appelle de plus en plus les « stimuli », que je subdivise en stimuli « statiques » (p. ex. photos et desseins ; voir p. ex. Bouquiaux & Thomas 1971) et stimuli « dynamiques » (p. ex. jeux et vidéos, voir p. ex. Chafe 1980). Parmi les nombreux outils d'élicitation employés de nos jours, figurent ceux que met à disposition l'institut Max Planck, dont ceux du groupe *Language and Cognition* qui sont regroupés dans *L&C Field Manuals and Stimulus Materials* (<http://-fieldmanuals.mpi.nl/>). Un outil fréquemment utilisé en linguistique comparative et historique est la liste de *Swadesh*, qui contient un ensemble de concepts primaires (lexique de base) permettant de comparer des variétés génétiquement éloignées dans un contexte extra-linguistique relativement « neutre ». Cette liste a été accommodée à plusieurs reprises pour des contextes linguistiques et extralinguistiques diversifiés, comme par exemple le contexte asiatique pour lequel elle a été développée par Matisoff (1978) et qu'on connaît mieux sous le nom de CALMSEA (*Culturally Appropriate Lexicostatistical Model for South East Asia*). La liste de Swadesh a été aussi utilisée sous format ajusté en berbère dans le cadre de l'ALR (Lafkioui 2007).
- 16 Pour les questionnaires de l'ALR, je me suis aussi inspirée de nombreuses études, telles celles de Basset (1929, 1936, 1939, 1942), Chambers & Trudgill (1980) et Gilliéron & Edmont (1902-1910), outre les outils et le savoir-faire professionnel de l'enregistrement et de l'observation acquis durant mes autres expériences d'enquête de terrain et d'étude en linguistique et en ethnographie (depuis 1992).
- 17 Au fur et à mesure que j'ai progressé dans la collecte et l'exploitation des matériaux, des adaptations et des affinements de ces questionnaires se sont imposés. Car, dans une région comme le Rif, très peu explorée à l'époque (surtout le rifain occidental et méridional), des situations inattendues – linguistiques, mais aussi géographiques et démographiques – ont sensiblement altéré le contenu et la forme du questionnaire de base (voir aussi Pop 1950 : 1075). Par exemple, il a fallu développer davantage la section du questionnaire consacrée aux verbes statifs car plusieurs particularités ont été dégagées en rifain occidental, notamment au niveau de leur conjugaison et de leur rapport supplétif avec le nom adjectival.
- 18 Il va sans dire que le passage par une langue seconde (arabe dialectal, français ou espagnol) a présenté une utilité considérable (surtout pour l'élicitation par traduction), dans la mesure où il a permis de minimaliser l'influence (linguistique et autre) possible de

l'enquêteur sur l'enquêté (*observer's paradox*). Ainsi, l'arabe marocain a été d'une grande aide dans ces zones du Rif qui sont encore bilingues (berbère-arabe), bien que le berbère y soit menacé par l'arabe marocain qui le remplace à une vitesse vertigineuse, notamment chez les Ghomara et dans les zones frontalières ouest (Figure 2 : n° 1 à 13), est (Figure 1 : n° 26-27) et sud (Figure 1 : n° 30 à 32) du continuum rifain. L'espagnol a également été d'intérêt pour ce genre d'enregistrement parce que le rifain est jusqu'à nos jours en contact étroit avec cette langue surtout par le biais de ses locuteurs résidant ou fréquentant les enclaves espagnoles de Melilla (*\*(ta)mlilt* ou *mrič* 'blanche' en rifain) et de Ceuta (*septem fratres* en latin) – ayant le statut de villes autonomes depuis 1995 – qui forment des centres de commerce transfrontaliers importants pour la région du Rif et même pour le reste du Maroc. Les nombreux immigrés résidant légalement ou clandestinement en Espagne ainsi que les touristes espagnols visitant régulièrement le Rif contribuent activement au plurilinguisme qui caractérise les locuteurs rifains, dont notamment ceux habitant les zones urbaines ou périurbaines.

- 19 *Enregistrements semi-dirigés*. Ceux-ci sont réalisés par le biais des interviews dans lesquelles des questions ouvertes traitant de thèmes divers sont posées. Ces questions peuvent être aussi soutenues par des images statives et dynamiques. Le but principal de ce genre d'interview est de laisser plus d'espace aux informateurs pour l'élaboration narrative ou conversationnelle tout en préservant le gabarit de l'entretien, ce qui permet d'obtenir des données relativement ciblées et donc aptes à la comparaison intralinguistique et interlinguistique.
- 20 Deux paramètres sont particulièrement importants pour l'usage des stimuli, et notamment pour ceux de nature visuelle et dynamique (vidéos, jeux) : 1) il convient d'accommoder le stimulus présélectionné au contexte extralinguistique en question (contexte naturel et culturel). Dans certains cas, il vaut mieux le créer sur place avec des locuteurs natifs dans un contexte ethnographique approprié (« ad hoc stimuli », Lüpke 2010). Ainsi pour le Rif, il était parfois judicieux de filmer des images sur place, surtout quand il s'agissait de la description des actions, comme par exemple la cueillette des figues. 2) Il importe aussi de bien spécifier le mode de passation des stimuli afin que les informateurs comprennent comment utiliser les images et les paroles. Ces explications sont préférablement données dans la langue des informateurs. Une familiarisation aux supports d'enregistrement utilisés est également essentielle au bon déroulement de la collecte des données.
- 21 *Enregistrements non dirigés*. Dans ce type d'enregistrement, la primauté est donnée aux expressions verbales et non verbales spontanées, tout en tenant compte de leur diversité formelle et fonctionnelle. L'observation-participante dans des contextes sociolinguistiques variés (familles, associations, espaces publics) fait également partie des principaux outils ethnographiques employés pour ces enregistrements. L'ethnographie – au sens que lui attribue Hymes (1972) – a l'avantage majeur de prendre en compte un grand nombre de variables par le biais d'une corrélation entre les variables sociales et les variables linguistiques dans un cadre multidimensionnel. Bien qu'il soit difficile de classifier le langage humain sans perdre certains aspects linguistiques et socioculturels de l'événement langagier en question, la classification de Hymes (1972) qu'il présente sous l'acronyme *SPEAKING* (*Setting, Participants, End, Act, Key, Instrumentalities, Norms, Genre*) a certainement son utilité. Elle permet non seulement de comprendre et d'aménager l'interaction (monolocution et interlocution) – ne serait-ce que de façon préliminaire –

mais elle facilite aussi des futures recherches au sein du corpus, pourvu que ses éléments ethnographiques soient soigneusement insérés dans la rubrique des métadonnées.

- 22 C'est la combinaison de ces trois méthodes d'enquête – ayant prouvé leur utilité à différents niveaux (Lafkioui 2007, 2011 ; Lüpke 2010) – qui a fourni un corpus représentatif et précis nécessaire pour comprendre les phénomènes géolinguistiques étudiés dans l'ALR. Les données obtenues ont été contrastées, tout d'abord entre elles au plan rifain et, par la suite, aussi au plan berbère et interlinguistique.
- 23 Les enquêtes de terrain de l'ALR ont été effectuées en quatorze mois étalés sur la période été 2001/été 2003. Elles ont été réalisées en plusieurs phases (trois au minimum), dont la sélection des points d'enquête accompagnée d'enregistrements préliminaires constitue la première phase. Dans la seconde, tout le programme d'enregistrements – incluant les différents types d'enquête – a été accompli. La dernière phase a servi de vérification des données collectées afin de les compléter et de les affiner.
- 24 C'était important d'essayer d'obtenir, à chaque séance, le programme d'enquête entier du même informateur. Mais ce n'était pas évident car cela dépendait fortement des capacités mentales et physiques de l'informateur, de sa disponibilité et du contexte d'enregistrement. Ainsi, les différentes parties du programme ont été parfois répandues sur plusieurs jours. Une version réduite du programme a été aussi utilisée dans les séances préliminaires et les séances de vérification ou de complétion.

### 2.3 Critères d'enregistrement

- 25 Les paramètres auxquels il conviendrait de prêter une attention particulière pendant les enregistrements concernent les informateurs<sup>7</sup> – employés ici au sens de « consultants », soulignant l'aspect interactif entre l'enquêteur et l'enquêté – et le contexte d'enquête extralinguistique.
- 26 Pour l'ALR, il était primordial de veiller non seulement à ce que les deux sexes, les différents âges et les différents statuts socio-économiques des informateurs soient représentés pour chaque point d'enquête, mais aussi à ce que leur origine native soit garantie. Outre leur ascendance, leur mobilité extra-locale a été vérifiée afin d'exclure au mieux les effets de contact et donc d'assurer la représentativité (voir aussi Basset 1959 : 109 ; Lafkioui 2007 ; Pop 1950 : 1159 à ce propos). Il importe de signaler que la réalisation des enquêtes n'aurait pas réussi autant sans répertoire linguistique commun – c.-à-d. compétences linguistiques et savoir-faire interactif – entre l'enquêteur et les enquêtés.
- 27 En ce qui concerne le contexte d'enquête, la différence s'est surtout située sur le plan formel (p. ex. école, travail, association) versus informel (p. ex. contexte familial, entre amis), bien qu'il soit dans certains cas difficile de trancher nettement entre ces deux plans. Car le degré de formalité/informalité est fort variable suivant le contexte en question. Ainsi, un simple dîner de famille, en présence du père, peut être considéré et vécu comme un contexte fort formalisé auquel chaque membre de famille participe de manière spécifique.

### 2.4 Sélection des points d'enquête

- 28 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<sup>8</sup>. 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 mes connaissances empiriques et scientifiques des différentes variétés parlées au Rif<sup>9</sup>.

- 29 D'autres approches ont aidé à trouver sur le terrain des points d'enquête indicatifs et à vérifier les conclusions premières tirées des inventaires linguistiques. L'une des plus centrales est la conscience de la variation linguistique des locuteurs, appelée aussi « conscience dialectale ». Elle n'a toutefois pas servi d'outil méthodologique de base pour la démarcation dialectale, comme en font certains dialectologues tels Daan (1999 : 20) car elle prend principalement appui sur l'expérience subjective des locuteurs dont les valeurs quantitatives – étant difficiles à mesurer – dépendent fortement de leurs dispositions individuelles (idéologiques, sociales, historiques). Même si les locuteurs se considèrent comme appartenant au même groupe linguistique – en se dénommant par le même anthroponyme par exemple – leur attitude linguistique est souvent motivée par le statut social accordé aux variétés en question et par le degré de solidarité à l'intérieur du groupe (Lafkioui 1998, 2006b). Par ailleurs, il n'est pas sans importance de mentionner que l'usage d'un anthroponyme berbère n'implique pas nécessairement une pratique linguistique de la langue berbère dans la région en question.
- 30 Un autre concept extralinguistique qui a été employé, avec précaution tout de même, comme critère subsidiaire dans la sélection des points d'enquête, est l'« intercompréhension », concept valorisé par Ameur (1990) mais rejeté par Durand (1992) dans leur essai de classification du berbère. Son inconvénient réside principalement dans le fait que sa validité dépend essentiellement de la volonté consciente ou inconsciente des locuteurs, ce qui ne permet pas toujours de savoir si elle découle réellement de leurs seules compétences linguistiques. Il n'est d'ailleurs pas rare de rencontrer des locuteurs qui prétendent ne pas comprendre, parce qu'ils refusent de comprendre, comme le signalent Chambers & Trudgill (1980 : 4) en disant « People, it seems, sometimes do not understand because, at some level of consciousness, they do not want to ».
- 31 Les points d'enquête présentés finalement dans l'ALR correspondent aux localités où la variation est la plus accusée et où elle permet la comparaison linguistique de la meilleure manière. Les *cent quarante et un* points géoréférencés – appartenant à trente-deux groupes rifains (appelés encore « tribus » dans le Rif) – ont été répertoriés avec leur code de localisation systématique au-dessous de la carte qui les présente (Lafkioui 2007 : 15 ; voir aussi ici Figure 3).

## 2.5 Systématisation des données

- 32 Dans le souci permanent de représentativité et d'exhaustivité, il était important de prendre en considération, autant que peut se faire, le *caractère oral* (multimodal) des textes lors de la transcription (et éventuellement de la traduction) des matériaux recueillis ; c'est-à-dire, toutes les dimensions (verbales, para-verbales et non verbales) accompagnant leur production effective, même si elles s'avèrent souvent difficiles ou laborieuses à rendre en graphie.
- 33 A cette fin, j'ai utilisé des systèmes de transcription et d'annotation existants, dont j'ai ajusté les applications aux besoins de mon corpus, tels p. ex. ELAN (*Eudico Linguistic*

*Annotator*, MPI) qui permet de réaliser plusieurs tâches de systématisation, comme l'annotation, le glosage interlinéaire et la conversion des données. L'outil Toolbox (la suite de Shoebox) est particulièrement efficace pour créer des textes interlinéaires. Des logiciels comme PRAAT (Université d'Amsterdam) pour l'analyse instrumentale des données sonores ont l'avantage d'être non seulement fort opérationnels mais aussi entièrement gratuits.

- 34 Tous les matériaux collectés dans le cadre de l'ALR ont été digitalisés. Un grand nombre a été systématisé selon les besoins d'analyse (p. ex. descriptifs/explicatifs, synchroniques/diachroniques, qualitatifs/quantitatifs). La systématisation en classes grammaticales et listes lexicales issues des analyses synchroniques et diachroniques était primordiale. Afin de mieux appréhender les phénomènes linguistiques rifains et de justifier certains choix de présentation, des analyses comparatives avec les données berbères d'autres régions nord-africaines se sont imposées. Plusieurs textes transcrits ont été sélectionnés et accompagnés d'une glose.

## 2.6 Documentation et archivage

- 35 Comme il a été expliqué au début de cet article, les métadonnées contribuent de manière fondamentale à une meilleure documentation des phénomènes étudiés ainsi qu'à une meilleure préservation des données collectées. De plus, leur sauvegarde de manière systématique (notamment en format numérique) facilite la comparaison synchronique et diachronique, en raison de la nature cumulative du processus concentrique et continu qui forme le corpus empirique (voir Figure 1).
- 36 Un des outils les plus usuels pour gérer numériquement les métadonnées est IMDI de Max Planck Institute (<http://www.mpi.nl/IMDI><sup>10</sup>), que j'ai utilisé pour certains sous-corpus de l'ALR et que j'utilise actuellement dans d'autres projets, ensemble avec un autre outil couramment employé en linguistique, à savoir OLAC Metadata Standard du réseau international OLAC (*Open Language Archives Community*). Un atout principal de ces deux outils est qu'ils sont conçus en format standard, ce qui les rend plus aptes à la procédure d'archivage, étape indispensable dans la constitution d'un corpus numérique. En fait, il est important d'employer des formats standards – formats bien diffusés et ayant prouvé leur validité – afin de garantir au mieux l'accessibilité des données numériques, leur exploitation ainsi que leur perpétuité. Mais bien qu'il existe des formats standards d'archivage, rien n'assure pour l'instant leur continuité en raison des problèmes de « portabilité » (Bird & Simmons 2003), c'est-à-dire l'aptitude d'un format numérique à être exécuté dans des logiciels différents et sur des ordinateurs de types différents sans modification considérable des données archivées.

## 3. Comment réaliser une carte géoréférencée digitale ?

- 37 Afin de pouvoir rendre de manière représentative la variation géolinguistique accusée en rifain, il était indispensable de concevoir une carte géoréférencée exacte des groupes (tribus) et villages berbérophones du Rif permettant, en outre, des traitements automatiques, car aucune carte de ce genre n'existait ou, du moins, n'était accessible au public<sup>11</sup>. Pour ce faire, j'ai utilisé des cartes topographiques et j'ai dessiné des nouvelles cartes par le biais des logiciels informatiques d'image, tels Qcad (actuellement LibreCad), CorelDRAW et InkScape. Ensuite, j'ai manipulé ces cartes afin d'y intégrer les faits

linguistiques analysés (procédé de « mapping ») moyennant des logiciels spécialisés en applications géographiques et des logiciels de traitement de données (systématisation et conversion) dont plusieurs étaient faits sur mesure (Lafkioui 2007). Ainsi, l'ALR a été entièrement conçu selon les méthodes techniques géolinguistiques les plus avancées, toutes soutenues par des logiciels spécialisés et adaptés aux exigences linguistiques spécifiques du berbère.

- 38 Sur cette carte numérique du Rif, j'ai délimité géographiquement les groupes berbérophones et je les ai indexés, comme le laisse voir la reproduction adaptée qui suit (Lafkioui 2007 : 14) :

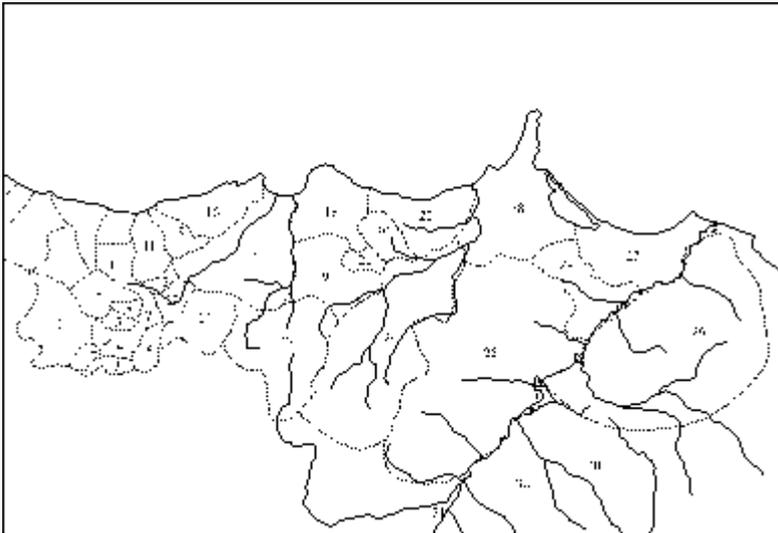


Figure 2. Carte des groupes berbérophones du Rif (d'après Lafkioui 2007)

39 **Noms des groupes berbérophones du Rif**

- 40 1 Ktama 12 Targist 23 Igzennayen  
 41 2 Tağzut 13 Ayt Mezduy 24 Ibḍalsen  
 42 3 Ayt Bušibet 14 Ayt Ğammart 25 Ayt Buyeḥya  
 43 4 Ayt Ḥmed 15 Ayt Iṭṭeft 26 Iznasen  
 44 5 Ayt Bunsar 16 Ibeqquyen 27 Ikebdanen  
 45 6 Ayt Bšir 17 Ayt Weryaḡel 28 IqelĖiyen  
 46 7 Zerqet 18 Ayt Tamsaman 29 Wlad Settut  
 47 8 Ayt Ḥennus 19 Ayt Tuzin 30 Ayt Buzeggu  
 48 9 Ayt Seddat 20 Ayt Wlišek 31 Gersif  
 49 10 Ayt Gmil 21 Tafersit 32 Tawrirt  
 50 11 Ayt Bufraḡ 22 Ayt SĖid  
 51 J'ai également reporté la sélection des points d'enquête sur cette carte numérique, comme le montre la figure suivante extraite de l'ALR (Lafkioui, 2007 : 15) :

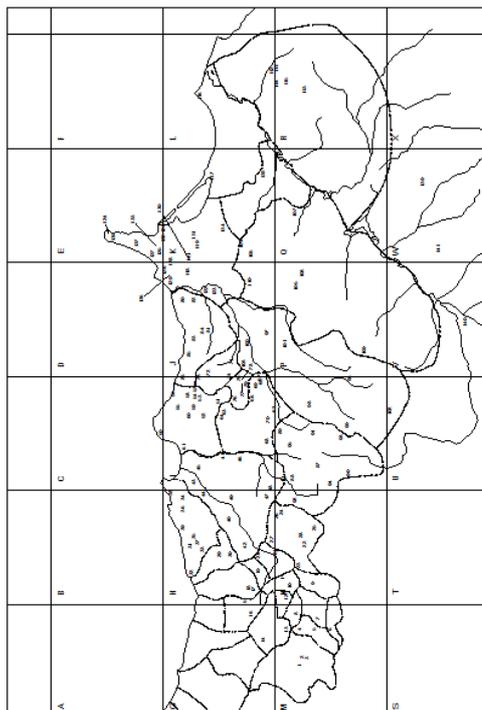


Figure 3. Carte des points d'enquête géoréférencés du Rif (d'après Lafkioui 2007)

## 4. Le procédé de « mapping »

- 52 Après avoir réalisé la carte des groupes des variétés berbérophones (Figure 2) et celle des points d'enquête (Figure 3) – toutes deux accompagnées de leur index systématique –, a suivi le procédé de « mapping », qui consiste à articuler les faits linguistiques sur les points géographiques choisis et pour lequel le logiciel de « Mapinfo Professional » a été employé.
- 53 La première étape de ce procédé était de numériser les matériaux suivant le format du logiciel « Mapinfo Professional », ce qui a nécessité leur systématisation et conversion moyennant Excel, Access et de nombreux logiciels informatiques sur mesure, tels par exemple :

ceux conçus en Mapbasic pour générer les cartes systématiquement et pour la représentation (symbole et couleur) des phénomènes linguistiques ;  
ceux conçus en Delphi pour l'édition des légendes (UlegendEditor) et pour la création des workmaps (1 workmap = 1 carte ; GenWorkMaps).

- 54 Ceci a abouti à un ensemble de matrices des données en format numérique. En voici un échantillon de l'ALR (format de Mapinfo Professional) dans le tableau suivant :

Tableau 3. Extrait des données numériques de l'ALR

Image 20000009000043AA00003CD172172D23.wmf

- 55 Le résultat final de cette systématisation numérique des données et de leur articulation cartographique est un ensemble de 972 cartes géolinguistiques présentant des phénomènes divers de la linguistique berbère (rifain). J'ai sélectionné 356 cartes pour la publication de l'ALR (Lafkioui 2007).
- 56 Chacune de ces cartes porte un titre qui introduit le phénomène traité dans la carte en question. La notation en majuscule dans l'exemple, présenté comme cas de figure dans le titre, renvoie à la position du phénomène analysé. Par exemple :
- 57 Carte 51 : Spirantisation dynamique de la vélaire simple /g/, le cas de /aGuGil/ 'orphelin' (Lafkioui 2007 : 55 ; voir ici-même Figure 4).
- 58 La majuscule « G » indique que cette forme de spirantisation traitée dans la carte 51 concerne uniquement les deux vélares du lexème /aGuGil/. Tout autre phénomène compris dans cet exemple n'est pas examiné, sauf s'il est en rapport direct avec la spirantisation. Dans le présent cas, il est aussi question de la mutation consonantique de la liquide /l/, pour laquelle je renvoie au chapitre de l'ALR (Chapitre 2, Section 4, p. 69-76) portant sur cette problématique. Les formes mutées du /l/ ont toutefois été insérées dans le corps du texte durant l'analyse de ce phénomène de spirantisation.
- 59 A propos de la spirantisation des deux vélares de /-aGuGil/, la première peut faire l'objet d'une atténuation de sa tension articulaire, d'une palatalisation ou d'une labialisation, tandis que la seconde est généralement palatalisée en /ž/ ([ʒ]). Ainsi, on observe les variantes suivantes en rifain<sup>12</sup> :

**Tableau 4. Variantes géolinguistiques du lexème /aGuGil/ 'orphelin' en rifain**

Variantes	Variétés géographiques
/agugil/ - /igigil/	Ayt Bušibet
/agužil/	Ktama, Tağzut, Ifeṭṭušen (Ibḍalsen)
/aməḥžur/	Ayt Seddat, Ayt ḥennus, Ayt ḥmed, Ayt Bunsar, Zerqet
/abužil/	Ayt Bšir
/abužir/ - /aḅužir/	Ayt Mezduy
/abužil/ - /abužir/	Ayt Bufraḥ
/abužir/ - /ayužir/	Targist
/agužir/	Ayt Gmil
/aḅužir/	Ayt Iṭteft, Ayt Weryaḡel, Ayt Ḥammart
/aḅužir/ - /aḅužiṛ/	Ibeqquyen
/ayužir/ - /ayužiṛ/	Igzennayen du Nord, Ayt Wlišek, certaines parties des Ayt Sēid de l'Est, Iqelēiyen
/ayužir/	Ayt Temsaman, Ayt Tuzin, certaines parties des Ayt Sēid, Tafersit

/ayužiř/	Certaines parties des Igzenayen
/agužiř/	Ibdalsen, certaines parties des Ayt Buyeħya
/agužiř/ et /ayužiř/	Certaines parties des Ayt Buyeħya, Gersif
/ayužil/	Ikebdanen, Zayyu, Iznasen, Tawrirt, Ayt Buzeggu

- 60 Notons l'emploi exclusif de l'emprunt arabe *aməħzur* ([{m@ħZU4 dans plusieurs variétés du rifain occidental (voir aussi Figure 4). Une comparaison de ces données avec celles d'autres variétés berbères d'Afrique du Nord permet d'avancer l'hypothèse selon laquelle la deuxième vélaire /g/ – qui n'apparaît que sous forme de /g/ ([g]) ou de /ž/ ([Z]) en rifain – aurait connu les transformations diachroniques suivantes :

\*-/gVgil/ → \*/-gVžil/ (règle de dissimilation)  
 \*/-gVžil/ → \*/-gVgil/ (règle d'assimilation)

- 61 La règle de dissimilation aurait été appliquée à toutes les variétés rifaines, alors que celle d'assimilation qui rétablit l'occlusive /g/ n'aurait été appliquée qu'ultérieurement dans certaines variétés du rifain occidental (Lafkioui 2007 : 55).

## 5. Présentation cartographique de la variation linguistique

- 62 Pour la présentation cartographique de la variation linguistique, j'ai opté dans l'ALR pour des cartes « display » sur lesquelles chaque signe symbolique correspond à une variante précise ou à un groupe de variantes défini. Cette méthode a plusieurs avantages dont notamment 1) l'aptitude de représenter la variation linguistique pour des localités exactes, 2) celle de rendre les cartes directement parlantes et 3) celle d'indiquer les rapports entre les différentes aires géolinguistiques. Afin d'optimiser l'intelligence des cartes, j'ai utilisé des symboles variables en forme et en couleur, et ce en fonction du phénomène linguistique en question, de sa variation (notamment diatopique), de sa complexité et de sa connexion avec d'autres phénomènes. Les variantes qui accompagnent ces signes dans la légende au sein de la carte-même ont été, si besoin est, complétées (p. ex. par des formes idiosyncrasiques) et commentées au plan linguistique (synchronique et diachronique) et sociolinguistique (variation diastratique et diaphasique) en dessous de la carte. De la sorte, les cartes de l'ALR sont également de type « interprétatif » ou « thématique » dans la mesure où elles sont le résultat d'analyse linguistique des faits observés et de leur variation dans l'espace. L'ALR englobe donc à la fois des cartes « sources » géoréférencées et des cartes « thématiques », ce qui est rendu possible grâce au *corpus empirique concentrique numérique* sur lequel il est fondé, permettant en outre un accès aux données et leur mise à jour de façon continue.
- 63 La carte géolinguistique, traitant de la spirantisation dynamique de la vélaire /g/ (voir § 4 ici-même) et extraite de l'ALR (Lafkioui 2007 : 55), en est un échantillon (Figure 4 ci-dessous).

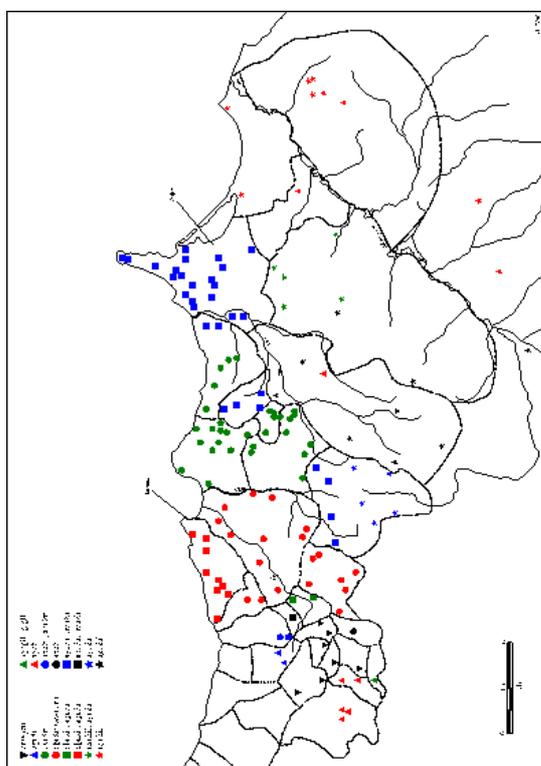


Figure 4. Spirantisation dynamique de la vélaire simple /g/, le cas de /aGuGil/ 'orphelin' (Lafkioui 2007 : 55)

## 6. Conclusion

64 Les modalités méthodologiques présentées dans cet article montrent clairement la nécessité, de plus en plus pressante, d'employer les techniques numériques et les nouvelles technologies de l'information dans les projets géolinguistiques, et ce dès le tout début de la recherche envisagée. Le modèle de corpus que j'appelle « corpus empirique concentrique » (voir Figure 1) et que j'ai présenté dans cette contribution, moyennant des données et des analyses géolinguistiques berbères, en porte la preuve. Durant les stades dynamiques et cumulatifs dont ce type de corpus est composé, le chercheur s'efforce de collecter et de systématiser les données linguistiques, de les accompagner d'une documentation linguistique et extralinguistique appropriée, et de les archiver de manière durable. Ce processus continu permet de réaliser des recherches diversifiées, d'ordre fondamental et appliqué, d'ordre synchronique et diachronique, et d'ordre quantitatif ainsi que qualitatif. Il permet en outre de contribuer au débat épistémologique de la linguistique, et notamment à ses domaines qui prennent appui sur les corpus, tels la géolinguistique.

## BIBLIOGRAPHY

- Ameur M. (1990). « A propos de la classification des dialectes berbères », *Etudes et documents berbères* 7 : 15-27.
- Basset A. (1929). *Etudes de géographie linguistique en Kabylie*. Paris : Leroux.
- Basset A. (1936). *Atlas linguistique des parlers berbères, Algérie, Territoires du nord*, fasc. I, *Equidés*. Alger.
- Basset A. (1939). *Atlas linguistique des parlers berbères, Algérie, Territoires du nord*, fasc. II, *Bovins*. Alger.
- Basset A. (1942). « Etudes de géographie linguistique dans le Sud marocain », *Hespéris* : 3-22.
- Basset A. (1959). *Articles de dialectologie berbère*. Paris : Librairie Klincksieck.
- Bird S. & Simons G. (2003). « Seven Dimensions of Portability for Language Documentation and Description », *Language* 79 : 557-582.
- Bouqiaux L. & Thomas J. (éd.) (1971). *Enquête et description des langues à tradition orale*. Paris : SELAF.
- Bouqiaux L. & al. (1980). *Problèmes de comparatisme et de dialectologie dans les langues africaines*. Paris : SELAF.
- Brugman H., Broeder D. & Senft G. (2003). « Documentation of Languages and Archiving of Language Data at the Max Planck Institute for Psycholinguistics in Nijmegen », Paper presented at the « Ringvorlesung Bedrohte Sprachen », *Sprachenwert – Dokumentation – Revitalisierung*, Fakultät für Linguistik und Literaturwissenschaft, Universität Bielefeld, (05/02/2003).
- Camps G. & Vignet-Zunz J. (1998). « Ghomâra », *Encyclopédie berbère* 20 | Gauda – Girrei, Aix-en-Provence : Edisud, 3110-3119.
- Chafe W. L. (éd.) (1980). *The pear stories : Cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ : Ablex Publishing Corporation.
- Chambers J. K. & Trudgill P. (1980). *Dialectology*. Cambridge : Cambridge University Press.
- Clements N. & Rialland A. (2008). « Africa as a phonological area », in B. Heine & D. Nurse (éd.) *A Linguistic Geography of Africa*. Cambridge : Cambridge University Press, 36-87.
- Colin G.S. (1929). « Le parler berbère des Ghmara », *Hespéris*, 173-208.
- Daan J. C. (1999). « Dialects », in D. R. Preston (éd.) *Handbook of perceptual dialectology*. Vol. 1, 9-30.
- Dalbera J. Ph. (2013). « La trajectoire de la dialectologie au sein des sciences du langage. De la reconstruction des systèmes dialectaux à la sémantique lexicale et à l'étymologie », *Corpus* 12 : 173-200.
- Durand O. (1992). « L'enchevêtrement des parlers berbères », *Rivista di Studi Orientali*, LXV, 3-4 : 185-194.
- Frei H. (1953). *Le livre des deux mille phrases*. Genève : Librairie Droz.
- Gilliéron J. & Edmont E. (1902-1910). *Atlas linguistique de la France*. 9 vol., Paris : Champion.

- Gilliéron J. & Edmont E. (1902). *Notice servant à l'intelligence des cartes*. Paris : Champion.
- Grinevald C. & Bert M. (2010). « Proposition de typologie des locuteurs de LED », *Faits de Langues* (Linguistique de terrain sur langues en danger : Locuteurs et linguistes) 35/36 : 117-132.
- Hymes D. (1972). « Models of the interaction of language and social life », in J. Gumperz & D. Hymes (éd.) *Directions in sociolinguistics : The ethnography of communication*. New York : Holt, Rinehart and Winston, 35-71.
- Ibáñez E. (1944). *Diccionario español-rifeño*. Madrid.
- Kurath H. (1972). *Studies in Area Linguistics*, Bloomington / London : Indiana University Press.
- Kurath H. & Bloch B. (1939). *Handbook of the Linguistic Geography of New England*, Providence : Brown University Press.
- Lafkioui M. (1998). « Les Berbères et leur langue : le cas des immigrés berbères en Belgique », in C. Canut (éd.) *Attitudes, Représentations et Imaginaires en Afrique*. Paris : L'Harmattan-Langues O', 119-130.
- Lafkioui M. (1999). *Syntaxe intégrée de l'énoncé non-verbal berbère*. Thèse de doctorat, Paris : INALCO.
- Lafkioui M. (2001). « Typologie des assertions affirmatives non-verbales du rifain. Approche de syntaxe intégrée », *Frankfurter Afrikanistische Blätter* 13 : 97-120.
- Lafkioui M. (2006a). « Complémentarité syntactico-énonciative entre morphèmes et intonèmes. Le cas du berbère », *Faits de Langues* 27 : 141-149 + références.
- Lafkioui M. (2006b). « Pratiques et représentations linguistiques en contexte multilingue. Le cas des Berbères en Belgique », *Quaderni del Dipartimento di Linguistica (Unical)* 24 : 73-84.
- Lafkioui M. (2007). *Atlas linguistique des variétés berbères du Rif*. Köln : Köppe.
- Lafkioui M. (2008). « Dialectometry Analyses of Berber Lexis », *Folia Orientalia* 44 : 71-88.
- Lafkioui M. (2009). Analyses dialectométriques du lexique berbère du Rif, in D. Ibrizimow, R. Vossen & H. Stroomer (éds.) *Studien zur Berberologie-Etudes Berbères 4 - Essais lexicologiques et lexicographiques*. Köln : Köppe, 133-150.
- Lafkioui M. (2011). *Études de la variation et de la structuration linguistiques et sociolinguistiques en berbère du Rif*. Köln : Köppe.
- Lafkioui M. (à paraître). « Dialectology and Linguistic Geography », in R. Vossen (éd.) *The Oxford Handbook of African Languages*. Oxford : Oxford University Press.
- Laoust E. (1926). « Le dialecte berbère du Rif », in *Rif et Jbala, Bulletin de l'Enseignement Public du Maroc* 71, Paris : Emile Larose, 79-81.
- Le Dù J. (2001). *Nouvel Atlas linguistique de Basse-Bretagne (NALBB)*. Brest : CRBC.
- Lüpke F. (2010). « Research methods in language documentation », in P. K. Austin (éd.) *Language Documentation and Description*, Vol. 7. London : SOAS, 55-104.
- Martinet A. (1967). *Éléments de linguistique générale*. Paris : Armand Colin.
- Matisoff J. A. (1978). « Variational Semantics in Tibeto-Burman : the 'organic' approach to linguistic comparison ». *Occasional Papers of the Wolfenden Society on Tibeto-Burman Linguistics*, Vol. VI. Philadelphia : Publication of the Institute for the Study of Human Issues (ISHI), xviii.
- Möhlig W. J. G. (1980). « L'atlas linguistique du Kenya : méthodes et résultats », in L. Bouqiaux & al. (éd.) *Problèmes de comparatisme et de dialectologie dans les langues africaines*, 57-83.

- Nathan D. & Austin P. K. (2004). « Reconceiving meta-data : Language documentation through thick and thin », in P. K. Austin (éd.) *Language documentation and description*, Vol. 2. London : SOAS, 179-187.
- Orton H. (1960). « An English dialect survey : Linguistic Atlas of England », *Orbis* 9 : 393-397.
- Orton H. (1962). *Survey of English dialects : introduction*. Leeds : E. J. Arnold.
- Pop S. (1950). *La dialectologie. Aperçu historique et méthodes d'enquêtes linguistiques. Dialectologie non romane*. Louvain/Gembloux : Chez L'auteur/J. Duculot.
- Renisio A. (1932). *Etude sur les dialectes berbères des Beni Iznassen, du Rif et des Senhaja de Srair*. Paris.
- Saussure F. de (1973 [1916]). *Cours de linguistique générale*. De Mauro T. (éd.), Paris : Payot.
- Trudgill P. (1975). « Linguistic geography and geographical linguistics », in C. Board & al. (éd.) *Progress in Geography*, Vol. VII. London : Edward Arnold.
- Trudgill P. (1994). *Dialects*. London : Routledge.

## NOTES

1. Ce travail a bénéficié d'une subvention de FWO (Research Foundation- Flanders) et d'une aide de l'Etat gérée par l'Agence Nationale de la Recherche au titre du programme « Investissements d'Avenir » portant la référence ANR-10-LABX-0083.
2. D'autres méthodes de présentation cartographique, en lien direct avec la dialectométrie, ont été présentées dans Lafkioui (2008, 2009).
3. A propos de l'importance des métadonnées, voir p. ex. Nathan & Austin (2004).
4. Outre le *Nouvel Atlas Linguistique de la France* (ou la collection *Les atlas linguistiques de la France par régions*).
5. C'est surtout depuis la reconnaissance nationale du berbère au Maroc que ce terme néologique connaît une diffusion plus large auprès des rifainophones (dont notamment les scolarisés), bien que son usage dans le Rif semble être plus ancien (Laoust 1926 : 80).
6. En revanche, la variété de Ghomara – assez distincte du rifain linguistiquement parlant – est entièrement détachée de ce continuum par des variétés arabes des Jbala, appelées *tajebliit* en berbère ou *jebliyya* en arabe dialectal.
7. En ce qui concerne la typologie des locuteurs (notamment des langues en danger), voir entre autres Grinevald & Bert (2010).
8. A ce sujet, voir entre autres Basset (1929, 1936, 1939, 1942), Bouquiaux *et al.* (1980), Frei (1953), Gilliéron & Edmont (1902-1910) dont leur *Notice servant à l'intelligence des cartes* (1902), Kurath (1972), Kurath & Bloch (1939), Möhlig (1980), Orton (1960, 1962), Pop (1950) et Trudgill (1975, 1994).
9. Cette connaissance a été obtenue de mes nombreuses recherches de terrain depuis 1992.
10. Pour de plus amples informations sur cet outil, voir Brugman, Broeder & Senft (2003).
11. Les seules cartes des tribus rifaines disponibles – non détaillées cependant – étaient celle de Renisio qui remonte à 1932 et celle de Ibáñez de 1944.
12. Voici la notation API des graphèmes particuliers utilisés dans les exemples et les variétés berbères présentés dans cet article : b ([B]), d (d-), ħ ([ĕ]), ħ (x), ġ ([G]), ř ([•], [4], [r]), š (S), t (t-), q ([q]), ε ([ə]), y ([j]), ž ([Z]).

---

## ABSTRACTS

A methodology for conducting geolinguistic research

This article looks at the principal methods that geolinguistic studies require. Geolinguistics is a large research field dedicated to analysing linguistic variation, change and diffusion. Firstly, the article gives a systematic overview of the methodological fieldwork concepts and tools needed in order to construct a corpus of oral (and visual) data which allow for geolinguistic research, including comparative analysis. A number of these concepts could also be applied to other disciplines of Humanities because of their general nature. Secondly, it shows which cartographic methods are best suited to present and represent geolinguistic variation. Data from the Atlas linguistique des variétés berbères du Rif (Lafkioui 2007) are used to illustrate the discussed methods.

Cet article se veut une présentation des méthodes principales que nécessitent les études en géolinguistique, un vaste domaine de recherche consacré à l'analyse de la variation, du changement et de la diffusion linguistiques. Je présenterai d'abord les outils méthodologiques d'enquêtes de terrain indispensables à la constitution d'un corpus de matériaux oraux (et visuels) qui permettent des analyses géolinguistiques, dont également celles de type comparatif. Plusieurs de ces concepts pourraient être appliqués à d'autres disciplines des sciences humaines en raison de leur caractère général. Ensuite, je montrerai quelles sont les méthodes cartographiques les plus appropriées pour présenter et représenter la variation géolinguistique. Des données de l'Atlas linguistique des variétés berbères du Rif (Lafkioui 2007) seront utilisées pour illustrer les méthodes examinées.

## INDEX

**Keywords:** geolinguistics, dialectology, methodology, corpus, linguistic atlas, berber

**Mots-clés:** géolinguistique, dialectologie, méthodologie, corpus, atlas linguistique, berbère

## AUTHOR

**MENA LAFKIOUI**

Université Sorbonne Paris Cité

LLACAN-UMR 8135 du CNRS/Inalco