



HAL
open science

Description des variétés berbères en danger du Sud-Oranais (Algérie) - Étude dialectologique, phonologique et phonétique du système consonantique

Mohamed El Idrissi

► To cite this version:

Mohamed El Idrissi. Description des variétés berbères en danger du Sud-Oranais (Algérie) - Étude dialectologique, phonologique et phonétique du système consonantique. Linguistique. Université Sorbonne Paris Cité, 2017. Français. NNT : 2017USPCF038 . tel-02079768v2

HAL Id: tel-02079768

<https://theses.hal.science/tel-02079768v2>

Submitted on 8 Apr 2019

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.

Institut National des Langues et Civilisations Orientales

École doctorale N°265

Langues, littératures et sociétés du monde

Langues et Cultures du Nord de l'Afrique et Diasporas - EA 4092

Centre de Recherche Berbère

THÈSE

présentée par :

Mohamed EL IDRISI

soutenue le : 08 décembre 2017

pour obtenir le grade de **Docteur de l'INALCO**

Discipline : Sciences du langage : linguistique et didactique des langues

Description des variétés berbères en danger du Sud-Oranais (Algérie) - Étude dialectologique, phonologique et phonétique du système consonantique

Thèse dirigée par :

Kamal NAÏT-ZERAD

Professeur des Universités, INALCO

RAPPORTEURS :

Mohamed LAHROUCHI

Chargé de recherche HDR, CNRS/SFL

Mohamed YEOU

Professeur des Universités, Université Chouaib Doukkali

MEMBRES DU JURY :

Mohamed LAHROUCHI

Chargé de recherche HDR, CNRS/SFL

Kamal NAÏT-ZERAD

Professeur des Universités, INALCO

Brigitte RASOLONIAINA

Maître de conférences HDR, INALCO

Rachid RIDOUANE

Chargé de recherche HDR, CNRS/LPP

Mohamed YEOU

Professeur des Universités, Université Chouaib Doukkali

Cette œuvre est mise à disposition selon les termes de la [Licence Creative Commons : Attribution - Pas d'Utilisation Commerciale - Pas de Modification 3.0 France](https://creativecommons.org/licenses/by-nc-nd/3.0/fr/).



Remerciements

J'aimerais avant tout remercier tous mes informateurs sans qui cette thèse n'aurait pas pu voir le jour. Ils ont accepté gentiment et gracieusement de me consacrer du temps pour élaborer mon corpus. Ils m'ont accueilli les bras ouverts et m'ont considéré comme un des leurs. Je ne pourrai pas malheureusement tous les citer et je m'excuse auprès d'eux si j'ai omis quelqu'un.

À Asla : je remercie tout particulièrement Mustapha Bouzid — président de l'association Ighermawen — qui a été d'un soutien capital lors de notre première enquête de terrain. Mes remerciements vont également en direction de Cheikh Abdenbi, Boujemaa Samadi, Ahmed Bouzid, Abdelqada Bouzid, Tahar Usaid, Abdelkrim Mourid, Mohamed Benhellel, Abdelhadi Mourid, Hmed Mourid, Bachir Abdendi, Mustapha Uhellel, Mohamed Ajaji, Jellul Bekkai et de tous les Imaziyeen n Easla,

À Beni Ounif : j'exprime ma reconnaissance à Hassi Hmed, Hamoda Mohamed, Ismael Slimani, Lhacen Hachlef, Mohamed Houbach, Sasi et à tous les Imaziyeen n At Winifen,

À Berrebi : je tiens à remercier Fodil Benni, Khelifa Zoui, Abdelaziz Bekki, Sliman Moqadem, Mohamed Rehho et tous les Imaziyeen n Berrebi,

À Boukais : je dédie ces remerciements à Abdelallah Alawi, Haj Mbarek, Hocin Kiromi et à tous les Imaziyeen n Boukayis,

À Boussemgoun : je souhaite gratifier Zoulikha Touati, Baci Oulhaj, Abdel Wahab, Ali Saadi et tous les Imaziyeen n Bussemyun,

À Chellala : je tiens à saluer Bachir Bennanan, Ali Hamuliya, Belqasem Mbarki, Lahbib Jebri, Tahar Jebri et tous les Imaziyeen n Cellala,

À Fendi : je rends grâce à Belgourari Aissa et Belgourari Abdelkarim pour leur hospitalité, je n'oublie pas non plus tous les Imaziyeen n Fendi,

À Igli : je suis très reconnaissant à Ali Boulanouar, Abdelqacem Saneba, Othmane Laarbi, Nasser et à tous les Imaziyeen n Igli,

À Kalaat Cheikh Bouamama "Moghrar tahtani" : je suis gré à tous les Imaziyeen n Moghrar n wadday et particulièrement à Moustapha Cherfaoui, Fatmi Tayyeb, Bensallam Mohamed mmi-s n Bachir de leur patience,

À Lahmar : j'exprime toute mon estime à Mohamed Rehho, Mohamed Taybi, Lahcen Taybi et à tous les Imaziyeen n Lahmar,

À Mazzer : je tiens à faire part de ma profonde gratitude à Omar Ghabia, Rabeh Ghanem et à tous les Imaziyen n Mazzer,

À Moghrar "fougani" : j'adresse ma reconnaissance à Qada Aisawi, Qada Tlidji, Baghdadi Aarki, Abdelqader Ziani, Abdelsellem Benacur, Rachid Ziyanni, Smahi Aisawi, Abdellah Tayba et à tous les Imaziyen n Mograr n nnej,

À Mogheul : je suis redevable à Jwima Benaissa, Mohamed Abidi, Boujida Sliman et à tous les Imaziyen n Muyil,

À Ouakda : je témoigne de ma reconnaissance à Anis Abbawi, Bachir, Boutkhil Abbawi, Abdelwahab, Sliman Ubuanan et à tous les Imaziyen n Wagda,

À Sfissifa : je dis merci aux membres de l'Association Tnant et tout particulièrement à Tahar Bouzar, Smain Usaid, Lhaj et à tous les Imaziyen n Tnant,

À Tiout : j'exprime ma gratitude à Hmed Boutrad, Jilali, Guerba Maamar et à tous les Imaziyen n Tyut.

Ces remerciements s'adressent aussi à M. Nait-Zerad qui m'a enseigné la linguistique berbère pendant cinq longues années et qui m'a par la suite encadré pour mon mémoire et pour cette thèse. Je lui suis reconnaissant pour la confiance et la disponibilité qu'il m'a accordées.

Je tiens aussi à remercier les institutions qui ont rendu possible le développement de cette recherche. En premier lieu l'INALCO et l'école doctorale pour avoir financé ce projet.

Du côté algérien, je suis reconnaissant aux responsables du Département de Langue et Culture Amazighes de l'Université de Tizi-Ouzou pour m'avoir invité à plusieurs reprises en Algérie.

Cette reconnaissance s'adresse également aux autorités locales des wilayas de Béchar, Naâma et El-Bayadh.

Je salue chaleureusement le jury qui consultera ce mémoire et je les remercie d'avoir accepté d'en faire partie.

Enfin, je remercie tous ceux qui de près ou de loin ont pu contribuer à la réalisation de ce travail, *ula s wul*.

Table des matières

	Pages
Table des figures	ix
Liste des tableaux	xi
Liste des cartes	xiii
Liste des photographies	xv
Liste des abréviations	xvi
Liste des gloses	xviii
Liste des notations mathématiques	xx
Liste des notations linguistiques	xxii
Transcription usuelle	xxiv
Convention et mise en forme rédactionnelles	xxv
Introduction générale	1
I.1 L'unité d'analyse	2

I.1.1	Le terrain d'étude	3
I.1.1.1	Époque contemporaine	3
I.1.1.1.1	Considérations sur le mode de vie	5
I.1.1.2	Aperçu historique	7
I.1.1.2.1	Période pré-musulmane	7
I.1.1.2.2	Période musulmane - Moyen-âge	9
I.1.2	Population d'étude	11
I.1.2.1	Situation démographique	11
I.1.2.2	Situation ethnique	13
I.1.2.3	Situation sur la démographie linguistique	14
I.1.2.3.1	Groupe septentrional	16
I.1.2.3.2	Groupe orientalo-central	17
I.1.2.3.3	Groupe occidentalo-central	18
I.1.2.3.4	Groupe méridional	19
I.1.3	L'objet d'étude	20
I.1.3.1	Considérations générales sur la langue berbère du Sud-Oranais	20
I.1.3.1.1	Le statut du berbère du Sud-Oranais	20
I.1.3.1.2	Les dénominations endogènes et exogènes du berbère du Sud-Oranais	21
I.1.3.1.3	La phylogénie du berbère du Sud-Oranais	22
I.1.3.2	État de la question des études linguistiques sur le Sud-Oranais	22
I.1.3.3	Délimitation de l'objet d'étude	23
I.1.3.4	Plan général	25
 Étude des variétés linguistiques : Dialectologie et Phonologie		27
II.1	Introduction	28
II.1.1	Terrain d'étude	29
II.1.1.1	Les points d'enquête	30
II.1.1.2	L'enquête de terrain	33
II.1.2	L'objet d'étude	34
II.1.3	Plan	35
II.2	État de l'art	38
II.2.1	La géolinguistique	39
II.2.1.1	La géolinguistique dans les études berbères	40

II.2.1.1.1	André Basset et la Berbérie	40
II.2.1.1.2	Lionel Galand et le Chleuh	42
II.2.1.1.3	Mena Lafkioui et le Rif	43
II.2.1.1.4	Géolinguistique et Kabylie	44
II.2.1.1.5	Géolinguistique et Aurès	45
II.2.1.1.6	Géolinguistique et Sud-Oranais	45
II.2.1.2	La géolinguistique dans le monde	46
II.2.1.2.1	La Variation linguistique et les Variables linguistiques	46
II.2.1.2.2	La Variation linguistique et la Distance géographique	48
II.2.1.2.3	La Variation linguistique et la Politique linguistique	48
II.2.2	Variétés linguistiques et Sciences des données	49
II.2.2.1	Dialectométrie	51
II.2.2.1.1	Travaux dans les études berbères	52
II.2.2.1.2	Travaux dans les autres aires linguistiques	52
II.2.2.2	Méthodes en Sciences des Données	53
II.2.2.2.1	Des Données linguistiques aux Données numériques	54
II.2.2.2.2	Matrice de similarité et Mesure de Distance	55
II.2.2.2.3	Partition des données	56
II.2.2.2.3.1	Partition des données : Apprentissage Non Supervisé	57
II.2.2.2.3.1.1	Non Supervisé et Hiérarchisé	57
II.2.2.2.3.1.2	Non Supervisé et Non Hiérarchisé	58
II.2.2.2.3.2	Partition des données : Apprentissage Supervisé	58
II.2.2.2.4	Réprésentation des partitions	59
II.2.2.3	Géostatistique linguistique	60
II.3	Méthodologie	62
II.3.1	Données et Corpus	63
II.3.2	Population d'étude	65
II.3.3	Matériels de collecte de données	67
II.3.4	Conditions de collecte de données	68
II.3.5	Le protocole	69
II.3.6	Validation des données	73
II.3.7	Métadonnées	73
II.3.8	Traitement des données	75
II.3.8.1	Tableur	75
II.3.8.2	Rstudio	76

II.3.8.2.1	Carte géolinguistique	77
II.3.8.2.1.1	Fonds de carte et Ressources cartographiques	78
II.3.8.2.1.2	Modèle graphique de nos cartes géolinguistiques	79
II.3.8.2.2	Dialectométrie	81
II.3.8.2.2.1	Méthodes de calcul numérique	82
II.3.8.2.2.2	Réprésentation des partitions	83
II.4	Résultats	84
II.4.1	Géolinguistique	86
II.4.1.1	Inventaire phonético-phonologique	86
II.4.1.2	Les variables linguistiques	92
II.4.1.3	Consonnes simples	96
II.4.1.3.1	Le matériau d'analyse	97
II.4.1.3.1.1	La variable G_i : Occlusion et Frication	97
II.4.1.3.1.2	La variable K_i : Occlusion et Frication	102
II.4.1.3.1.3	La variable T : Aspiration et Affrication	107
II.4.1.3.2	Discussion	109
II.4.1.4	Consonnes géminées	119
II.4.1.4.1	Le matériau d'analyse	119
II.4.1.4.1.1	La variable $^{CC}_i$: Frication et Affrication	120
II.4.1.4.1.2	La variable $^{GG}_i$: Frication, Affrication et Occlusion	124
II.4.1.4.1.3	Les variables $^{SS}_i$ et $^{ZZ}_i$: Frication et Affrication	130
II.4.1.4.1.4	Les variables DD , BB et GG : Sonorisation et Assourdissement	133
II.4.1.4.1.5	La variable $^{QQ}_i$: Sonorisation et Assourdissement	135
II.4.1.4.2	Discussion	138
II.4.1.5	Synthèse	145
II.4.2	Dialectométrie	148
II.4.2.1	Les variables dialectales	148
II.4.2.1.1	Problème terminologique	149
II.4.2.1.2	Modèle théorique	151
II.4.2.1.3	Application au berbère du Sud-Oranais	153
II.4.2.2	Apprentissage Non Supervisé	154
II.4.2.2.1	Le matériau d'analyse	154
II.4.2.2.1.1	Partition hiérarchisée ascendante	156
II.4.2.2.1.2	Partition non hiérarchisée	161
II.4.2.2.2	Discussion	165

II.4.2.3	Apprentissage Supervisé	170
II.4.2.3.1	Le matériau d'analyse	170
II.4.2.3.2	Discussion	173
II.4.2.4	Synthèse	175
II.5	Appendice	178

Étude des Rhotiques alvéolaires : Phonétique expérimentale et Phonologie **182**

III.1	Introduction	183
III.1.1	Terrain d'étude	184
III.1.2	Objets d'étude	186
III.1.3	Plan	187
III.2	État de l'art	190
III.2.1	Les rhotiques alvéolaires en berbère	191
III.2.2	Les rhotiques alvéolaires dans le monde	193
III.2.2.1	Problème de classement et de dénomination : Rhotique / Liquide / Tremulant	193
III.2.2.2	Caractéristiques phonétiques des rhotiques alvéolaires	196
III.2.2.2.1	Tap	197
III.2.2.2.1.1	Articulation	197
III.2.2.2.1.2	Durée du contact	198
III.2.2.2.2	Trill	198
III.2.2.2.2.1	Articulation	199
III.2.2.2.2.2	Aérodynamique	200
III.2.2.2.2.3	Nombre et durée des contacts	201
III.2.2.2.2.4	Difficultés d'articulation et Rhotiques alvéolaires	202
III.2.2.2.3	Pharyngalisation et Rhotiques alvéolaires	203
III.2.2.2.4	Différences phonétiques entre les taps et les trills	204
III.2.2.2.5	Accentuation et Rhotiques alvéolaires	205
III.2.2.2.6	Articulation et Contexte	205
III.2.2.2.7	Formants et Rhotiques alvéolaires	207
III.2.2.2.8	Perception et Rhotiques alvéolaires	208
III.2.2.3	Caractéristiques phonologiques des rhotiques alvéolaires	209

III.2.2.3.1	Mutation et Rhotiques alvéolaires	209
III.2.2.3.1.1	Diachronie	209
III.2.2.3.1.2	Dialectologie	210
III.2.2.3.1.3	Allophonie et Neutralisation	210
III.2.2.3.2	Contrastivité entre les trills et les taps	211
III.2.2.3.3	Sonorité et Rhotiques alvéolaires	212
III.2.2.3.4	Homorganicité et Rhotiques alvéolaires	213
III.2.2.3.5	Phonotactique et Rhotiques alvéolaires	213
III.2.2.3.6	Gémination et Taps	214
III.2.2.4	Voisement et Rhotiques alvéolaires	215
III.2.2.5	Vocoïdes et Rhotiques alvéolaires	216
III.2.2.6	Modélisation, Synthèse sonore, Simulation et Rhotique alvéolaire	218
III.3	Méthodologie	219
III.3.1	Corpus	220
III.3.2	Population d'étude	221
III.3.3	Matériel expérimental	222
III.3.3.1	Microphone à électret	224
III.3.3.1.1	Fonctionnement technique	224
III.3.3.1.2	Avantage et inconvénient	225
III.3.3.1.3	Le microphone à électret MM-701	225
III.3.3.2	Enregistreur numérique	226
III.3.4	Conditions d'enregistrement	227
III.3.4.1	La réverbération	228
III.3.4.2	Le rapport signal sur bruit	230
III.3.5	Le protocole	231
III.3.6	Traitement des données	232
III.3.6.1	L'étiquetage et les mesures : PRAAT	232
III.3.6.1.1	L'étape de segmentation	232
III.3.6.1.1.1	Les frontières de la rhotique alvéolaire	234
III.3.6.1.2	Paramétrages pour le traitement du signal	235
III.3.6.1.3	Mesure des propriétés physiques	237
III.3.6.1.4	Présentation des résultats	237
III.3.6.2	Le traitement statistique : RStudio	238
III.3.6.2.1	Pré-traitement statistique des données	238
III.3.6.2.2	Traitement statistique des données	239

III.4 Résultats	240
III.4.1 Analyse phonologique	242
III.4.1.1 Distribution des Rhotiques Alvéolaires	242
III.4.1.1.1 Le matériau d'analyse	243
III.4.1.1.1.1 Tap simple	243
III.4.1.1.1.2 Tap pharyngalisée	246
III.4.1.1.1.3 Trill simple	249
III.4.1.1.1.4 Trill pharyngalisée	252
III.4.1.1.2 Discussion	254
III.4.1.2 Variantes allophoniques	257
III.4.1.2.1 Variantes contextuelles	257
III.4.1.2.1.1 Tap simple et Tap pharyngalisée	258
III.4.1.2.1.2 Tap et Trill	259
III.4.1.2.1.3 Trill simple et Trill pharyngalisée	265
III.4.1.2.2 Variantes libres	267
III.4.1.3 Synthèse	270
III.4.2 Analyse phonétique	272
III.4.2.1 L'articulation primaire	272
III.4.2.1.1 Mode d'articulation de l'articulation primaire	272
III.4.2.1.1.1 Le matériau d'analyse	273
III.4.2.1.1.1.1 Mode d'articulation de la tap	273
III.4.2.1.1.1.2 Mode d'articulation de la trill	276
III.4.2.1.1.2 Discussion	278
III.4.2.1.2 Durée du contact de la tap	280
III.4.2.1.2.1 L'analyse statistique	280
III.4.2.1.2.2 Discussion	282
III.4.2.2 Les vocoïdes du tap	283
III.4.2.2.1 L'environnement des vocoïdes	283
III.4.2.2.1.1 Le matériau d'analyse	283
III.4.2.2.1.1.1 En position initiale	284
III.4.2.2.1.1.1.a Unité phonique : Tap simple	284
III.4.2.2.1.1.1.b Unité phonique : Tap pharyngalisée	285
III.4.2.2.1.1.2 Médiane	285
III.4.2.2.1.1.2.c Unité phonique : Tap simple	286
III.4.2.2.1.1.2.d Unité phonique : Tap pharyngalisée	288
III.4.2.2.1.2 Discussion	288

III.4.2.2.2	Durée des vocoïdes	291
III.4.2.2.2.1	L'analyse statistique	292
III.4.2.2.2.2	Discussion	293
III.4.2.3	Synthèse	296
III.5	Appendice	297
	 Conclusion générale	 303
	 Glossaire	 310
	Index	312
	Bibliographie	320

Table des figures

II.4.1	Partition des variables dialectales — méthode MDS	156
II.4.2	Comparaison des méthodes de partition hiérarchisée ascendante	157
II.4.3	Échec de la partition des variables dialectales — méthode CART	171
II.4.4	Partition des variables dialectales — méthode CART	172
III.3.1	Schéma électrique simplifié du microphone à électret	224
III.3.2	Extrait d'un passage du corpus présentant un chant d'oiseau	228
III.3.3	Exemple de la segmentation de l'item : @[ɜbred]@ \route\	234
III.4.1	Spectrogramme de @[ɜrgɜz]@ \homme\	243
III.4.2	Spectrogramme de @[ɜɜres]@ \froid\	244
III.4.3	Spectrogramme de @[ɜre]@ \écrire{2.SG.IMP}\	244
III.4.4	Spectrogramme de @[ɜr]@ \nouer{2.SG.IMP}\	245
III.4.5	Spectrogramme de @[røz:ɛɜ]@ \chercher{1.SG.AI}\	245
III.4.6	Spectrogramme de @[im:ɜrɜdɜs]@ \étrangler{3.M.SG.PRET}\	246
III.4.7	Spectrogramme de @[tɜɜrɜ]@ \récepteur\	247
III.4.8	Spectrogramme de @[iɜrɜf]@ \geler{3.M.SG.PRET}\	247
III.4.9	Spectrogramme de @[ɜrɜn]@ \poignée\	248
III.4.10	Spectrogramme de @[rɜb:ɛ]@ \dieu\	248
III.4.11	Spectrogramme de @[et:ɜr]@ \demander{3.M.SG.PRET}\	249
III.4.12	Spectrogramme de @[rɜl]@ \mariage\	250
III.4.13	Spectrogramme de @[idre]@ \pousser_énergiquement{3.M.SG.IMP_INT}\	250
III.4.14	Spectrogramme de @[tɜrest]@ \argile_blanche\	251

III.4.15	Spectrogramme de @[ɜ̃ɦrdɜ̃n]@ \uromastyx\	251
III.4.16	Spectrogramme de @[d ^ɛ ʌr ^ɛ]@ \pied\	252
III.4.17	Spectrogramme de @[r ^ɛ z ^ɛ ʌ]@ \casser{1.SG.PRET}\	252
III.4.18	Spectrogramme de @[et:ɔ̃ɔr ^ɛ ʌ]@ \tromper{3.M.SG.AI}\	253
III.4.19	Spectrogramme de @[ʌɦr ^ɛ d ^ɛ ʌn]@ \esclave_affranchi\	253
III.4.20	Spectrogramme de @[ʌzr ^ɛ ɔ̃d ^ɛ]@ \sorte_de_citrouille\	256
III.4.21	Spectrogramme de @[iqr ^ɛ d ^ɛ]@ \fracturer{3.M.SG.PRET}\	262
III.4.22	Comparaison des spectrogrammes de @[ɜ̃ɦ r dɜ̃n]@ \uromastyx\	263
III.4.23	Spectrogramme de @[tisɔ̃z:ɜ̃r/ɛ̃]@ \fourches\	268
III.4.24	Spectrogramme de @[edɔ̃r/ɛ̃:]@ \tourner{3.M.SG.PRET}\	269
III.4.25	Spectrogramme de @[jɔ̃rs]@ \être_créditeur_de{2.SG.IMP}\	274
III.4.26	Spectrogramme de @[ɜ̃brɛ̃ʒn]@ \noir\	275
III.4.27	Spectrogramme de @[ɛ̃ɔ̃ʌt:ern]@ \épinés_de_palmier\	277
III.4.28	Spectrogramme de @[jɔ̃rɔ]@ \rendre{3.M.SG.PRET}\	278
III.4.29	Boîte à moustache de la durée de contact des taps	281
III.4.30	Spectrogramme de @[rɛ̃ʃəm]@ \mouiller{2.SG.IMP}\	284
III.4.31	Spectrogramme de @[r ^ɛ s ^ɛ ən]@ \être_calme{3.M.PL.PRET}\	285
III.4.32	Spectrogramme de @[frfr]@ \s'envoler{2.SG.IMP}\	286
III.4.33	Spectrogramme de @[is:rkos]@ \mentir{3.M.SG.AI}\	287
III.4.34	Spectrogramme de @[ɛ̃rɛ̃]@ \attacher{2.SG.IMP}\	287
III.4.35	Spectrogramme de @[tis:fr ^ɛ t ^ɛ]@ \balai\	288
III.4.36	Boîte à moustache de la durée du Vocoide	293

Liste des tableaux

I.1.1	Nombre d’habitants par Wilaya	11
II.1.1	Comparaison de l’enquête de terrain d’André Basset avec la nôtre.	32
II.4.1	Tableau du système phonético-phonologique	88
II.4.2	Les représentations phonologiques des variables	116
II.4.3	Les formes des variables [^] BB et [^] GG	135
II.4.4	Typologie des formes de surface des géminées	139
II.4.5	La nouvelle taxonomie des variétés linguistiques.	153
II.4.6	Les matrices de confusion des méthodes CART et Random Forest	173
II.5.1	Les items des variables	180
II.5.2	Les gloses des items	181
III.2.1	Durée en ms des cycles des trills dans différentes langues	201
III.2.2	Exemple d’évolution phonétique de la rhotique	209
III.3.1	Profils de nos informateurs.	222
III.3.2	Caractéristiques techniques du microphone à électret.	226
III.3.3	Caractères non X-SAMPA utilisés pour la transcription des items	233
III.3.4	Les paramétrages du spectrogramme utilisés sur PRAAT.	235
III.4.1	Groupe consonantique de taps et de consonnes homorganiques	254
III.4.2	Trills pharyngalisées précédées par une consonne homorganique	255

III.4.3	Les contextes phonétiques de [r ^s]	258
III.4.4	Contraste phonologique entre tap et trill	260
III.4.5	Contextes d'apparition de la trill pharyngalisée	266
III.4.6	Les types de syllabe en médiane	286
III.4.7	Effectif des variables retenues pour la durée des vocoïdes	292

Liste des cartes

I.1.1	Situation géographique et administrative de la région dite du Sud-Oranais	4
I.1.2	Répartition de la population par point d'enquête	12
I.1.3	Répartition en groupe des points d'enquête	15
I.1.4	Répartition entre berbérophones et arabophones - Groupe septentrional	16
I.1.5	Répartition entre berbérophones et arabophones - Groupe orientalo-central	17
I.1.6	Répartition entre berbérophones et arabophones - Groupe occidentalo-central	18
I.1.7	Répartition entre berbérophones et arabophones - Groupe méridional	19
II.1.1	Les points d'enquête	34
II.2.1	Exemple de carte linguistique réalisée par André Basset	42
II.3.1	Le fond de carte des cartes linguistiques	81
II.4.1	La variation de $^G_1 : aG_1em \backslash \text{puiser}\{2.SG.IMP}\backslash$	98
II.4.2	La variation de $^G_2 : G_2ar \backslash \text{entre}\{PREP}\backslash$	99
II.4.3	La variation de $^G_3 : arG_3az \backslash \text{homme}\backslash$	100
II.4.4	La variation de $^G_4 : azG_4za \backslash \text{vert}\backslash$	101
II.4.5	La variation de $^K_1 : ameK_1li \backslash \text{déjeuner}\backslash$	103
II.4.6	La variation de $^K_2 : aK_2sum \backslash \text{viande}\backslash$	104
II.4.7	La variation de $^K_3 : rK_3e\grave{d} \backslash \text{danser}\{2.SG.IMP}\backslash$	105
II.4.8	La variation de $^K_4 : K_4al \backslash \text{terre}\backslash$	106
II.4.9	La variation de $^T : aTbir \backslash \text{pigeon}\backslash$	107
II.4.10	La représentation phonologique de K_2	113
II.4.11	La variation de $^CC_1 : uCC_1um \backslash \text{couscous}\backslash$	120
II.4.12	La variation de $^CC_2 : (tt)reCC_2em \backslash \text{mouiller}\{2.SG.IMP_INT}\backslash$	121

II.4.13	La variation de $^{\wedge}CC_3 : i/aCC_3er \backslash ongle \backslash$	122
II.4.14	La variation de $^{\wedge}CC_4 : eCC_4 \backslash manger\{2.SG.IMP\} \backslash$	123
II.4.15	La variation de $^{\wedge}GG_1 : iGG_1en \backslash un\{NUM\} \backslash$	125
II.4.16	La variation de $^{\wedge}GG_2 : (tt)meGG_2er \backslash moissonner\{2.SG.IMP_INT\} \backslash$	126
II.4.17	La variation de $^{\wedge}GG_3 : tteGG_3 \backslash faire\{2.SG.IMP_INT\} \backslash$	127
II.4.18	La variation de $^{\wedge}GG_4 : GG_4i \backslash guérir\{2.SG.IMP\} \backslash$	128
II.4.19	La variation de $^{\wedge}GG_5 : eGG_5 \backslash laisser\{2.M.SG.IMP\} \backslash$	129
II.4.20	La variation de $^{\wedge}SS_1 : (ke/i)SS_1i \backslash prendre\{2.SG.IMP\} \backslash$	131
II.4.21	La variation de $^{\wedge}ZZ_1 : r(V)ZZ_1a \backslash chercher\{2.SG.IMP_INT\} \backslash$	131
II.4.22	La variation de $^{\wedge}SS_2 : SS_2en \backslash savoir\{2.SG.IMP\} \backslash$	132
II.4.23	La variation de $^{\wedge}ZZ_2 : aZZ_2el \backslash courir\{2.SG.IMP\} \backslash$	133
II.4.24	La variation de $^{\wedge}DD : miDDen \backslash gens \backslash$	134
II.4.25	La variation de $^{\wedge}QQ_1 : meQQ_1er \backslash grandir\{2.SG.IMP_INT\} \backslash$	136
II.4.26	La variation de $^{\wedge}QQ_2 : neQQ_2 \backslash tuer\{2.SG.IMP_INT\} \backslash$	137
II.4.27	Répartition géographique des haplolectes de la Partition Hiérarchisée Ascendante (PHA) — Carte à point	159
II.4.28	Répartition géographique des haplolectes de la PHA — Carte choroplèthe	160
II.4.29	Répartition géographique des haplolectes de la méthode k-mode — Carte à point	162
II.4.30	Nouvelle répartition géographique des haplolectes de la méthode k-mode — Carte à point	163
II.4.31	Répartition géographique des haplolectes de la méthode k-moyenne — Carte à point	164
II.4.32	Nouvelle répartition géographique des haplolectes de la méthode k-moyenne — Carte à point	165
II.4.33	Répartition géographique des haplolectes de la méthode ROCK — Carte à point	167
II.4.34	Relief et variation	169
III.1.1	Les localités parlant le Tamazight de l'Atlas Saharien	185

Liste des photographies

I.1.1	Désert rocheux — Lahmar —	5
I.1.2	Désert sableux — Taghit —	6
I.1.3	<i>L'ayerm</i> de Ouakda	7
I.1.4	Gravure rupestre située à Mograr n wadday	8
I.1.5	Inscription libyco-berbère située à Moghrar n wadday	9
II.1.1	Les notes d'André Basset — Fonds Basset —	30
II.3.1	Séance de travail avec un informateur à Moghrar n wadday	68

Liste des abréviations

- 3D** 3 Dimensions. 60, 79
- A** Aoriste. 71, 104, 138, 145, 260
- AI** Aoriste Intensif. 71, 104, 115, 125, 138, 145, 260
- ALR** Atlas Linguistique des variétés du Rif. 43, 44
- API** Alphabet Phonétique International. xxv, 142, 215, 233, 237, 275, 276, 298
- BULAC** Bibliothèque Universitaire des Langues et Civilisations. 22, 30, 42
- CART** Classification and Regression Trees. 82, 83, 170, 172, 173, 174, 175
- INALCO** Institut National des Langues et Civilisations Orientales. 22, 45, 230
- MDS** Multidimensional scaling. ix, 58, 155, 166, 172
- ONS** Office National des Statistiques. 11, 12, 13
- PHA** Partition Hiérarchisée Ascendante. xiv, 82, 83, 154, 156, 158, 159, 160, 161, 169
- PRET** Prétérit. 71, 104
- SDD** Sciences des Données. 23, 29, 36, 49, 50, 53, 59, 77, 149, 155, 170, 305
- SDL** Sciences du Langage. xxvii, 25, 29, 39, 50, 59, 77, 167
- SHS** Sciences Humaines et Sociales. xxvi, 3, 220
- SIG** Sciences de l'Information Géographique. 23, 29, 39, 60, 305
- SNR** Signal Noise Rate. 230

TmAS Tmaziyt de l'Atlas Saharien. 184, 186, 188, 221, 230, 253, 255, 257, 260, 261, 266, 275, 278, 283, 289, 294, 307

UPSID UCLA Phonological Segment Inventory Database. 197, 198, 199

Liste des gloses

1 1^{ème} personne. ix, x, 124, 245, 252

2 2^{ème} personne. ix, x, xiii, xiv, 95, 97, 105, 121, 123, 125, 126, 127, 128, 130, 131, 132, 133, 135, 137, 138, 145, 244, 254, 255, 258, 260, 265, 266, 273, 284, 286, 287, 290

3 3^{ème} personne. ix, x, 130, 246, 247, 249, 250, 252, 258, 261, 265, 266, 268, 277, 285, 286, 289

ADV Adverbe. 98, 254

AI Aoriste Intensif. ix, x, 245, 252, 265, 286, 289

DEM_DIST Démonstratif distal. 301

DEM_PROX Démonstratif proximal. 98

IMP Impératif. ix, x, xiii, xiv, 95, 97, 105, 123, 127, 128, 130, 132, 133, 138, 145, 244, 254, 255, 258, 260, 265, 266, 273, 284, 286, 287, 290

IMP_INT Impératif intensif. ix, xiii, xiv, 121, 125, 126, 131, 135, 137, 145, 250, 260

M Masculin. ix, x, xiv, 128, 246, 247, 249, 250, 252, 258, 261, 265, 266, 268, 277, 285, 286, 289

NUM Numéral. xiv, 124

PL Pluriel. x, 285

POT Particule du potentiel. 302

PREP Préposition. xiii, 99, 260

PRET Prétérit. ix, x, 124, 130, 246, 247, 249, 252, 258, 261, 265, 266, 268, 277, 285

SG Singulier. ix, x, xiii, xiv, 95, 97, 105, 121, 123, 124, 125, 126, 127, 128, 130, 131, 132, 133, 135, 137, 138, 145, 244, 245, 246, 247, 249, 250, 252, 254, 255, 258, 260, 261, 265, 266, 268, 273, 277, 284, 286, 287, 289, 290

Liste des notations mathématiques

- " Seconde d'arc. 79
- / Division. 230, 232
- % Pourcentage. 191, 215
- ° Degré. 79
- Ω Ohm. 225
- b** Bit. 226
- cm** Centimètre. 227
- cmH₂O** Centimètre d'eau. 200
- dB** Décibel. 225, 230, 235
- dB/Hz** Décibel par hertz. 235
- dB/Oct** Décibel par octave. 235
- dm³/s** Decimètre cube par seconde. 200
- Hz** Hertz. 194, 206, 225, 226, 235, 237, 274
- kHz** Kilohertz. 197, 225, 226
- km** Kilomètre. 5
- km²** Kilomètre carré. 11
- log** Logarithme. 230

m Mètre. 79

ms Milliseconde. 198, 201, 204, 217, 228, 235, 261, 269, 279, 281, 293

Pa Pascal. 237

s Seconde. 235, 237, 267, 294

V Volt. 225

Liste des notations linguistiques

!! Racine. 114, 115

\ \ Glose sémantique (métalangage en français). 91, 93, 95, 98, 99, 100, 104, 105, 110, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 144, 145, 234, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 258, 259, 260, 261, 263, 266, 267, 268, 273, 274, 276, 277, 284, 285, 286, 287, 289

/ / Phonème. 90, 107, 112, 113, 114, 115, 116, 139, 140, 141, 144, 256, 261, 262, 270

@ @ Item linguistique. 91, 93, 95, 98, 99, 104, 110, 112, 113, 114, 115, 124, 125, 129, 130, 134, 135, 137, 138, 140, 144, 145, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 258, 259, 260, 261, 263, 264, 266, 267, 268, 273, 274, 276, 277, 284, 285, 286, 287, 289

[[]] Unité sonore. 124, 268, 269, 271, 275

[] Unité phonique. 90, 91, 93, 94, 95, 97, 98, 99, 101, 104, 106, 107, 108, 109, 110, 111, 112, 113, 114, 116, 117, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 143, 144, 145, 234, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 258, 259, 260, 261, 263, 264, 266, 267, 268, 273, 274, 276, 277, 284, 285, 286, 287, 290

Frontière morphologique. 124, 130, 260, 284, 293, 294

^X Cognat phonologique. 94, 95, 96, 97, 98, 99, 100, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 116, 119, 120, 121, 122, 124, 126, 127, 128, 129, 130, 131,

132, 133, 134, 135, 136, 138, 140, 141, 142, 144, 145

|| Classe d'unités phoniques. 122, 132, 209, 241, 242, 254, 255, 258, 260, 261, 262, 263, 264, 267, 268, 269, 272, 275, 278, 279, 282, 286, 288, 289, 290, 291, 294, 295, 296

{ } Glose grammaticale (métalangage en français). 95, 98, 99, 105, 121, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 144, 145, 244, 245, 246, 247, 249, 250, 252, 254, 255, 258, 260, 261, 266, 268, 273, 277, 284, 285, 286, 287, 289

| | Archiphonème. 259, 264

• Frontière syllabique. 242, 256, 284, 291, 292, 293, 294

C Consonne générique. 44, 205, 217, 242, 284, 286, 291, 292, 294

V Voyelle générique. 44, 205, 242, 286

X,Y,Z Sons génériques. 90, 139, 140, 143

Convention et mise en forme rédactionnelles

Rédaction

Orthographe

Les conventions orthographiques adoptées au cours de cette thèse correspondent à la réforme de l'orthographe révisée en 1990¹.

Alphabet et transcription

Nous aurons recours à plusieurs alphabets au cours de cette thèse. Nous utiliserons l'**Alphabet Phonétique International (API)** pour tout ce qui concerne la transcription précise des items linguistiques. Lorsque ces derniers seront mentionnés dans le développement du corps de notre thèse, c'est cet alphabet qui fera office d'alphabet type. Dans le cas contraire, nous ferons usage de la transcription dite usuelle. Les items mis en appendice sont dans cette situation.

D'autres transcriptions seront utilisées. Celles-ci sont celles en usage chez les orientalistes français. Il en existe plusieurs, c'est pour cette raison que d'un mot à un autre — comme des noms propres, des toponymes, ... — nous ne retrouvons pas les mêmes transcriptions. Nous avons repris tels quels ces termes utilisés dans nos sources sans homogénéiser les alphabets.

1. Le texte de la réforme : [Les Rectifications de l'orthographe](#).

Langue et dénomination

Le français est la langue de rédaction de cette thèse, mais parfois certains anglicismes se sont glissés dans notre texte. Nous avons pris le parti de ne pas traduire systématiquement les anglicismes en français. Car, d'une part, certaines terminologies n'ont pas d'équivalent en français. Donc, nous ne pouvons nous permettre de créer des néologismes. Et, d'autre part, aujourd'hui la culture scientifique est une culture anglo-saxonne. Cette dernière domine le monde de la recherche et pour pouvoir communiquer ses travaux sans équivoque au sein de cette communauté de chercheurs internationaux, il est nécessaire de faire usage de concepts utilisés par ces derniers. Pour s'assurer que nous parlons du même concept, il est préférable d'utiliser les étiquettes retrouvées dans les littératures scientifiques anglaises.

Nous avons, par ailleurs, fait le choix d'utiliser les dénominations onomastiques berbères quand cela nous a été possible de le faire. Les Berbères étant dominés par la culture arabo-musulmane, les toponymes ont été arabisés. Et ils sont souvent enregistrés dans leur prononciation arabe dans les registres administratifs ou dans la littérature. Donc, ces dénominations, que nous allons parfois restituer, sont celles que nous ont signalées les habitants de ces contrées. Nous avons ressenti que cela leur tenait à cœur que l'usage autochtone soit utilisé, donc, nous ne voulons pas trahir leur mémoire.

État de l'art, Bibliographie et Citation

Notre bibliographie reprendra le style American Psychological Association (APA)² qui est devenu en quelque sorte le standard en **Sciences Humaines et Sociales (SHS)**. Les ouvrages cités dans notre thèse, pour mettre en exergue nos sources, reprendront un style qui se rapproche de l'APA. Ces citations seront accompagnées ou non par une numérotation paginale. Pour les références correspondant à des sections d'ouvrages — revues, ouvrages collectifs — nous n'indiqueront pas la page de renvoi. Ce genre de papiers traite dans son ensemble du point qui nous intéresse et ne se trouve pas isolé quelque part dans une sous-section. Donc, il n'est pas pertinent d'indiquer précisément la page. De plus, lorsqu'un ouvrage a déjà été cité avec une mention du numéro de page et que nous réitérons ce référencement, alors lors de la nouvelle citation aucune indication paginale ne sera ébauchée. Il existe aussi le cas de thèses mal paginées, donc il nous a été

2. Pour de plus amples informations : www.apastyle.org (Consulté le 25/09/2017).

impossible de nous y référer. Dans tous les autres cas, le numéro de page sera indiqué.

Le choix de nos sources s'est fait en fonction de plusieurs critères. D'abord, notre objectif n'est pas de faire une historiographie des idées linguistiques. Donc, cela implique que le choix de nos sources ne s'est pas porté nécessairement vers les fondateurs de tel ou tel concept linguistique, ni en direction de supposées principales figures d'un domaine de recherche. Notre objectif a été de présenter les idées en cours et toujours débattues en linguistique. En somme, ce qui est à la pointe de la recherche et qui constitue les récents apports scientifiques. Ce point explique pourquoi nos principales sources sont assez récentes. D'autre part, comme nous nous inscrivons dans le champ des études berbères, donc nous nous sommes appuyé sur des berbérissants pour justifier nos propos quand cela était possible.

Ainsi, notre état de l'art s'est axé sur l'essentiel et pour tous les axes de recherche secondaire de notre thèse, nous n'avons pas pris la peine de faire une revue de la littérature. Par conséquent, ce sont les résultats de travaux scientifiques qui nous ont importé et non pas les cadres théoriques. Enfin, signalons également que notre bibliographie ne comportera que les travaux qui ont été cités au cours de notre thèse.

Terminologie

Du fait que nous réalisons une thèse de doctorat, alors nous nous devons de faire usage de terme scientifique. Sauf qu'en **Sciences du Langage (SDL)**, il n'y a pas de consensus terminologique. Par conséquent, nous avons eu recours à plusieurs termes pour désigner une même idée. Nous avons également fait le choix quand cela était nécessaire d'adopter une nouvelle terminologie scientifique. Car, nous avons considéré qu'en la matière il y avait une carence. Par ailleurs, nous considérons que ce travail doit être à la portée de spécialistes comme du grand public. Ainsi, nous avons réalisé un glossaire pour rendre accessible à tous le jargon linguistique qui est utilisé dans cette thèse.

Date

Nous emploierons au cours de cette thèse le calendrier grégorien. Les dates hégiriennes retrouvées dans la littérature ne seront pas reprises.

Mise en forme et typographie

Texte

Cette thèse a été mise en forme par le langage \LaTeX (X_YTeX).

Taille	12
Police	Cambria
Interligne	1.5
Format	A4
Marge	2,5 cm

Illustration

Comme le dit l'adage, une image vaut mille mots. Nous rajouterons qu'une image demande un temps de mise en œuvre bien plus important que de rapporter la même chose par des mots. La raison à cela est dans le processus de mise en forme des données, qu'il faut réaliser pour qu'elles soient exploitables informatiquement, qui est, disons-le, pénible et long.

D'autre part, toutes les images figurant dans notre thèse ne sont pas notre propriété. Pour celles qui ne le sont pas, nous avons systématiquement indiqué la source. Donc, les images où n'y figurent aucune mention indiquant qui en est le propriétaire ont été réalisées par nous-même.

Les images non propriétaires

Ainsi, un certain nombre d'images qui seront utilisées au cours de cette thèse ne sont pas le produit de notre travail. Elles proviennent de travaux d'autres chercheurs et nous les avons utilisés pour illustrer nos propos notamment au cours de nos revues de la littérature. Nous n'avons pas obtenu d'autorisation pour utiliser ces images, donc si certains chercheurs ou maisons d'édition voient un inconvénient à ce que leurs illustrations soient affichées, nous leur demandons de nous contacter. Celles-ci seront écartées de notre thèse si les propriétaires de ces images en font la demande. D'un point de vue technique, toutes ces images sont des images bitmap et au format JPEG (Joint Photographic Experts Group). Elles ont été extraites en réalisant une impression d'écran. Ce qui implique que le rendu n'est pas de bonne qualité et l'image peut paraître pixelisée.

Le choix de ce format s'est imposé à nous. Pour que notre thèse puisse respecter le standard PDF/A et être archivable, nous nous devons d'opter pour ce format.

Les images propriétaires

C'est ce type d'images, dont nous faisons mention précédemment, qui demande un temps de réalisation conséquent. Celles-ci correspondent à toutes les cartes linguistiques et les figures des signaux acoustiques qui ont été présentées au cours de cette thèse. Nous indiquerons par la suite quels ont été les outils utilisés pour créer ces images. Ici même, nous mentionnerons seulement que ces images sont des images vectorielles au format EPS (Encapsulated PostScript). L'avantage de ce type d'images est leur qualité de rendu.

Cependant, pour des questions techniques, l'encodage EPS ne gère pas un certain nombre de diacritiques comme le symbole de la pharyngalisation [ʔ]. Donc, si nous voulons garder la technique de l'image vectorielle, il est nécessaire que ces illustrations soient en PDF (Portable Document Format). Donc, c'est ce qui a été entrepris avec quelques images qui figurent dans notre thèse.

Enfin, nous avons intégré d'autres types d'images qui n'ont pas le format EPS. Ces dernières sont au format JPEG et correspondent notamment aux quelques photographies que nous avons publiées.

Introduction générale

CHAPITRE I.1

L'unité d'analyse

De manière générale, en SHS, cette unité d'analyse s'articule en trois parties et toute recherche doit en définir les contours. Celles-ci correspondent à : l'objet d'étude, le terrain d'étude et la population d'étude¹. Ces sous-unités formant l'unité d'analyse sont différemment abordées en fonction des disciplines. Les distinctions se font au niveau des variables choisies pour chacune de ces sous-unités.

Délimiter les variables qui seront soumises à analyse ne suffit pas pour définir l'unité d'analyse. Il y a également la question quantitative. Ce dernier point doit être défini tout aussi bien au niveau du nombre de variables que de la taille d'échantillons retenue pour représenter ces variables.

Ce dernier aspect ne sera pas traité dans le cadre de cette introduction générale. Il le sera dans les différentes parties de notre thèse. Chacune de ces parties ayant leur propre spécificité, alors il était plus évident d'évoquer cette question au sein de ces parties. Ainsi, cette partie sera consacrée à présenter les variables que nous avons choisies. Celles-ci seront structurées selon l'approche que nous venons de définir.

Par la suite, ces différentes sous-unités pourront être redéfinies pour se conformer à nos objectifs de recherche. Donc, ce chapitre définit l'unité d'analyse générale. Ainsi, nous allons en premier lieu présenter notre terrain d'étude, puis suivra l'exposition de notre population d'étude et nous aborderons enfin notre objet d'étude.

I.1.1 Le terrain d'étude

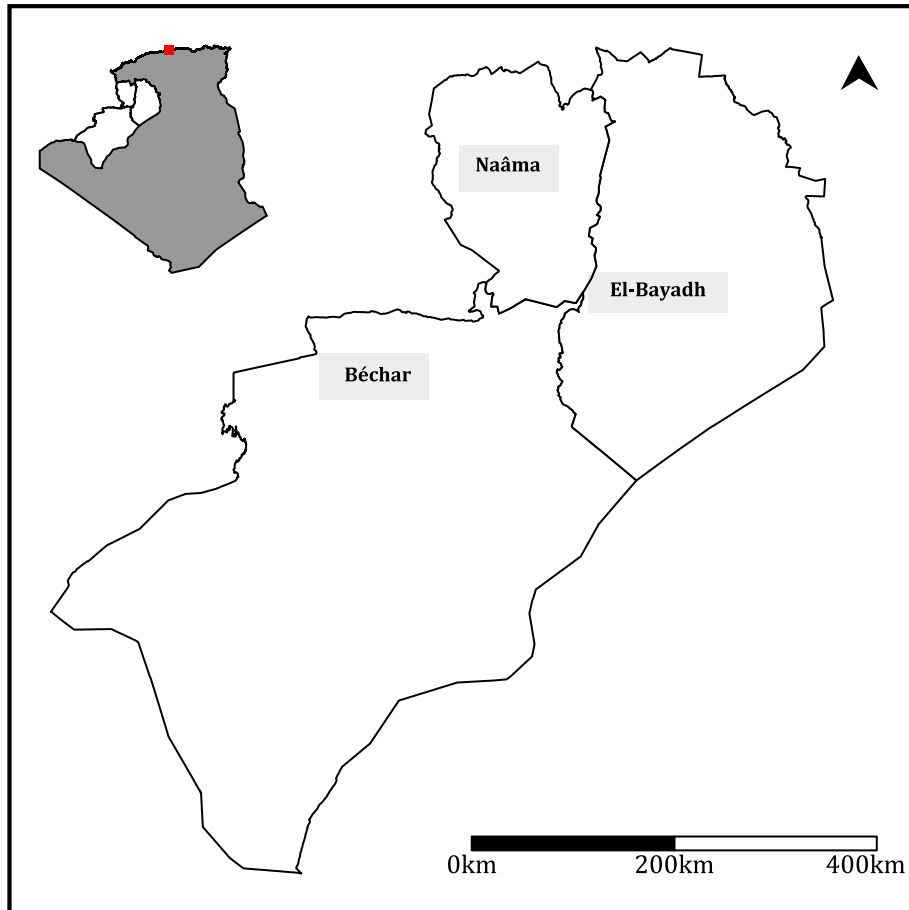
Dans un premier temps, nous allons évoquer la situation qui prévaut actuellement dans notre terrain d'étude. Nous parlerons principalement du découpage administratif et du mode de vie de cette contrée. Puis, nous ferons un saut dans l'histoire de cette région. Nous traverserons les différentes couches historiques de notre terrain d'étude.

I.1.1.1 Époque contemporaine

Notre terrain d'étude se situe au sud-ouest algérien, il borde la frontière algéro-marocaine — voir la carte I.1.1 ci-dessous —. Cette région est appelée plus généralement dans la littérature française : les « Ksours du Sud-Oranais » ou tout

1. Une quatrième unité est à inclure, c'est l'unité temporelle. Cette dernière ne sera pas développée, car nous nous situons en synchronie. Nous faisons l'hypothèse que le système linguistique est stationnaire.

simplement Sud-Oranais².



Carte I.1.1 : Situation géographique et administrative de la région dite du Sud-Oranais

Cette appellation provient de l'époque coloniale, mais elle continue toujours à être employée dans certaines littératures scientifiques comme c'est le cas, à titre d'exemple, dans le livre *L'encyclopédie maghrébine* [Cheurfi, 2007, p.97]. Cependant, depuis l'indépendance de l'Algérie, cette dénomination n'est plus utilisée dans l'administration algérienne. Il y a d'autres termes et découpages administratifs qui sont en vigueur de nos jours, comme nous allons le voir tout de suite.

Ce territoire est divisé en trois grands ensembles administratifs appelés wilaya. Cette division territoriale comprend à l'intérieur d'autres découpages administratifs. Mais nous passerons sur ces autres considérations administratives pour évoquer la question

2. Elle est aussi parfois dénommée : Ksours Oranais, Monts des Ksours. L'orthographe peut également varier, nous retrouvons parfois le mot « ksour » écrit de la manière suivante « kçour ».

de la répartition territoriale de nos points d'enquête. Car cet immense territoire n'est pas entièrement habité. Les populations qui nous intéressent sont concentrées à certains endroits. Et ce sont ces emplacements que nous appelons des points d'enquête. Par conséquent, ces berbérophones ne sont pas regroupés dans une même circonscription administrative et du même coup une même circonscription électorale. D'où la nécessité de mentionner ces subdivisions territoriales, car elles ont inévitablement un impact sociologique.

Ainsi, nous trouvons plus exactement ces parlers berbères dans le nord de la Wilaya de Béchar, et dans le sud de la wilaya de Nâama et de la wilaya d'El-Bayadh. Au point de vue géographique, ces populations vivent dans des oasis et l'étendue de cette zone berbérophone est assez vaste, puisque entre le point le plus au nord et le point le plus au sud il y a environ 500 km de distance. À noter qu'une partie de cette langue berbère s'étend aussi du côté marocain : Figuig, Ain Chair, Ich, ... etc³. Ces variétés marocaines ne font pas partie de notre terrain d'étude, donc nous n'évoquerons pas ces derniers au cours de cette thèse. Notre étude sera strictement limitée à la partie algérienne.

I.1.1.1.1 Considérations sur le mode de vie

Le climat de cette région est aride, toute la zone est touchée par la désertification. La partie la plus septentrionale de notre terrain d'étude est une contrée montagneuse et le sol y est donc plutôt rocheux. Cette chaîne de montagne est appelée l'Atlas saharien. Quant à la partie la plus méridionale, le terrain y est moins accidenté et se trouve au bord de l'Erg occidental qui est le plus grand désert sableux d'Algérie. Cet endroit est dénommé la vallée de la Saoura.



Photographie I.1.1 : Désert rocheux — Lahmar —

3. Notre étude bibliographique et une enquête de terrain nous ont permis de relever qu'il existe d'autres variétés de cette langue à d'autres endroits du côté marocain. Plus précisément dans la région de Bouanane qui se trouve au sud de la province de Figuig.

Pour compléter cette brève description géologique de cette région, mentionnons les principaux cours d'eau parcourant la région. Ce point est important puisque toutes les localités qui nous concernent se trouvent à proximité d'un cours d'eau. Ainsi, ces lits d'eau sont la Zousfana et le Guir pour la wilaya de Béchar. Ces deux derniers prennent leurs sources tous du côté marocain et se croisent pour donner l'oued Souara. Plus au nord, la région connaît une hydrogéologie plus complexe du fait que cela soit un pays de montagnes. Nous allons par conséquent seulement faire référence aux principaux cours d'eau qui sont la Zoubia, la Namous et la Gharbi.



Photographie I.1.2 : Désert sableux — Taghit —

Concernant l'habitat, ces populations sont toutes sédentaires. S'agissant de leur résidence, ils vivent dans de petites zones urbaines. La ville a toujours constitué leur espace de vie traditionnel. Ces centres urbains ancestraux sont nommés en berbère *ayerm* et en arabe *ksar*. Ces mots peuvent être traduits en français par cité fortifiée. Une étude du Père François Cominardi [Cominardi, 1991] sur la cité fortifiée de Chellala donne une bonne description de l'architecture de ces bâtiments. Toutefois, de nos jours les habitants ont déserté ces anciens *ayerm* pour vivre dans des maisons bâties avec des matériaux plus modernes. Soulignons par ailleurs qu'entre ces villes, la région est désolée et est pratiquement inhabitée, quelques rares bédouins parsèment ici et là l'espace.



Photographie I.1.3 : L'ayerm de Ouakda

D'autre part, l'activité économique de cette région est principalement tournée vers le secteur tertiaire et primaire. Traditionnellement, mais c'est aussi le cas actuellement, ces berbères pratiquaient la culture maraichère et la culture du palmier dattier. Dans l'ensemble des oasis que nous avons visité, nous y avons trouvé des palmeraies. Toutefois, ces dernières ne sont plus exploitées dans la majorité de ces oasis et sont donc à l'abandon.

I.1.1.2 Aperçu historique

Nous allons poursuivre la description de notre terrain d'étude en évoquant la dimension historique de cette région. Nous ne fournirons qu'un bref aperçu des différentes périodes relatives à l'histoire du Sud-Oranais. Nous avons divisé cette présentation en deux parties : la période pré-musulmane et la période musulmane. Cette dernière englobe l'époque allant du moyen-âge à la colonisation française.

I.1.1.2.1 Période pré-musulmane

Cette région est fameuse pour son histoire préhistorique. Elle regorge de traces historiques de cette époque lointaine. Ainsi, les gravures rupestres retrouvées dans le Sud-Oranais sont parmi les plus nombreuses d'Algérie [Lhote, 1970]. Les datations

données pour ces artefacts et cet art rupestre varie entre le paléolithique et le néolithique.



Photographie I.1.4 : Gravure rupestre située à Mograr n wadday

Entre cette période préhistorique riche en matériaux historiques et le moyen-âge, il n'y a quasiment rien pour ainsi dire. En effet, il n'existe pas d'évocations précises sur cette région. Ainsi, les auteurs de l'antiquité d'expression latine parlaient de manière indistincte des régions qui se trouvaient en dehors de leur autorité. Nous trouvons dans la littérature historique des mentions de peuples qui sont taxés de différents noms : Gétules, Autololes, Vesunes, Bavares [Desanges, 1990] [Camps, 1991] [Desanges, 1998]. Les territoires occupés par ces derniers ne sont pas déterminés avec exactitude. De plus, certains de ces peuples antiques sont considérés comme étant apparentés. Ou alors, ils peuvent ne correspondre à aucune réalité anthropologique de l'époque et n'être qu'un taxon général. Quoi qu'il en soit, leur lieu de vie était situé dans les contrées se rapportant au Sahara d'aujourd'hui. En conséquence, nous pouvons en tirer comme information qu'à ces époques, il était fort probable que cette région fût habitée.

Au demeurant, il existe quand même des traces matérielles de cette activité humaine. Nous savons qu'une civilisation berbère existait dans la région. Ce fait est attesté par

les nombreuses nécropoles retrouvées. Ces dernières portent le nom de Bazinas ou tumulus [Camps, 1991]. Mais nous le savons surtout par rapport aux différents écrits libyco-berbères qui ont été découverts ici et là. Ces derniers dans un premier temps furent dénommés inscriptions rupestres, voire Hadjra Mektouba [Hachid, 1992]. Tout ceci atteste que le Sud-Oranais connaissait une forte présence humaine depuis des temps reculés. Celle-ci s'est poursuivie jusqu'à nos jours comme nous allons le voir.



Photographie I.1.5 : Inscription libyco-berbère située à Moghrar n wadday

I.1.1.2.2 Période musulmane - Moyen-âge

Concernant cette période de l'histoire, nous avons là aussi peu d'éléments. La consultation d'ouvrages historiques comme ceux d'Al Bekri ou d'Al Idrissi n'a pas été fructueuse. De manière effective, les véritables témoignages historiques concernant cette région apparaissent à partir du XIV^{ème} siècle. Avec la mention de Figuig par Ibn Khaldoun [Ibn Khaldun, 1968a, p.240] — retranscrit par De Slane : Fadjidj — et de Bousseghoun [Ibn Khaldun, 1968b, p.459]. Avant cette période, les auteurs arabes ne mentionnent jamais cette région. Nous savons, par contre, avec cette certitude qu'un état existait à proximité du Sud-Oranais. Celui-ci était dénommé le royaume de Sijilmasa. Même si la capitale de ce royaume ne correspond pas exactement à notre terrain d'étude, il est probable que son influence s'y étendait. Mais tout ceci ne sont que des hypothèses, car, à la même époque, plus au nord, il existait d'autres états souverains comme le royaume Ifrenides ou Rostémides. Il est même possible d'envisager que cette région était politiquement indépendante. Il n'y a aucune certitude en la matière.

Cela équivaut aussi pour les siècles suivants. Nous n'avons pas d'informations sur les liens politiques que pouvaient entretenir cette terre berbère avec les empires berbéro-musulmans qui surgirent à divers endroits de l'Afrique du nord occidental — Almoravides, Almohades, Mérinides, Zianides, ... etc —. Peut-être avaient-ils fait allégeance à ces pouvoirs? Ou peut-être avaient-ils échappé à l'emprise de ces califats? Le doute est total à ce sujet.

Si nous poursuivons l'évocation des événements historiques, il faut nous rendre directement au XVI^{ème} siècle. Car, à partir de cette époque, des ouvrages évoquent un peu plus précisément le Sud-Oranais. C'est ainsi que Léon l'Africain et Marmol dans leur description géographique de l'Afrique du nord mentionnent cette région. Figuig est toujours cité, mais d'autres localités sont décrites comme Wakda [Léon L'Africain, 1898, p.240]. Par la suite, divers auteurs, du XVII^{ème} siècle, d'expression arabe [Ahmed and El-Aiachi, 1846] rapportent par écrit leur voyage à la Mecque pour le Hajj. Le Sud-Oranais était une des étapes de leur périple. Ainsi, nous y trouvons citées différentes localités qui sont les mêmes que celles d'aujourd'hui. Cela s'inscrit dans un récit évoquant des pèlerins venus de régions plus à l'ouest qui passaient par le Sud-Oranais. Ces fervents croyants stationnaient dans ce qui constitue pour nous actuellement des points d'enquête, avant de poursuivre leur chemin.

D'autre part, il apparaît, de manière plus claire, que cette région jouissait d'une certaine autonomie politique par rapport aux grands empires régionaux de l'époque que sont les Alaouites et les Ottomans. Ils sont l'objet de différentes expéditions pour les soumettre à chaque fois. Il semble que ces deux empires se querellaient pour dominer le Sud-Oranais. Mais leur emprise n'était jamais totale, puisque les campagnes punitives se renouvelaient à plusieurs reprises [Lacoste, 1960, p.400] [Cour, 2004].

Le XIX^{ème} siècle est plus riche en informations relatant la réalité de ce temps-là et qui constituent toutes des moments d'histoire. Nous devons cela à la colonisation française qui archivera les faits de cette époque et documentera sur nombre d'aspects relatifs à cette région. Ainsi, la France colonisera ce territoire à partir de la fin du XIX^{ème} siècle non sans mal. Puisque, elle essuyera des résistances, notamment celles de Cheikh Bou 'Amâma [Berque, 1978, p.495]. D'autres événements tumultueux se déroulèrent au XX^{ème} siècle comme la guerre d'Algérie entre 1954 et 1962 ou la guerre des Sables en 1963 qui opposa le Maroc et l'Algérie.

Après cette courte évocation historique, nous revenons à notre époque contemporaine pour décrire un peu plus la situation générale du Sud-Oranais.

I.1.2 Population d'étude

Cette deuxième sous-unité d'analyse sera dédiée à cerner notre population d'étude. Trois points seront examinés. Nous évoquerons la situation démographique du Sud-Oranais. Puis, nous mentionnerons la réalité ethnique de cette région. Enfin, le dernier point sera consacré à la démographie linguistique.

I.1.2.1 Situation démographique

Au plan démographique, cette région est peu peuplée, car c'est une région désertique. La densité de population fluctue entre 3 et 30 habitants au km². Selon le site de l'[Office National des Statistiques \(ONS\)](#)⁴ qui est un organisme étatique algérien, les trois wilayas comprendraient, à elles toutes, environs 700 000 habitants⁵. La wilaya de Béchar étant la plus grande et la plus peuplée des trois.

Wilaya	Nbr. d'habitants
Béchar	270 061
Naâma	192 891
El-Bayadh	228 624

TABLE I.1.1 : Nombre d'habitants par Wilaya

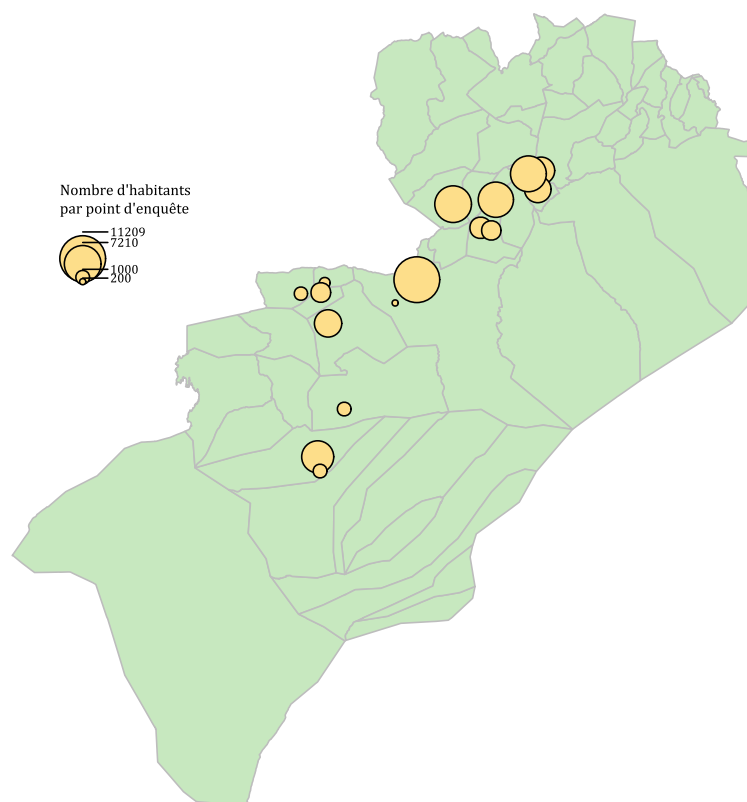
Mais ce n'est pas l'ensemble de ce territoire, ni l'ensemble de ces populations qui ont été l'objet de notre attention. Notre population d'étude est seulement circonscrite à une partie du Sud-Oranais. Cette population est en majorité concentrée dans quelques localités. Ces derniers constituent donc nos points d'enquête. Ainsi, cette population d'étude se trouve répartie sur l'ensemble du Sud-Oranais et comprend 16 localités exactement. Les noms de ces localités sont, par ordre alphabétique : Asla, Beni Ounif, Berrebi, Boukais, Boussemgoun, Chellala, Fendi, Igli, Kalaat Bouamama - Moghrar Tahtani, Lahmar, Mazzer, Mogheul, Moghrar - Moghrar Fougani, Ouakda, Sfissifa, Tiout⁶.

4. <http://www.ons.dz> (Consulté le 05/02/2006).

5. Chiffres du recensement de 2008.

6. La localisation géographique de ces points d'enquête sera vue par la suite (Partie I.1.2.3, page 14 Partie II.1.1.2, page 33).

Répartition de la population par point d'enquête (Source: ONA)



Carte I.1.2 : Répartition de la population par point d'enquête

Nous pouvons voir sur la carte I.1.2 que le nombre d'habitants par point d'enquête est relativement faible. La plus petite localité comprend environ 200 habitants, alors que la plus grande environ 10 000 habitants. Mais tous ces chiffres sont à prendre avec des pincettes. Car ils ne correspondent pas exactement aux nombres d'habitants qui figurent précisément à ces endroits. Ces valeurs se rapportent en réalité aux communes — les lignes grisées sur la carte I.1.2 — et celles-ci englobent dans leur circonscription plusieurs localités urbaines dispersées sur ce territoire. Donc, si nous souhaitons avoir une estimation précise du nombre d'habitants pour chaque point d'enquête, alors il faut revoir à la baisse les chiffres que nous affichons. Le site de l'ONS ne propose pas de tels renseignements, par conséquent, nous ne pouvons fournir plus d'informations rigoureuses.

I.1.2.2 Situation ethnique

Il faut tout de même préciser que tous ces habitants ne sont pas berbérophones, il n'y a qu'une infime minorité qui a comme langue maternelle une langue berbère. Cela est dû au fait que cette région est habitée en majorité par des arabophones. Donc, il existe dans cette région deux ethnies différentes qui se côtoient et cela depuis des siècles.

Du fait des péripéties de l'histoire, le sud-ouest algérien et de manière générale l'Afrique du nord a connu une "migration" de populations arabes et une arabisation de sa population autochtone [Camps, 1983]. Et au cours des siècles, les dialectes arabes se sont développés au profit des dialectes berbères. Les groupes berbères, ainsi dépassés par le nombre, se sont concentrés dans quelques localités et le reste de l'espace était occupé par des nomades arabes/arabophones. Cette situation a persisté jusqu'à la colonisation française au XIX^{ème} siècle, et à partir de cette période, même ces endroits exclusivement berbérophones ont vu l'installation d'arabophones et cela s'est encore plus accéléré après l'indépendance de l'Algérie en 1962⁷.

Cette conjoncture complique par conséquent le recensement de berbérophones d'autant plus que la question linguistique ne fait pas partie des points inventoriés par l'état algérien lors de ces campagnes de recensement. Donc, nous avons dû nous débrouiller par nous même pour comptabiliser le plus objectivement possible cette catégorie de populations. Cela a été fait avec nos propres observations sur le terrain, des témoignages des habitants et des chiffres statistiques donnés par l'ONS sur le nombre d'habitants dans ces localités. De cette façon, nous sommes parvenu à estimer le nombre de berbérophones à environs 50 000 personnes⁸ — ce chiffre est plutôt à considérer comme étant la fourchette haute —. C'est bien sûr un chiffre qui ne reflète pas de manière précise la situation démographique de cette région, mais qui donne un ordre d'idée sur la disparité démographique entre berbérophones et arabophones. Ainsi, ces premiers ne représentent finalement qu'environ 7% de la population des trois wilayas. Ce chiffre a une importance capitale pour les points que nous allons évoquer par la suite.

Par ailleurs, si nous comparons notre chiffre avec celui qui a été avancé par Edmond Doutté [Doutté and Gautier, 1913, p.127-129 & p.139], alors nous pouvons nous rendre compte à quel point le berbère a perdu du terrain. En un siècle, la démographie de ces Berbères n'a quasiment pas évolué. Ainsi, en 1913, pour l'ensemble du Sud-Oranais,

7. Nous tenons ces informations de témoignages recueillis sur le terrain.

8. Sur le site www.Ethnologue.com, le nombre de 58 000 personnes est avancé (Consulté le 05/02/2006).

les Berbères étaient estimés à environ 30 000 personnes. Et déjà à cette époque, le déséquilibre entre Arabes et Berbères, à l'avantage de ces premiers, était très important.

I.1.2.3 Situation sur la démographie linguistique

Nous allons présenter succinctement certaines caractéristiques liées à la démographie linguistique de nos points d'enquête. L'une des particularités de ces parlers est qu'ils sont à considérer comme étant des "langues en danger", comme il est maintenant d'usage de les dénommer. Ce fait est basé sur notre propre constat sur le terrain et il existe d'autres personnes qui l'ont évoqué dans certaines études [Mouili, 2012] [Guemide and Benachiba, 2012] [Kossmann, 2010]⁹.

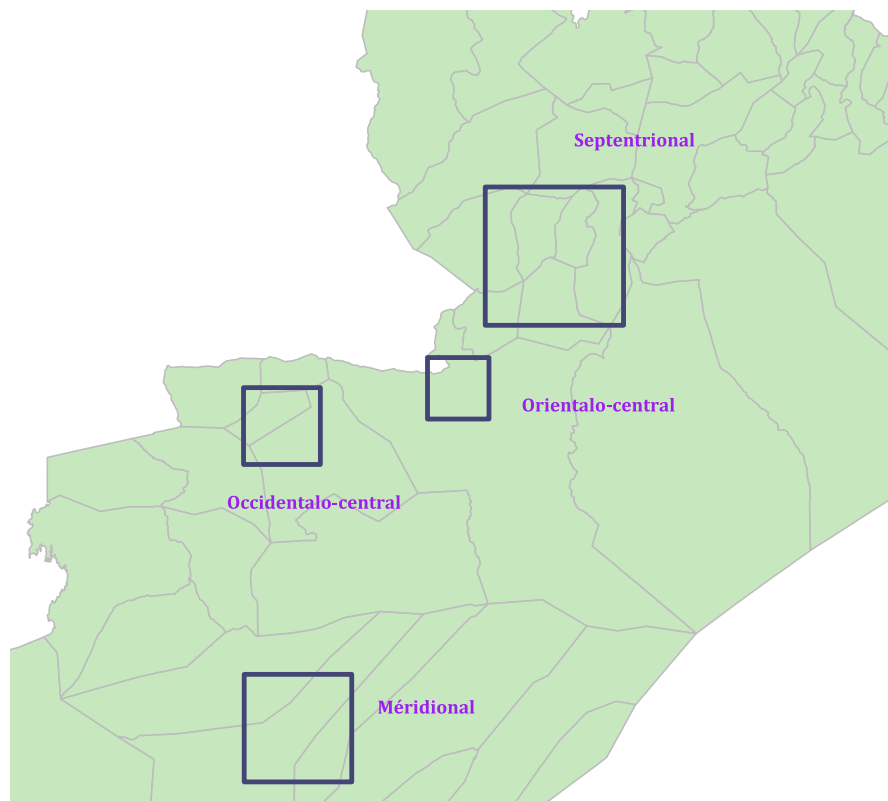
L'extinction du berbère dans ces contrées s'est amorcée depuis longtemps. Edmond Doutté [Doulté and Gautier, 1913, p.127-129 & p.46] signale déjà au début XX^{ème} siècle que le berbère dans cette région régresse. Sur environ une vingtaine d'oasis d'origine berbère, dans la quasi-totalité de ces endroits, il n'y a plus de transmission aux nouvelles générations. Ce processus s'est accéléré approximativement dans les années 70. Les raisons de cette "désaffection" sont multiples. L'une des principales est le fait que ces groupes berbères sont en nombre très réduit proportionnellement à la population arabophone. Cette disproportion numérique a exercé inévitablement une pression qui a poussé ces locuteurs du berbère à vouloir s'intégrer aux particularismes identitaires de la masse qui les entourait. Mais cette disparition du berbère ne touche pas de la même manière toutes les oasis, il y en a qui sont plus touchées que d'autres. Certaines oasis se maintiennent très bien alors que d'autres ont déjà disparu et d'autres sont en voie, à des degrés divers, de disparition.

Après ce tour d'horizon, nous allons voir plus en détails l'ensemble de ces parlers un par un et indiquer leurs situations respectives. Nous n'évoquerons dans cette partie que les variétés linguistiques qui en synchronie, donc au moment présent où nous menons ce travail, connaît encore la présence effective de locuteurs berbères. Donc, les variétés qui ont disparu ne seront pas recensées au cours de cette thèse. En cela,

9. Le site www.Ethnologue.com indique que la langue serait « menacée ». Sur le site [Atlas interactif de l'UNESCO](#) sur les langues en danger dans le monde, il n'y a par contre aucune donnée pour les parlers du côté algérien, mais il mentionne la situation de Figuig. S'agissant de cet endroit, il considère que c'est une langue « vulnérable ». Nous ne débattons pas dès à présent de la question des critères d'évaluation et de la terminologie utilisée par ces institutions. Ce point n'est pas à l'ordre du jour. Une application de ces indices de vitalité des langues a été réalisée par Naït-Zerrad [Naït-Zerrad, 2012], nous renvoyons à son travail pour de plus amples informations.

nous faisons référence au cas de Béchar ou Taghit qui au début du XX^{ème} siècle était encore berbérophone, mais à l'heure actuelle, les habitants autochtones ne parlent plus berbère¹⁰. La présentation de ces variétés se fera en quatre grandes phases. Chacune d'elles se focalise sur une portion territoriale de Sud-Oranais où sont concentrés géographiquement les différents points d'enquête.

Regroupement géographique des points d'enquête



Carte I.1.3 : Répartition en groupe des points d'enquête

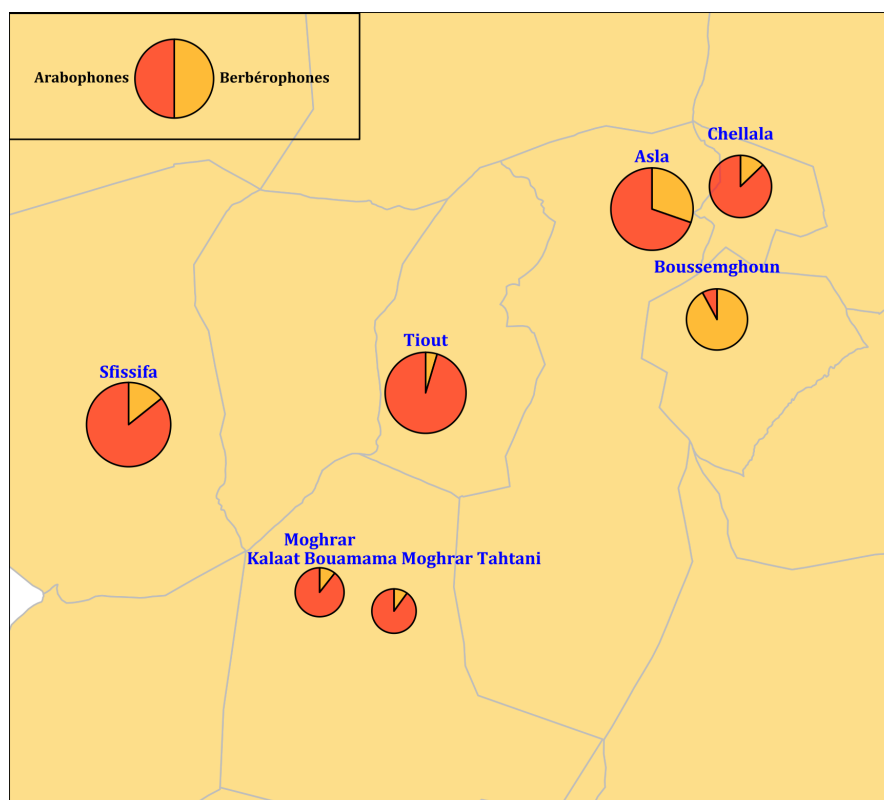
Ainsi dans la partie suivante, nous allons discuter d'un certain nombre de questions de démographie linguistique. Le déroulement de cet examen suivra le schéma d'organisation qui est affiché ci-dessus (Carte I.1.3). Mais nous tenons à préciser que nous n'avons pas réalisé d'enquête de démographie linguistique approfondie. Les résultats qui seront affichés sont approximatifs, ils ne sont là que pour donner un ordre d'idée.

10. Pour plus d'informations sur ces variantes qui ont disparues, voir Partie II.1.1, page 29.

I.1.2.3.1 Groupe septentrional

Commençons par la première région, celle-ci se situe dans la wilaya de Naâma et d'El-Bayadh. Ce bloc de parlers comprend exactement sept localités berbérophones qui sont représentées sur la carte (I.1.4) ci-dessous. Le berbère se maintient encore à Boussemghoun et à Asla. Dans ce dernier lieu, bien que le berbère soit encore présent chez toutes les générations, les personnes ayant l'arabe comme langue maternelle constituent la majorité des habitants d'Asla. Ce qui n'est nullement le cas à Boussemghoun où les berbérophones de cette ville vivent isolés et n'ont pas connu l'installation en leur sein d'une importante communauté arabophone. Cela sous-entend que pour les autres villes que nous n'avons pas citées, mais visibles sur la carte (I.1.4), le berbère y est seulement parlé par les anciennes générations et sont, de plus, en minorité dans leur propre ville.

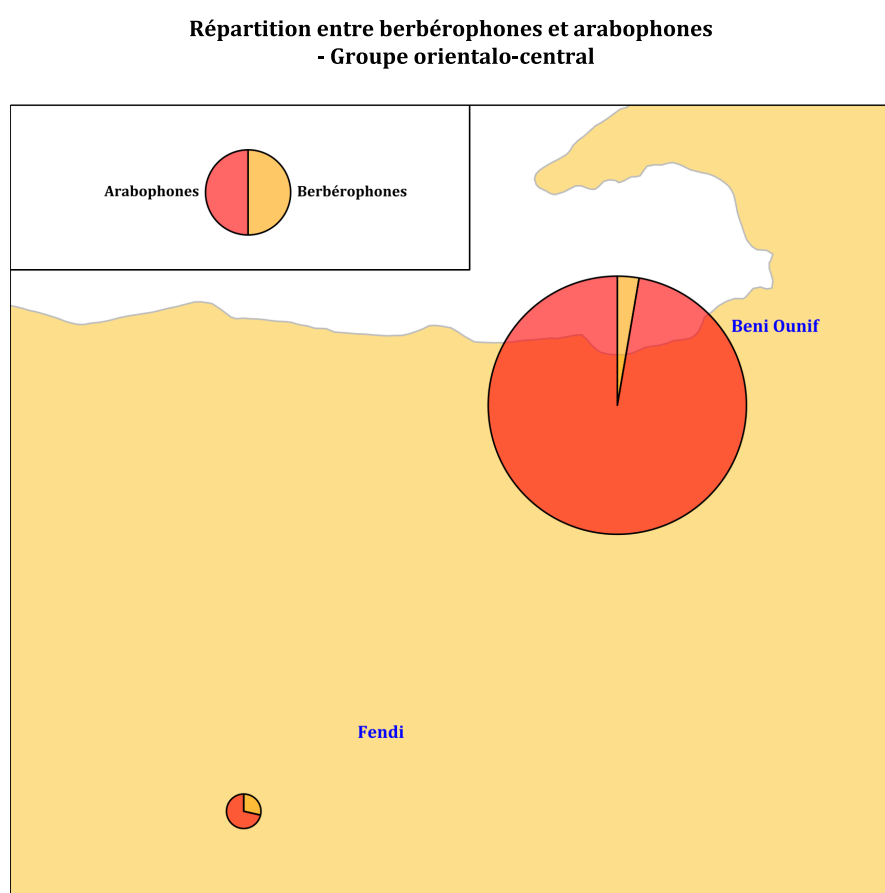
Répartition entre berbérophones et arabophones
- Groupe septentrional



Carte I.1.4 : Répartition entre berbérophones et arabophones - Groupe septentrional

I.1.2.3.2 Groupe orientalo-central

La deuxième variété du Sud-Oranais est parlée dans le nord de la wilaya de Béchar. À deux endroits, le berbère y est parlé. Ceux-ci correspondent à Beni Ounif et Fendi. En ces deux localités, le berbère n'est plus que l'affaire de personnes âgées. La transmission chez les jeunes générations s'est interrompue. De plus, les populations arabophones non autochtones sont largement majoritaires. À titre d'information, Beni Ounif jouxte la ville de Figuig qui est juste après la frontière entre l'Algérie et le Maroc.

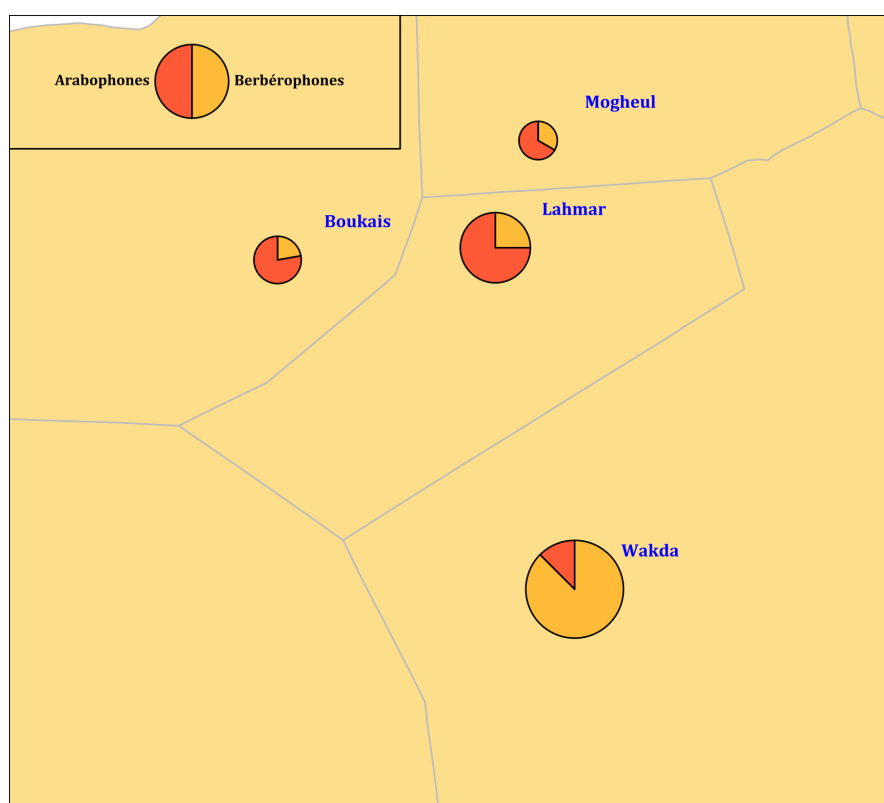


Carte I.1.5 : Répartition entre berbérophones et arabophones - Groupe orientalo-central

I.1.2.3.3 Groupe occidentalo-central

La deuxième variété du Sud-Oranais est parlée dans le nord ouest de la wilaya de Béchar. Nous y trouvons dans ces lieux quatre zones berbérophones qui sont Ouakda, Boukais, Lahmar et Mogheul. De ces quatre villes, seule Ouakda est encore berbérophone et ne connaît pas un recul de la pratique du berbère. Alors que ce n'est nullement le cas dans les trois autres où le berbère est seulement parlé par des personnes âgées. À Ouakda, les berbères vivent au milieu d'une forte population arabophone du fait qu'il se trouve à proximité de Béchar qui est la plus grande ville de tout le sud-ouest algérien. Mais cela n'a rien entaché à leur attachement à leur langue maternelle. À l'intérieur des trois autres villes, une forte population arabe vit également aux côtés des autochtones berbères.

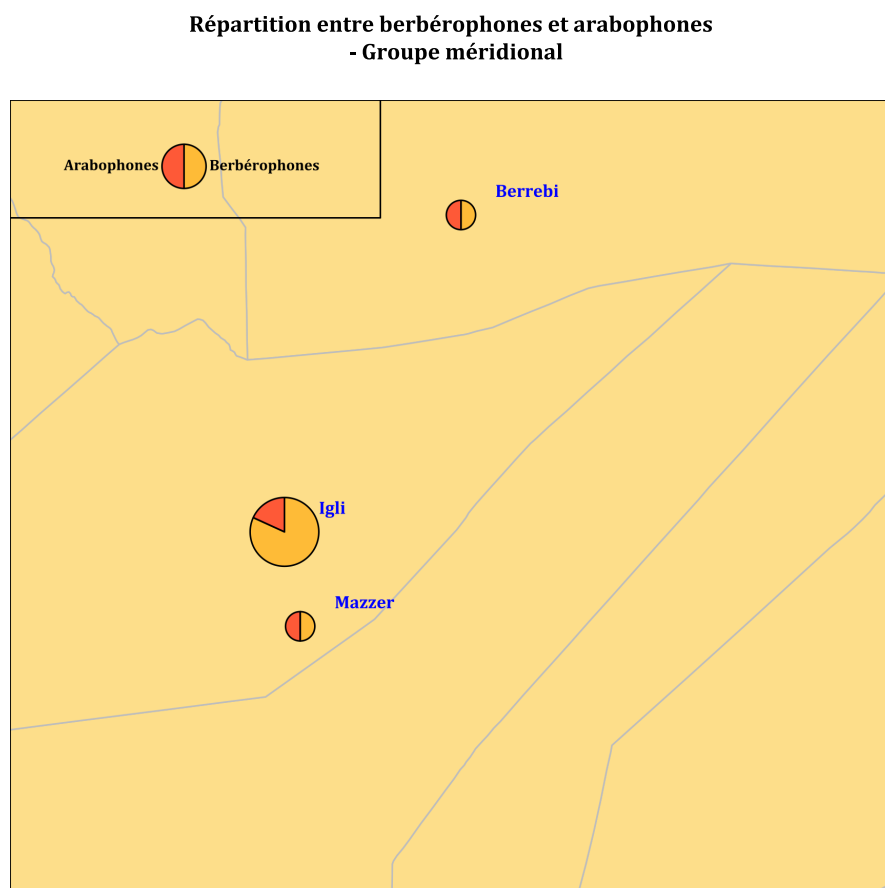
Répartition entre berbérophones et arabophones
- Groupe occidentalo-central



Carte I.1.6 : Répartition entre berbérophones et arabophones - Groupe occidentalo-central

I.1.2.3.4 Groupe méridional

Le dernier dialecte se trouve dans le sud de la wilaya de Béchar. Dans cette région, il existe trois aires de peuplement berbère. Ces trois endroits se nomment : Berrebi, Igli et Mazzer. Ici encore, le berbère est menacé de disparaître. À Berrebi et à Mazzer, le berbère est bien plus en danger qu'à Igli, où il semble que la non transmission du berbère est plus récente que dans les deux autres. Bien que cette région soit à l'écart et donc n'a pas connu réellement la venue d'un peuplement arabe, il n'en reste pas moins que le berbère perd de ses marques dans ces trois localités.



Carte I.1.7 : Répartition entre berbérophones et arabophones - Groupe méridional

I.1.3 L'objet d'étude

Cette thèse s'inscrit dans une étude descriptive des variétés berbères présentes au niveau du terrain d'étude que nous venons de décrire. Au sein de ce territoire délimité, le berbère peine à se maintenir. Donc, il est plus qu'urgent d'en faire la description. Ainsi, ce travail est un apport qui s'inscrit dans cette démarche. Une telle entreprise a déjà été réalisée avant nous et c'est que nous allons présentement mettre en avant. Puis, nous allons présenter les limites de notre objet d'étude, car il est inconcevable de pouvoir faire une description exhaustive de ces variétés. Mais nous allons d'abord entamer cette partie en évoquant quelle est la place des variétés berbères du Sud-Oranais au niveau de l'état algérien.

I.1.3.1 Considérations générales sur la langue berbère du Sud-Oranais

Cette partie sera consacrée à présenter les dimensions extra-linguistiques qui cernent les variétés linguistiques du Sud-Oranais. Nous toucherons un mot sur le statut politique du berbère du Sud-Oranais. Puis, nous évoquerons les dénominations en usage pour parler du berbère du Sud-Oranais. Enfin, nous parlerons brièvement de la branche linguistique à laquelle se rattache historiquement le berbère de cette région.

I.1.3.1.1 Le statut du berbère du Sud-Oranais

Comme nous venons de le voir la présence du berbère remonte à très longtemps et qu'elle est en train de disparaître doucement mais sûrement. Toutefois, récemment le statut du berbère a changé en Algérie. Depuis 2016, une langue dénommée Tamazight ou Langue Amazighe est consacrée comme langue officielle¹¹. Cette dénomination renvoie en réalité aux langues de la famille berbère. De plus, il y est précisé que "L'Etat œuvre à sa promotion et à son développement dans toutes ses variétés linguistiques en usage sur le territoire national (Art. 4)". Si nous prenons au mot ces écrits, alors cela intègre les variétés berbères du Sud-Oranais.

Un premier pas a été entrepris dans ce sens. L'année scolaire 2016/2017 a vu des ouvertures de cours de berbère dans la wilaya de Béchar. Mais cela n'a été engagé

11. La constitution de la République Algérienne : www.joradp.dz/HFR/Index.htm.

que dans cette wilaya. Les autres wilayas de notre terrain d'étude n'ont pas connu d'instauration de classes de berbère. D'ailleurs, au moment où nous rédigeons cette thèse, nous n'avons aucune idée sur ce qu'il en est en 2017/2018. Est-ce que cet enseignement perdure? Est-ce qu'il s'est étendu aux autres wilayas?

Cependant, cet enseignement du berbère était restreint à un niveau scolaire, la quatrième année de l'école primaire algérienne. En outre, seulement quelques localités connaissaient cette nouvelle matière. Nous savons avec certitude que pour les villes de Béchar, Berrebi, Beni Abbes, Igli et Ouakda un enseignant a été affecté. Par contre, dans les autres localités, le berbère n'y était pas enseigné. Soulignons que certaines de ces villes ne sont pas des localités où le berbère est une langue vernaculaire. D'autre part, ces professeurs de berbère étaient tous d'origine kabyle et par conséquent, ils enseignaient la langue kabyle "standard" à ces élèves. Ce fait nous a été relaté par l'enseignant d'Igli. Ce point est important à souligner, car pour ceux qui ne sont pas familiers de la diversité des langues berbères. Il est à relever que le kabyle n'est pas une variété du berbère proche des variétés berbères du Sud-Oranais.

Complétons ce tableau en évoquant le fait qu'en dehors de cette ouverture timide, le berbère du Sud-Oranais est totalement absent des autres réalités de la vie publique. Au niveau administratif, juridique, médiatique¹² et dans bien d'autres domaines, le berbère n'a pas droit de cité. Peut-être que cette situation changera dans les années futures et que probablement la vitalité du berbère prendra une autre tournure.

I.1.3.1.2 Les dénominations endogènes et exogènes du berbère du Sud-Oranais

Quant à la langue de ces populations, selon les auteurs, différentes dénominations sont employées. Le premier à avoir enquêté sur les parlers de cette région est René Basset, il désigne l'ensemble de ces parlers par le terme de "chelha", mais également par le terme de "dialectes du Sud-oranais" [Basset, 1886]. Quant à André Basset qui a enquêté dans cette région des décennies plus tard, il reprend le même type de dénomination vague de son prédécesseur "parlers du Sud-oranais" [Basset, 1959]. Il faut dire d'emblée que ces dénominations ne sont en rien celles employées par les populations elles-mêmes, il s'agit d'appellations exogènes. En réalité, sur le terrain nous ne retrouvons pas qu'une seule dénomination endogène utilisée, mais plusieurs.

Cette région dite du Sud-Oranais comprend trois grandes dénominations endogènes.

12. Il existe quelques minutes — 45 minutes exactement — accordées par semaine au berbère dans des antennes locales de radio régionale, comme celle de Naâma et d'El-Bayadh.

Si nous reprenons nos aires régionales, chacune d'elles utilise une appellation lui étant propre. Les désignations en question sont : tmaziyt /tmɛziɪt/, tacelhit /tɛʃəlhit/ et tabeldit /tɛbɛldit/. Le groupe septentrional est celui qui utilise la dénomination de tmaziyt. Le nom tabeldit est employé uniquement par les habitants d'Igli et de Mazzer. Berrebi qui appartient au même groupe méridional utilise plutôt la dénomination tacelhit. C'est la même chose pour les groupes centraux. Toutefois, d'après les notes d'André Basset (voir Partie II.1.1.1, page 30), il se pourrait que cela n'ait pas été toujours le cas. Ce dernier note que ses informateurs de Lahmar appellent leur langue tmaziyt. Donc, il apparaît que ce vocable de tacelhit est probablement d'un usage récent.

I.1.3.1.3 La phylogénie du berbère du Sud-Oranais

D'un point de vue phylogénique, comme nous l'avons déjà dit, cette langue fait partie de la famille des langues berbères, mais il y a un autre palier qui est important à mentionner à leur sujet est que celle-ci est une langue zénète. Les langues zénètes sont présentes sur un vaste territoire, mais sont principalement situées au Maghreb central et notamment en l'Algérie.

Dans la littérature scientifique berbère, cette région est souvent présentée comme faisant partie de la grande famille des langues zénètes [Naït-Zerrad, 2013, p.5802]. Parmi celles-ci, nous pouvons citer le rifain, le chaoui, le mozabite, le ouarsenis, ... etc. Le zénète est une des branches de la langue berbère, comme le sont le touareg, le zénaga, le kabyle, les parlers berbères orientaux — Siwa, Sokna, ... — et le chleuh [Aikhenvald, 1988, p.43].

Comme nous l'avons mentionné précédemment, il existe plusieurs dialectes dans cette région. Bien que dans la littérature berbérissante le berbère du Sud-Oranais soit mentionné comme homogène dialectalement [Naït-Zerrad, 2013], ce qui est vrai jusqu'à une certaine mesure, il paraît plus raisonnable tout de même de distinguer à l'intérieur plusieurs grandes variétés. Ces affirmations que nous tenons sont le fruit d'une étude dialectologique (Partie II.4.2, page 148) que nous avons menée dans le cadre de cette thèse.

I.1.3.2 État de la question des études linguistiques sur le Sud-Oranais

Les études sur le berbère du Sud-Oranais sont relativement peu nombreuses comparativement à ce qui a été fait pour le chleuh, le kabyle et le touareg par exemple.

Parmi les oasis qui jouxtent la frontière algéro-marocaine, il existe une oasis appartenant à cette famille de langue berbère qui fut étudiée assez sérieusement par plusieurs chercheurs. Cet endroit, qui s'appelle Figuig, se trouve au Maroc, tout près de la frontière avec l'Algérie. Les travaux de description linguistique qui lui ont été consacrés sont ceux de Kossmann [Kossmann, 1997] et de Saa [Saa, 1995]. Mais s'agissant des autres oasis, pour certaines d'entre elles, les données les plus récentes datent de plus d'un demi-siècle. Ce travail n'a jamais été publié, c'est un ensemble de notes recueillies par André Basset et son équipe qui sont allés relever ces informations sur le terrain en 1932 et 1939 [Kossmann, 2010, p.69]. Le tout est entreposé à l'INALCO et à la BULAC (Paris) dans le fond « André Basset ». Ces notes ont été exploitées par Kossmann [Kossmann, 2010] sous la forme d'un article publié dans *Études berbères V – Essais sur des variations dialectales et autres articles*. Mais ce chercheur a seulement traité une partie du corpus qui se trouvait dans ce fond et a axé son travail sur le parler d'Igli. Il existe également des documents réalisés par René Basset sur ces parlers, mais ceux-ci remontent à la fin du XIX^{ème} siècle. S'agissant de travaux de phonétique plus pointus, il est à relever que le parler de Figuig a fait l'objet de deux récentes études réalisées par Yeou¹³ [Yeou and Honda, 2011] [Yeou and Honda, 2015] portant sur les consonnes coronales géminées occlusives [d̪d̪], [t̪t̪^h] et [t̪t̪^ɿ]. Ce sont les seuls travaux de phonétique existant pour ce parler. Concernant les autres variétés du Sud-Oranais aucune étude de phonétique expérimentale n'a été réalisée sur leur langue.

I.1.3.3 Délimitation de l'objet d'étude

Notre objet d'étude est bien sûr la langue, mais l'étude d'un système linguistique peut englober et s'étendre dans bien des directions. Si nous devons réaliser une description complète du berbère du Sud-Oranais, alors chaque point d'enquête suffirait en lui-même pour être l'objet d'étude d'une thèse. De ce fait, si nous souhaitons avoir une vision globale de toutes les variétés linguistiques de cette région, il nous faut délimiter notre objet d'étude. Par conséquent, notre ambition se limitera à étudier quelques faits linguistiques. Ainsi, notre étude se focalisera sur le système consonantique. D'autre part, il ne s'agira pas non plus d'étudier chaque consonne. Là encore, nous nous sommes limité à l'étude de quelques consonnes.

Cette restriction s'explique par notre volonté de mener une étude qui se situe

13. Ce chercheur projete d'autres projets de recherche qui concernent toujours Figuig [Yeou, 2011].

à l'intersection de plusieurs domaines disciplinaires. Cette transdisciplinarité nous impose de réduire notre champ d'étude à quelques points. Les questions de faisabilité sont inflexibles que nous devons nous soumettre à cet impératif de temps et de moyens. Ainsi, pour rendre compte des particularités linguistiques des variétés du Sud-Oranais plusieurs approches seront appliquées. Ce travail comporte une dimension de linguistique descriptive et de documentation linguistique sur des variétés berbères qui ont été très peu étudiées. D'autre part, nous ferons appel aux **Sciences des Données (SDD)** et à la statistique pour quantifier et évaluer nos données. Par ailleurs, nous aurons recours à des méthodes graphiques d'exploration visuelle utilisées en **Sciences de l'Information Géographique (SIG)** et en statistiques pour présenter les résultats. En outre, la phonétique et la phonologie seront la colonne vertébrale de cette thèse. Enfin, l'informatique fait partie aussi des domaines qui sont à l'origine de ce travail académique.

De plus, nos objets d'étude ne seront pas identiques en fonction des parties de notre thèse. Suivant les approches que nous adopterons, celles-ci conditionneront le choix de nos objets d'études et le nombre de variables retenues. Ce nombre variera, mais il ne sera jamais important. Le fait d'avoir peu de faits linguistiques à étudier est nécessaire, si nous voulons parvenir à des résultats probants. C'est pour cette raison qu'il est impératif de découper les problèmes en plus petits problèmes. Cette division en sous-problèmes rend la démarche scientifique plus rigoureuse. Ainsi, dans notre première grande partie, nous cernerons un certain nombre de problèmes phonologiques. Dans la deuxième grande partie, ces derniers feront toujours l'objet de nos considérations. Mais nous diviserons encore plus notre objet d'étude pour nous calquer sur nos objectifs de recherche.

Cela étant dit, nous allons rendre compte de quels sont nos objectifs de recherche. Ceux-ci sont doubles. Il y a en premier lieu une volonté descriptive. C'est-à-dire de faire jour des particularités linguistiques du Sud-Oranais. Ce premier travail se fera de manière transversale et verticale. Transversal dans la mesure où notre objet d'étude portera sur l'ensemble du Sud-Oranais. Cet axe correspond à notre étude dialectologique qui constituera la première grande partie de cette thèse (Partie I, page 28). Et vertical en raison du fait que notre étude sera circonscrite à un point d'enquête. Ce dernier coïncide avec notre deuxième grande partie sur l'étude des rhotiques alvéolaires (Partie II, page 183).

Puis, le second objectif est de discuter de questions de linguistique générale. Cet aspect sera transversal à nos deux grandes parties. À l'intérieur de ces derniers, nous

questionnerons les représentations phonologiques de nos divers objets d'études. Cela nous amènera à discuter de questions théoriques. Le but étant de déterminer la validité scientifique de ces derniers au regard des données sur le berbère du Sud-Oranais.

I.1.3.4 Plan général

Cette thèse s'articule en deux grandes parties. Et dans chacune d'elle, nous avons en ligne de mire la réalité cognitive du langage. Notre perspective est à partir des données du berbère du Sud-Oranais de questionner cette dimension tout en les confrontant aux théories linguistiques en cours en **SDL**.

Les études berbères depuis les années 70 ont fait un pas dans ce sens. Cela est surtout le fait de chercheurs anglicisants. Ces chercheurs de culture scientifique anglaise ont produit des travaux en adoptant des théories linguistiques bâties sur des descriptions linguistiques de langues indoeuropéennes ou sémitiques principalement. Ces études ont notamment comme objet d'étude le système phonético-phonologique [**Guerssel, 1977**], mais d'autres faits linguistiques sont aussi concernés [**Ouhalla, 1988**]. Donc, nous nous inscrivons dans cette continuité, mais sans nous affilier à un cadre théorique hypothétique.

Comme notre thèse est intégralement dédiée au système consonantique, alors notre étude ne s'intéressera qu'à la réalité psychique de ces consonnes. L'étude de ces dernières se fera en cherchant à déterminer quelle est l'**unité morphémique** et le **phonème** qui sont sous-jacent au **morphème** et à l'**unité phonique**. D'autre part, notre analyse se limitera à une étude de premier ordre. Cela signifie que notre intérêt se portera essentiellement sur deux aspects : l'entrée — réalité cognitive — et la sortie — réalité physique —. En d'autres mots, ce qui permet de passer de l'entrée à la sortie ne sera pas un point que nous aborderons.

Cette boîte noire ne sera pas l'objet de notre étude. Nous questionnerons l'hypothèse qui depuis les premiers travaux de Chomsky [**Chomsky, 1965**] [**Chomsky and Halle, 1968**] assertent que le rapport entre l'entrée et la sortie est dissymétrique.

It might be supposed that surface structure and deep structure will always be identical. In fact, one might briefly characterize the syntactic theories that have arisen in modern structural(taxonomic) linguistics as based on the assumption that deep and surface structures are actually the same (cf. Postal, 1964a, Chomsky, 1964). The central idea of transformational grammar is that

they are, in general, distinct and that the surface structure is determined by repeated application of certain formal operations called "grammatical transformations" [Chomsky, 1965, p.16]

Real progress in linguistics consists in the discovery that certain features of given languages can be reduced to universal properties of language, and explained in terms of these deeper aspects of linguistic form. [Chomsky, 1965, p.35]

Ainsi, il y a une entrée et plusieurs sorties. Nous tenterons de faire la démonstration empirique que ce rapport dyssémétrique et biunivoque n'est pas suffisant pour expliquer le fonctionnement du langage. Plusieurs faits linguistiques résistent à cet axiome. Mais nous discuterons avant tout des méthodes adoptées pour déterminer la réalité phonologique d'un système linguistique. Autrement dit, sont-elles à même de rendre compte de cette dimension sans équivoque ?

Pour mettre en œuvre ce programme, nous aurons recours à différentes approches : La dialectologie et la phonétique.

— Partie I : Dialectologie

La Partie I est une étude dialectologique. Au cours de celle-ci, nous subdiviserons notre travail en plusieurs parties. Nous débuterons par une étude géolinguistique qui présentera les différences phonétiques qui existent entre les variétés du Sud-Oranais. Cette dernière sera suivie par une étude dialectométrique qui tentera de mesurer les distances linguistiques qui existent entre les variétés du Sud-Oranais.

— Partie II : Rhotique alvéolaire

La Partie II est prévue pour mener une étude plus minutieuse sur des aspects phonétiques. Ce point constituera la deuxième partie. Avant cela, une autre étude sera entreprise et sera axée sur des questions exclusivement phonologiques, mais en faisant appel à la phonétique expérimentale.

Première partie

Étude des variétés linguistiques : Dialectologie et Phonologie

CHAPITRE II.1

Introduction

Nous ne pouvons pas dire que la dialectologie a en linguistique le vent en poupe. Elle aurait plutôt une image de discipline désuète. D'ailleurs, cela ne date pas d'aujourd'hui. Voilà ce qu'écrivait Sever Pop, il y a plus d'un demi-siècle :

Dans bien des pays pour obtenir un grade universitaire nécessaire à leur carrière, les étudiants sont tenus de présenter une contribution scientifique. Très souvent dans le domaine de la dialectologie, le travail une fois publié, l'activité des diplômés cesse à jamais [Pop, 1950, p.XI].

Ce fait n'a malheureusement pas changé. Pourtant, nous avons tout de même eu le désir de mener une étude dialectologique. Toutefois aujourd'hui la dialectologie demande bien plus de compétences techniques qu'à l'époque. C'est une discipline qui est de nos jours à la croisée de plein de domaines : les SDD, les SIG et bien évidemment les SDL (voir Partie II.2, page 38).

Mais avant de rentrer dans le vif du sujet, nous allons établir un certain nombre d'informations générales nécessaires à la compréhension de notre étude. Nous allons évoquer notre terrain d'étude ainsi que notre objet d'étude. La population d'étude sera décrite dans la partie méthodologie (Partie II.3.1, page 68).

II.1.1 Terrain d'étude

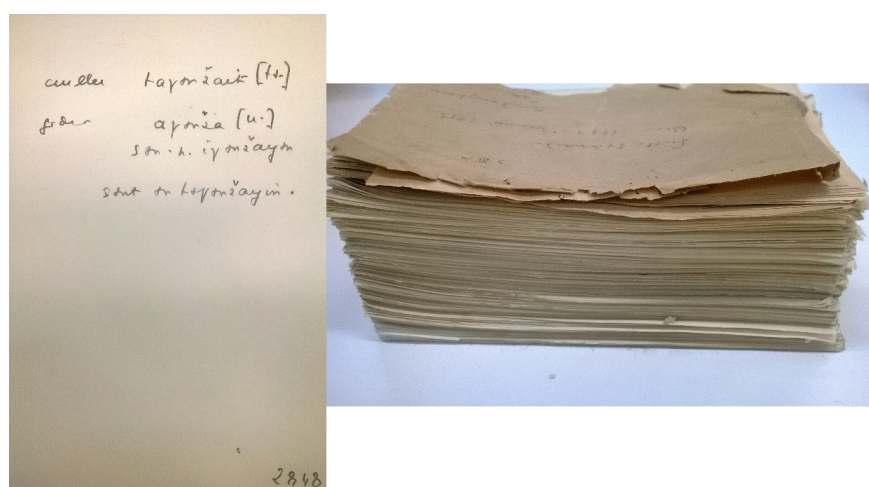
Nous n'allons pas représenter notre terrain d'étude, puisque nous l'avons déjà fait. Par contre, nous allons davantage parler de nos enquêtes de terrain dans cette région. Ainsi, une étude dialectologique suppose de comparer les faits linguistiques collectés en différents endroits d'un espace géographique, alors il y a nécessité de définir des points d'enquête. Bien que notre terrain d'étude soit très vaste, il s'avère que nous n'avons pas à nous poser la question du maillage de notre territoire. Car, les territoires berbérophones sont limités à quelques localités comme nous l'avons indiqué.

Nous aurions pu choisir au sein de ce nombre réduit de localités quelques points d'enquête. Mais nous avons préféré faire de toutes ces localités nos points d'enquête. Car, au delà de la question linguistique, il y a également la dimension sociologique. Nous voulions faire un état des lieux de la réalité de la langue berbère. Où est-elle présente et où elle ne l'était plus ? C'est ce qui nous a poussé à parcourir de bout en bout le Sud-Oranais de Chellala — le point le plus au nord — à Mazzer — le point le plus au sud —. Cela correspond à des milliers de kilomètres parcourus.

II.1.1.1 Les points d'enquête

Nous nous sommes rendu dans chacune des localités où le berbère est encore présent. Pour y constater de visu la réalité de la vitalité du berbère en ces lieux. Tous ces endroits n'étaient pas forcément accessibles aisément, mais nous nous sommes tout de même donné du mal pour nous y rendre — à pied ou en auto-stop —. Pour certains points d'enquête, nous n'avons pas eu à séjourner sur place pour rencontrer des informateurs. Les informateurs de Mazzer, par exemple, ont été rencontrés à Igli.

Nous allons prendre le temps aussi d'évoquer le travail d'André Basset sur cette région. Les notes issues de ces enquêtes de terrain dans le Sud-Oranais sont entreposées dans le Fonds André Basset détenu par la BULAC¹. Ce dernier comprend des travaux qui concernent différentes régions de l'Afrique du nord, donc toutes ces données ne concernent pas exclusivement le Sud-Oranais. Celles qui se rapportent à notre terrain d'étude sont cataloguées dans des boîtes bien étiquetées. Les boîtes en question sont numérotées de 90 à 100. Et les documents sont répertoriés d'après une côte : 3/III-4 et 3/II-3. Ces derniers portent le titre de *Ksours du Sud-Oranais*.



Photographie II.1.1 : Les notes d'André Basset — Fonds Basset —

Donc, avant nous, André Basset a été le dernier linguiste à se rendre sur le terrain². Nous avons par conséquent en quelque sorte repris la relève. L'enquête de terrain de ce

1. Description de ce fonds Basset : <http://bulac.hypotheses.org/4235>.

2. Nous avons eu oui-dire sur le terrain que d'autres chercheurs — algériens — se sont rendus dans le Sud-Oranais pour faire des enquêtes de terrain pendant la période de notre thèse, mais aussi apparemment avant. Pour l'instant en matière de publications rien n'a été publié. Ce qui a été publié a été recensé dans l'Introduction générale (Partie I.1.3.2, page 22).

berbérisant a certaines particularités. Donc, nous allons mener une comparaison entre l'enquête de terrain d'André Basset et la nôtre.

Dans le tableau II.1.1 ci-dessous, nous avons mis en évidence les diverses localités où la présence du berbère était attestée à l'époque d'Edmond Doutté. Cette liste est aussi basée sur nos propres observations sur le terrain. Sfissifa — Béchar — n'est pas cité dans l'enquête d'Edmond Doutté, mais il fut recensé par André Basset. À partir de là, nous avons indiqué les endroits qui ont été prospectés par André Basset et nous-même.



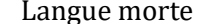
Une des différences qui existe entre notre enquête de terrain et celle d'André Basset, c'est que nous avons pu nous rendre physiquement dans toutes ces localités. Ce qui n'est pas le cas de ce berbésisant qui opérait différemment. André Basset avait pour méthode de se rendre dans les marchés régionaux pour y rencontrer des informateurs venus de différentes localités de la région. Ce qui lui permettait d'optimiser ses enquêtes. Mais il lui arrivait également de se rendre dans les lieux de vie de ces berbérophones.

D'autre part, en matière de données, André Basset n'en a pas collectées dans toutes les localités. Certaines oasis n'ont pas été retenues par André Basset. Nous n'avons trouvé aucune donnée pour toutes les oasis de la région de Taghit. Nous aussi, nous n'avons pas de données pour toutes les oasis. Cela s'explique par le fait que le berbère n'est plus en usage dans ces lieux. Ainsi, dans les localités de Béchar, Chellala gueblia, Taghit et Zaouia Tahtania nous n'y avons pas trouvé de berbérophones. S'agissant de Béchar, grâce aux données d'André Basset, des traces sur la langue de ces habitants ont pu être conservées. Mais pour le berbère des autres localités, nous ne saurons jamais quels étaient leurs aspects particuliers.

Sfissifa de la wilaya de Béchar — à ne pas confondre avec celle de la wilaya de Naâma — est l'autre lieu où nous n'avons pas de données. Bien que nous nous soyons rendu sur place, nous n'avons pas pu rencontrer de locuteurs. Des habitants sur place nous ont asserté qu'il existait une poignée de très vieilles personnes — cinq personnes tout au plus — qui parlaient le berbère.

Localité (Wilaya)	Basset .A	El Idrissi M.
Boussemgoun (El Bayadh)	■	■
Chellala "dahrania" (El Bayadh)	■	■
Chellala gueblia (El Bayadh)		M
Asla (Naâma)	■	■
Moghrar "fougani" (Naâma)	■	■
Kalaat Bouamama "Moghrar tahtani" (Naâma)	■	■
Sfissifa (Naâma)	■	■
Tiout (Naâma)	■	■
Beni Ounif (Béchar)	■	■
Fendi (Béchar)	■	■
Mogheul (Béchar)	■	■
Sfissifa (Béchar)	■	■
Lahmar (Béchar)	■	■
Boukais (Béchar)	■	■
Ouakda (Béchar)	■	■
Béchar (Béchar)	■	M
Zaouia Tahtania (Béchar)		M
Taghit (Béchar)		M
Berrebi (Béchar)	■	■
Igli (Béchar)	■	■
Mazzer (Béchar)	■	■

TABLE II.1.1 : Comparaison de l'enquête de terrain d'André Basset avec la nôtre.

Visité 
 Non visité 
 M  Langue morte

Enfin en matière quantitative, les données d'André Basset pour certaines localités sont peu fournies. La localité d'Igli est la seule oasis où il y a collecté des données de manière conséquente. Il a été même jusqu'à recueillir de courts textes.

II.1.1.2 L'enquête de terrain

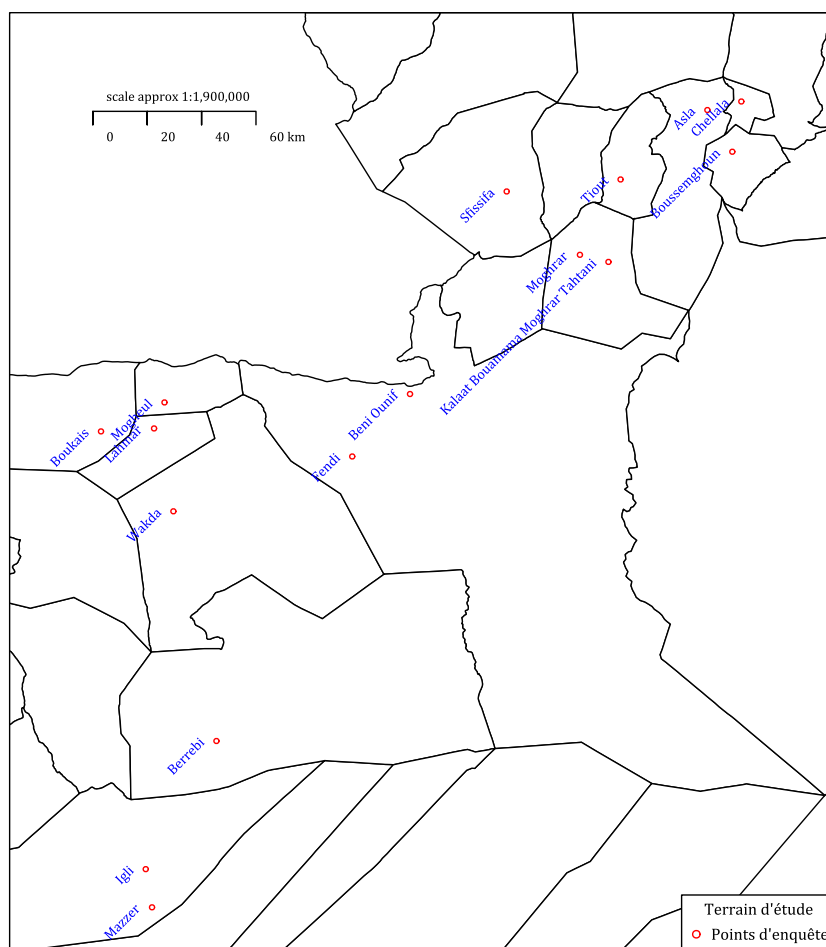
Nous avons eu l'occasion de nous déplacer plusieurs fois sur le terrain. Plus exactement, nous avons été à trois reprises dans le Sud-Oranais. Nous avons séjourné sur place à chaque fois environs deux mois. Notre première enquête de terrain a eu lieu aux mois de février-avril 2015. Puis, l'année suivante, nous sommes retourné dans la région aux mois d'avril-juin 2016. La dernière fois que nous avons été sur notre terrain d'étude, ce fut aux mois d'avril-juin 2017.

Au cours de ces périodes, nous ne nous sommes pas rendu systématiquement dans toutes les localités. Par exemple, nous avons pu aller à Fendi que lors de notre dernière enquête de terrain. Mais les lieux ne présentant pas de difficultés d'accès ont pu être visités à chacun de nos séjours dans le Sud-Oranais.

Les déplacements sur place se faisaient en transport en commun la plupart du temps. Tous ces lieux ne sont pas desservis par une ligne régulière, donc il faut recourir à d'autres moyens, comme les transports clandestins, pour atteindre nos points d'enquête. L'auto-stop a aussi été une autre façon de nous déplacer entre ces localités et nos lieux de résidence.

Effectivement, rarement nous séjournions dans la localité même. Il fallait systématiquement faire l'aller-retour entre nos points d'enquête et notre hôtel. Nous avons pu à de rares occasions séjourner chez nos informateurs qui nous ont hébergé³.

3. Qu'ils trouvent ici toute ma gratitude, notamment Moustapha Bouzid pour m'avoir accueilli chez lui à Asla. Mais aussi toutes les autres personnes qui m'ont proposé l'hospitalité, bien que j'aie dû décliner leur offre.



Carte II.1.1 : Les points d'enquête

II.1.2 L'objet d'étude

Une étude dialectologique peut se faire de bien des manières. Si nous entendons par là que la dialectologie correspond à l'étude de la variation linguistique. Alors, nous pouvons concevoir d'allier à l'envie les variables possibles pour rendre compte des liens qu'entretient un système linguistique avec des dimensions extra-linguistiques.

Ainsi, nous avons décidé de faire une étude géolinguistique qui lie le terrain d'étude — représenté par des coordonnées spatiales — et l'objet d'étude — représenté par des variables linguistiques —. L'originalité de cette étude se présente à divers niveaux. D'abord, ce travail constituera la première véritable étude de géolinguistique portant

sur le Sud-Oranais. Nous employons le terme de "véritable", car André Basset avait intégré le Sud-Oranais dans certaines de ces cartes linguistiques, mais cela n'est resté que superficiel. Ensuite, dans le cadre de notre thèse, nous avons exploré de nouveaux outils de conception de cartes linguistiques. Ces outils sont très utilisés dans d'autres domaines, en revanche, en linguistique berbère, à notre connaissance, ce travail constituera la première intégration de ces méthodes.

Par ailleurs, le second aspect de notre recherche est un travail de classification. L'objectif étant de classer les différentes variétés du Sud-Oranais qui partagent des propriétés communes. La classification interne des variétés du Sud-Oranais n'ayant jamais été réalisée, notre travail est précurseur en la matière. D'autre part, certaines des méthodes qui seront utilisées pour réaliser cette tâche s'avèrent être nouvelles. Sauf erreur de notre part, elles n'ont jamais fait figure d'outils de classement dans des études dialectologiques.

II.1.3 Plan

La première grande partie de cette thèse se décompose en trois parties. La première étape sera consacrée à présenter l'état de l'art en matière de dialectologie. Nous nous intéresserons à ce qui a été réalisé dans les études berbères de manière générale. Puis, nous poursuivrons en élargissant notre revue de la littérature à d'autres régions du monde. Cela permettra de déterminer quels sont les acquis scientifiques de la dialectologie.

Cette étape sera suivie d'un autre chapitre relatif à la méthodologie. Nous y détaillerons tout le processus qui nous a amené à nos résultats. Nous évoquerons également les outils techniques que nous avons employés pour traiter nos données. Mais aussi de quelle manière ces dernières ont été obtenues. Nous chercherons à rendre compte également des limites auxquelles nous avons été confrontés sans travestir la réalité. Cette volonté de ne pas idéaliser le travail de chercheur est essentielle pour ne pas tomber dans des interprétations erronées.

Ensuite, nous ferons étalage de nos résultats. À l'intérieur de cette partie, nous prévoyons de présenter plusieurs points. Chacun de ces points évoquera certains faits. Toutefois, nous n'avons pas la prétention en étudiant ces faits d'avoir épuisé toutes les questions de recherche possible. Bien au contraire, car volontairement, nous nous sommes limité à l'analyse de points particuliers. Effectivement, il existe plein d'aspects

concernant la variation linguistique qui peuvent être étudiés. Dans le cadre de cette thèse, nous allons seulement nous concentrer sur quelques-uns pour une question de faisabilité. C'est ainsi que nous avons décidé d'aborder en deux étapes l'étude de la variation linguistique. À la fin de chacune de ces étapes, nous discuterons de nos résultats. Nous allons ici présenter quels sont les points que nous allons traiter :

— Géolinguistique

L'étude de géolinguistique constituera la première étape. Nous nous attacherons à analyser la variation régionale de différentes variables linguistiques. Des cartes linguistiques nous serviront de support pour exemplifier ces variations linguistiques. Concrètement, nous étudierons les points suivants :

1. Rendre compte de la distribution géographique des variables linguistiques.
Cette étude jette un regard sur la répartition des variables linguistiques sur le territoire du Sud-Oranais.
2. Caractériser les représentations phonologiques des variables linguistiques.
Nous prêterons une oreille attentive à la réalité cognitive des variables linguistiques. Le but est de déterminer si parmi celles-ci certaines ne sont pas des allophones et quelles sont celles qui ont été neutralisées.
3. Analyser les méthodes qui sont mobilisées en phonologie.
Les moyens expérimentaux ne se sont pas encore généralisés en phonologie, alors la méthode déductive basée sur différents postulats — unicité morphémique, équivalence de qualité entre phonèmes de quantités différentes — est utilisée pour accéder à ces représentations. Ainsi, nous ferons usage de ces raisonnements pour évaluer leur pertinence.

— Dialectométrie

Dans un deuxième temps, nous présenterons nos résultats de dialectométrie. Nous comptons déterminer quels sont les **nucléolectes**⁴ qui forment des **haplolectes**.

1. Comparer les techniques d'analyse des données.
Ce travail exploratoire aura pour objectif d'évaluer un certain nombre de méthodes utilisées en **SDD**.

4. Cette nouvelle terminologie sera introduite à la Partie II.4.2.1.2, page 151.

2. Classer les variétés linguistiques.

Notre tâche sera de partitionner les différents nucléoctes en fonction de leurs caractéristiques linguistiques.

En tout dernier lieu, nous finirons cette première grande partie en résumant nos principaux résultats. Et nous évoquerons les pistes futures de recherche en matière de dialectologie.

CHAPITRE II.2

État de l'art

Les premières études de dialectologie à proprement parler naissent en Europe. OÙ des savants vont se charger de décrire les faits linguistiques de variétés linguistiques parlées dans divers territoires d'Europe [Pop, 1950]. Sever Pop fait remonter ses premiers travaux à la fin du XVIII^{ème} siècle. Cependant, le terme de dialectologie a à cette époque un sens restrictif. Toute publication dialectologique comprend uniquement l'étude des variétés linguistiques non littéraires. De fait, les variétés linguistiques prestigieuses étaient exclues.

Dorénavant, cette perception n'a plus cours, mais c'est également le cas au niveau de la méthodologie. Les voies pour étudier la variation linguistique sont multiples. Pour notre part, nous avons opté pour une étude géolinguistique et dialectométrique. Ainsi, par la suite, nous allons rendre compte des spécificités de ces champs disciplinaires en réalisant une revue de littératures. Cela sera entrepris d'abord en évoquant les travaux en géolinguistique, puis viendront ceux en dialectométrie.

II.2.1 La géolinguistique

Des études qui évoquent des variations régionales sont vieilles comme le monde, mais celles qui associent ces variations linguistiques à des cartes géographiques sont plus récentes dans l'histoire de la linguistique. Ce genre de travaux s'insère dans le sous champ disciplinaire dénommé géolinguistique ou linguistique géographique [Chaker, 1998]. C'est un champ disciplinaire qui combine les SDL et les SIG.

La géographie linguistique en elle-même débute à la fin du XIX^{ème} siècle. Toujours selon Sever Pop, celui-ci affirme que le fondateur de cette discipline est Jules Gilliéron. Et qu'à partir de ses travaux, d'autres de la sorte verront le jour un peu partout en Europe et dans le monde. Les motivations pour constituer ces atlas étaient multiples. Elles pouvaient être d'ordres scientifiques comme patrimoniales. La finalité de cette discipline, entre autres, est d'aboutir à des atlas linguistiques.

Rares sont les pays développés qui n'ont pas ce genre d'ouvrages qui décrit la variation linguistique se trouvant sur leur territoire. En revanche, dans les pays en voie de développement, ce genre d'entreprise n'est pas commun. Des chercheurs ont recensé ces travaux^{1 2}, nous renvoyons vers leurs papiers pour qui souhaiterait s'enquérir sur

1. Bibliographie de Joachim Grzega [p.17] : www1.ku-eichstaett.de/SLF/EngluVglSW/0n0n-4.pdf (Consulté le 07/07/2017).

2. Recensement d'atlas linguistique disponible en ligne : www.geolinguistics.com/atlas.htm (Consulté le

ces atlas linguistiques — les liens sont en note de bas de page —.

Donc, nous ne mentionnerons au cours de cette revue de littératures que les travaux qui associent variables linguistiques et informations géographiques.

II.2.1.1 La géolinguistique dans les études berbères

Nous allons de manière succincte présenter les divers travaux menés dans ce domaine. L'ordre de présentation de cette revue de littératures se fera de manière chronologique, par auteur et par aire linguistique berbère. Les études de géolinguistique sont l'une des disciplines des études berbères les plus anciennes, mais l'une des moins foisonnantes. En plus d'un demi-siècle, nous recensons peu d'études publiées et les aires linguistiques concernées se répartissent de manière inégale. Par exemple, les variétés berbères du moyen-atlas n'ont jamais fait l'objet de tels travaux.

Une autre particularité des études berbères est que l'enquêteur est le chercheur lui-même. La collecte des données ne se fait pas par correspondance ou par le biais de plusieurs enquêteurs comme c'est le cas pour certaines études géolinguistiques dans le domaine indoeuropéen [Pop, 1950]. Cette limite en matière de moyens — humains et financiers — explique le fait qu'il n'y ait pas de nombreux travaux en géolinguistique. Et quand ces derniers existent, l'envergure de ces travaux n'atteint pas ce qui a cours ailleurs.

Par ailleurs, notre revue de littératures ne se penchera que sur les aires linguistiques berbères qui ont fait l'objet d'études géolinguistiques approfondies. Donc, nous ne mettrons pas en exergue des régions comme le Chélif où il existe des études de géographie linguistique, mais qui se limite aux quelques cartes réalisées par André Basset. D'ailleurs, nous allons débiter cet état de l'art en évoquant les travaux de ce berbérisant.

II.2.1.1.1 André Basset et la Berbérie

André Basset a été le précurseur des études géolinguistiques en berbère. Ses travaux innovateurs portaient sur l'ensemble du domaine berbère, avec toutefois comme terrain principal l'Algérie. Toutes les principales aires linguistiques berbères de ce pays sont concernées. Sauf erreur de notre part, seuls les îlots berbères du nord ouest algérien —

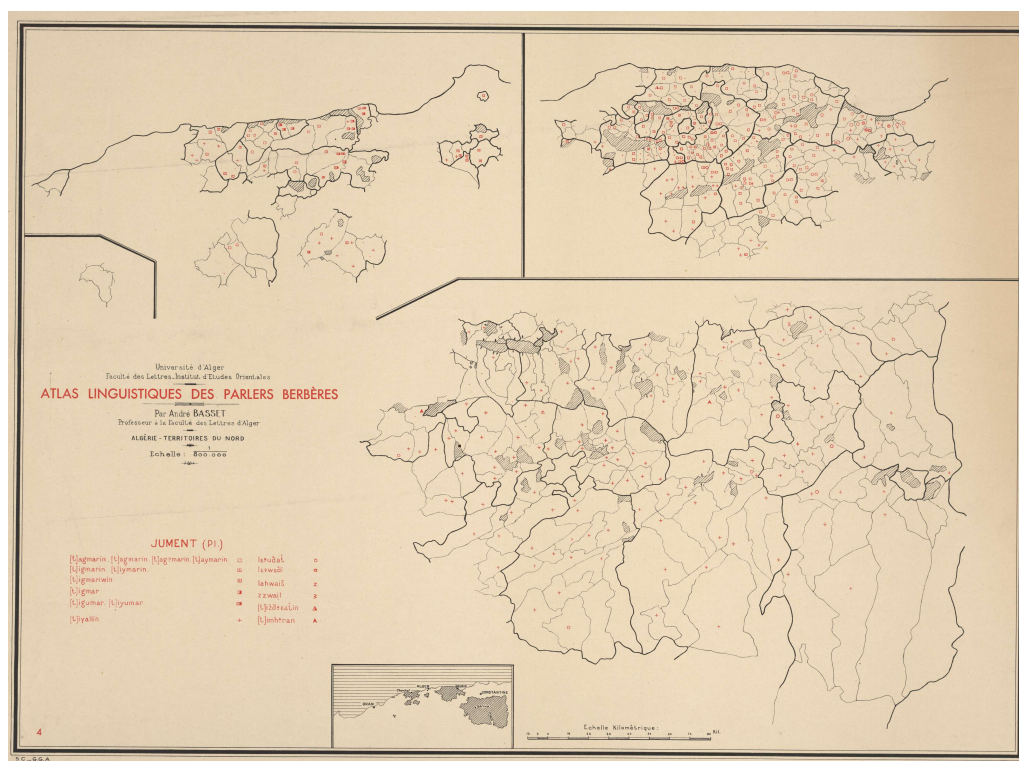
07/07/2017).

l'Oranais — échappent aux études géolinguistiques menées par André Basset. Le quart nord-est de l'Algérie reste toute de même son terrain de prédilection.

Ainsi, les premiers travaux publiés dans ce domaine débutent en 1929 et se poursuivent à intervalle régulier jusqu'en 1959 [Basset, 1929] [Basset, 1935] [Basset, 1936] [Basset, 1938] [Basset, 1939] [Basset, 1948] [Basset, 1951, p.1069-1081] — republication de celle de 1939 — [Basset, 1959] — cette dernière est une republication postume de celles de 1948 —. Bien que les principales publications concernent l'Algérie, il s'intéresse à d'autres aires linguistiques comme l'aire chleuh au Maroc [Basset, 1942]. Ses travaux n'ont jamais fait l'objet d'un recueil de cartes linguistiques. Celles-ci sont éparpillées dans diverses publications. Ces publications correspondent soit à des articles où sont insérées dans le corps du texte des cartes. Soit, ce sont des atlas linguistiques où figurent seulement des cartes accompagnées de commentaires³.

Une remarque d'ordre esthétique et typographique peut être formulée au sujet des cartes réalisées par André Basset. Elles sont de qualités diverses. Certaines sont de factures professionnelles — voir carte II.2.1 ci-dessous —, alors que d'autres sont clairement réalisées à la main et de manière rudimentaire [Basset, 1941]. Nous n'avons pas pu déterminer qui étaient les auteurs de ces cartes, était-ce André Basset en personne ou d'autres individus spécialisés en cartographie? Il semble que les cartes en elles-mêmes proviennent du Gouvernement Général de l'Algérie, mais nous n'en savons pas plus.

3. Certaines de ces cartes en haute résolution sont libres d'accès sur le site de l'université de Bordeaux Montaigne qui a numérisé ses archives : <http://1886.u-bordeaux-montaigne.fr> (Consulté le 15/09/2017).



Carte II.2.1 : Exemple de carte linguistique réalisée par André Basset

Le vœu d'André Basset était de réaliser un atlas linguistique du berbère, malheureusement il n'a pas pu mener ce projet à terme. Il est décédé avant de réaliser cet ouvrage. Une grande partie des données qu'il a collectées n'a pas été publiée. Donc, comme nous avons pu le voir ce chercheur a pu couvrir diverses régions berbères et il reste jusqu'à ce jour le seul à avoir mené des études géolinguistiques sur un aussi vaste territoire.

II.2.1.1.2 Lionel Galand et le Chleuh

Ce chercheur a aussi réalisé une courte étude dans ce domaine [Galand, 1964]. Nous employons le qualificatif de court, car il ne s'agit que d'un article et peu d'items furent étudiés. De plus, le terrain était réduit à une sous aire linguistique du chleuh. L'endroit étudié par Lionel Galand ne faisait pas partie des points d'enquête pris en compte par André Basset. Au moment où nous rédigeons cette thèse, le travail de Galand et de Basset sont les seules études recensées de géolinguistique qui existent pour le domaine chleuh ⁴.

4. D'après Lionel Galand, dans son article de géolinguistique, celui-ci signale qu'André Basset avait réalisé d'autres enquêtes de dialectologie dans le sud du Maroc. Ces derniers seraient entreposés dans le fonds André Basset qui est géré par la BULAC (voir Partie II.1.1.1).

L'autre fait à souligner, nous constatons que le visuel des cartes linguistiques publiées est de qualité professionnelle. Donc, nous en déduisons qu'à cette époque, les techniques de cartographie étaient à la portée des chercheurs et étaient passées définitivement dans les arcanes des méthodes des chercheurs. Cependant, ce travail cartographique n'a pas été mené par Lionel Galand lui-même, mais par une équipe de l'Institut de géographie national. De plus, ces cartes étaient encore réalisées manuellement. Cela n'a rien de surprenant, car l'informatique n'avait pas encore vu le jour. Depuis, les outils ont changé comme nous allons le voir.

II.2.1.1.3 Mena Lafkioui et le Rif

Puis, il faudra attendre presque un demi-siècle pour revoir d'autres études de géolinguistique. Mena Lafkioui [Lafkioui, 2007] a relancé cette discipline dans les études berbères⁵. Son terrain d'enquête était la région du Rif au nord du Maroc. Son travail a débouché sur la publication d'un atlas linguistique complet. Le premier du genre dans les études berbères. Nous le qualifions de complet, car si nous le comparons au travail d'André Basset, cette publication est beaucoup plus dense et riche.

Toutefois, cet atlas se démarque des atlas linguistiques traditionnels. Ces derniers, comme l'*Atlas linguistique de la France* [Gilliéron and Edmont, 1912], ne sont qu'une compilation de cartes linguistiques qui se comptent par centaines. Dans le cas de l'*Atlas Linguistique des variétés du Rif (ALR)*, l'approche est différente. Ce qui différencie le travail de Mena Lafkioui est que son travail comporte également une étude linguistique qui sont plus que de simples notes.

De plus, son atlas se base sur un échantillonnage des variables d'étude. Les variables choisies se réfèrent à toutes les principales composantes d'un système linguistique. Toutefois, elle ne retient que celles qui sont pertinentes. Donc, cette chercheuse laissera de côté l'aspect quantitatif pour se centrer sur les variables linguistiques qui font l'objet d'une variation à l'intérieur de l'aire linguistique rifaine.

En outre, son travail constitue le premier type d'études géolinguistiques intégrant l'informatique. Pour ce faire, elle s'est basée sur tout un ensemble d'outils informatiques allant de la constitution de cartes, à l'élaboration d'un maillage, en passant par l'archivage numérique des données [Lafkioui, 2015]. Toutes ces étapes ont été menées par l'auteur elle-même. Le maillage du territoire rifain est dense pour ce qui est du cœur du Rif, mais

5. Par équité, il faut signaler le travail de Khellaf Madaoui en 1995 dans le cadre d'un mémoire de Diplôme d'Étude Appliqué.

pour les territoires à la marge, les points d'enquête ne sont pas autant parsemés.

Autre fait à signaler, les cartes réalisées par Mena Lafkioui reproduisent le découpage territorial traditionnel de la société rifaine⁶. Cette division spatiale correspond à l'unité tribale. Cette façon de présenter l'aménagement du territoire rifain n'est pas commune. Car, dans les précédentes études — à l'exception du travail d'André Basset sur le sud-ouest marocain — et celles qui suivront, les dialectologues se sont attachés à plaquer sur les cartes linguistiques les divisions administratives des états coloniaux ou post-indépendances.

II.2.1.1.4 Géolinguistique et Kabylie

Pour le kabyle, il existe plusieurs travaux qui portent sur ce domaine. Il y a d'abord les travaux d'André Basset qui furent les premiers dans le genre (Partie II.2.1.1.1). Puis, il faut attendre des décennies après pour trouver d'autres travaux qui ont pour axe de recherche la variation linguistique d'ordre géographique. Nous évoquerons ici même que les principales études et nous laisserons de côté les études mineures — mémoires de licence —. Ces dernières années deux thèses ont intégré dans leur recherche la dialectologie. Il y a la thèse de Said Guerrab [Guerrab, 2014] et celle de Samir Ben Si Said [Ben Si Said, 2014].

Dans la première de ces thèses, une partie des travaux de ce chercheur comprend une étude géolinguistique. Son travail est semble-t-il inspiré par la méthodologie adoptée par Mena Lafkioui. Nous retrouvons la même démarche. Cependant, contrairement à l'ALR, plusieurs pans du domaine kabyle n'ont pas fait l'objet d'enquête, comme le signale d'ailleurs l'auteur lui-même. Par ailleurs, le questionnaire de Said Guerrab comprend des unités phrastiques, alors que cela semble ne pas être le cas pour le corpus de Mena Lafkioui.

Le travail académique de Samir Ben Si Said comprend également un ensemble de cartes linguistiques. Mais faire une description géolinguistique n'est pas l'axe central de ses recherches. Ses travaux sont plutôt orientés vers la morpho-phonologie. Il s'attache à déterminer, à partir de la théorie CVCV, pour quelles raisons certaines variétés du kabyle ont des gabarits syllabiques différents pour des mêmes cognats. Il ne tiendra compte que de cette seule variable linguistique. De plus, son terrain est relativement restreint et ne

6. Ce découpage territorial correspond aussi à un niveau d'organisation territoriale marocain dénommé cercle. Mais la chercheuse lors de la réalisation de ses cartes a à l'esprit, la division traditionnelle et non pas la circonscription administrative.

concerne que très peu de points d'enquête. Cependant, ses données mériteraient d'être comparées avec celles de Said Guerrab. C'est d'ailleurs le seul travail dans le champ des études berbères qui aborde la question des représentations phonologiques sous l'angle de la dialectologie.

II.2.1.1.5 Géolinguistique et Aurès

À l'inverse du domaine kabyle, l'aire linguistique chaouie est très peu pourvue de travaux linguistiques. Mais ce manque a été rattrapé ces dernières années avec l'existence de plusieurs travaux de description linguistique. Parmi ces études, nous comptons des travaux de dialectologie. Cela dit, les premières études géolinguistiques sont le fait d'André Basset qui a intégré la région de l'Aurès dans ses publications de *l'Atlas linguistiques des parlers berbères* [Basset, 1936] [Basset, 1939].

Commençons par évoquer le premier travail réalisé après celui d'André Basset. Celui-ci correspond à l'étude de Salim Lounissi [Lounissi, 2010] qui est menée dans le cadre d'un magistère à l'université de Tizi Ouzou en Algérie. Donc, c'est le premier travail de géolinguistique qui n'est pas encadré par une institution européenne. Ce travail ne prend pas en compte l'ensemble du territoire chaoui. Néanmoins, il englobe plusieurs variables linguistiques.

Encore plus récemment, une thèse a été soutenue à l'INALCO [Boudjellal, 2015]. Le travail de Malek Boudjellal étend son terrain d'enquête à l'ensemble du territoire de l'Aurès. En revanche, comme l'auteur le signale, le maillage n'est pas réalisé de manière uniforme. Certains espaces de l'Aurès sont bien quadrillés, alors que dans d'autres, les points d'enquête sont restreints. Toutefois, ce travail constitue une avancée substantielle par rapport aux enquêtes de Salim Lounissi. Outre l'extension territoriale, davantage de variables linguistiques sont prises en compte.

II.2.1.1.6 Géolinguistique et Sud-Oranais

Le Sud-Oranais a également fait l'objet d'études géolinguistiques, mais cette région faisait partie d'un terrain d'enquête plus vaste comprenant tout le Sahara algérien et les territoires touaregs — Mali et Niger inclus —. Sa collecte s'étendait même jusqu'à l'ouest libyen et le sud tunisien. Concernant le Sud-Oranais de ce que nous pouvons nous rendre compte, les résultats publiés par André Basset sont en partie différents des nôtres — Exemple la carte 4 [Basset, 1948] à comparer avec II.4.11 —. Certaines données

correspondent bien à certains points d'enquête, mais ne peuvent être généralisées à l'ensemble des variétés du Sud-Oranais. Deux hypothèses peuvent être imaginées quant à ces divergences. Soit, il y a eu une méprise de la part d'André Basset. Ou bien, l'autre interprétation possible, c'est de considérer qu'entre son enquête de terrain et la nôtre, il y a eu une évolution linguistique.

II.2.1.2 La géolinguistique dans le monde

Ce que nous souhaitons faire ressortir au cours de cette partie est les enseignements apportés par la géolinguistique. Il existe actuellement des atlas linguistiques de par le monde, donc ces matériaux devraient pourvoir donner un certain nombre d'indications sur la variation linguistique. Mais, en réalité, il y a très peu de résultats de portée générale qui sont tirés des études menées en géolinguistique.

Ces études ressemblent à une énumération de données, sans que celles-ci soient discutées pour savoir ce qu'elles nous disent par rapport à la réalité géographique. Comment cette dernière modèle la réalité linguistique — si elle le modèle!—?. Par ailleurs, autre limite à mettre sur le compte des études géolinguistiques, le terrain d'étude est cantonné aux zones rurales. Les villes sont délaissées et peu d'études [Williams and Merwe, 1996] ont lié variation linguistique et visualisation géographique en zone urbaine.

II.2.1.2.1 La Variation linguistique et les Variables linguistiques

Notre revue de littératures nous a permis de constater qu'un large panel de variables linguistiques étaient choisies pour établir la variation linguistique — qu'elle soit illustrée ou non par des cartes linguistiques par la suite —. Les deux principales variables linguistiques prises en compte par les dialectologues sont celles appartenant aux unités significatives — toutes catégories syntaxiques confondues — et aux systèmes phonético-phonologiques. La variation syntaxique quant à elle joue un rôle mineur, de même que la variation paradigmatique [Scherrer and Stoeckle, 2016] [Rabanus, 2011]. La dimension syntaxique commence ces dernières décennies à être l'objet d'un certain intérêt et plusieurs projets existent dans ce sens⁷.

7. Pour de plus amples informations sur ces projets, ce site internet recense les divers projets académiques qui ont prévu d'étudier la variation syntaxique de différentes aires linguistiques : www.dialectsyntax.org (Consulté le 07/07/2017).

Par ailleurs, les unités morphémiques ou phonologiques sont également très peu prises en considération [Nerbonne and Krandschmar, 2006]. Les travaux de Weinreich [Weinreich, 1968] ont pourtant tenté de pousser les dialectologues dans ce sens, mais la réalité cognitive n'est pas un objet de recherche qui a passionné la dialectologie. Dans les années soixante et soixante-dix apparaissent des approches théoriques — generative dialectology, structural dialectology [Chambers and Trudgill, 1998, p.33-42] — qui essaient de décrire la variation linguistique sous l'angle de la phonologie. Malgré cela, elles n'ont pas essaimé auprès des nouvelles générations de dialectologues comme nous l'avons signalé.

Il n'y a également pas de prise en compte d'unités plus abstraites comme la sémantique. Ainsi, à notre connaissance, la variation sémantique n'est pas considérée. Si elle l'est, elle l'est de manière éparse, comme c'est le cas dans les études berbères. Il existe nombre d'études réalisées par des berbérissants, dans une démarche descriptive, mentionnant les différences sémantiques de tel ou tel cognat dans tel ou tel lieu. En revanche, ces travaux ne s'inscrivent pas dans une visée dialectologique qui serait la recherche systématique et régulière de ces variations sémantiques. Sinon, il existe des travaux considérant la variation sémantique [Glynn and Fischer, 2010], mais qui ne se basent pas sur une visualisation spatiale des données, ni qui cherche à réaliser une segmentation des variétés linguistiques. Ainsi, nous pouvons dire que la morphologie est réellement la variable par excellence qui est retenue par les dialectologues.

Ensuite, il semble que le choix de ces variables linguistiques et la collecte de données se fassent d'après une méthodologie qui est peu ou prou la même. Cette dernière est bien établie de par le monde chez les dialectologues. Le choix des unités lexicales se portera au niveau du lexique fondamental. Le nombre d'informateurs par points d'enquête ne dépasse guère les chiffres des unités. L'élicitation et le questionnaire sont la norme. Le nombre de variables retenues fut un temps très important, cependant la tendance actuelle est de choisir un échantillon représentatif. De par leur expérience, les dialectologues ont déterminé que le nombre de variables devaient être réduit et qu'il fallait éviter la recherche d'exhaustivité. Mais comment est déterminé cette représentativité, c'est une question qui n'est pas explicitée. Il semble que chaque chercheur aborde son terrain par tâtonnement. Plusieurs étapes sont nécessaires pour mettre sur pied un questionnaire comme l'indique plusieurs chercheurs [Basset, 1959] [Rasoloniaina, 2014] [Lafkioui, 2015].

Malgré tout, l'ensemble de ces aspects ne nous informent pas sur quels sont les

fait linguistiques qui ne font pas l'objet de variation ou qui le sont moins par rapport à d'autres faits linguistiques. Il semble toutefois que certains chercheurs ont mis en évidence que les faits linguistiques qui se distinguaient des autres avaient une aire d'extension qui se propageait à partir d'un centre de diffusion régionale. Ce genre de modèle est couramment dénommé en français la Théorie des vagues [Labov, 2007]. D'autre part, la gradualité de ces variations diffère d'un fait linguistique à l'autre. Un fait linguistique peut connaître de nombreuses variations, alors que pour un autre fait, il n'y aura pas autant de différences. Tous ces constats sont amplement illustrés dans l'ouvrage de référence de Sever Pop [Pop, 1950].

II.2.1.2.2 La Variation linguistique et la Distance géographique

Un des enseignements tiré par les études géolinguistiques est le rapport qu'entretient la variation linguistique et l'espace. Ainsi, plusieurs chercheurs ont corrélié la distance linguistique à la distance géographique [Séguy, 1971]. Plus une variété est distante de l'autre, plus des différences linguistiques se feront sentir. Pour rendre compte de cet aspect, il est nécessaire de réaliser une étude dialectométrique — ce point sera abordé plus en détail par la suite (Partie II.4.2) —. Le dialectologue Hans Goebel a même proposé une technique de visualisation à partir de ce qu'il appelle une matrice de corrélation pour rendre compte d'autres phénomènes [Goebel, 2012]. Ses résultats montrent, à partir des variétés linguistiques françaises, que la diffusion d'une variété se fait de manière inégale. Certaines auront une étendue très vaste, alors que d'autres sont circonscrites à portion congrue.

D'autre part, grâce à la géographie linguistique, il a été constaté qu'il existait des continuums linguistiques. La variation se fait de proche en proche, elle est rarement abrupte [Chambers and Trudgill, 1998, p.5]. Mais cette continuité ne se fait pas de la même façon pour tous les faits linguistiques. Ce qu'une variété peut partager avec une autre, ne sera pas nécessairement partagé avec d'autres variétés. C'est pour cette raison que la notion d'isoglosse a vu le jour.

II.2.1.2.3 La Variation linguistique et la Politique linguistique

Cette dénomination de politique linguistique englobe d'autres sous ensembles disciplinaires comme l'aménagement linguistique, les droits linguistiques et la sociologie du langage. Une des dimensions qui est en jeu est le principe de territorialité ou territoire

linguistique [Viaut, 2007, p.69]. Ce principe commande à ce qu'une langue présente sur un territoire donné doit pouvoir jouir d'un statut officiel. Pour qu'une action soit possible dans ce sens, cela implique d'avoir des données statistiques sur ces populations et sur l'extension de l'aire linguistique de cette variété. Cette dernière correspond le plus clair du temps à des minorités linguistiques [Auer and Schmidt, 2010, p.315].

L'autre aspect où la géolinguistique trouve écho concerne l'aménagement linguistique. La caractérisation des variétés linguistiques — leurs particularités linguistiques — et de leur répartition spatiale sont capitales pour mener une standardisation. Cet inventaire des variétés linguistiques a trouvé une application au niveau des politiques linguistiques. Bien que plusieurs études signalent l'étroitesse des liens qui existe entre la géolinguistique et l'aménagement linguistique, nous n'avons pas trouvé de cas concrets où explicitement la géolinguistique a joué un rôle décisionnel dans un processus d'aménagement linguistique. C'est certainement pour cette raison que Mena Lafkioui a publié un article [Lafkioui, 2007] pour rappeler l'importance que revêt la géolinguistique pour agir sur des variétés linguistiques.

II.2.2 Variétés linguistiques et Sciences des données

La variation linguistique peut être traitée en se limitant à des analyses purement linguistiques ou bien de manière plus transdisciplinaire. Cette dernière ouvre des champs d'investigation considérables pour la compréhension du langage. Ainsi, les faits linguistiques peuvent aussi être considérés comme des données après qu'elles soient collectées par des linguistes. L'une des disciplines qui peuvent aider à traiter ces données sont les mathématiques.

Ce domaine qui lie mathématiques et traitement des données se dénomme par bien des noms. Il tend ces dernières années à être regroupé sous une seule dénomination qui est les **SDD** — en anglais Data Science —. Sous cette dernière classification est regroupée tout ce qui se rapporte à la Fouille de données — Data Mining —, au Machine Learning, à l'Intelligence artificielle et au Deep learning. Des champs qui se distinguent par leur objet d'étude, mais aussi par les calculs mathématiques utilisés. Toutefois, il partage en commun l'utilisation de l'informatique. Ces disciplines ne pourraient exister sans l'utilisation de logiciels dédiés à leur pratique. Ce dernier est le troisième pilier de cette discipline.

Du point de vue des calculs, les techniques mathématiques à la base des **SDD**

fonctionnent toutes sur une modélisation mathématique qui correspond à l'Espace vectoriel orthonormé. Celui-ci est une sorte d'espace physique idéalisée avec un référentiel et des axes — si tridimensionnel : x,y,z —. Cela implique que chaque variable est caractérisée par cet espace à partir de ses coordonnées spatiales qui sont la projection d'un élément géométrique — un point généralement — sur des axes. Il existe plusieurs moyens pour déterminer une valeur qui assignera une place à cette variable au sein de cet espace. Ainsi, le fait que ces variables soient dans un espace vectoriel, des distances peuvent être postulées entre ces variables. Et par conséquent grâce à ces distances de pouvoir manipuler ces variables en indiquant celles qui sont plus ou moins proche l'une de l'autre. Quand ces variables sont nombreuses les matrices sont utilisées comme outil mathématique pour réaliser des calculs [Steele and Chandler, 2016, p.14].

Le fait que ces variables soient projetées dans un espace a son importance. Car dans le cas des variables linguistiques, il n'est pas évident de caractériser des concepts abstraits en terme de distance. Ceci pose un problème pour ce type de variables qui ne sont pas des objets inscrits dans un espace continu et qui sont seulement des réalités discrètes. Car, comme l'on soulèvait plusieurs linguistes [Goebel, 2012], comment déterminer par exemple la distance d'une unité lexicale par rapport à une autre ? De plus, les variables qualitatives — ce que sont les variables linguistiques — ont la facheuse tendance à encoder différents niveaux de compréhension. Une variable qualitative en cache toujours une autre ou plusieurs. À l'heure actuelle, les machines sont incapables de décoder ces différentes couches. Donc, l'Homme, s'il veut que son étude sur ces données aboutisse à des résultats pertinents, doit traiter lui même ses données avant de les présenter aux programmes informatiques. Ainsi, le traitement de ces données peut se faire de manière fine ou grossière et cela induit inévitablement des résultats qui ne seront pas les mêmes. Par conséquent, ces limites que nous venons de soulever donnent une idée des approximations qui peuvent résulter de ces études.

D'autre part, en lien avec cette branche des mathématiques, les calculs statistiques classiques sont aussi utilisés pour valider la pertinence des résultats obtenus ou les comparer. Voire certaines techniques en SDD sont conditionnées aux valeurs statistiques. Tous les domaines des statistiques ne sont pas nécessairement concernés. Seul une partie de ce champ disciplinaire a à faire avec les SDD. Ainsi, de manière synthétique cela correspond aux calculs mesurant la distribution et aux calculs de probabilité.

Après ce court résumé introduisant les SDD et les fondations mathématiques qui sous-tendent cette discipline, évoquons maintenant comment celles-ci ont pu être

utilisées en SDL⁸ et notamment en dialectologie. Pour cela, il faut avoir à l'esprit que la finalité des SDD est de déterminer les rapports qu'entretiennent les données entre elles. La perspective est d'établir quelles sont les données qui ont des liens étroits — en anglais le mot utilisé pour qualifier ces rapports, dans toutes les disciplines confondues, est *patterns* [Waller and Fawcandt, 2013] —. Et pour mettre en saillance ce type de données, ces dernières seront regroupées pour former des partitions. C'est cet aspect qui est exploité par la linguistique. En se basant sur cette définition générale, les données linguistiques peuvent être analysées en prenant en considération différents axes de recherche. Mais si nous limitons notre champ de recherche aux analyses recoupant la géographie et la linguistique. Alors, nous distinguons deux directions de recherche en dialectologie qui lient les SDD avec des variables linguistiques. Cela correspond à la dialectométrie et la géostatistique linguistique. Nous allons commencer par faire une revue de littératures par ordre chronologique d'apparition de ces disciplines.

II.2.2.1 Dialectométrie

Depuis les premiers travaux de lexicostatistique à nos jours, les études qui lient mathématiques et linguistique ont quelque peu évolué. La dialectométrie s'inscrit dans ce mouvement. Les spécialistes en dialectométrie s'accordent pour dire que le fondateur de cette discipline est Jean Séguy dans les années soixante-dix [Goebel, 2012]. Mais l'utilisation de calcul pour indiquer des distances entre des variétés linguistiques remonte à bien plus longtemps, avant même les travaux de Swadesh dans les années cinquante [Métoz, 2006].

La lexicostatistique et la dialectométrie se distinguent sur deux points. Le premier utilise ces calculs pour déterminer l'âge et le moment où des variétés linguistiques voient le jour [Gudschinsky, 1956]. Le second utilise plus ces calculs pour mettre en évidence les distances linguistiques qui existent entre les variétés. En plus, des méthodes graphiques comme des cartes linguistiques sont utilisées pour visualiser cette variation. Les travaux de cette époque étaient réalisés manuellement. L'outil informatique ne vient que bien après, avec les premiers travaux de Hans Goebel [Goebel, 2012].

Précisons par ailleurs que nous ne mentionnerons au cours de cet état de l'art que les travaux se rattachant à la dialectométrie et laisserons tous les travaux précurseurs

8. Il existe d'autres disciplines qui ont recours à ces méthodes comme le Traitement automatique des langues ou la Fouille de textes.

en lexicostatistique. De même, nous n'évoquerons pas tous les travaux de classification réalisés sur les langues berbères. Pour ce dernier point, nous renvoyons à l'article de Meftaha Ameer [Ameer, 1990] pour de plus amples informations.

II.2.2.1.1 Travaux dans les études berbères

Dans les études berbères, nous ne recensons que trois chercheurs qui ont conduit des études poussées en dialectométrie. Le premier à avoir initié cette méthode est Kamal Naït-Zerad. Il publiera quatre articles dans ce sens. Les deux premiers ont comme terrain d'étude la Kabylie [Naït-Zerad, 2005] [Naït-Zerad, 2009] et le troisième concerne plusieurs aires linguistiques berbères [Naït-Zerad, 2006]. Les données de Kamal Naït-Zerad pour ces études dialectométriques sont de seconde main. Ses données sont fondées sur des collectes menées par d'autres berbérissants. La dernière étude concerne le sud marocain. Cette fois-ci ses données sont de première main [Naït-Zerad, 2004]. Mais l'objectif de ses travaux est exploratoire, ses articles sont à valeur pédagogique. Le but étant de sensibiliser la communauté des berbérissants sur les potentiels de recherche que représente la dialectométrie, car comme il le stipule, dans les études berbères ce genre de travaux "n'en est qu'à ses débuts" [Naït-Zerad, 2009, p.352].

Le second travail est celui de Mena Lafkioui qui réalisera un article sur la question [Lafkioui, 2009]. Le Rif est toujours son terrain d'étude de prédilection. Enfin, le dernier travail sur le sujet a été réalisé par Said Guerrab [Guerrab, 2014]. Il prolongera le travail de Naït-Zerad sur la Kabylie en intégrant plus de points d'enquête et plus de variables linguistiques. L'auteur indique que les résultats auxquels il aboutit sont sensiblement identiques à ceux de Naït-Zerad.

Ces trois chercheurs utiliseront les mêmes outils informatiques et les mêmes calculs statistiques pour leurs études dialectométriques. Ces derniers ont été mis en place par une équipe de chercheurs de l'université de Groningen — pour plus de renseignements sur ces outils voir Partie II.3.8.2 à la page 76 —.

II.2.2.1.2 Travaux dans les autres aires linguistiques

Les travaux de dialectométrie ont essaimé de par le monde. Nous n'évoquerons pas la teneur de ces travaux, car elles ne sont pas le cœur de notre sujet. Nous signalerons seulement quelques aires linguistiques où nous trouvons des études dialectométriques. Nous regrouperons les variétés linguistiques par grandes aires linguistiques ou

géographiques.

L'aire germanique est l'aire linguistique qui est à la pointe en matière de dialectométrie. Nous dirons même que sans ces chercheurs d'origine germanique — allemands, autrichiens, hollandais, ... etc —, probablement la dialectométrie serait restée mort-née. Quoi que qu'il en soit nombre d'études sont apparues dans ce domaine (*Etat-Unis* : [Lee and Krandzschmar, 1993]; *Angleterre* : [Szmrecsanyi, 2012]; *Hollande* : [Nerbonne, 1999]; *Norvège* : [Heeringa and Johnson, 2009]; *Suisse* : [Scherrer and Stoeckle, 2016]; *Allemagne* : [Schiltz, 1996])

L'aire romane a aussi connu son lot de travaux dialectométriques. Hans Goebel [Goebel, 2012] a exploité les différents atlas linguistiques pour réaliser des études sur les variétés linguistiques dérivées du latin. Toujours en Europe, nous retrouvons des études dialectométriques qui concernent le basque [Aurrekoandxea, 2016].

L'Afrique subsaharienne, pour sa part, connaît également un certain nombre d'études dialectométriques [Guarisma, 1986] [Kluge, 2005]. Le continent asiatique recèle plusieurs études géolinguistiques, mais il semble que des études de dialectométrie ne soient pas encore en vogue. Notre étude bibliographique a seulement recensé quelques études dans le domaine (*Chine* : [Yang, 2010]; *Iran* : [Cathcart, 2015], *Japon* : [Kumagai, 2016]). Soulignons que les études concernant ces deux aires géographiques utilisent des techniques statistiques et d'explorations visuelles limitées contrairement à ce que nous pouvons retrouver dans les études germaniques ou romanes.

II.2.2.2 Méthodes en Sciences des Données

Les problèmes soulevés en dialectométrie restent des questions ouvertes, il n'existe pas de chercheurs qui vous stipuleront que c'est de cette façon que les choses se réalisent. Le potentiel offert par les méthodes qui ont cours en SDD sont loin d'avoir toutes été explorées en dialectométrie, alors même que de nouvelles méthodes voient constamment le jour. Par conséquent, nous allons élargir notre revue de la littérature aux autres méthodes existantes.

Certaines d'entre elles ont été employées dans les études que nous avons citées et nous les évoquerons sans prétendre à l'exhaustivité. Mais nous mentionnerons également les autres possibilités non encore explorées. Ces dernières correspondent seulement à celles qui sont implantées informatiquement — dans R, Python, Scala ou Java entre autres —. Car toutes les méthodes qui existent ne sont pas nécessairement informatisées.

L'analyse des données passe par plusieurs étapes. La présentation de notre état de l'art se déroulera dans l'ordre de succession de ce processus. Chaque stade que nous évoquerons présentera une indication sommaire de l'intérêt que revêt cette étape. Nous avons distingué quatre grandes étapes.

II.2.2.2.1 Des Données linguistiques aux Données numériques

La transformation des faits linguistiques en valeurs numériques est la première des étapes. Les machines ne comprenant que les chiffres, les faits linguistiques représentant les variables linguistiques doivent être transformés en valeurs numériques. Mais cette étape n'est pas nécessairement obligatoire. Cependant, certaines méthodes que nous verrons par la suite ne peuvent être utilisées si ces observations qualitatives ne sont pas transformées en observations numériques. Ensuite, il y a lieu de distinguer les méthodes qui produisent *in fine* des matrices de similarité et celles qui changent uniquement les faits linguistiques en chiffres.

Pour cette dernière, différentes méthodes existent. Citons en premier lieu la méthode Dummy qui transforme les observations qualitatives en valeur numérique binaire [Bandyopadhyay and Saha, 2012, p.60]. Ainsi, une variable comprenant trois différentes observations sera éclatée en trois autres variables. Et les observations figurant dans ces nouvelles variables seront des 1 ou des 0. Le désavantage de cette méthode est qu'elle peut aboutir à un nombre démesuré de variables. Une autre méthode qui donne des résultats semblables à la méthode Dummy est la méthode Bipartite spectral graph [Wieling and Nerbonne, 2011]. Puis, l'autre possibilité est de se baser sur la fréquence de distribution d'une observation qualitative et d'affecter à cette dernière sa fréquence de distribution. Le risque de cette méthode est d'avoir une fréquence de distribution identique et donc de neutraliser les différences entre les observations. La plus simple des méthodes est pour chaque différente observation qualitative de donner un chiffre de manière arbitraire. Le souci de cette méthode est que nous perdons des informations sur les liens possibles qu'entretenaient les observations entre elles.

L'autre méthode détermine des distances entre les observations qualitatives de deux points d'enquête différents. Cette méthode est dénommée Mesure de similarité. Au sein de cette méthode, il faut distinguer celles se basant sur les chaînes de caractères — en anglais Edit distance — et celles qui ne tiennent pas compte de ce paramètre. La première de ces méthodes est intuitivement plus proche de la réalité linguistique. Pour ce cas, il existe également beaucoup de possibilités. Plusieurs chercheurs ont évalué

et recensé une partie de ces techniques [Kruskal, 1983] [Heeringa and Kleiweg, 2006] [Naït-Zerrad, 2009]. La distance entre ces caractères peut se faire en prenant en compte un seul caractère ou plusieurs — N-gramme méthode —. Ou bien le caractère peut être divisé en unités plus petites correspondant aux traits phonétiques. Une prise en compte ou non de l’alignement des caractères est aussi possible. Pour chaque type de différences existant entre ces caractères une valeur différente peut être affectée — poids —. Et la totalité de ces valeurs pourra être normalisée suivant le nombre de caractères existant dans le morphème. Nous venons de résumer qu’un bref aperçu de certaines des différentes techniques pouvant être utilisées. En somme, plus les caractères sont identiques et plus la distance sera faible et inversement — Des caractères identiques auront comme valeur 0 —.

L’autre méthode de Mesure de similarité se dénomme en anglais Simple Matching Distance [Gan and Ma, 2007, p.76]. Cette méthode affecte une valeur 1 si les chaînes de caractères d’un fait linguistique entre deux variétés linguistiques sont identiques en tout point et 0 autrement. Cette dernière méthode est celle qui fut employée pour la première fois en dialectométrie [Séguy, 1971].

Arrivé à ce stade, nous obtenons plusieurs valeurs pour chaque variable et chacune de ces valeurs correspond à la mesure de similarité entre deux points d’enquête. L’étape suivante est d’obtenir une seule valeur entre deux points d’enquête en cumulant les valeurs de chacune de ces variables.

II.2.2.2 Matrice de similarité et Mesure de Distance

En fonction de la méthode choisie précédemment et du type de partition choisie (voir ci-dessous Partie III.3.6.1.1), il est nécessaire de créer une matrice de similarité — ou de dissimilarité —. Donc, cette étape peut être négligée, à la condition seulement que les faits linguistiques soient transformés en données numériques. Situons nous dans le cas où nous avons besoin d’établir une matrice de similarité.

Ainsi, pour ce faire, les calculs pouvant être utilisés sont divers. Le choix du calcul dépend de la méthode choisie pour créer une mesure de similarité. C’est-à-dire si les valeurs obtenues l’ont été par la méthode d’Edit Distance ou par la méthode de Simple Matching Distance. Si la méthode d’Edit Distance est utilisée, alors la distance linguistique entre deux points d’enquête peut être mesurée en ayant recours à différentes formules mathématiques [Bandyopadhyay and Saha, 2012, p.60]. La plus connue est la plus utilisée en dialectométrie est la distance euclidienne.

Ces méthodes sont normalement destinées pour mesurer une distance entre deux valeurs — ou deux points si nous le projetons dans l'espace —. Mais dans le cas présent, après avoir effectué une mesure de similarité, il n'y a plus qu'une seule valeur. Donc, la question peut se poser pourquoi utiliser dans ce cas une mesure de distance par la suite. Toujours est-il que c'est de cette manière que certains en dialectométrie procèdent.

Ensuite, si la méthode retenue est la Simple Matching Distance, alors il y a aussi plusieurs calculs qui peuvent être employés. Par exemple, citons : Simple Matching Coefficient, Positive Matching Index, Tripartite Similarity Index, Similarity Coefficient de Gower et d'autres — voir [Mitchell, 2012, p.142] [Bren, 1998] —. Ces calculs ont l'avantage de normaliser les valeurs puisqu'elles sont toutes divisées par la suite par le nombre de variables linguistiques.

La matrice de similarité peut aussi être obtenue en se basant sur des calculs statistiques comme la métrique du χ^2 .

II.2.2.2.3 Partition des données

C'est au cours de cette étape que le regroupement des données se fait. Le but est de réaliser une partition des données réunissant les variétés linguistiques qui se ressemblent. Cette partition des données peut se faire de bien des manières. Mais il est courant de distinguer deux méthodes différentes. Ces dernières sont dénommées Apprentissage Non Supervisé et Apprentissage Supervisé.

Bien que ces deux méthodes réalisent une partition des données. Le point de départ de leur problématique n'est pas le même. Dans la première méthode, nous partons du principe que n'avons aucune idée des partitions possibles qui pourraient exister entre les données. Alors que pour la deuxième méthode, nous savons de quelle manière nous voulons regrouper nos données. Ces deux objectifs opposés entraînent l'utilisation de méthodologies et de techniques différentes.

Brièvement, nous allons présenter ces deux méthodes. Nous allons dans un premier temps voir comment la partition des données s'opère dans le cas d'un Apprentissage Non Supervisé. Puis, nous évoquerons les méthodes par Apprentissage Supervisé. Il ne s'agira pas pour nous de faire état de toutes les méthodes qui existent. Notre intention est plutôt de faire l'état des lieux des méthodes utilisées en dialectométrie tout en mettant en lumière les autres méthodes qui n'ont pas été prises en compte par les chercheurs de cette discipline. Les calculs mathématiques sous-jacents ces méthodes ne seront pas non plus évoqués. Pour une vision globale des méthodes en cours, nous renvoyons à ces

ouvrages qui présentent de manière assez exhaustive les différents calculs utilisés par ces deux méthodes : Apprentissage Non Supervisé [Gan and Ma, 2007] et Apprentissage Supervisé [Alpaydin, 2014].

II.2.2.3.1 Partition des données : Apprentissage Non Supervisé

Pour parvenir à cette étape, il n'est nul besoin de passer par la réalisation d'une matrice de similarité. Mais certaines méthodes ne peuvent être concrétisées sans cette matrice. Celles qui font usage de matrices de similarité peuvent permettre — cela dépend de la méthode choisie — de créer une distance hiérarchique entre les variétés linguistiques. Alors que les méthodes sans matrice de similarité se contentent de regrouper les variétés linguistiques en groupes sans les hiérarchiser.

Ainsi, nous avons donc distingué deux grandes méthodes d'Apprentissage Non Supervisé. Celles qui réalisent une hiérarchisation des données et celles qui n'y ont pas recours. Cette distinction a son importance en dialectométrie dans la mesure où la hiérarchisation revêt une importance considérable en dialectologie. Pour des histoires de phylogénie linguistique, les dialectologues ont une préférence pour les méthodes hiérarchisées.

II.2.2.3.1.1 Non Supervisé et Hiérarchisé

Cette méthode est de loin la plus utilisée en dialectométrie. Toutes les études que nous avons pu citer précédemment utilisent la partition hiérarchique pour rendre compte des rapports qui existent entre les variétés linguistiques. Pour cela, plusieurs techniques peuvent être utilisées. Toutefois, deux types de méthode sont discernables. Il y a la méthode ascendante et la méthode descendante.

Ces dernières dépendent toutes des calculs qui ont pu être utilisées pour parvenir à établir la matrice de similarité, d'une part. Et d'autre part, d'autres calculs sont nécessaires pour la méthode ascendante pour regrouper les variétés entre elles. Ces méthodes d'agrégation — ou linkage — portent le nom de méthode Ward, Average, Single, Complete entre autres [Gan and Ma, 2007, p.109].

Cependant, ces méthodes de partitions hiérarchiques donnent des résultats fluctuants. Le nombre de variables linguistiques considérées, l'ordre d'agencement de ces variables dans la base de données, l'algorithme utilisée vont changer la partition et l'agrégation des données.

II.2.2.3.1.2 Non Supervisé et Non Hiérarchisé

Abordons cette fois-ci l'autre méthode Non Supervisée. Celle-ci se distingue par le fait que la partition des données ne se fera pas en ayant recours à une agrégation hiérarchique. Il n'y a qu'une dimension de partition. Une nouvelle fois, les techniques pouvant réaliser de telle partition sont nombreuses. Commençons par citer les techniques qui demandent qu'il soit spécifié le nombre de partitions souhaitées. Celles-ci comprennent les méthodes k-moyennes — Wieling [[Wieling and Nerbonne, 2011](#)] utilise cette méthode —, k-médoïdes, k-médiane entre autres. Ces derniers se distinguent par les méthodes d'agrégation utilisées — voir ci-dessus —.

D'autres méthodes se basent sur une réduction des dimensions en déterminant mathématiquement les valeurs propres et l'espace propre d'une matrice. Cela revient à déterminer quelles sont les variables qui ont réellement une influence sur les partitions. Les méthodes appartenant à cette catégorie sont Principal Component Analysis, Factor Analysis [[Aurrekoetxea, 1995](#)] [[Nerbonne and Krantzschmar, 2006](#)] [[Péronnet, 1998](#), p.572], MDS entre autres. Cette dernière méthode a été très employée en dialectométrie — voir les travaux publiés dans les études berbères (Partie [II.2.2.1.1](#) page 52) —. Car cette méthode fait usage de la matrice de similarité et par conséquent cela facilite le repérage des partitions.

Toujours dans le cas d'une partition non hiérarchisée, nous pouvons encore citer les méthodes qui font usage des calculs statistiques. Il existe plusieurs méthodes qui peuvent être utilisées. À titre d'exemple, citons Gaussian mixture et il en existe encore bien d'autres.

II.2.2.3.2 Partition des données : Apprentissage Supervisé

Toute donnée ayant pu être traitée par la méthode par Apprentissage Non Supervisé peut connaître le même traitement en Apprentissage Supervisé. Cette méthode est même parfois considérée comme l'étape devant suivre un Apprentissage Non Supervisé. Après que des connaissances aient été tirées de l'Apprentissage Non Supervisé, il faut s'assurer que les partitions obtenues soient adéquates et non pas dues au hasard. Pour valider ces partitions, l'Apprentissage Supervisé peut servir de méthode pour réaliser cet objectif.

L'idée est de s'assurer que toutes les données se répartissent dans les partitions choisies. Pour ce faire, ce travail se déroule en deux étapes dénommées Training et Validation/Test. Ainsi, les données brutes sont divisées en deux, une partie servira

pour le Training et l'autre partie pour la Validation. De cette manière, si les résultats concordent, il y a une prédiction qui est formulée et donc cela permet de valider les partitions. Cette description générique se réfère à ce qui est appelée une Cross-validation. Ainsi, pour réaliser cette phase, plusieurs techniques peuvent être utilisées — K-fold, Leave-P-Out, Holdout Method — qui sont dans le principe identique comme nous l'avons expliqué, mais qui vont se différencier sur certains aspects.

Enfin, signalons qu'à notre connaissance, aucun chercheur-e en dialectométrie s'est aventuré-e à utiliser cette méthode. C'est pour cette raison que nous n'avons pas cité d'études. Cependant, il existe d'autres champs des **SDL** qui y ont recours. Par exemple, les chercheurs en Fouille de textes se servent de cette méthode, entre autres, pour classer les textes sur une base thématique.

II.2.2.2.4 Représentation des partitions

La dernière étape est ensuite de présenter les résultats obtenus concernant les partitions. Les méthodes visuelles qui peuvent être utilisées sont nombreuses. Certaines sont communes à tous les domaines des **SDD**. Alors que d'autres sont propres aux dialectologues. Nous faisons référence aux représentations prenant en compte la réalité géographique. Nous nous attarderons uniquement sur ce type de représentations et laisserons de côté les autres représentations plus classiques — dendogramme, heatmap, ... etc —.

Une exploration visuelle des données est ainsi réalisée en liant partitions des données et géographie. D'un coup d'œil, il est possible de voir l'extension d'une variété linguistique et d'en tirer des enseignements. Pour parvenir à ce type de cartes linguistiques, souvent les dialectologues ont recours à une segmentation territoriale de leur terrain d'étude. Ou alors aucune structuration de l'espace n'est réalisée et ils se contentent de représenter les partitions en utilisant des symboles qui pointent sur les points d'enquête.

Les méthodes employées pour parvenir à ce parcellement sont multiples. Goebel [Goebel, 2010] a été le principal innovateur dans le domaine en utilisant ce qu'il appelle des beam map — ou dénommé par d'autres network map [Nerbonne and Heeringa, 1997] [Kumagai, 2016] — et des honeycomb map. La segmentation territoriale de ces cartes linguistiques a été obtenue par la méthode Delaunay-Voronoi. Ces deux cartes ont donc la même segmentation territoriale, elles se distinguent seulement au niveau des réalités linguistiques qu'elles décrivent. La honeycomb sert à décrire les grandes lignes de fractures entre les variétés, celles-ci correspondent donc aux frontières linguistiques.

Alors que la beam map met en avant la proximité linguistique entre différents points d'enquête.

Par ailleurs, ces cartes linguistiques ainsi segmentées correspondent à des cartes thématiques. En conséquence, deux types de cartes thématiques peuvent être employées, soit une carte choroplèthe, soit une carte isoplèthe [Pi, 2006]. Pour la première carte, qui est la plus utilisée, la couleur est nécessaire. Un des avantages tiré de cette colorisation, c'est qu'elle peut être graduable. Ainsi, elle peut varier en proportion du degré de similarité entre les variétés. Un seul type de couleurs sera affecté, mais le degré de teinte variera pour marquer la distance ou la proximité linguistique entre deux territoires. La dernière des cartes se contente d'entourer avec des lignes les variétés linguistiques sans se préoccuper nécessairement de leur surface. L'école de dialectométrie de Groningen a mis en place dans ses outils informatiques des cartes isoplèthes — voir Partie II.3.8.2 page 76 —.

II.2.2.3 Géostatistique linguistique

Cette dernière direction de recherche est un champ d'investigation récent en dialectologie. Donc, très peu d'études ont été menées dans ce domaine. Donc, nous ne nous pencherons pas longuement sur ce sujet. Les méthodes de géostatistique sont loin d'avoir été exploitées. Ces travaux ne sont encore qu'au balbutiement. Notons par ailleurs que les études que nous avons pu lire ont essentiellement comme terrain d'étude l'aire germanique.

Les initiateurs de la géostatistique linguistique sont donc des chercheurs d'origines germaniques. Ces terrains d'étude ont des topologies géographiques diverses. Cela peut correspondre à des plaines comme au Pays-Bas ou des espaces montagneux comme en Suisse. Ce dernier aspect peut être exploité pour mener des études plus complexes. Ainsi, le fait que la Suisse soit une région montagneuse, par conséquent, les points d'enquête ne sont pas situés sur un même plan spatial, a du probablement poussé certains dialectologues à explorer cette nouvelle piste. Ainsi, grâce aux techniques des SIG des visualisations en 3D sont possibles [Sibler et al., 2012] et peuvent être appliqués pour étudier la variation linguistique de populations enclavés dans des montagnes.

Cela ne sont que des aspects graphiques, en réalité, l'avantage certain de la géostatistique par rapport à une étude dialectométrique, c'est que le terrain n'est plus qu'une surface de projection pour visualiser des distances linguistiques. Il devient

II.2.2. VARIÉTÉS LINGUISTIQUES ET SCIENCES DES DONNÉES

aussi un des facteurs d'analyse de la variation linguistique [Rumpf and Pickl, 2009]. Les calculs utilisés par ces chercheurs sont divers. Nous recensons des études qui utilisent des méthodes d'Autocorrélation Spatiale [Lee and Krantzschmar, 1993] [Grieve, 2014] [Stoeckle, 2016] [Bart et al., 2013]. D'autres études emploient plutôt des techniques statistiques comme les méthodes d'Interpolation Spatiale [Rumpf and Pickl, 2009] [Wattel and Reenen, 2010] [Sibler et al., 2012].

CHAPITRE II.3

Méthodologie

Cette partie sera consacrée à présenter le processus de collecte de données. Nous y évoquerons de quelle manière nous nous sommes procuré ces données. De quoi est constitué ces données ? Par quel biais ces données ont été collectées ? Auprès de qui nous nous sommes procuré ces données ? Comment avons-nous traité ces données ? Toutes ces questions seront abordées au cours de ce chapitre.

II.3.1 Données et Corpus

Notre intention ici sera de présenter les données que nous avons en notre possession. Ces données comprennent celles qui seront analysées au cours de cette thèse — cela correspond à notre corpus — comme celles qui n'ont pas encore été traitées.

D'abord, il y a lieu de distinguer d'une part le corpus oral du corpus écrit. Car, bien que le berbère de cette région soit une langue orale, il existe, comme nous l'avons signalé, des travaux sur cette région qui constituent tout aussi bien des données. Ces dernières, bien que de seconde main, nous ont servi d'informations d'appoint. Un apport d'intérêt que sont venues compléter nos données, notamment d'un point de vue diachronique. Mais aussi, elles furent essentielles en nous aidant à nous familiariser avec ces variétés avant même de nous rendre sur le terrain. Elles nous ont également aidé à formaliser nos axes de recherche au début de notre thèse. Toutefois, ce corpus écrit reste relativement restreint au vue de notre ambition de décrire l'ensemble des variétés de cette région. Donc, il nous fallait collecter d'autres données, c'est ce que nous avons fait.

Les données ainsi collectées ne sont constituées que d'un corpus oral de première main, c'est-à-dire non traité par d'autres en les formalisant sous forme écrite. La collecte de données peut se faire directement ou indirectement par un procédé de prise de son (voir ci-dessous Partie II.3.3, page 67). Directement, cela signifie qu'entre l'acquisition des données et la retranscription de ces dernières seuls nos sens et notre mémoire sont en jeu. Cette transcription suit instantanément la production des données. Alors que par indirect, nous spécifions par là que la transcription est faite après coup à l'écoute des données enregistrées. Donc, cela ne sous-entend pas nécessairement qu'il faille considérer la prise de son de données orales non obtenues par nous-même comme n'étant pas de première main. Elles le sont tant que nous n'accédons pas à elles sous forme rédigée.

Cette distinction est notable, car nous avons pu collecter sur le terrain ou sur internet des corpus vidéos ou audios comme des émissions de radio où figurent bien évidemment

du berbère. Nous considérons ces données comme une richesse dans la mesure où les données sur cette région se font rares et celles où discourent des personnes encore plus. Et nous les cataloguons comme étant aussi de première main puisqu'elles n'ont jamais été exploitées au niveau académique. Cette première forme de corpus oraux n'est pas la seule que nous ayons en notre possession.

L'autre type de corpus que nous ayons correspond à des unités significatives. Prises pour la plupart d'entre elles en dehors de toutes considérations syntaxiques ou phrastiques. Il serait plus exact de dire que nous n'avons pas inséré systématiquement dans un contexte syntaxique ces unités significatives. Seules certaines unités grammaticales et verbales ont été mises en lien invariablement avec d'autres unités significatives comme les pronoms affixes de personne ou les prépositions. Autre limite à signaler, ces unités significatives collectées sont bien entendues polysémiques, mais nous n'avons pas collecté l'ensemble de cette richesse sémantique. Cela étant dit, nous allons présenter de quoi est constitué ce corpus. Il y a d'abord le corpus issu d'un questionnaire. Ce corpus que nous venons d'évoquer a été collecté dans les différentes localités que nous avons présentées dans la partie introductive de notre Introduction générale (Partie I.1.2, page 11). Puis, ce dernier est enrichi d'un lexique qui est plus ou moins étendu selon les points d'enquête. Igli et Asla restent quand même les localités où nous avons le plus de données comparativement aux autres.

Donc, d'un point de vue méthodique, pour mettre en exergue la variation linguistique de cette région, nous avons eu recours à un questionnaire. Ce dernier a été élaboré bien avant de nous rendre sur le terrain et il se compose d'un lexique en français et en arabe dialectal. Le but étant d'avoir l'équivalent de ces unités significatives dans le berbère de ces oasis. Ce lexique fut inspiré de la liste Swadesh [Swadesh, 1950] et de la liste Chaker [Chaker, 1984, p.216] qui est elle-même dans le principe inspirée de la liste Swadesh. Puis ensuite, elle fut agrémentée de diverses autres unités significatives pan berbères connues comme présentant des variations phonétiques [Kossmann, 1999]. Notre objectif était d'avoir un questionnaire multifonctionnel. Nous comptons avec ce lexique déterminer le degré de parenté et les différences entre les variantes berbères du Sud-Oranais. Cette liste brute au cours de notre enquête a connu des changements pour être améliorée au fur et à mesure de nouvelles variations rencontrées sur le terrain. Concrètement, chaque nouveau fait linguistique rencontré à Asla où dans une autre oasis était ensuite intégré dans ce questionnaire. Puis, nous revenions pour constater ce qu'il en était à Asla où dans les autres oasis.

Ainsi, le questionnaire que nous avons utilisé a été concocté en prenant en compte plusieurs aspects. Nous le voulions représentatif du système linguistique de ces différentes variétés. Par conséquent, nous avons fait en sorte que les principales catégories syntaxiques soient présentes. Cela comprend les verbes, les noms et quelques unités grammaticales. Pour chacune de ces catégories, nous avons pris soin qu'elles englobent des champs sémantiques divers. Pour cela, nous avons pris en compte la *realia* propre à cette région. Ainsi, c'est une région traditionnellement orientée vers l'agriculture maraichère et la culture des palmiers. Leur mode de vie est sédentaire et urbain. Donc, nous avons essayé de coller à cette réalité. De plus, nous avons choisi nos items parmi le vocabulaire fondamental. Ce choix s'explique par le fait que généralement ce type de vocabulaires est le moins substitué par des emprunts. Cela permet d'assurer une comparaison fiable entre les variétés. Nous pourrions nous étendre davantage sur l'élaboration de ce questionnaire comme la prise en compte de la transivité des verbes ou des différentes catégories de morphologie verbale. Mais étant donné que ce questionnaire ne sera pas dans sa totalité exploité et que nous avons seulement l'intention de traiter la dimension phonético-phonologique, alors nous préférons ne pas développer plus longuement ce point. Nous rajouterons dans ce sens que la somme de nos items est censée contenir l'ensemble des unités phoniques qui puisse être trouvé dans le système phonético-phonologique de ces variétés.

Nous ne présenterons pas l'ensemble du corpus que nous avons collecté. D'une part, une grande partie de celui-ci n'a pas encore été traitée. Seules les données du questionnaire ont fait l'objet d'un traitement, mais là encore nous présenterons seulement les données en rapport avec les axes de recherche inclus dans cette partie. La totalité du questionnaire ne sera pas exploitée dans le cadre de cette thèse. Nous figurerons uniquement les données qui ont fait l'objet d'une analyse. Ces dernières ont été placées à la fin de la Partie I (Appendice, page 178). Signalons par ailleurs que ces données collectées ont servi à l'établissement du corpus qui a été mis en place pour l'étude des rhotiques alvéolaires qui sera abordée dans la deuxième partie de cette thèse (Partie II, page 183).

II.3.2 Population d'étude

Nous nous sommes rendu dans 16 points d'enquête, dans chacun d'eux nous avons pu rencontrer au moins un informateur. Parfois dans certaines oasis, nous n'avons

pas pu échanger avec d'autres personnes. Il n'y a vraiment que dans les localités de Bousseghoun, Asla, Sfissifa (Naâma), Igli et Berrebi où nous avons eu plusieurs informateurs.

Nous ne pouvons affirmer que le choix de nos locuteurs est basé sur un échantillonnage prenant en compte les dimensions sociologiques. Nos informateurs sont principalement des hommes dépassant la quarantaine, nous avons que très peu d'informatrices féminines. Ces dames que nous avons rencontrées, elles aussi, étaient principalement très âgées. Les femmes dans cette région sont mises à l'écart des hommes n'étant pas des membres de leur famille. La mixité dans l'espace public comme privé est évitée pour des raisons religieuses. Si dans certaines oasis, nous n'avons pas pu avoir un nombre plus important d'informateurs, c'est tout bonnement parce que le berbère dans ces oasis est en danger, donc le nombre d'informateurs y est très réduit.

De plus, l'accès difficile à certaines localités a rendu aussi compliqué la possibilité de rencontrer d'autres informateurs. Il y a également le paramètre temps et étendu géographique qui a nous compliqué la tâche. Car comme nous souhaitons nous rendre dans toutes ces oasis, donc nous ne pouvions nous éterniser dans l'une d'elle indéfiniment dans l'espoir de rencontrer un autre informateur.

Le choix de ces informateurs s'est ainsi imposé à nous dans la plupart des cas. Étant donné que nous avons à notre disposition une disponibilité limitée en matière d'informateurs. Nous n'avons pas de critères très stricts visant à avoir une représentativité exhaustive en termes de classes sociales. Nous avons comme seul critère d'aller à la rencontre de personnes maîtrisant le berbère, si possible matures.

Ce critère de l'âge n'est pas en soi inexorable. Un critère que nous ne pouvons passer outre. Bien loin de nous l'idée d'une telle intransigeance. Nous avons aussi parmi nos informateurs de jeunes personnes. En réalité, ce choix pour des personnes âgées est motivé par trois raisons. La première, à l'inverse de ce qui se passe chez les jeunes, le dialecte des plus vieilles personnes est moins sujet aux influences des autres variétés linguistiques et le recours aux innovations linguistiques y est très réduit, donc nous avons à faire à un système linguistique assez stable. L'autre raison, comme les évolutions ne se font pas seulement par des innovations, mais aussi par l'occultation de faits linguistiques. Alors même si ces individus âgés ne font plus usage dans leur quotidien de ces faits linguistiques oubliés, appartenant à un système linguistique qui les avaient précédés, il en reste toujours quelques choses dans leurs esprits et peuvent en faire mention. Ce genre de données est capital pour l'histoire du dialecte, surtout dans le cas

de ces variétés orales sur lesquelles nous n'avons aucune donnée historique. La dernière raison est tout bêtement que dans certaines oasis, il n'y a plus de jeunes qui maîtrisent le berbère, donc la question ne se pose plus.

II.3.3 Matériels de collecte de données

La collecte des données a été réalisée par le biais de différents supports. Dans un premier temps, pour recueillir notre corpus, nous avons opté pour la transcription écrite sur des calepins. Au premier contact avec les informateurs, c'était cet outil que nous utilisions pour noter ce que nous répondaient les informateurs. Si nous avons fait usage d'un carnet, le but était seulement de ne pas effrayer les informateurs. L'utilisation au premier abord d'un matériel informatique pouvait risquer de les rebuter. Puis, lorsque la confiance était établie, nous faisons usage d'appareil numérique pour enregistrer l'échange. L'enregistrement s'est fait avec un dictaphone numérique — marque : Olympus WS-813 — et/ou un appareil photo-vidéo numérique — marque : Canon EOS —.

Mais la plupart du temps, nous avons eu recours à ces deux moyens pour collecter nos données. L'intérêt de continuer à faire usage de la transcription manuelle est que cela nous sert d'exercices. Une manière d'éveiller nos sens et d'affiner notre oreille aux particularités phonétiques de ces variétés. Lors de l'audition de nos enregistrements cela nous permet de confronter ces données qui ont été prises différemment. Et ensuite s'il existe un écart, nous avons un retour et nous pouvons nous rendre compte où nous avons péché. Ce passage par l'écrit puis par une écoute à posteriori des mêmes données est capitale pour la validité des données acquises (voir ci-dessous), mais aussi pour avoir une maîtrise fine de ces variétés qui est aussi l'un de nos objectifs.

Si nous n'aboutissons pas à cette maîtrise, comment alors retranscrire des heures d'enregistrement si nous n'arrivons à décrire qu'un quart des propos tenus par nos informateurs. Car, de retour du terrain, nous sommes seul face à ces données et il est impossible de demander à ces informateurs de traiter pour nous ces données. Donc, pour ne pas que celles-ci soient gâchées et restent inexploitées, nous devons passer par cette étape d'apprentissage.

II.3.4 Conditions de collecte de données

Nos séances de travail n'ont été menées que très rarement dans des locaux. Le plus clair du temps, celles-ci ont été réalisées à l'extérieur. Nous n'avons aucun contrôle possible sur le choix du lieu. Nous enquêtons là où nos informateurs étaient présents. Cela pouvait être assis dans une rue à même le trottoir. Dans la palmeraie. Dans un véhicule de transport. Ce genre de conditions a des implications quant à la qualité de l'environnement sonore. Il va de soi qu'ainsi nous travaillons dans un environnement bruyant. Cette nuisance a pour conséquence d'altérer la réception des réponses émises par nos informateurs. Cela implique une plus grande vigilance de notre part quant à la validité de nos données.



Photographie II.3.1 : Séance de travail avec un informateur à Moghrar n wadday

II.3.5 Le protocole

La collecte de ces données s'est faite de plusieurs manières. Notre corpus oral est constitué d'un corpus élicité et d'un corpus spontané. Tous deux ont été exploités pour obtenir nos résultats. Le corpus spontané est constitué de différents types de structures discursives comme des discours, des dialogues ou des récits. Celui-ci a été obtenu soit en ayant recours à des entretiens non directifs, c'est-à-dire qu'aucun thème n'est imposé à l'informateur, il s'exprime de son propre arbitre sur le sujet que lui même a décidé d'aborder. Soit par un entretien directif, cette fois-ci nous demandons à l'informateur de réagir à une thématique bien particulière. Ou bien encore, notre corpus ne s'inscrit pas dans le cadre d'un entretien. Cette fois, nos informateurs s'adressent mutuellement la parole, nous sommes seulement l'observateur passif de cet échange que nous collectons.

L'autre corpus, qui est élicité et d'où sont tirées les principales données qui seront présentées par la suite, a été obtenu de deux façons différentes. Une partie de ce corpus est élicitée sans guide, c'est-à-dire sans qu'il ait fait l'objet de notre part d'une préparation réfléchie en amont. Dans le vif de l'entretien, nous saisissons l'occasion pour interroger notre informateur sur comment un élément se dirait dans sa langue. Cela est fait donc de manière improvisée. Puis, il y a également l'élicitation guidée qui lui a été élaborée avant même de nous rendre sur le terrain. Ce dernier est la colonne vertébrale de notre étude dialectologique. En cela que les mêmes questions, figurant dans ce questionnaire, ont été posées dans tous nos points d'enquête. Ce qui garantit une comparaison robuste entre les variétés de notre terrain d'enquête.

Enfin pour faire réagir nos informateurs afin qu'ils nous répondent, il faut les stimuler. Par conséquent, nous avons eu recours à différents types de stimuli. Nous commencerons par évoquer ceux que nous avons utilisés de manière inconstante pour des cas précis. Parmi ces stimuli en question, il y a les stimuli visuels basés sur la monstration — geste corporel — de l'élément à traduire ou sur des images. Nous nous sommes beaucoup appuyé sur ce type de stimuli par exemple pour tous ce qui concerne les parties du corps. D'ailleurs, nous commençons toujours notre questionnaire en interrogeant nos informateurs sur ce thème. Cela permet d'évaluer les compétences linguistiques de l'informateur et le cas échéant de déterminer s'il est nécessaire ou non de continuer l'entretien. Ou tout du moins d'avoir à l'esprit qu'il faut prendre avec des pincettes les futures réponses de nos informateurs. Dans le cas bien évidemment où nous nous rendrions compte de lacunes au niveau de la maîtrise de la langue.

Puis, il y a aussi les stimuli écrits. Ce genre de stimuli correspond à une présentation sous forme rédigée — manuscrite ou numérique — de l’item qui nous intéresse. Cette situation n’a eu lieu qu’exceptionnellement. Généralement, lors d’une élicitation non guidée. La langue utilisée fut l’arabe classique la plupart du temps. Par ailleurs, nous sommes aussi passé par ce biais lors de communications électroniques. Car nous avons continué à échanger à distance avec certains de nos informateurs.

Le dernier type de stimuli utilisé correspond aux stimuli oraux. L’oral a été notre principal vecteur de communication lors de nos entretiens. Les langues dont nous nous sommes servis furent le français, l’arabe dialectal et le berbère. Toutes les trois ont été utilisées, l’une plus que l’autre suivant les compétences linguistiques de nos informateurs. Donc, les combinaisons employées sont : d’une langue tierce — arabe ou français — vers le berbère, du berbère vers une langue tierce ou du berbère vers le berbère. Ces configurations dépendent du processus utilisés pour soutirer des données à nos informateurs. Ces processus sont dénommés par nous même *processus de performance* et *processus de connaissance*. Le premier consiste à demander à l’informateur : “comment dis-tu pain en berbère?” Quant au second, la question sera : “connais-tu le mot *ayrum*?” ou “que veut dire le mot *ayrum*?”.

Le choix de l’un ou de l’autre de ces processus dépend de ce que nous souhaitons obtenir. D’abord, il faut dire que le processus de performance est le processus par défaut, celui que nous utilisons d’emblée lors de nos entretiens. Toutefois, il arrive régulièrement que lors de cette phase la réponse obtenue ne corresponde pas à nos attentes. Par exemple, notre informateur nous donnera un emprunt à l’arabe. Donc, nous souhaitons nous assurer s’il existe un synonyme dans leur variété pour établir des cognats¹. Autre cas de figure où nous utilisons le processus de connaissance, c’est pour nous assurer de la validité de la réponse fournie par notre informateur. Cela s’inscrit dans une démarche de validation scientifique de nos données et ce moyen est l’un de ceux que nous utilisons pour vérifier chacune de nos données (voir ci-dessous Partie II.3.6, page 73).

La quasi totalité des stimuli a été produite par nous même. À de très rare cas, une tierce personne est intervenue pour expliquer à notre informateur la question qui lui était posée. Car notre maîtrise de l’arabe dialectal, notamment celle du sud-ouest algérien, est rudimentaire. Mais grâce à la bienveillance de nos informateurs, ils ont pu supporter nos

1. L’établissement de ces cognats est essentiel pour déterminer des cognats phonologiques et *in fine* des haplolectes (voir Partie II.4.1.2, page 92).

lacunes dans cette langue.

Ces exemples nous amènent à évoquer le choix de la formulation de nos stimuli. Nous avons eu recours encore une fois à différentes formulations. Ces formulations dépendent inévitablement de la langue de communication. Mais de façon générale, elles sont structurées de manière identique. Ces formulations sont fonction de la catégorie syntaxique de nos items. Nous pouvons diviser ces catégories en trois grandes parties : les noms, les verbes et les unités grammaticales. La première de ces catégories est formulée comme dans l'exemple précédent. C'est une question ouverte et directe où pour chaque concept nominal nous demandons le singulier et le pluriel². Au fur et à mesure de l'énumération des items, ces questions deviennent moins formelles et il suffit d'évoquer l'item pour que l'informateur sache ce que nous attendons de lui.

Pour les verbes, les choses se déroulent différemment. Chaque concept verbal est déployé aux trois aspects positifs du berbère : **aoriste (A)**, **prétérit (PRET)** et **aoriste intensif (AI)**. Pour ce faire, pour chacun de ces aspects nous conjugons le verbe. Cette conjugaison comprend un indice de personne ou plusieurs — morphème obligatoire en berbère — et un adverbe temporel — nous incluons la particule d'orientation sous cette étiquette —. De manière concrète cela donne : “demain il” prononcé par nous même. Et nous demandons à l'informateur de compléter et lui dira : “demain il boira”. Parfois, nous contextualisons avec un énoncé si l'informateur a du mal à trouver l'aspect désiré. Cela arrive souvent pour l'**AI**, car ce thème peut exprimer soit l'aspect progressif ou l'aspect itératif. Cette méthode aide l'informateur à se rappeler des formes correspondant à ces aspects. Il faut avoir à l'esprit que le berbère de cette région contrairement à l'arabe n'est pas une langue qui est enseignée. Et donc nos informateurs ne connaissent pas les rudiments de grammaire de leur langue, ils n'ont pas de “ tiroirs verbaux ” tout fait dans leur mémoire. Nous ne pouvons leur demander directement de nous donner par exemple l'aoriste du concept verbal manger. Même en utilisant les équivalents de ces aspects en français — présent, passé, futur — ou en arabe — **المُسْتَقْبَلُ، المَاضِي، المَضَارِعُ** — ils ne sont pas en mesure de donner convenablement les aspects pour chaque verbe. Car, en berbère bien que les formes verbales soient aspectuelles, il existe tout de même des temps qui se construisent avec des auxiliaires temporels. Donc, lorsque nous demandons le passé d'un verbe très souvent la réponse correspond à ce genre de construction. De ce fait pouvoir guider l'informateur est nécessaire pour s'assurer d'obtenir les formes aspectuelles que nous souhaitons.

2. Chacun d'eux constitue un item et une question.

Enfin pour la dernière catégorie, les unités grammaticales, nous avons opté pour une formulation avec une question ouverte. Cette fois-ci, la question ne s'ouvre pas sur une unité significative, mais sur une unité phrastique. À l'intérieur de cette phrase se trouve notre unité grammaticale. Pour une même classe d'unités grammaticales, nous reprendrons la même structure phrastique. Par exemple pour les prépositions, notre question sera : "comment dis-tu il est X sur la table en berbère?" où X correspond à une préposition — sur, sous, derrière, ... —. D'autre part, ces unités grammaticales ont souvent des allomorphes, donc nous avons fait en sorte de changer leur distribution ou bien de changer les unités qui leur étaient adjacentes pour obtenir les autres formes.

Donc, comme vous l'aurez compris, le questionnaire était parfaitement segmenté entre ces catégories et était évoqué dans l'ordre de présentation que nous venons de dérouler. Cette manière de procéder s'est élaborée progressivement, l'expérience acquise au cours de nos entretiens avec nos informateurs nous a permis d'affiner nos séances de travail. C'est ainsi que nous sommes arrivés bon an, mal an à cette méthode où nous trouvons que nous obtenons de bons résultats. Toutefois, certains biais peuvent être introduits du fait d'utiliser cette méthode. La répétitivité d'une morphologie — par exemple le pluriel externe — dans des items peut pousser l'informateur à reproduire cette même morphologie sur d'autres items quand il a un doute. Parfois, pour contrecarrer cet effet de contamination, il n'est pas inutile de reprendre l'informateur et de lui demander si la réponse n'est pas plutôt celle-ci, cela formulé sous forme de questions alternatives : dites-vous X ou Y?. Cette suggestion ne fausse en rien les données, puisque s'il réagit positivement à notre proposition, nous lui faisons que lui rappeler ce qu'il avait omis. Cela n'est ni plus ni moins qu'une défaillance de la mémoire de la part de l'informateur. Néanmoins, s'il considère sa première réponse comme étant celle qui convient, alors nous avons tout le loisir de confirmer ou d'infirmer sa réponse avec d'autres informateurs.

Nous pourrions choisir pour neutraliser ce biais, de présenter nos items dans un ordre aléatoire — ce que nous faisons, mais à l'intérieur de chacune de nos catégories — cependant, nous perdrons en efficacité en terme de temps de réaction de nos informateurs. Si nous devions adopter un ordre aléatoire cela aurait pour conséquence d'allonger les séances de travail. Car, pour répondre à ce questionnaire, il faut bien entre 3h et 4h — sans compter les digressions — en suivant notre méthode. Parfois, avec un même informateur cela peut être réalisé d'un tenant ou alors cela peut se répartir sur plusieurs jours. Ce temps qui peut paraître relativement court s'avère en réalité dur à

tenir. Car il n'est pas évident de trouver des informateurs qui peuvent nous consacrer autant de temps et qui n'éprouve pas de lassitude à nous répondre. Donc, la variable temps est importante, nous ne pouvons nous permettre d'allonger nos séances de travail. C'est en raison de cela qu'un compromis est effectué entre efficacité et fiabilité. Nous rajouterons qu'un questionnaire doit pouvoir être vivant et dynamique. Cela passe par conséquent par un enchaînement des items qui se fait de manière rythmée. Cette organisation en catégorie permet de conserver cette rythmique.

II.3.6 Validation des données

Comme nous venons de l'évoquer, les données obtenues font systématiquement l'objet d'une démarche de validation scientifique. Nous avons recours pour cela à plusieurs moyens. L'emploi de l'un ou de l'autre de ces moyens est fonction des circonstances. Si les conditions ne nous permettent pas d'employer l'un de ces moyens, alors nous tentons par ces autres moyens. Notre objectif est pour chaque donnée de mener une double vérification. La double vérification peut se faire à plusieurs niveaux. Il y a d'abord la double vérification au niveau du récepteur, c'est-à-dire nous même. Vérifier que la réception des données ait bien été réalisée. Pour cela la double vérification se déroule par l'écoute des enregistrements à plusieurs reprises. Puis, il y a aussi la vérification au niveau de l'émetteur. Notre exigence est d'avoir au minimum une deuxième source pour chaque item. Il arrive que la validité des données ne peut pas être réalisé en les recoupant auprès de deux locuteurs. Comme à Fendi où nous n'avons qu'un informateur. Mais nous considérons que si les données fournies par ce seul informateur sont corroborées par celles des localités avoisinantes, alors nous estimons qu'elles sont valides.

II.3.7 Métadonnées

Ces métadonnées concernent essentiellement des informations sur nos informateurs. Disons de prime abord que cette entreprise n'a pas été menée avec le plus grand soin. Nous n'avons pas fait preuve de rigueur dans la collecte de ces métadonnées. Cela s'explique par différentes raisons. Il y a d'abord les informateurs, que nous qualifierons de secondaire, avec lesquels nos entretiens ont été brefs et qui ne se sont pas renouvelés.

Pour ce genre d'informateurs à cause du peu de temps passé avec eux, des circonstances des rencontres, des usages de politesse, tout cela ne nous a pas poussé à les questionner sur leur identité.

Parce que, nous considérons qu'un entretien ne doit pas se dérouler de manière abrupte. Car à aucun moment nous ne rémunérons ces informateurs. Nous les approchons toujours dans le respect de leur personne, cela ne doit pas leur faire penser à un interrogatoire. Donc, pour cette catégorie d'informateurs nous n'avons aucune métadonnée. Tout du moins aucune d'ordre sociologique, la seule métadonnée que nous avons est qu'ils sont originaires de l'un des points d'enquête. Donc, nous savons que les données obtenues auprès d'eux proviennent du lieu en question. La provenance de ces données est préservée.

S'agissant de nos informateurs principaux, nous avons un certain nombre de métadonnées. La première des choses est la connaissance de leur état civil. Nous avons noté leur nom et prénom. Toutefois, nous avons aucune certitude quant à l'orthographe de ces derniers. Parfois, au lieu de leur prénom, nous avons le surnom ou un diminutif — qada pour abdelqader —. Puis, nous nous sommes fixé comme objectif de savoir si dans leur entourage familial — époux/épouse, enfants — la langue de communication était le berbère ou l'arabe dialectal. Si leur épouse ou époux est d'origine berbère ou arabe. Ainsi que leur ascendance, d'où sont originaire leurs parents. Puis, nous avons cherché à connaître leur situation socioprofessionnelle et leur niveau d'étude. D'autres métadonnées comme le moment et le lieu exact de la collecte n'ont pas été retenues. Car elles ne sont pas des métadonnées pertinentes pour les axes de recherche que nous nous sommes fixé.

L'ensemble de ces informations est reporté dans des fiches signalétiques. Cependant, du fait que cette thèse pourra être disponible auprès du grand public, alors nous préférons de pas faire figurer ces documents dans ce travail académique. Toutefois, nous rendons disponibles ces fiches signalétiques à la communauté des chercheurs. Pour quiconque envisagerait de travailler sur le berbère du Sud-Oranais.

Tout ceci implique également que nos données numériques ne sont pas décrites par des métadonnées. Nous n'avons pour l'instant mené aucune action pour caractériser ces données même d'un point de vue de l'étiquetage³. Donc, les normes de description pour métadonnées comme le ISLE Meta Data Initiative (IMDI)⁴, qui servent pour l'archivage de

3. À l'exception de notre corpus de phonétique (voir Partie III.3.6.1.1, (page 232)).

4. Pour plus d'informations, ci-contre un lien redirigeant vers de la documentation sur cette norme :

données, n'ont pas été appliquées. L'axe central de notre recherche n'était pas de réaliser de la Documentation de langue — en anglais Language documentation —. Donc, pour des questions de cohérence et de faisabilité nous n'avons pas entrepris ce travail. Nous avons remis à plus tard ce projet, puisque nous avons à faire à des variétés en danger, donc il y a une nécessité d'archiver ces données.

II.3.8 Traitement des données

Les données de nos jours ne peuvent plus être traitées manuellement comme cela était le cas à l'époque d'André Basset. À notre époque, le traitement de ces données passe inévitablement par leur numérisation. Les outils informatiques sont indispensables. Nous allons ainsi présenter quels furent les principaux outils informatiques utilisés pour traiter nos données.

II.3.8.1 Tableur

Le premier outil informatique que nous avons utilisé est le tableur. Comme sa dénomination l'indique, son interface graphique fait rappeler un tableau à grille avec des colonnes et des lignes. Le principal intérêt de ce tableur est qu'il peut servir de base de données. Nous pourrions utiliser d'autres outils informatiques pour stocker nos données, cependant, le tableur est simple d'utilisation et répond à nos besoins qui sont relativement basiques. Notre principal souci est d'avoir une interface graphique qui rende lisible les données du premier coup et qui garantisse une intégrité des données, en plus d'être accessible sur d'autres supports informatiques.

Donc, nos données manuscrites ou audios sont toutes retranscrites à l'intérieur de ce tableau. Le fait de stocker numériquement ces données permet de préserver l'intégrité dans le temps des données. Puis, surtout, du fait que les données sont structurées, elles peuvent être plus facilement manipulables par la suite pour réaliser d'autres choses comme nous le verrons ci-dessous. Et cela à contrario d'un stockage des données qui se ferait dans un traitement de texte.

La structuration de ces données s'est faite de plusieurs façons tout dépendait du traitement réalisé par la suite. Donc, nous avons différents tableaux présentant les mêmes données. Par exemple, le tableau utilisé pour réaliser des cartes géolinguistiques

tla.mpi.nl/imdi-metadata (Consulté le 05/08/2017).

est un tableau à double entrée qui est disposé de la manière suivante : en ligne nous avons les points d'enquête et en colonne nos données linguistiques.

Tous les tableurs de nos jours peuvent mener différentes tâches comme des graphiques ou des calculs statistiques. Mais la puissance de traitement des données reste tout de même limitée. D'autres outils informatiques permettent de réaliser ce travail, c'est ce que nous allons voir de ce pas.

II.3.8.2 Rstudio

Nous avons eu recours à ce logiciel pour obtenir un certain nombre de représentations graphiques. Ce logiciel qui est libre et open source est destiné au calcul statistique et permet de visualiser les données sous la forme de graphiques. Les calculs et les visualisations des données sont réalisés grâce au langage R. Rstudio n'est qu'une interface graphique — il en existe d'autres —, le langage R est celui qui permet de réaliser les analyses de données. Donc, cela implique que pour manipuler les données, il faut faire usage des lignes de commande. Cependant, des packages existent pour faciliter la manipulation de ces données à partir de l'interface en ligne de commande.

Effectivement, il existe tout un ensemble de packages mis à la disponibilité du grand public qui aident à traiter les données. Ces packages intègrent des fonctions informatiques qui réalisent les opérations nécessaires jusqu'à l'exécution de la tâche. L'utilisateur ne manipule pas directement ces fonctions, il lui suffit d'écrire quelques lignes de codes comportant des arguments et des paramètres qui appellent ces fonctions et le tout est exécuté automatiquement. Donc, ce langage est facile d'utilisation, même si les utilisateurs n'ont pas une connaissance approfondie en programmation.

Certains se demanderont, mais pourquoi passer par de la programmation si cela peut être fait avec un logiciel en manipulant des icônes. La réponse est tout simplement l'automatisation et la polyvalence. En rédigeant nos lignes de codes dans des scripts, nous pouvons reproduire les tâches désirées sans avoir à reprendre tout depuis le début pour créer nos représentations graphiques par exemple. Quasi instantanément, nous avons nos calculs ou nos illustrations. Cela a son importance, si nous avons à produire plusieurs graphiques comme c'est notre cas. Ensuite, par le biais de ce seul environnement, nous pouvons réaliser différentes choses allant du calcul statistique au graphique et bien d'autres choses ce qui est un gain de temps. Nous n'avons à nous familiariser qu'avec un seul outil. Et Rstudio a l'avantage de permettre cette polyvalence.

Ces dernières années ont vu l'apparition de logiciels spécialisés^{5 6 7} conçues pour la dialectologie qui permettent de faire de la géolinguistique et de la dialectométrie. Certains sont même totalement externalisés et sont accessibles sur internet^{8 9} sans installation de logiciels. Cependant, ce qu'ils permettent de réaliser reste limité. Ils n'offrent pas toute la puissance et la souplesse que peut proposer des logiciels d'analyse et de visualisation de données comme Rstudio ou Matlab. D'autant plus que la durée de vie de ces logiciels n'est pas garantie ni la question de la propriété intellectuelle des graphiques produits par ces applications. Si la mise à jour de ces programmes ne se fait plus et si le propriétaire n'est plus en activité, ce qui se produit fatalement, alors tout le travail mené avec ces logiciels se retrouve perdu.

Donc, toutes nos cartes géolinguistiques seront produites par le biais du langage R et de l'interface Rstudio. Il en sera de même pour les calculs mathématiques qui seront utilisés. Cette approche n'est pas nouvelle, mais elle n'est pas non plus très courante. L'utilisation de cet outil en **SDL** commence seulement à émerger. Nous avons trouvé quelques rares études qui font usage du langage R pour analyser et représenter géographiquement leurs données [Asnaghi and Geeraerts, 2016] [Grieve, 2016]. Ces deux précédents chercheurs considèrent également que leur recherche constitue une partie intégrante des **SDD**.

II.3.8.2.1 Carte géolinguistique

Après avoir importé nos données dans Rstudio, nous pouvons les traiter pour réaliser des cartes géolinguistiques. Pour ce faire, les données linguistiques ne sont pas suffisantes. Nous avons besoin également de données géographiques comme la géolocalisation de nos points d'enquête. Pour les géolocaliser, il nous faut leurs coordonnées géographiques qui sont la longitude et la latitude. Nous avons utilisé le site internet Google Maps¹⁰ pour obtenir ces coordonnées. Pour cela, il faut naviguer sur la carte jusqu'au lieu qui nous intéresse, puis de zoomer au maximum à un endroit de la localité, si possible le centre de la ville. Puis, dans la barre URL se trouve les

5. Le logiciel GeoLing : www.geoling.net (Consulté le 05/09/2017).

6. Le logiciel RG/L04 — basé sur le langage R — : www.let.rug.nl/kleiweg/L04/ (Consulté le 05/09/2017).

7. Le logiciel Visual DialectoMetry — Il n'y a pas de téléchargement en libre accès — : www.dialectometry.com (Consulté le 05/09/2017).

8. Le site internet : www.gabmap.nl (Consulté le 05/09/2017).

9. Le site internet : <http://eudia.ehu.es/diatech/index> (Consulté le 05/09/2017).

10. Le site www.maps.google.com (Consulté le 05/04/2015).

coordonnées géographiques du lieu en question qu'il suffit de copier et de coller dans notre base de données. Cependant, il est possible d'avoir l'ensemble de ces coordonnées en les récupérant dans des bases de données déjà établies. Ces bases de données sont dénommées Index géographique ou Gazetteer¹¹. Mais du fait que notre jeu de points d'enquête est réduit, nous avons préféré réaliser cette étape manuellement pour nous assurer que le référencement se fasse correctement. D'autre part, mentionnons que ces coordonnées géographiques correspondent au standard de géoréférencement dénommé World Geodetic System de 1984 (WGS84) et plus particulièrement au standard de l'European Petroleum Survey Group 3857 (EPSG :3857) qui sont ceux utilisés par Google Maps¹².

II.3.8.2.1.1 Fonds de carte et Ressources cartographiques

Cela étant fait, il faut par la suite se procurer des fonds de carte qui doivent être gratuite et open source. Pour cela, il existe des bases de données en ligne qui proposent de tels documents. Personnellement, nous avons utilisé le DIVA-GIS¹³. L'avantage de ce site internet, qui propose aussi un logiciel de SIG pour l'écogéographie et la biogéographie, est qu'il regroupe différentes données conçues pour les SIG issues de différentes autres bases de données. Par ailleurs, ces ressources cartographiques peuvent être choisies pays par pays sans avoir à télécharger l'ensemble des données de toutes les contrées du monde. De plus, ces cartes sont numériques et il en existe de différentes sortes. Celles qui nous intéressent correspondent aux cartes délimitant les frontières administratives et les cartes représentant la topologie géographique. Ces cartes sont aux formats Shapefiles et Grid. Ces deux derniers se distinguent par la technique utilisée pour modéliser ces cartes. Le premier est une image vectorielle et le second est une image pixélisée.

D'autre part, concernant les frontières administratives, différents découpages administratifs sont disponibles. Cela va de la commune à la wilaya, cependant, nous avons décidé de ne représenter que les frontières des wilayas pour ne pas surcharger nos cartes. Pour plus de détails sur le découpage administratif, nous renvoyons à l'Introduction générale (Partie I.1.2, page 11) où nous avons abordé cette question. À titre d'information, ces cartes proviennent à l'origine de la base de données : Global

11. Le site du National Geospatial-Intelligence Agency propose de telles données : <http://geonames.nga.mil/gns/html/> (Consulté le 05/04/2015).

12. Pour plus d'informations voir ce site de base de données sur l'EPSG : <https://epsg.io/3857> (Consulté le 05/04/2015).

13. Le site www.diva-gis.org (Consulté le 05/04/2015).

Administrative Areas (GADM).

S'agissant de la topologie géographique, elles ne seront pas utilisées dans le cadre de notre étude géolinguistique, mais plutôt dans l'autre partie de notre étude, en dialectométrie. Toutefois, nous mentionnons ici cet aspect pour une question d'ergonomie rédactionnelle. Parmi les données sur la topologie géographique, nous avons retenu les cours d'eau et l'altitude terrestre. Les cours d'eau présentés sont en réalité des lits d'eau asséchés pour la majorité d'entre eux. En outre, nous ne comptons pas tous les représenter, seuls les principaux seront illustrés. Les données proviennent de la base de données établie par le Digital Chart of the World (DCW).

Les autres données de topologie géographique dont nous nous sommes servies sont l'altitude terrestre. Plus précisément, c'est un modèle numérique 3D de la surface du globe terrestre. Ce genre d'illustration est plus communément appelé Modèle Numérique de Terrain ou Altitude du Sol. La particularité de ce type de données est qu'elle met en exergue les reliefs et donc les montagnes. Ces dernières données altimétriques correspondent à la forme topographique qui nous intéresse pour étude. Il va de soi que nous ne présenterons pas toutes les montagnes, hormis les chaînes de montagnes principales. Nous avons utilisé un éclairage fictif pour obtenir le rendu 3D dont voici les valeurs : angle d'élévation de 4° et azimuth de 277°. Ces dernières valeurs sont cependant de peu d'importance, puisque nous avons le fait choix d'appliquer un rendu couleur à ces données altimétriques. Par ailleurs, nous attirons l'attention sur le fait que la résolution des images fournies est de 30", ce qui revient à une précision de l'ordre de 926 m par pixel. D'autres résolutions plus précises existent, mais pour les besoins de notre recherche cette qualité de précision est suffisante. Enfin, soulignons que toutes ces données ont été obtenues par le Consultative Group for International Agricultural Research - Consortium for Spatial Information (CGIAR-CSI) auprès de la NASA Shuttle Radar Topographic Mission (SRTM).

II.3.8.2.1.2 Modèle graphique de nos cartes géolinguistiques

Dans cette section notre objectif est d'expliquer comment ces cartes sont organisées. Toutes nos cartes géolinguistiques sont centrées sur nos points d'enquête. Donc, cela implique que les territoires des différentes wilayas, étant en dehors de l'espace des points d'enquête, seront exclus des cartes géolinguistiques. D'autre part, nous avons fait figurer la ville de Béchar comme point d'ancrage visuel pour repérer les différents points d'enquête. Cet endroit sera représenté sur la carte par un symbole de couleur noire : +.

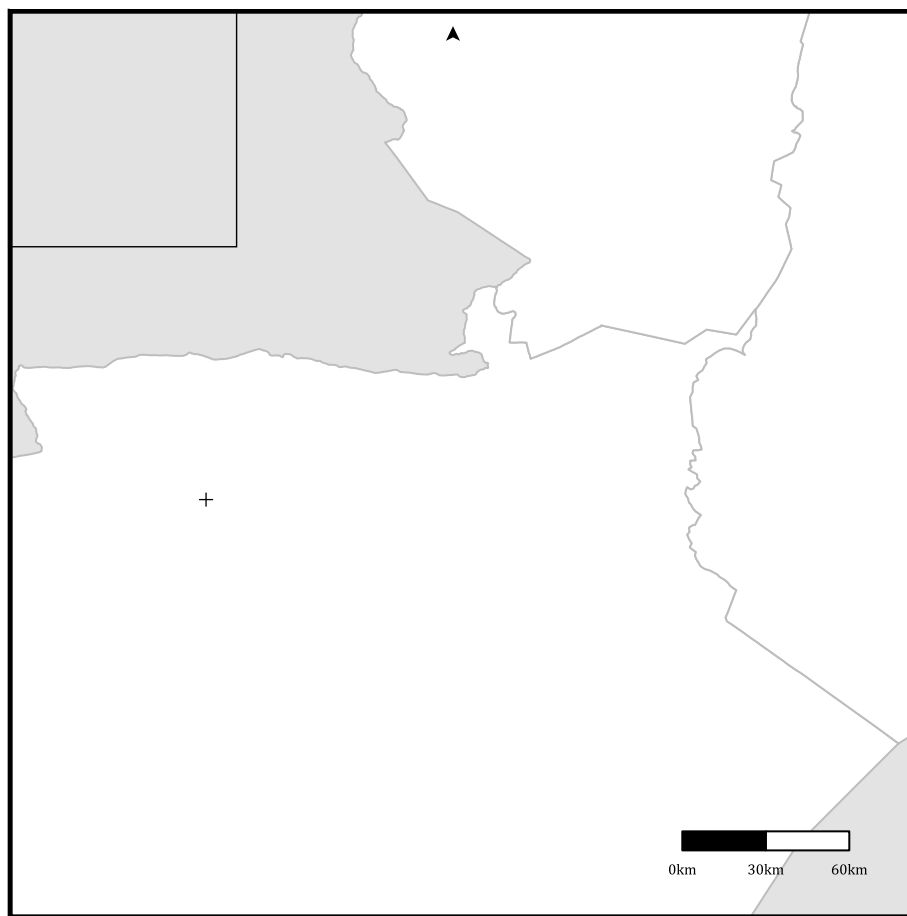
Le choix de cette ville s'explique par le fait que c'est la plus grande ville de cette région.

Nous n'indiquerons pas sur ces cartes les noms des différentes localités pour ne pas alourdir le rendu visuel. Cela dans le but que le lecteur se concentre sur les résultats sans qu'il soit surchargé par un foisonnement de métadonnées. Nous renvoyons à la carte II.1.1.2 (page 33) qui a été présentée dans l'Introduction si besoin est.

Sur chaque carte géolinguistique figure une rose des vents qui indique le nord — au centre et en haut de la carte —. Et une échelle kilométrique pour fournir un ordre de grandeur des distances qui séparent nos points d'enquêtes entre eux — à droite et en bas de la carte —. Quant à l'échelle de visualisation, elle n'y est pas présente. Mais toutes nos cartes à l'origine sont à une échelle d'environ 1:300 million. Toutefois, cette dernière dimension n'est plus respectée du fait du redimensionnement de nos cartes. Cette opération s'est révélée nécessaire pour que les cartes puissent respecter les dimensions de nos pages.

Puis, nous avons aussi rajouté une légende où sont indiquées les données étudiées — à gauche et en haut de la carte —. En titre figure la variable étudiée et l'item qui sert à exemplifier cette variable. Sous l'intitulé de la légende est indiqué les formes que prennent la variable dans les différents points d'enquête. À chacune de ces formes correspond un symbole. Chacun de ces symboles a une forme particulière — cercle, carré, losange, triangle — et une couleur particulière — respectant si possible les oppositions chromatiques — qui le distingue d'un autre symbole s'il y en a plusieurs. Les points d'enquête sont aussi aguichés par ces symboles, cela signifie qu'en ces lieux, c'est la forme représentée par ce symbole qui est utilisée au niveau de ce point d'enquête. D'autre part, le seul symbole qui a une signification particulière est le symbol : *. Ce symbole signifie que nous n'avons pas de données ou que les données ne correspondent pas à notre variable. Donc, les points d'enquête étant dans cette situation sont indiqués par ce symbole et dans la légende figurera en lieu et place d'un fait linguistique la mention NA (Non Applicable).

Enfin, le fond de carte en lui-même est parsemé de lignes vectorielles de couleur noire représentant les frontières des wilayas. L'intérieur des wilayas est d'un color blanc, alors que les territoires à l'extérieur des wilayas sont d'un color grisâtre. L'espace grisâtre à gauche des wilayas représente le Maroc et la ligne vectorielle séparant le Maroc et l'Algérie est bien évidemment la frontière entre ces deux pays.



Carte II.3.1 : Le fond de carte des cartes linguistiques

II.3.8.2.2 Dialectométrie

Nous allons ici même présenter la méthodologie adoptée pour obtenir nos résultats de dialectométrie. D'abord, évoquons les données qui seront exploitées. Les données collectées ne seront pas totalement utilisées pour cette étude dialectométrique. Nous nous limiterons aux variables linguistiques étudiées dans la partie Géolinguistique (Partie II.4.1, page 86). Les sous variables ne présentant pas de différences entre les variétés linguistiques seront amputées de notre corpus. Donc, ce dernier est seulement composé de variables linguistiques en rapport avec le système consonantique.

II.3.8.2.2.1 Méthodes de calcul numérique

Pour déterminer les partitions de nos variétés linguistiques, nous allons explorer plusieurs méthodes. Ainsi, nous emploierons la méthode d'Apprentissage Non Supervisé et la méthode d'Apprentissage Supervisé. Nous utiliserons les méthodes classiques d'analyse de données, mais notre objectif est aussi d'utiliser des méthodes de calculs numériques qui n'ont pas encore été prospectées. Comme nous l'avons vu dans l'état de l'art (Partie II.2.2.2, page 53), les méthodes potentielles sont très nombreuses. Par conséquent, pour des questions de faisabilité, il nous a fallu réduire notre travail exploratoire à quelques méthodes.

Concernant la méthode d'Apprentissage Non Supervisé, nous avons élaboré deux types de matrice de similarité. Nous avons utilisé la méthode Simple Matching Distance en utilisant le calcul de Gower. L'autre matrice de similarité utilisée correspond en réalité à une matrice de corrélation. Pour obtenir cette dernière, nous avons utilisé le test de corrélation de Kendall — métrique du χ^2 — qui nous a fourni une distance entre les variétés linguistiques. Cependant, avant d'arriver à cette matrice de corrélation, il nous a fallu transformer nos faits linguistiques en valeur numérique. Pour ce faire, nous avons employé la méthode Dummy.

Par la suite, ces dernières données numériques et nos matrices de similarité nous ont servi pour réaliser nos partitions. Pour y parvenir, nous avons adopté plusieurs méthodes. Ainsi, notre choix s'est porté dans un premier temps sur une PHA. Nous avons choisi comme méthode d'agrégation la méthode Complete. Nous avons également soumis nos données à d'autres méthodes de partition. Nous avons utilisé pour cela la méthode k-moyenne et la méthode k-mode. Cette dernière méthode est censée être l'équivalent de la méthode k-moyenne pour les données qualitatives. La mesure de similarité est réalisée par la méthode Simple Matching Distance et la mesure de distance s'obtient par un calcul propre à la méthode k-mode. Celui-ci, à notre connaissance, n'a jamais été utilisé en dialectologie pour réaliser une étude dialectométrique. Quant à la méthode k-moyenne, il est nécessaire de reprendre les données numérisées par la méthode Dummy. Ensuite, nous avons choisi d'utiliser l'algorithme Hartigan-Wong pour réaliser nos partitions avec la méthode k-moyenne.

En dernier lieu, nous réaliserons un Apprentissage Supervisé. Cette méthode n'a pas encore fait l'objet d'exploration par les dialectologues. Donc, ce travail servira à sonder dans quelle mesure ces algorithmes peuvent apporter à la dialectométrie. Ainsi, les

méthodes qui seront appliquées à nos données se rapportent à la méthode **Classification and Regression Trees (CART)** et à Random Forest. Nous avons décidé d'appliquer plus exactement la méthode classification tree qui est normalement dévolue aux données qualitatives. Les paramètres utilisés sont ceux qui sont établis par défaut. Pour pouvoir utiliser ces techniques, notamment Random Forest, il est nécessaire de reprendre les données obtenues par la méthode Dummy.

II.3.8.2.2 Représentation des partitions

Pour représenter et analyser nos partitions plusieurs techniques existent. Nous avons choisi une exploration visuelle pour rendre compte de nos résultats. Une partie de notre étude dialectométrique sera présentée sous forme d'arbre et de dendrogramme. Cette présentation des données convient à la **PHA** et à la méthode **CART**.

Mais, nous avons aussi l'intention de reproduire nos résultats sur des cartes linguistiques. Ce fond de carte est celui que nous avons défini précédemment (Partie **II.3.8.2.1**). D'autre part, nous avons envisagé d'utiliser en parallèle soit des cartes à point ou des cartes choroplèthes. Ce dernier sera envisagé pour représenter notamment la variation graduelle entre les variétés du Sud-Oranais. Nos cartes choroplèthes ont été obtenues en réalisant un diagramme de Voronoï.

CHAPITRE II.4

Résultats

Cette partie sera consacrée à la présentation de nos résultats. Cette étude dialectologique ne s'intéressera qu'au système consonantique et pour ce faire, nous avons sélectionné un certain nombre de consonnes qui nous paraissait intéressant d'analyser. Nous avons prévu dans un premier temps de mener une étude géolinguistique. Nous observerons les différences qui existent entre les différentes variétés. Puis, cela sera aussi l'occasion de discuter de questions de phonologie.

Nous poursuivrons l'étude dialectologique avec une étude dialectométrique. Nous verrons lesquels de ces nucléoles sont proches ou éloignés les uns des autres. Nous utiliserons des méthodes d'Apprentissage Supervisé et des méthodes d'Apprentissage Non Supervisé pour parvenir à partitionner les variétés dialectales.

II.4.1 Géolinguistique

L'étude de la distribution géographique des faits linguistiques sera l'unité d'analyse de cette partie. Nous avons sélectionné pour cela quelques consonnes pour illustrer cette situation. Ces unités phoniques correspondent à des consonnes simples et à des consonnes géminées. Pour chacune d'elles, nous en avons là aussi retenu certaines. Mais avant d'entamer ce travail, nous allons commencer par faire état de l'inventaire phonético-phonologique des variétés berbères du Sud-Oranais. Puis, nous préciserons le cadre conceptuel qui sera employé pour mettre en exergue cette variation linguistique.

II.4.1.1 Inventaire phonético-phonologique

Pour déterminer le système phonétique des variétés du berbère de cette région, nous avons choisi de nous concentrer sur un point d'enquête. Le système phonétique des différentes variétés étant quasiment identique, alors la mise en évidence du système d'une seule variété peut valoir de représentant à l'ensemble des autres variétés. D'autant plus que par la suite, nous allons présenter les différences phonétiques qui existent entre ces différentes localités. Donc, une présentation générale du système phonétique d'un point d'enquête complétée par une étude géophonétique doit pouvoir donner une image des systèmes phonétiques de l'ensemble des oasis du Sud-Oranais.

Le système phonétique que nous avons fait ressortir n'est pas à proprement parler un inventaire de toutes les particularités phonétiques possibles. Nous n'avons pas pris la peine de nous appesantir sur les détails phonétiques de toutes sortes qui peuvent exister, notamment les variations idiomatiques. Ce système est en quelque sorte un "système moyen", c'est pour cette raison que nous dénommons le système que nous allons présenter par système phonético-phonologique. Nous cumulon les deux dénominations pour indiquer que le système que nous présentons n'est ni le système phonologique, ni le système phonétique. Nous venons d'expliquer pourquoi, nous ne le considérons pas comme un véritable système phonétique.

Mais nous ne le considérons pas non plus comme un système phonologique. Car la détermination du système phonologique de ces variétés du berbère est problématique. Certains aspects de leur système s'avèrent difficiles à trancher entre ce qui relève du système phonétique ou phonologique. Nous nous trouvons accolé à des difficultés et leur analyse résiste aux différentes théories actuelles. Les méthodes connues pour déterminer

la phonologie d'une variété linguistique ne sont pas non plus d'une aide pour régler cette problématique. Quoi qu'il en soit, nous avons pris le parti de dégager un système qui est entre les deux dimensions.

Ce système servira de première approche et nous assumons le choix de simplifier la réalité. Le point d'enquête que nous avons retenu pour élaborer ce système est la localité d'Asla dans la wilaya de Naâma. Cet endroit est le lieu où nous avons le plus de données. Ce fut aussi le premier point d'enquête visité. Ainsi, les unités lexicales ou grammaticales collectées à Asla sont en nombre conséquent pour nous donner une représentation fidèle de leur système phonético-phonologique. Ce lieu servira donc comme une sorte d'archétype du système consonantique de l'ensemble de ces variétés du berbère.

Comme nous pouvons le voir sur le tableau II.4.1.1 ci-dessous, le système phonético-phonologique d'Asla n'est pas très différent de ce qui est retrouvé dans les autres langues berbères. D'une part, les variétés berbères de cette région ne sont pas des langues spirantes. Nous dirons qu'elles sont partiellement spirantes. Cette distinction qui est faite concernant la spirantisation ou non des langues berbères est ancrée dans les études berbères depuis fort longtemps. Ainsi, les langues berbères sont classées généralement en deux grandes catégories. Il y a d'une part les langues qui seraient occlusives et celles qui seraient fricatives. C'est comme ça que la tradition berbérissante distingue les langues berbères en se fondant sur leur système phonétique.

Dans la littérature, les termes "parlers spirants / occlusifs" ou "dialectes spirants / occlusifs" sont ceux qui sont les plus usités. Nous pourrions citer des auteurs, mais ils sont tellement nombreux que nous ne pouvons tous les citer et cela ne rimerait à rien de mettre en avant un chercheur plus qu'un autre pour évoquer cette question terminologique. Il serait plus pertinent de se poser la question pourquoi les berbérissants sont si nombreux à employer ce classement dans leur description des langues berbères. En tout état de cause, il convient de souligner que cette distinction est très largement répandue dans les études berbères.

Consonnes	bilabiale	labioden.	alvéolaire	post-alvéo.	palat.	vélaire	labio-vél.	uvulaire	pharyngale	glottale
occlusive	b	t	d					q		
	<i>pharyngalisée</i>		d ^f							
	b:	t:	d:		k:	g:	g ^w :	q:		
	<i>pharyn. gémi.</i>	t: ^f								
nasale	m		n							
	<i>géminée</i>		n:							
roulée	m:		r							
	<i>pharyngalisée</i>		r ^f							
battue			r							
	<i>pharyngalisée</i>		r ^f							
fricative		f	s	z	ʃ	ʒ		χ	ɣ	h
	<i>pharyngalisée</i>		s ^f	z ^f		ʒ ^f			ħ	ʕ
	<i>géminée</i>	f:	s:	z:	ʃ:	ʒ:		χ:	ħ:	ʕ:
	<i>pharyn. gémi.</i>		z: ^f							
spirante / approx.									j	w
	<i>géminée</i>								j:	
latérale									l	
	<i>géminée</i>								l:	
	<i>pharyn. gémi.</i>								l: ^f	

TABLE II.4.1 : Tableau du système phonético-phonologique

Concrètement, la spirantisation ou la non spirantisation concerne les consonnes bilabiales, dentales et vélares¹. On dira que les langues berbères qui ont un mode d'articulation constrictif pour ces trois types de consonne qu'elles sont des langues spirantes. C'est le cas par exemple pour le kabyle et le rifain. Donc, les langues qui ne sont pas dans ce cas sont cataloguées comme étant des langues occlusives comme par exemple le touareg.

Cependant, cette division des langues berbères sur la base de ce critère n'est pas pertinente pour toutes les langues berbères, comme c'est le cas dans le Sud-Oranais. Pour les deux premiers types de consonnes, toutes les variétés du Sud-Oranais ne connaissent pas de spirantisation². Le mode d'articulation est occlusif. Pour le dernier type de consonnes, les choses ne sont pas uniformes. Il existe des variétés qui, pour les vélares, ont un mode d'articulation spirant et d'autres un mode d'articulation occlusif. Voire certaines variétés font une distinction à ce niveau entre les vélares voisées et non voisées³ qui pour l'une sera fricative et pour l'autre occlusive. Cette présentation sommaire des faits montre qu'il existe une diversité de réalisations du mode d'articulation entre ces phonèmes. Ainsi, ces unités phoniques ne se recoupent pas au niveau de leur obstruction. Par conséquent, un tel inventaire phonétique ne peut rentrer dans cette classification binaire des langues berbères qui est passée en usage chez les berbérissants.

De plus, si nous considérons l'ensemble de l'espace berbère, le qualificatif de spirantisation n'est pas suffisant pour décrire la diversité des réalisations phonétiques. Car, lorsque les berbérissants parlent de spirantisation, ils englobent sous cette dénomination diverses réalisations phonétiques qui ont un mode d'articulation identique, mais des points d'articulation différents. Ainsi, dans cette opposition entre les langues "occlusives" et les langues "spirantes", il n'y a que le critère de l'obstruction qui est pris en compte. Pourtant au sein de cet ensemble de langues "spirantes", la spirantisation n'implique pas qu'une différence au niveau du mode d'articulation, cela aussi intègre le point d'articulation. La spirantisation s'accompagne pour certaines variétés du berbère également par une position plus antérieure des spirantes⁴ par rapport aux occlusives.

1. Le point d'articulation n'est pas exactement à ces niveaux-là selon les variétés du berbère, mais elles se situent toutes dans les environs de ces points d'articulation.

2. Sauf une, voir à la page 107 (carte II.4.9).

3. Ce point est traité et illustré à la page 103 (carte II.4.5).

4. Par la suite, nous emploierons le mot "spirant" dans le sens où la consonne spirante ne se différencie de la consonne occlusive que par une différence au niveau de l'obstruction. Cela afin de distinguer, les variétés qui connaissent un phénomène de palatalisation de celles qui ignorent de telle caractéristique

En somme, il y a un phénomène de palatalisation. Donc, dans le monde berbère, nous pouvons retrouver plusieurs cas de figure. Des langues totalement occlusives, ou totalement spirantes. Des langues partiellement occlusives et partiellement spirantes. Des langues partiellement spirantes et partiellement palatalisées. Des langues partiellement occlusives et partiellement palatalisées. En résumé, la réalité est plus complexe que ce qui est affirmé habituellement tout particulièrement concernant l'évolution des vélaires⁵.

En dehors de ce que nous venons de signaler, ces variétés du berbère sont en osmose avec les autres langues berbères. Elles ont ainsi toutes dans leur inventaire des consonnes géminées. Au niveau phonétique, n'importe quel point d'articulation et mode d'articulation peut avoir en contrepartie une géminée. Cette gémination présente dans un morphème peut être soit lexicale, soit morphologique. En revanche, comme nous pouvons le constater, certaines colonnes n'ont pas de consonnes géminées. Nous constatons ce fait particulièrement pour certaines consonnes gutturales, approximantes et alvéolaires. Cela ne signifie pas que ces consonnes simples ne connaissent pas de géminées, il s'avère que tout simplement leur réalisation phonétique est articulatoirement différente de leur contrepartie simple.

De manière abstraite, un phonème simple /X/ aura pour unité phonique [X], alors que la contrepartie géminée de ce phonème /X:/ aura pour unité phonique [Y:]. Pour les cas de cette sorte, nous avons choisi de présenter l'unité phonique [Y:] et non le phonème /X:/ de la contrepartie /X/. C'est de cette manière que la tradition berbérissante présente le système phonologique des différentes langues berbères décrites. Alors même que nous n'avons pas à faire vraisemblablement au système phonologique, mais probablement plus à la réalité phonétique. Nous parlons au conditionnel, car il n'est pas certain que les choses soient aussi simples entre les représentations et les formes de surface des géminées⁶. Donc, c'est aussi en raison de ce fait que nous ne nous sommes pas appliqué à démarquer nettement le système phonétique du système phonologique, du reste aussi pour respecter la tradition berbérissante.

Par ailleurs, d'autres phénomènes nous ont poussé à ne pas prétendre avoir cerné la question du système phonétique et phonologique de ces variétés. Ainsi, l'autre aspect phonético-phonologique caractéristique des langues berbères est la pharyngalisation. Ce

phonétique à l'exemple du kabyle.

5. Cette question sera abordée à la page 97 (partie II.4.1.3.1).

6. La gémination sera vue plus en détail à la partie II.4.1.4 (page 119).

dernier en berbère touche, normalement, exclusivement les unités phoniques coronales. Cependant, ce ne sont pas tous les phonèmes coronaux qui sont concernés. La consonne occlusive /t/ ne connaît pas de contrepartie pharyngalisée et il en est de même pour /d:/. Toutefois, cela est hypothétiquement vrai pour la réalité phonologique. Mais au point de vue phonétique, il arrive de rencontrer des consonnes pharyngalisées étant au niveau des points d'articulation des consonnes précédentes. Ceci est également le cas pour toutes les autres consonnes coronales qui n'ont pas de pendant pharyngalisé. Si nous les avons écartées de notre inventaire, c'est qu'elles présentaient une fréquence d'apparition très faible. Donc, nous ne pouvions écarter le fait que ces types de phonèmes fussent la conséquence de phénomènes idiomatiques.

C'est pour cette même raison que nous n'avons pas intégré d'autres unités phoniques. Le fait qu'elles soient peu usitées rend leur caractérisation phonologique difficile. Ainsi, d'autres unités phoniques sont dans le même cas. Donc, il faut ajouter à ces catégories manquantes aussi les consonnes labiales et/ou nasales "pharyngalisées". Nous mettons ce dernier mot entre crochets, car nous ne sommes pas certain que le second point d'articulation se trouve dans la zone pharyngale. Probablement qu'articulatoirement, nous avons à faire à une uvularisation — la vélarisation n'est pas à exclure non plus —, mais cela reste à confirmer. Les cas auxquels nous pensons sont les items comme @wet[n:^x]a@ \ma_sœur\. La question est : l'exposant x est-il à remplacer par ^ɣ, ^ʁ ou ^ʁ ?

En outre, évoquons la situation de certaines consonnes labio-vélaires. Ainsi, parmi elles, quelques-unes ont été écartées et d'autres figurent dans l'inventaire. Une seule se trouve dans l'inventaire, toutefois comme nous l'avons déjà affirmé, il n'est pas certain qu'il s'agisse d'un phonème. Malgré tout, elle s'y trouve, cela étant dit nous allons laisser de côté ce point pour évoquer celles qui ne s'y trouvent pas. En fait, il existe bien d'autres labio-vélaires qui se comptent sur les doigts d'une main telles que @aj[m:^w]un@ \parcelle_de_terre\, @tace[k:^w]art@ \cabas\, @[b:^w]a@ \père\. Il est facile de se rendre compte que celles-ci ne surgissent que dans le cas de consonnes géminées et quand le conduit oral connaît une phase d'occlusion. Donc, ce contexte limité expliquerait pourquoi nous les retrouvons pas en abondance.

D'autre part, comme vous l'aurez remarqué dans le tableau II.4.1.1, Asla n'a pas dans son système phonético-phonologique de consonnes dento-alvéolaires fricatives — Non-sibilantes —⁷ — comme nous avons pu l'illustrer sur la carte II.4.9 —, ni

7. Nous ne savons pas de manière certaine si le point d'articulation est, par exemple pour la consonne sourde, dental [θ] ou alvéolaire [θ], donc c'est pour cette raison que nous employons le terme

de consonnes bilabiales fricatives. Mais pourtant cette variété comme celles qui lui sont apparentées emploient plus largement des consonnes palato-alvéolaires ou des consonnes spirantes que de consonnes vélaires. Nous soulignons ce fait, car les descriptions des langues berbères nous avaient habitué à ce que celles qui présentaient des consonnes palato-alvéolaires au lieu de consonnes vélaires aient aussi dans leur inventaire phonologique des dento-alvéolaires non-sibilantes et des bilabiales fricatives.

Après avoir commenté le système phonético-phonologique d'Asla, nous allons présentement discuter de la variation phonético-phonologique retrouvée dans l'ensemble du Sud-Oranais. Cela permettra de compléter le précédent tableau sur le système phonétique et phonologique. Ce dernier correspond au système d'Asla et en nous basant dessus l'objectif est de voir si ces unités phoniques présentent des variations ou non dans les autres oasis berbères. C'est de cette manière, de proche en proche, que nous comptons déterminer le système phonético-phonologique de l'ensemble des **nucléoles** de cette région. Sans que nous ayons besoin de collecter un corpus conséquent dans chacune de ces oasis. Même s'il est vrai que chacun de ces nucléoles a son propre système phonétique et phonologique, il s'avère qu'ils partagent plus d'unités phoniques en commun que d'unités phoniques différentes.

Pour rendre compte de cette variation, nous avons utilisé notre questionnaire. Rappelons que ce questionnaire est conçu de telle manière à contenir l'ensemble des unités phoniques consonantiques.

II.4.1.2 Les variables linguistiques

Nous avons établi un certain nombre de variables linguistiques. Celles-ci, comme nous l'avons déjà annoncé, seront restreintes au système phonotico-phonologique et plus spécifiquement aux consonnes. Ces variables correspondent donc à des cognats phonologiques. Cette notion de cognat phonologique est un néologisme. Le cognat en linguistique correspond à une paire de mots — ou plus — provenant de différentes langues qui partagent une morphologie et un sens qui sont approximativement les mêmes. Le cognat phonologique s'inscrit dans le même état d'esprit, mais cette fois l'unité n'est pas le morphème, mais l'unité phonique. Cela étant dit, il ne faut pas considérer ces deux types de cognat comme totalement indépendant l'un de l'autre. Car cette unité phonologique se caractérise d'abord en mettant en évidence un cognat

dento-alvéolaire.

— appelons-le dorénavant cognat morphémique —. Ce n'est que par la suite lorsque le cognat morphémique est déterminé qu'une possible caractérisation du cognat phonologique est envisageable.

Le travail de mise en évidence de cognats n'est pas en soi le plus dur à déterminer. Un bon nombre de cognats morphémiques sont connus depuis les premières études sur le berbère. René Basset menait presque souvent des comparaisons entre les variétés du berbère dans chacune des études qu'il réalisait. Cela a toujours été une tradition chez les berbérissants de mettre en rapport les langues berbères. Cela s'est toujours poursuivi, nous en voulons pour preuve les derniers travaux portant spécifiquement sur ce genre de problématique : [Naït-Zerrad, 1998] [Naït-Zerrad, 1999] [Naït-Zerrad, 2001]. De plus, l'établissement de cognats morphémiques ne souffre pas de problèmes de variations morphologiques importantes. La morphologie du berbère est assez conservatrice comparée à ce qui peut exister en indoeuropéen. Puisque de bout en bout du monde berbère, nous pouvons trouver des morphèmes qui sont presque en tout point de l'Afrique du nord identiques. Alors si dans des points extrêmes de l'Afrique du nord nous pouvons trouver des cognats, cette tâche est rendue encore plus aisée si cela concerne des variétés proches géographiquement et linguistiquement.

Toutefois, il ne faut pas se précipiter en pensant que l'établissement d'un cognat suffirait à ce que chaque unité phonique de ces cognats soit liée. Loin de là. L'établissement d'un cognat morphémique n'est pas suffisant pour affirmer que l'ensemble des phonèmes présents dans les unités significatives sont de paire à paire liées et sont donc des cognats phonologiques. Cela n'est pas suffisant pour établir une correspondance phonologique. Il est possible qu'il puisse exister dans ces cognats morphémiques des unités phonologiques qui forment des cognats phonologiques, mais cela ne sera certainement pas le cas pour toutes les unités phonologiques. Ou à l'inverse, toutes les unités phonologiques d'un cognat morphémique peuvent bien être des cognats phonologiques. Par exemple, tous les pronoms indépendants de personne de chacune des variétés du Sud-Oranais forment un cognat morphémique. C'est le cas pour le pronom indépendant de la 2^{ème} personne du pluriel. Cependant, si nous prenons l'exemple de ce pronom pour la variété d'Asla — @fem:umi@ — et de Bousseghoun — @fek:^wmim@ —, il est peu probable que les unités phoniques [m:] et [k:] forment un cognat phonologique. Nous pouvons affirmer cela, car la correspondance phonologique n'a cours que pour ces deux unités significatives. Nous ne la retrouvons nulle part ailleurs dans d'autres morphèmes.

En plus de ce que nous venons de dire, un autre aspect est aussi à retenir. Ces cognats phonologiques ne sont pas exemptés par des phénomènes de coarticulation ou d'assimilation. Une correspondance phonologique n'étant pas répandue au niveau d'autres cognats morphémiques ou retrouvée dans des contextes limités est un indice que cette correspondance est seulement fortuite. C'est le cas dans les cognats morphémiques @[rɜ]@ et @[rɜ]/@[krɜ]@ signifiant \labourer\ ou \semer\. Le premier est prononcé dans les localités de Ouakda jusqu'à Boussemghoun et le second dans les autres localités du Sud-Oranais. Dans cet exemple, nous pourrions penser que les phonèmes en final [ɜ] et [z] forment un cognat phonologique, puisqu'il y a bien une correspondance phonologique. Mais une analyse plus détaillée des unités significatives de ces variétés linguistiques nous indique que ce n'est nullement le cas. Le constat que nous faisons est que dans ces localités les dentales fricatives dans l'environnement d'une palatale ont tendance à se palataliser. D'autres unités significatives présentent le même phénomène comme @[tɛzrest]@ \hiver\ qui est prononcé @[tɛzreft]@ à Ouakda — ce phénomène d'assimilation ne s'est pas répandu à Asla et ses environs pour toutes les unités significatives ayant le même contexte —. Nous pourrions nommer ce type d'unités phonologiques des *allocognats phonologiques* par analogie avec allophones. En revanche, il faut se garder de les considérer comme des cognats phonologiques à part entière dans la mesure où cette relation ne se manifeste que de manière contextuelle. Pour enfoncer le clou, un cognat phonologique doit survenir dans toutes les positions morphologiques ou tout du moins dans la majorité d'entre elles.

Donc, la distribution est un critère à prendre en compte. Ce dernier nous permet de déterminer si une variation phonologique qui apparaît entre différents systèmes linguistiques est de l'ordre du cognat phonologique ou non. Ainsi, les méthodes appliquées pour établir le cognat phonologique sont le comparatisme et la fréquence de l'observation d'événements semblables. Si aucune régularité au niveau des correspondances entre phonèmes n'est établie comme c'est le cas pour [k:] et [m:], alors il faut écarter l'hypothèse d'un cognat phonologique. Par conséquent, c'est en prenant en compte toutes ces considérations que nous avons établi un certain nombre de cognats phonologiques. Ces derniers nous serviront à établir la variation entre ces différentes variétés du berbère.

D'autre part, ces cognats phonologiques sont à distinguer des protophonèmes. Bien que l'établissement d'un cognat phonologique implique une dimension diachronique, nous ne prétendons pas établir une loi phonétique et encore moins la reconstruction

d'un prothophonème. Ce cognat phonologique est à considérer comme une variable, donc comme un concept nous aidant à caractériser la variation linguistique. Nous nous inscrivons totalement au niveau synchronique. Cela étant ces variables linguistiques, pour des questions de mode de présentation, ont besoin d'être étiquetées. Ce besoin d'une présentation synthétique nous a poussé à caractériser ces données qui correspondent à des cognats phonologiques par un symbole. Ainsi, chaque variable linguistique sera présentée de la manière suivante : \hat{X} . Le x majuscule représente notre variable et le circonflexe est là pour rappeler que c'est un concept. Cela est nécessaire parce que nos symboles sont définis par des lettres et le choix de ces lettres n'est pas innocent. L'adoption de ces lettres s'est faite en prenant en considération l'ensemble des variétés linguistiques du berbère. Les cognats morphémiques présents dans le Sud-Oranais existent aussi dans d'autres aires linguistiques du berbère. En conséquence, nous pouvons faire les mêmes rapprochement et voir comment ces cognats phonologiques se manifestent dans le monde berbère.

Ainsi, c'est en fonction d'une comparaison externe et interne au Sud-Oranais que nous avons décidé de quelle lettre représenterait nos variables. Nos choix se sont tournés souvent vers les variétés linguistiques qui pour une même variable linguistique présentaient une seule forme. Alors que dans le Sud-Oranais cette variable linguistique présente plusieurs formes. Par exemple, les cognats @a[g]em@ \puiser{2.SG.IMP}\ et @ar[g]az@ \homme\ qui sont ainsi énoncés en chleuh seront prononcés à Asla respectivement @a[j]em@ et @ar[g]az@. Cette mise en relation de ces unités significatives nous permet de nous rendre compte au premier abord de deux cognats phonologiques l'un a les formes [g] et [j] et l'autre a les mêmes formes [g]. Mais au lieu de considérer ces deux cognats phonologiques comme distincts, nous les réunissons sous une même étiquette qui sera notre variable et nous les symbolisons par la lettre G en considération des réalisations du chleuh. Cependant, précisons d'ores et déjà que ce polymorphisme n'est pas nécessairement conditionné par l'environnement phonétique comme nous le verrons par la suite. Après ce qui vient d'être énoncé, nous pouvons expliquer pourquoi en plus du symbole que nous avons précédemment décrit, nous avons rajouté un chiffre à certaines de nos variables. Cela a été réalisé dans le but d'identifier convenablement les cognats phonologiques. Finalement, nous obtenons le symbole final suivant : \hat{X}_i . Soulignons par ailleurs que cet élargissement à d'autres variétés linguistiques a l'avantage de pouvoir reproduire les mêmes comparaisons à une échelle plus large dans de futurs travaux.

Maintenant que notre approche est définie, nous allons à présent analyser les données. Les cognats phonologiques que nous avons choisis ne représentent pas l'ensemble des cognats phonologiques retrouvés dans le Sud-Oranais. D'abord, tous les cognats phonologiques ne présentant pas de variations n'ont pas été retenus comme variables. Donc, au sein de ces derniers seuls ceux qui connaissent une variation ont attiré notre attention. Cette fois-ci encore, nous n'avons sélectionné que ceux qui présentaient un intérêt pour les questions que nous voulions traiter. L'objectif de cette étude n'est pas seulement de montrer comment la variation linguistique se disperse au niveau géographique. Mais de tirer aussi, entre autres, un certain nombre d'enseignements d'ordre plus général sur les évolutions phonologiques ainsi que sur les représentations phonologiques. Nous avons en partie évoqué cette problématique dans la partie précédente, cette fois notre intention est d'en parler plus en détail. De ce fait, dans un premier temps les cartes linguistiques qui seront présentées illustreront les réalisations phonétiques de ces variables. Et les questions phonologiques seront abordées par la suite dans la partie Discussion (II.4.1.3.2 page 109).

Ensuite, nous traiterons à part la question des consonnes simples et des consonnes géminées. Car les phonèmes en jeu ne sont pas identiques et sont relatifs à des phénomènes phonologiques distincts. Et au sein de chacun de ces types de consonnes nous avons déterminé plusieurs variables. Une partie des variables à l'intérieur de ces catégories de consonnes traitent d'un même phénomène, alors que d'autres sont totalement indépendantes. D'ailleurs, certaines des variables des consonnes simples rentrent en résonnance avec certaines des variables des consonnes géminées. C'est ce que nous tâcherons de mettre en exergue dans un deuxième temps après l'analyse des matériaux.

D'autre part, comme nos variables spatiales ont été précédemment définies (Partie I.1.2.3, page 14), donc nous n'allons pas les évoquer à nouveau. Nous dirons seulement que celles-ci constituent nos variables indépendantes et que nos variables dépendantes sont dévolues à nos variables linguistiques.

II.4.1.3 Consonnes simples

La variation des consonnes simples sera traitée par trois variables. Les deux premières variables $\wedge G$ et $\wedge K$ sont liées, elles se distinguent en partie par leur mode de voisement. Cette différence de voisement n'est pas l'unique particularité qui distingue

ces deux variables. Ces variables sont déployées en plusieurs sous variables \hat{G}_i et \hat{K}_i . Les phénomènes qui touchent ces dernières sont en partie également différents. Quant à la dernière variable \hat{T} , il n'y a ni différence de voisement ni sous variables. Mais elle traite de phénomènes phonétiques qui sont rarement discutés dans les études berbères. Ces phénomènes en question ne sont pas sans relation avec les deux premières variables. Ils sont en lien avec la question de l'occlusion et de la frication qui serait le clivage qui marquerait les langues berbères.

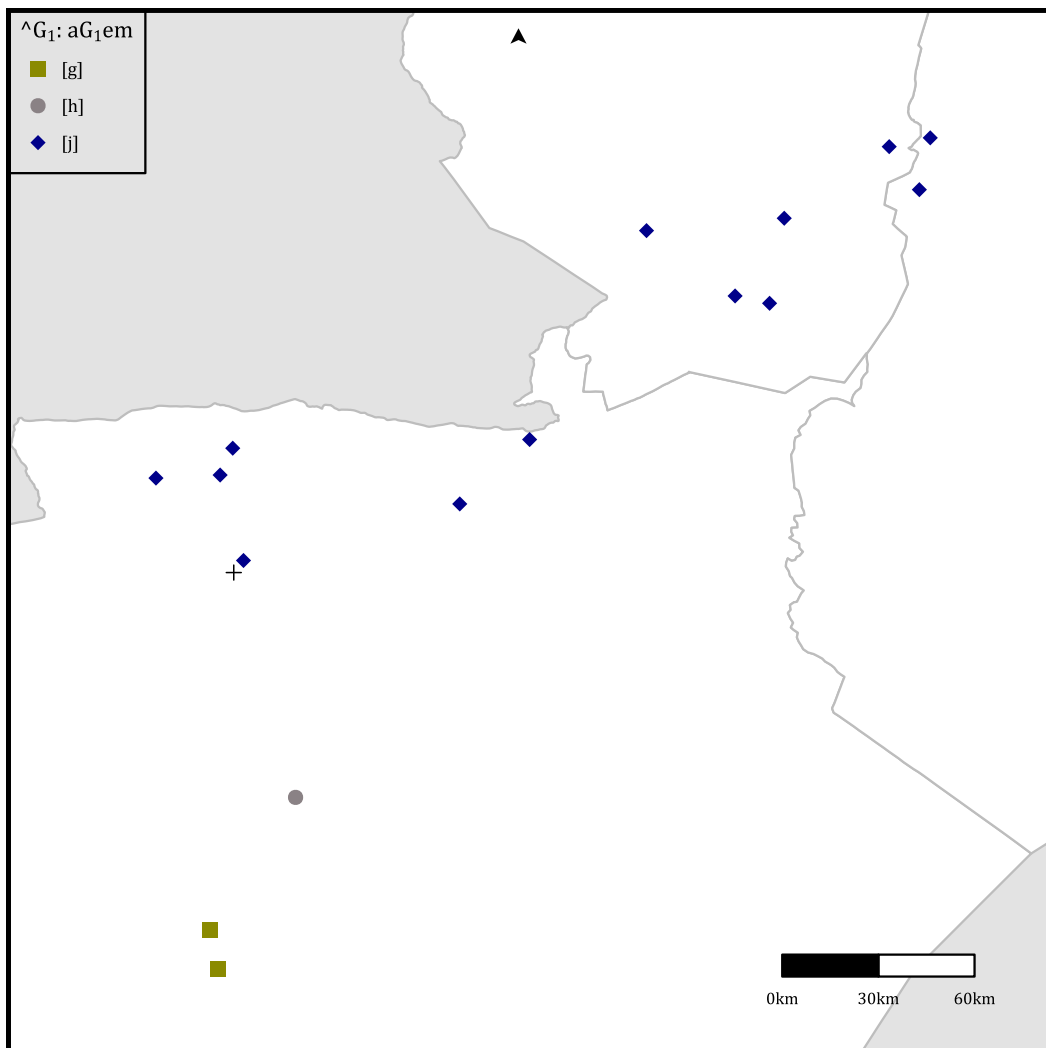
II.4.1.3.1 Le matériau d'analyse

Nous allons dans un premier temps évoquer les variables \hat{G}_i et \hat{K}_i . Ces dernières connaissent un développement différent suivant les localités. Elles alternent entre consonnes occlusives et consonnes fricatives. Cette alternance se manifeste soit entre différentes variétés linguistiques ou soit à l'intérieur d'un même système linguistique.

Les différentes formes de réalisations phonétiques qui sont regroupées sous ces variables se distinguent principalement par leur mode d'articulation. Il existe des variétés qui ont un mode d'articulation pour ces variables qui est plutôt occlusif et d'autres plutôt fricatif. Cependant, cette division n'est pas aussi tranchée que nous pouvons le croire en comparaison de ce que nous pouvons retrouver en rifain ou en chleuh par exemple. Présentons en premier lieu comment les choses surviennent avec la variable \hat{G} .

II.4.1.3.1.1 La variable \hat{G}_i : Occlusion et Frication

Cette première sous variable présente des réalisations phonétiques très diverses. Comme nous le disions en préambule, les formes se répartissent entre des occlusives et des fricatives. Donc, il y a la réalisation occlusive de cette variable qui est [g] et les réalisations fricatives [j] et [h]. La forme [g] n'est attestée que dans le sud du Sud-Oranais — Igli et Mazzer —. La forme [h] n'est attestée qu'à Berrebi. Le reste des localités utilise la forme [j].

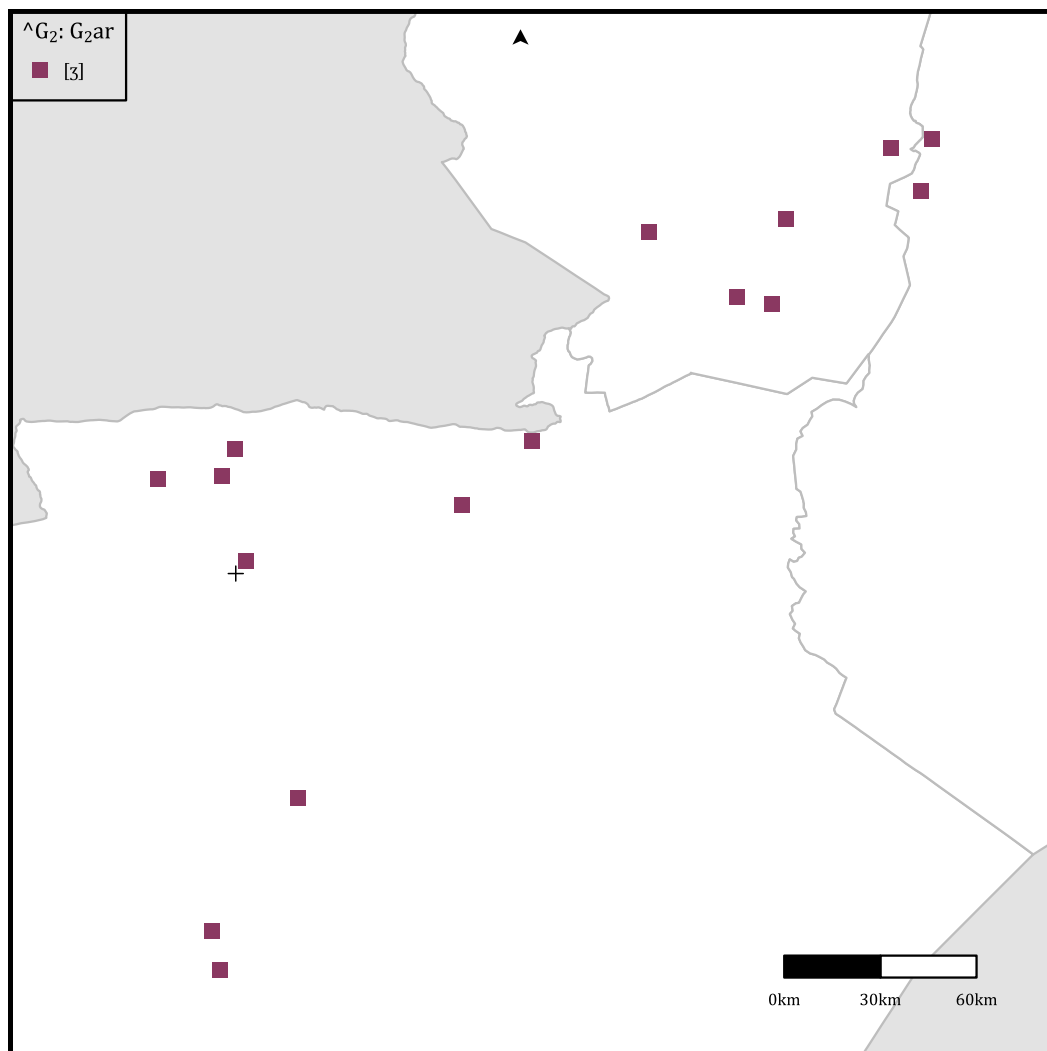


Carte II.4.1 : La variation de $\wedge G_1 : aG_1em$ \puiser{2.SG.IMP}

Cette présentation succincte peut être complétée en mettant le doigt sur le fait que le mode d'articulation n'est pas la seule particularité physique qui les différencie. Le point d'articulation est aussi en jeu. Ainsi, en plus de cela, il y a palatalisation, vélarisation et débucalisation — ces dénominations ne postulent aucune direction d'évolution sur le changement phonétique entre ces occlusives et ces fricatives. Il en sera de même systématiquement par la suite —. Mais pour cette même sous variable, le nucléote de Berrebi peut utiliser la forme [j] à l'instar des autres variétés du nord comme l'atteste cet exemple @[j]ud-u@ \maintenant{ADV-DEM_PROX}, et les localités du sud continueront à utiliser la forme occlusive et diront @[g]ud-u@.

D'autre part, cette disposition n'est pas aussi immuable que ce que nous pouvons

rencontrer dans d'autres aires linguistiques du berbère. Pour cette même variable, d'autres formes peuvent exister. La deuxième sous variable connaît pour forme [ʒ]. Cette dernière est distribuée de manière homogène sur le territoire. Aucun de nos points d'enquête présente une réalisation phonétique différente. Donc, mêmes les variétés qui avaient pour forme une occlusive pour la sous variable ^G_1 utiliseront pour ^G_2 une fricative.

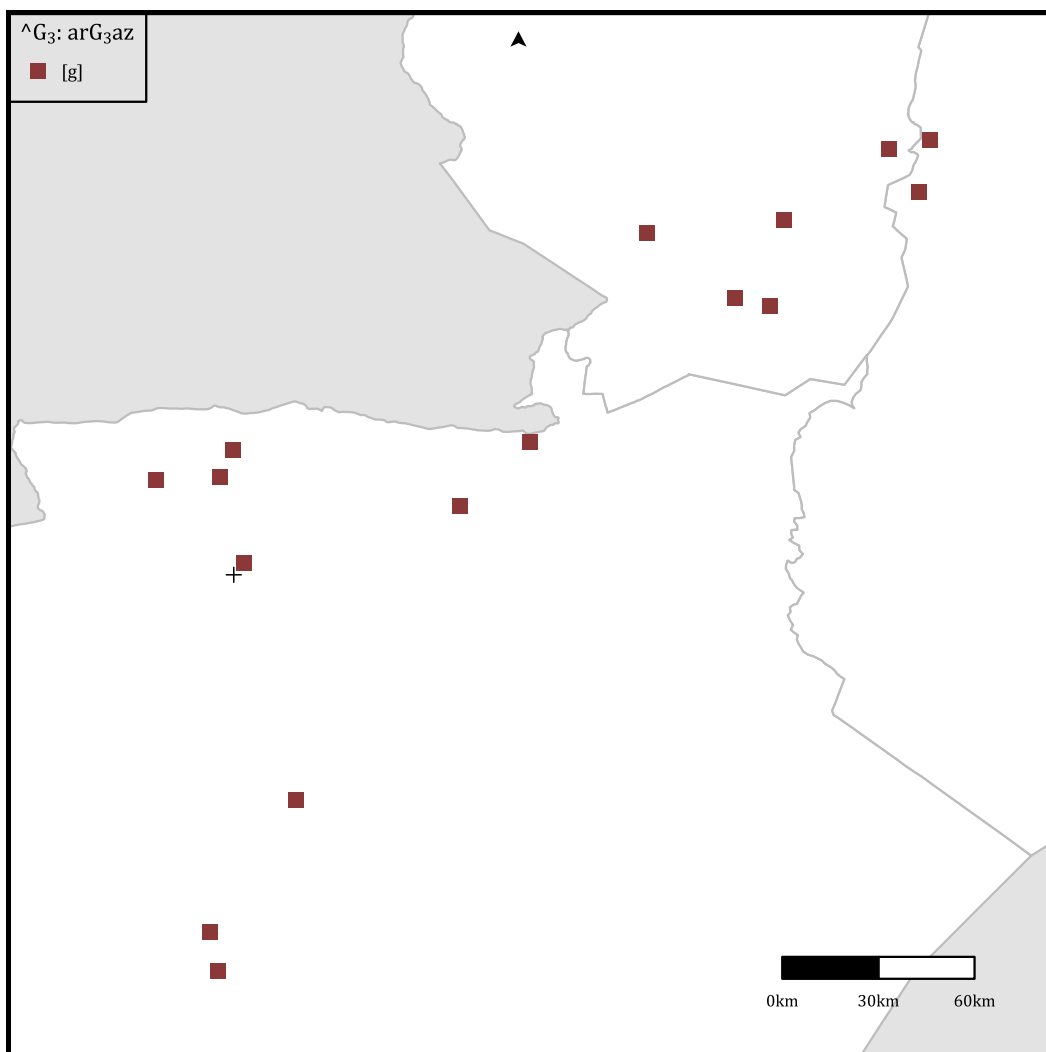


Carte II.4.2 : La variation de $\text{^G}_2 : G_{2ar}$ \ entre{PREP} \

Toutefois, la consonne fricative utilisée n'est pas la même que celles retrouvées pour ^G_1 . Le point d'articulation est plus antérieur. Ce tableau semble brosser une certaine harmonie au niveau des formes pour cette variable, mais cela n'est vrai qu'en apparence. Car, encore une fois, Berrebi se distingue des autres localités. Pour certaines unités

significatives, ce nucléolecte utilisera la forme [h] de la sous variable $\wedge G_1$, alors que les autres localités utiliseront tous la forme [ʒ] de la sous variable $\wedge G_2$. Comme le confirme cet exemple où nous aurons @a[h]ris@ \hiver\ pour Berrebi alors qu'ailleurs nous avons collecté les formes @a[ʒ]rif@ ou @a[ʒ]ris@.

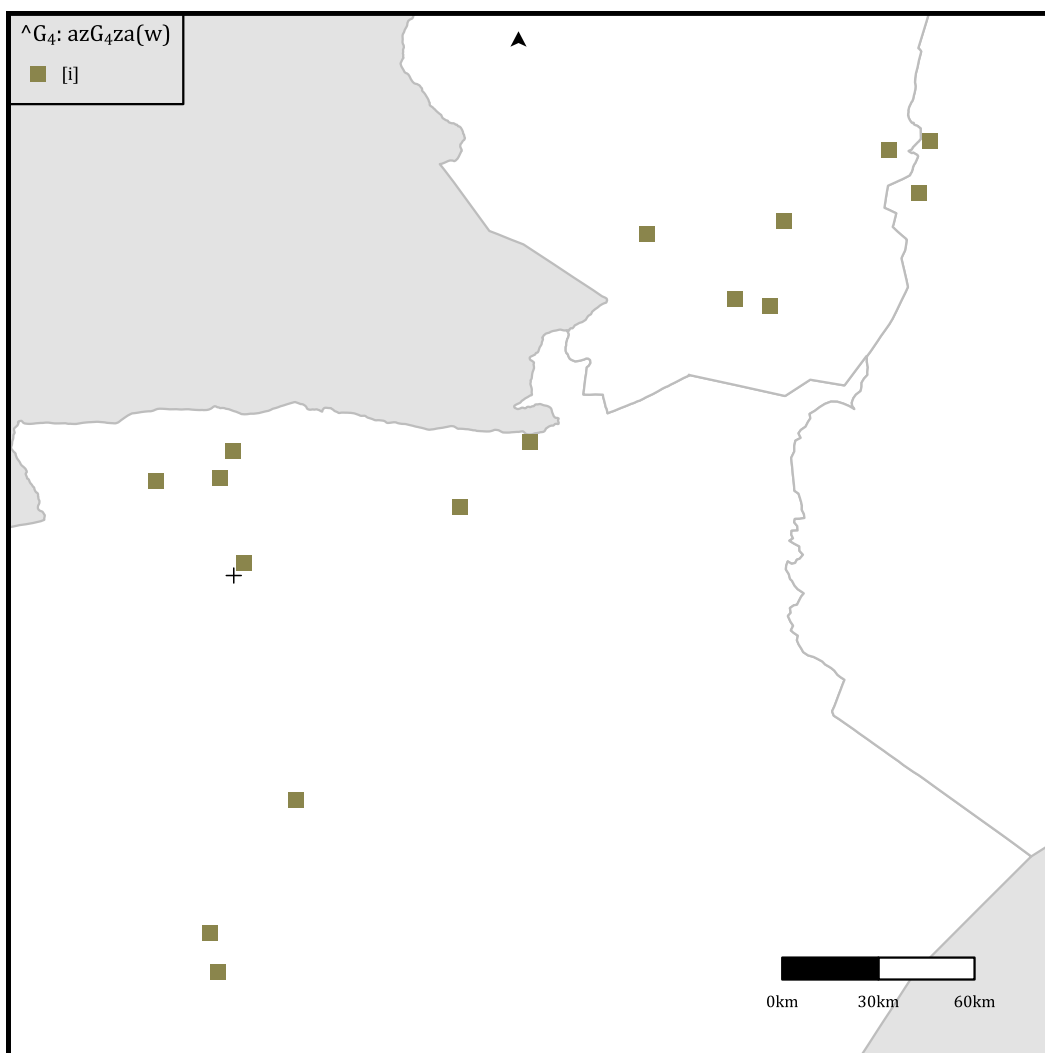
L'autre cas de figure rencontré connaît la même distribution que la sous variable précédente comme nous pouvons le voir ci-dessous. Toutefois, la sous variable $\wedge G_3$ n'est pas contrastive de la même façon. Les variétés qui avaient pour forme des fricatives pour la sous variable $\wedge G_1$ s'emploient à utiliser une occlusive pour cette sous variable.



Carte II.4.3 : La variation de $\wedge G_3$: arG3az \homme\

Pour Igli et Mazzer, cette forme est dans la continuité de la sous variable \hat{G}_1 . Ce qui n'est nullement le cas pour les autres variétés. Donc, pour ces dernières variétés se rajoutent la forme [g] à leur inventaire phonétique. Mais une fois encore, cette disposition n'est pas aussi unidirectionnelle. La situation inverse peut exister comme nous pouvons le voir sur la carte suivante.

Effectivement, cette fois-ci ce sont les autres localités qui se trouvent en adéquation avec la sous variable \hat{G}_1 . Pour cette sous variable \hat{G}_4 , ce sont les variétés d'Igli et de Mazzer qui rejoignent la forme utilisée par les autres localités.



Carte II.4.4 : La variation de $\hat{G}_4 : azG_4za \setminus \text{vert} \setminus$

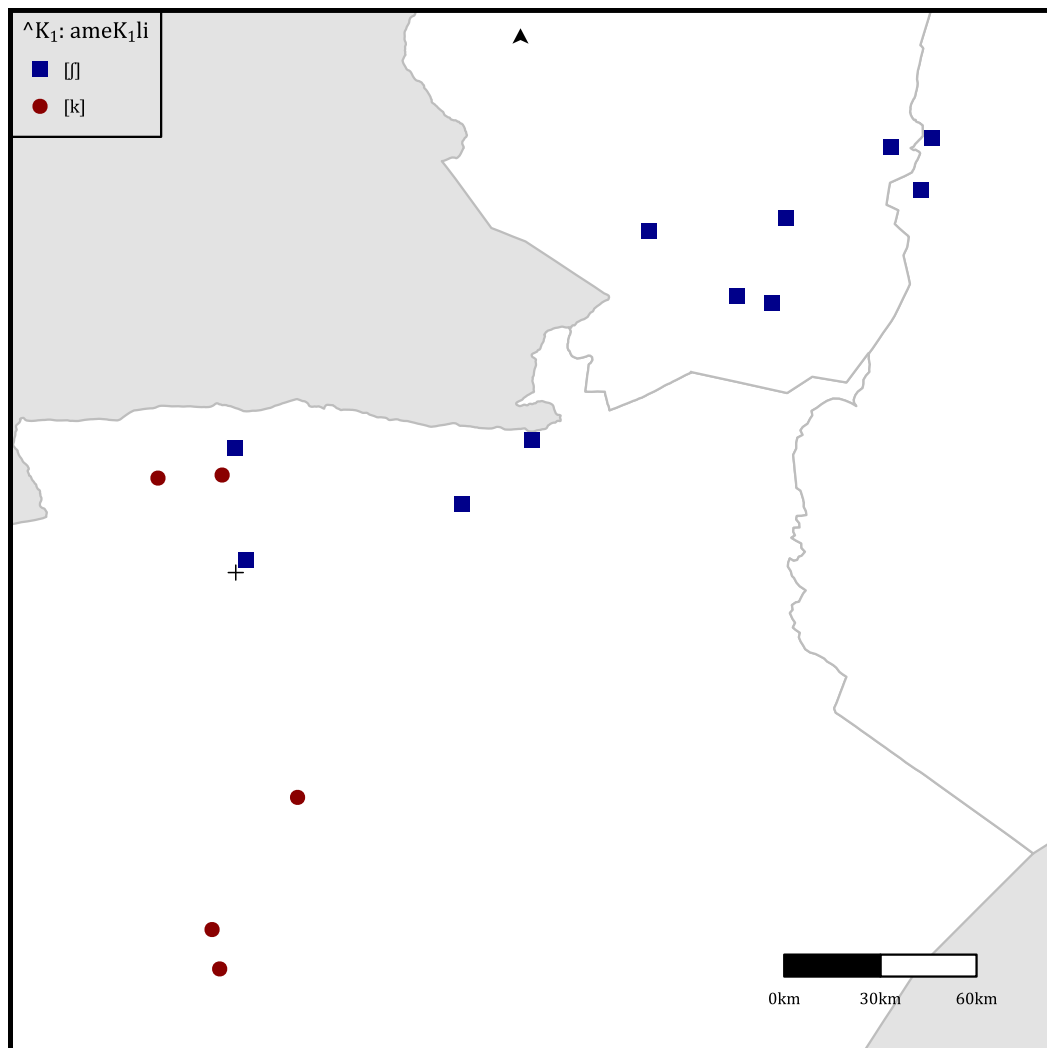
Donc, les autres variétés linguistiques maintiennent la forme utilisée dans la variable \hat{G}_1 . Précisons par ailleurs que cette sous variable a en réalité deux formes [j] et [i]. Bien que l'exemple⁸ utilisé n'ait pas pour forme une consonne — cette configuration est présente seulement pour cet exemple —, il s'avère que cette voyelle n'est que la vocalisation de la semi-voyelle *j. Par conséquent, nous l'avons tout de même retenue pour exemplifier cette sous variable.

Nous arrivons à la fin de la présentation des matériaux d'analyse pour la variable \hat{G} . Observons ce qui l'en est pour la variable \hat{K} .

II.4.1.3.1.2 La variable \hat{K}_i : Occlusion et Frication

Cette seconde variable connaît de même des sous variables. Elle présente des correspondances avec la variable \hat{G} , mais cela n'est pas le cas sur tous les plans. Si nous commençons avec la sous variable \hat{K}_1 , la même distinction est observée. La division entre réalisation occlusive et fricative est toujours maintenue. Nous avons celles, pour cette sous variable, qui ont pour forme une consonne occlusive vélaire et celles qui ont une consonne fricative alvéolo-palatale.

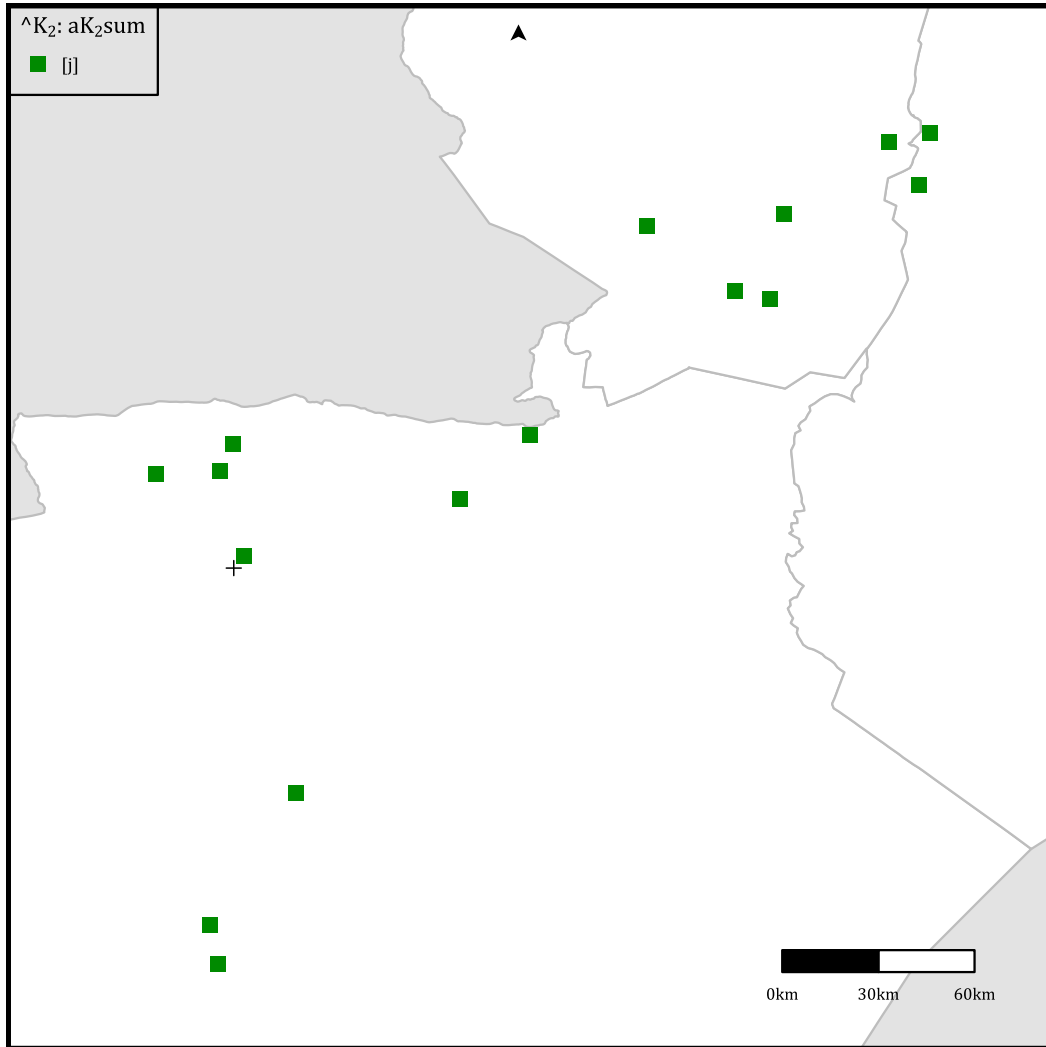
8. Malgré que toutes les variétés aient pour cet exemple la même forme pour \hat{G}_4 , ce n'est pas le cas pour tout le reste de la morphologie. Igli et Mazzer ont en plus la semi-voyelle [w] en final.



Carte II.4.5 : La variation de ^K_1 : ameK₁li \déjeuner\

De plus, la distinction au niveau du point d'articulation est aussi conservée entre les formes de cette sous variable. Cependant, contrairement à la contrepartie voisée, la sous variable ^K_1 ne connaît pas de troisième réalisation. Les habitants de Berrebi ne réalisent pas cette sous variable comme une consonne débucalisée comme c'est le cas pour la sous variable ^G_1 . D'autre part, les localités qui utilisent les fricatives et celles qui produisent les occlusives ne se segmentent pas de la même manière qu'avec la sous variable ^G_1 . Cette fois-ci, les localités de Lahmar et de Boukais utilisent la forme occlusive pour ^K_1 , alors que pour ^G_1 , ils utilisent la forme fricative. Pour les autres localités, les variétés qui utilisent une occlusive pour ^G_1 continuent à faire usage d'une occlusive pour ^K_1 . Le même constat peut être établi pour ceux qui utilisent des fricatives.

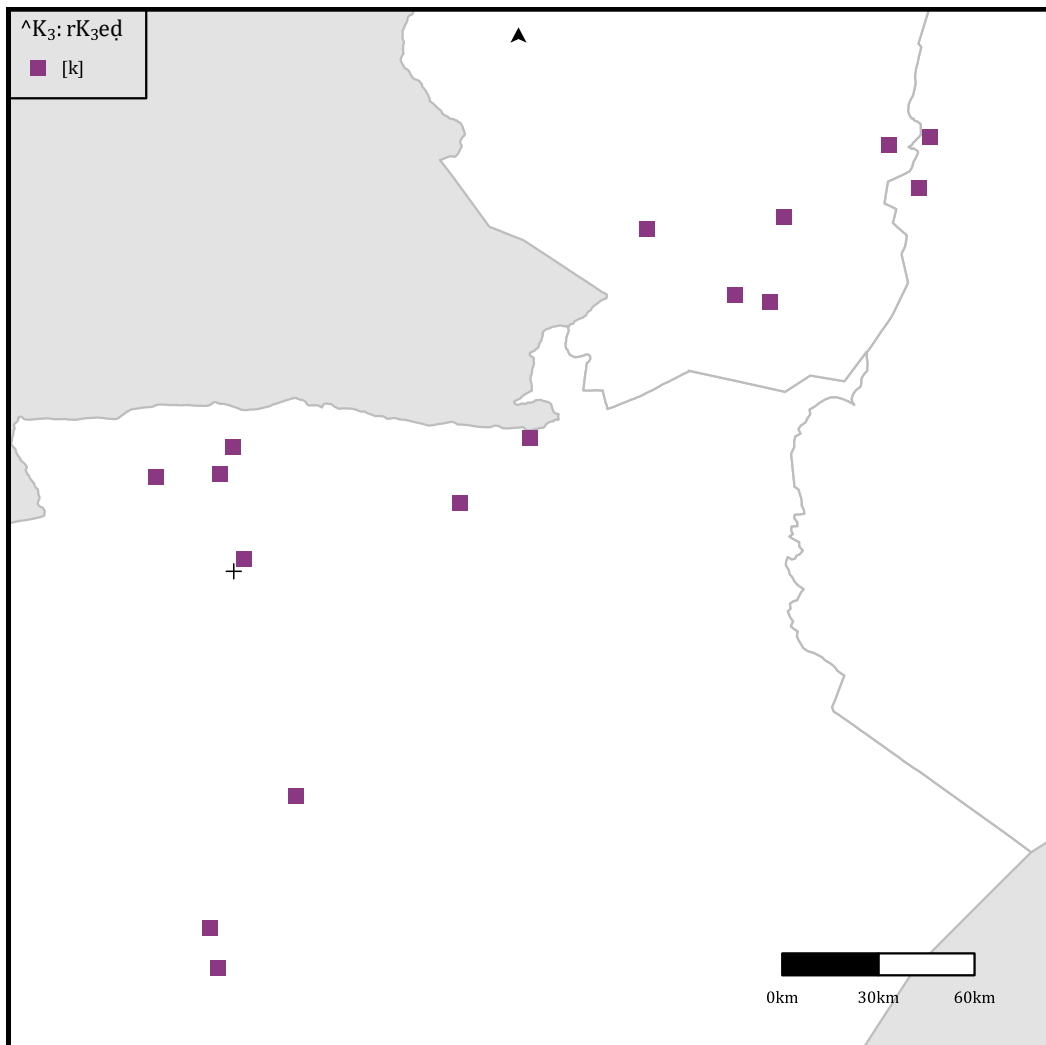
Comme pour la variable \hat{G} , les choses ne se présentent pas aussi régulièrement. Pour cette variable, nous constatons également des irrégularités. La sous variable suivante est dans ce cas.



Carte II.4.6 : La variation de \hat{K}_2 : aK₂sum \viande\

Cette sous variable a uniquement une forme. Toutes les variétés utilisent la forme [j]. Toutefois, les variétés de Mazzer et de Boukais se démarqueront des autres en réalisant cette sous variable comme un [k] dans certains cognats morphémiques. Ainsi, à Mazzer, le pluriel de @a[j]sum@ se dira @[k]esman@. Et pour un autre morphème, à Boukais, l'AI du verbe signifiant \prendre\ se prononcera @[k]essi@, alors que la forme du **PRET** ou de l'**A** se réalisera @isi@ (< *ysi). Tandis qu'à Berrebi, par exemple, pour les mêmes morphèmes utilisés à Mazzer et à Boukais, cela se dira respectivement : @isman@ (<

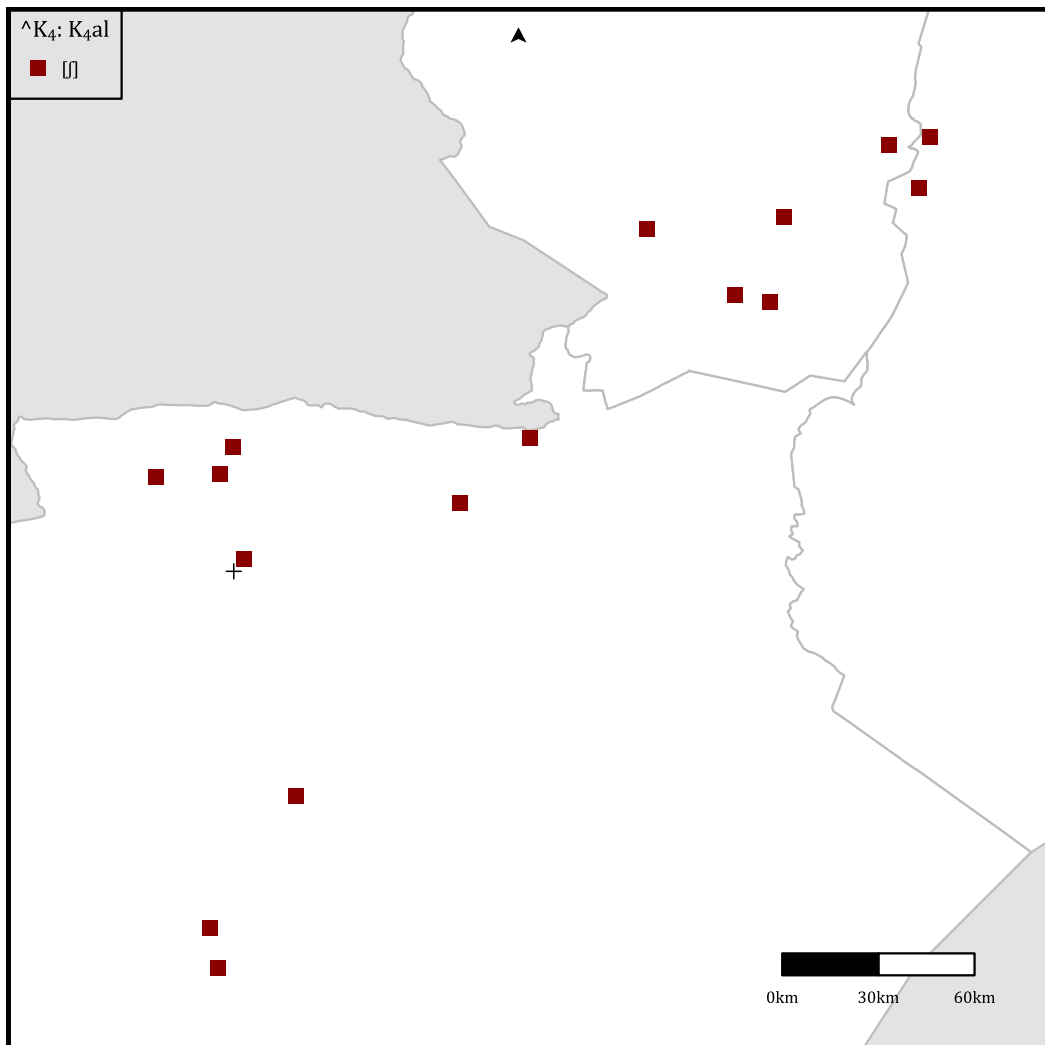
*iysman) et @issi@ (< *yessi). De plus, rappelons que pour la variable $\wedge G$ toutes les formes des sous variables sont des consonnes sonores. Il y a une homogénéisation au niveau du voisement. Cela est également le cas dans la sous variable précédente $\wedge K_1$, toutes les formes sont sourdes. Donc, nous devrions nous attendre à ce que les formes de $\wedge K_2$ soient aussi sourdes. Pourtant, rien de tel n'a lieu. Pour ce fait, l'alignement n'est pas respecté. Au lieu d'une consonne sourde, la sous variable $\wedge K_2$ est réalisée de manière voisée. En outre, elle ne se réalise pas comme une obstruente, mais comme une approximante. Passons à une autre sous variable qui elle aussi a une distribution identique sur tout le territoire du Sud-Oranais.



Carte II.4.7 : La variation de $\wedge K_3 : rK_3eɖ$ \ danser{2.SG.IMP} \

Cependant, cette troisième sous variable illustre une discontinuité entre les sous variables \hat{K}_3 et \hat{K}_1 . Ce déphasage ne concerne pas une grande partie des variétés de la wilaya de Béchar qui, elles, sont dans la continuité avec la sous variable \hat{K}_1 . Ce que nous évoquons concerne uniquement les variétés qui avaient pour forme [ʃ] et qui pour cette sous variable ont une occlusive.

Cette inversion dans le choix du mode d'articulation pour la variable \hat{K} n'est pas propre aux variétés septentrionales. Ce phénomène peut aussi concerner les autres variétés et dans cette situation, ces dernières adopteront la consonne fricative comme nous pouvons le constater ci-dessous.

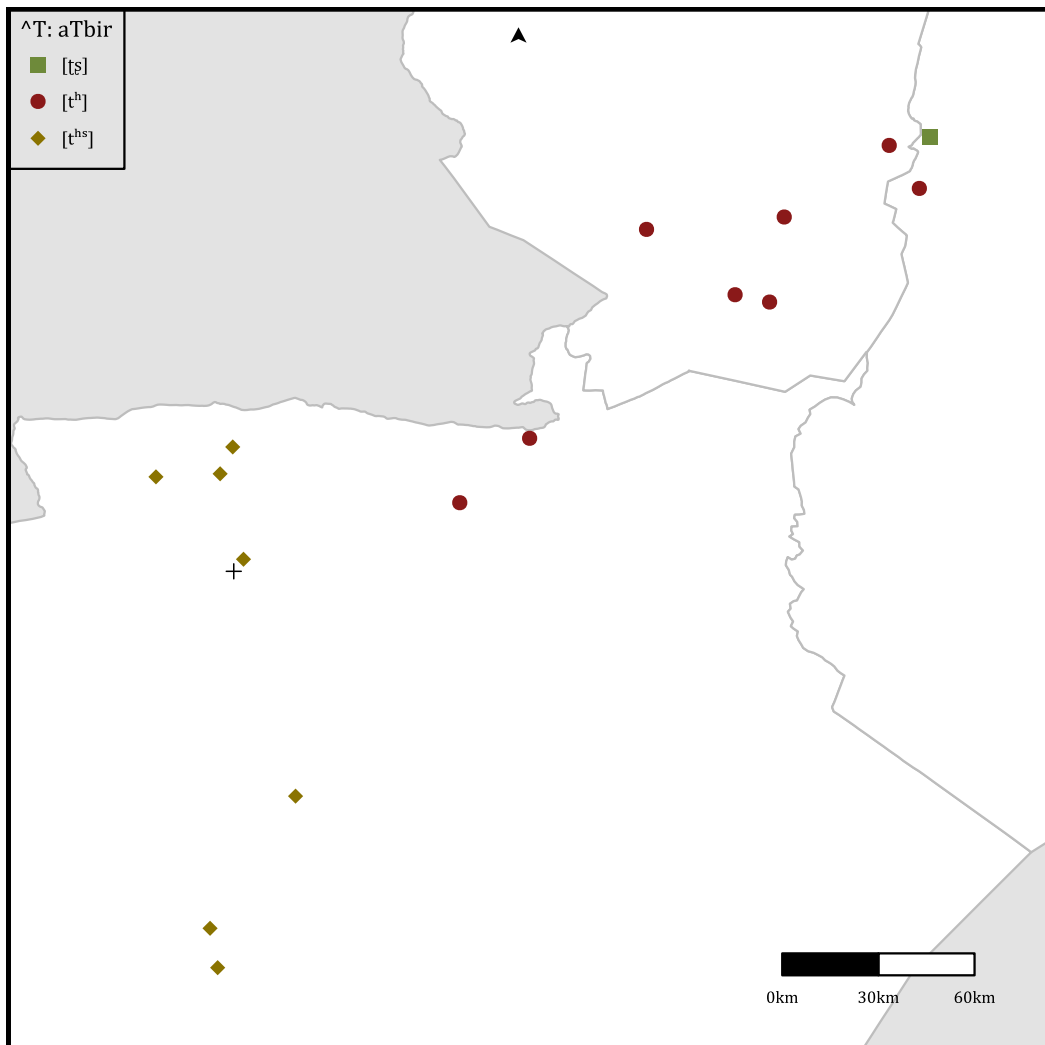


Carte II.4.8 : La variation de $\hat{K}_4 : K_{4al}$ \ terre \

Donc, à l'inverse de la sous variable précédente, celles qui avaient pour forme une fricative pour la variable \hat{K}_1 ont pour la sous variable \hat{K}_4 toujours une fricative. Par contre, cela ne sera nullement le cas pour les autres variétés. Ces dernières adopteront la forme des autres variétés et utiliseront la forme fricative.

II.4.1.3.1.3 La variable \hat{T} : Aspiration et Affrication

Les phénomènes que nous évoquerons touchent les consonnes simples comme les consonnes géminées, mais nous avons tout de même inclu ce point dans cette partie pour des questions d'optimisation rédactionnelle.



Carte II.4.9 : La variation de \hat{T} : aTbir \pigeon\

Comme nous pouvons le voir sur la carte II.4.9, la ^hT connaît trois formes différentes. La première forme [t^s] qui n'est propre qu'à un seul nucléote — celui de Chellala — correspond à une consonne rétroflexe affriquée. C'est la première fois qu'une telle consonne a été inventoriée pour le berbère.

Nous ne sommes pas sûr que cette unité phonique ait cette réalisation articulatoire, toutefois, il est évident que la réalisation phonétique retrouvée à Chellala est différente des réalisations produites dans les autres localités du Sud-Oranais au niveau du point d'articulation. Dans les autres localités, il fait peu de doute que le point d'articulation est le même. D'après l'étude articulatoire de Yeou [Yeou and Honda, 2011], même si elle ne concerne que les géminées, nous pouvons avancer l'hypothèse que le point d'articulation est dento-alvéolaire.

Après cette simple présentation des données, il y a d'autres aspects à relever. Les phénomènes en question peuvent être regroupés sous une même typologie phonétique. Ainsi, sous cette variable ^hT sont regroupées des consonnes coronales sourdes occlusives qui se distinguent par leur désocclusion. Cette dernière est suivie par une unité sonore non voisée, c'est ce point qui fait leur caractéristique commune. Toutefois, cette unité sonore au niveau articulatoire, comme phonologique, n'est pas identique suivant les différents nucléotes. À Chellala, après le relâchement, une sifflante est produite. Pour le cas de cette variété, cette friction est comprise dans le phonème. C'est pour cette raison que nous considérons cette consonne rétroflexe comme une affriquée.

Pour les autres variétés, les choses sont autres. Bien que, nous ayons affirmé que toutes les variétés du Sud-Oranais à l'exception de celle de Chellala avaient le même point d'articulation, elles se distinguent tout de même sur d'autres aspects. Ainsi, au niveau du relâchement de cette occlusive nous constatons une différence. Certaines variétés ne connaissent qu'une aspiration et d'autres qui en plus d'une aspiration présentent une friction dont le point d'articulation est au même niveau que lors de la phase d'occlusion. Cet aspect nous a longtemps fait hésiter sur la nature de cette unité phonique. Était-ce une affriquée [tθ] ou bien une consonne aspirée? L'unité sonore sifflante n'est pas systématiquement perçue. Donc, nous nous interrogeons sur le statut phonologique de cette unité sonore. Mais en comparaison, cela nous a jamais été donné d'entendre cette dernière dans les autres points d'enquête plus au nord. Nous en avons conclu qu'au point de vue articulatoire, il existait bien une différence. Et que les variétés du sud ne présentaient pas la même unité phonique que celles du nord.

Autre point que nous souhaitons soulever concerne le terme aspiration. Cette

aspiration est un phénomène phonétique qui est souvent caractérisée par une unité de mesure qui est le Voice Onset Time (VOT). La définition articulatoire la plus basique que nous pouvons fournir à son sujet est que les plis vocaux restent ouverts après le relâchement de l'occlusion. Et c'est cet aspect qui définit une consonne aspirée d'une autre consonne non aspirée. Toutefois, dans certaines langues du monde ce moment où l'entrée du conduit vocal et les plis vocaux sont ouverts, une constriction a lieu au niveau glottal comme c'est le cas en anglais [Hanson and Stevens, 2003]. Et cette dernière réalisation phonétique est parfaitement perceptible à l'ouïe. Une telle chose n'a pas lieu pour ces consonnes aspirées du Sud-Oranais et qui si elles manifestent une constriction est au niveau dento-alvéolaire.

II.4.1.3.2 Discussion

Après ce tour d'horizon de la variation phonético-phonologique de ces variétés, il apparaît qu'en réalité le système phonétique et phonologique de ces localités est quasiment le même concernant les consonnes simples. La différence entre ces variétés ne se situe pas au niveau de l'inventaire. Mais elles se distinguent par rapport à la distribution de ces unités phoniques au niveau des unités significatives. Donc, la fréquence de réalisations de ces unités phoniques sera également dissemblable. Ainsi, il va de soi que statistiquement les consonnes occlusives de nos variables seront retrouvées beaucoup plus fréquemment à Igli qu'à Asla par exemple. Et inversement, les fricatives seront davantage usitées dans le nord du Sud-Oranais qu'au sud. Toutes ces variétés ont par conséquent des consonnes fricatives et des consonnes occlusives. Les variétés du sud peuvent utiliser les occlusives [g] et [k] comme les fricatives [ʃ] et [ʒ] et pour celles du nord la situation est identique.

En revanche, le choix d'une réalisation fricative peut se porter sur plus d'une consonne fricative comme nous le montre le cas de Berrebi. Ce dernier peut utiliser pour la variable ^G_1 les formes [h] et [j] et pour la variable ^G_2 les formes [ʒ] et [h]. Bien que dans la majorité des variétés du Sud-Oranais les variables aient les mêmes formes et se distinguent de la même façon, à Berrebi, les choses ne sont pas linéaires. Le cas présent n'est attesté que pour la variable ^G , alors que pour la variable ^K , la variation se représente avec régularité. Ainsi, nous pouvons supposer que si les formes adoptées par ^K sont stables d'un point d'enquête à un autre, alors nous pouvons écarter l'hypothèse d'un phénomène de contact. Il n'y a pas de raisons objectives de s'imaginer que pour la variable ^G les habitants de Berrebi auraient eu recours à l'emprunt et se sauraient gardé

de le faire pour la variable \hat{K} . Par conséquent, il faut peut-être envisager de segmenter les sous variables \hat{G}_1 et \hat{G}_2 en deux autres sous variables représentant des cognats phonologiques différents. Mais comme à notre connaissance ce fait n'a jamais été recensé autre part dans le monde berbère. Alors, il peut paraître comme précipité d'envisager cette hypothèse plus sérieusement. D'autant plus que les exemples présentant cette distinction figurent seulement dans les exemples fournis. Nous n'avons pas dans notre corpus d'autres unités significatives ayant les mêmes configurations. Il faudrait d'autres faits pour envisager cette idée. C'est ce que nous allons voir maintenant.

Donc, à cela se rajoute aussi un autre cas que nous n'avons pas encore mentionné et qui rejoint ce que nous retrouvons à Berrebi. La sous variable \hat{G}_2 peut aussi dans d'autres localités avoir plusieurs formes. Nous ne rencontrons ce fait qu'au sein d'une unité significative. Le singulier de cette unité significative aura pour forme [ʒ] dans toutes les localités du Sud-Oranais. Alors que le pluriel de cette unité significative aura pour forme [g], [i] (< *y) suivant les variétés⁹. Cette unité significative en question est @am[ʒ]er@ \faucille\ et @im[i]ran@ (< *imeyran) ou @ime[g]ran@ \faucilles\ — ces dernières réalisations correspondent à la distribution géographique de la sous variable \hat{G}_1 —. Donc, le singulier de cette unité significative correspond à la sous variable \hat{G}_2 et le pluriel à la sous variable \hat{G}_1 . Cet exemple est très parlant dans la mesure où ces deux morphèmes sont étymologiquement liés et partagent le même radical. Mais qui, toutefois, vont connaître une alternance au niveau de l'une des unités phoniques de leur racine. Ainsi, cela soulève des interrogations puisque nous avons une unité morphémique qui a deux formes. Quel est le phonème sous-jacent aux unités phoniques de cette unité significative et plus largement des formes retrouvées pour la sous variable \hat{G}_2 ? Il n'est pas aisé de répondre à cette question, car il ne semble pas qu'il y ait de contextes particuliers. Puisque, l'environnement phonétique est le même.

Par conséquent, en additionnant la variation signalée pour Berrebi et ce que nous venons d'évoquer, ces données compliquent davantage l'analyse en introduisant une dose de confusion. En définitive, la seule manière de sortir de cette complication est de considérer que les sous variables \hat{G}_1 et \hat{G}_2 comprennent d'autres sous variables. Ainsi, les cognats phonologiques représentés par ces sous variables seraient au nombre de cinq. Si tel est le cas, il y aurait une sous variable \hat{G}_{1a} qui aurait pour forme suivant les variétés [g], [j] et [h]. Et une sous variable $\hat{G}_{1\beta}$ qui aurait les formes [g] pour Iqli et Mazzer et [j] pour les autres localités. Puis, s'agissant de la sous variable \hat{G}_2 , celle-ci

9. Pour cette unité significative, nous n'avons pas de données pour Berrebi.

se diviserait entre la sous variable \hat{G}_{2a} , comprenant les formes [h] pour Berrebi et [ʒ] pour les autres localités, et la sous variable $\hat{G}_{2\beta}$ qui aurait comme unique forme [ʒ]. Toutefois, pour cette dernière, il y a lieu de distinguer deux autres sous variables qui ont les mêmes formes de surfaces [ʒ], mais qui se distinguent au niveau des représentations $\hat{G}_{2\beta\phi}$ et $\hat{G}_{2\beta\gamma}$. Cette configuration permet de pouvoir poser comme hypothèse que la forme de $\hat{G}_{2\beta\phi}$ — représentée par @am[ʒ]er@ — est un allophone¹⁰ de \hat{G}_{1a} ou $\hat{G}_{1\beta}$ — représenté en fonction des variétés par @im[i]ran@ (< *imeyran) ou @ime[g]ran@ —. En considérant cette hypothèse comme probable, il reste la question de savoir quel est le phénomène à l'origine de cette distinction au niveau des surfaces de cette unité significative. À l'heure actuelle, nous sommes dans l'incapacité de proposer une hypothèse solide à ce problème au vu du nombre réduit de cas présentant cette variation.

De plus, une autre difficulté surgit comment déterminer à quelles sous variables appartiennent les formes [ʒ] et [j]? À savoir, est-ce que [ʒ] correspond à $\hat{G}_{2\beta\phi}$ ou à $\hat{G}_{2\beta\gamma}$? Et [j] coïncide-t-il avec la sous variable \hat{G}_{1a} ou $\hat{G}_{1\beta}$? Car, effectivement, à défaut d'avoir des cognats morphémiques, donc d'éléments qui permettent de faire une comparaison, il est impossible de savoir à quelles sous variables ils dépendent. Et *in fine* à quel phonème il faut les rattacher. Alors, il y a donc dans ce cas indétermination. Ainsi, si nous retrouvons un cognat morphémique présent uniquement à Boussemghoun et à Igli, mais pas à Berrebi, il apparaît difficile de trancher la situation. Cette lacune n'est pas exceptionnelle, elle tend à être la norme. Du fait de l'arabisation, il s'avère que les cognats morphémiques se rarifient. Car les morphèmes étymologiquement berbères sont remplacés par des unités significatives d'origine arabe. Donc, dans une telle situation pour ne pas avancer d'hypothèses extrapolées, il est nécessaire d'envisager un archiphonème pour les formes de ces sous variables.

Poursuivons l'analyse de ces matériaux en évoquant la question de la contextualité et de son corrolaire la distribution. Au sein des sous variables des variables \hat{G} et \hat{K} , n'y aurait-il pas certaines de ces variables qui ne seraient présentes que dans un contexte restreint. Si nous allons dans ce sens, alors nous observons que certaines sous variables ont une distribution limitée. Toutes les sous variables autres que \hat{G}_1 et \hat{K}_1 sont dans l'environnement de coronales. De surcroît, la position à l'intérieur a aussi son importance. Ainsi, si nous portons une attention plus accrue à la sous variable \hat{K}_2 , celle-ci n'apparaît qu'à l'initial d'un morphème. De plus, elle est systématiquement suivie par une consonne fricative alvéolaire. L'autre sous variable qui semble être dans le même cas est la sous

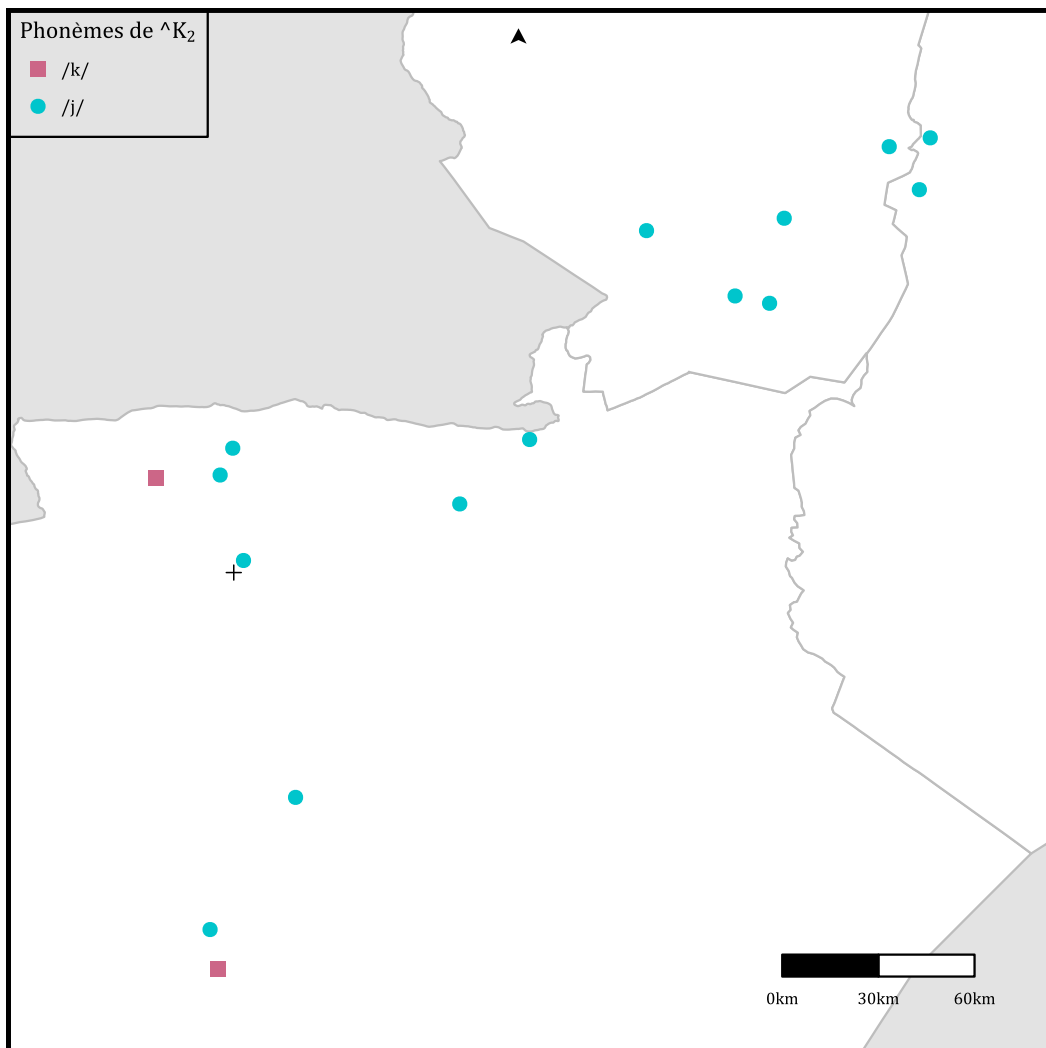
10. N'ayant pas d'exemple pour Berrebi, nous ne pouvons trancher.

variable \hat{G}_4 . Nous constatons que cette dernière est présente dans tous les cas après une consonne fricative alvéolaire. Elle figure par ailleurs toujours dans la première syllabe d'un morphème. Cependant, soulignons que nos observations sont fondées sur peu de cognats morphémiques. Donc, il est possible que ces données soient dues au hasard. Mais pour le moment retenons cette hypothèse comme étant probable.

Donc, en recoupant les distributions de ces deux sous variables, nous nous apercevons qu'elles apparaissent dans un même contexte. Toutes les deux sont réalisées à proximité d'une fricative alvéolaire et au début d'un morphème. Au vu de ces observations, une question se pose, est-ce qu'au niveau phonologique n'y a-t-il pas lieu de considérer la forme [j] de ces sous variables comme provenant d'un phonème qui ne serait pas la semi-voyelle /j/? Cette question se pose sérieusement, d'autant plus, comme nous l'avons vu à Mazzer, le pluriel de @a[j]sum@ se réalise avec un [k], de même à Boukais, mais pour une autre unité significative. Donc, nous pouvons envisager, tout du moins pour le cas de Mazzer et de Boukais, que la forme [j] de la sous variable \hat{K}_2 est la forme de surface d'un /k/ ¹¹. Ce qui implique une neutralisation de ce phonème dans les contextes susmentionnés. Pour les autres localités, nous ne pouvons que postuler que la forme de surface est identique à la forme sous-jacente. Puisqu'au plan synchronique, il serait extrapolé d'essayer de plaquer la réalité de Mazzer et de Boukais sur les autres points d'enquête. Ainsi, l'étape suivante de cette réflexion est d'envisager pour les cas de Mazzer et Boukais que ces deux variétés se distinguent de leurs consœurs au niveau phonologique.

Toutefois, un autre point épineux persiste, il n'y a pas de correspondances entre les faits de Mazzer et de Boukais. L'un dira @i[k]esman@ — Mazzer — et l'autre @[i]sman@ — Boukais — et les mêmes diront @[i]ssi@ — Mazzer — et @[k]essi@ — Boukais —. La question se pose pourquoi cette disjonction? La réponse n'est pas simple. La solution la plus simple serait de dire que nous avons deux cognats phonologiques différents comme nous l'avons avancé pour les données précédentes. Ainsi, il y aurait la sous variable \hat{K}_{2a} représentant @a[j]sum@ et la sous variable $\hat{K}_{2\beta}$ représentant @[i]si@.

11. Nous considérons [j] comme étant l'allophone au lieu de [k], c'est parce que nous postulons que [k] est l'unité phonique étymologique.

Carte II.4.10 : La représentation phonologique de \hat{K}_2

Par conséquent, nous avons repris la carte linguistique de la sous variable \hat{K}_2 pour montrer à quoi elle correspond au niveau phonologique. Nous avons fusionné sur une même carte les deux sous variables \hat{K}_{2a} et $\hat{K}_{2\beta}$, mais il ne faut pas considérer pour autant que leurs représentations phonologiques sont identiques.

Nous pourrions dire de même pour la sous variable \hat{G}_4 . Qu'elle est probablement un allophone. Néanmoins, dire de quel phonème tient la forme de cette sous variable est difficile à déterminer. Ce cas ne se prête pas à une analyse aisée. À l'inverse du cas précédent, le pluriel de ces morphèmes ne se réalise pas avec une forme différente et se dira @iz[i]zawen@ (< *izeyzawen). Nous serions tenté de rapprocher cette sous variable de la sous variable \hat{G}_3 . Par déduction, c'est la seule sous variable qui partage la même

distribution géographique. Toutefois, s'agissant des représentations, aucune indication ne laisse à penser que la forme de $\wedge G_4$ pourrait correspondre à /g/. Probablement au niveau diachronique la situation devait être différente, mais, dorénavant, la forme de $\wedge G_4$ n'est pas déphasée par rapport à sa représentation. Mais pouvons-nous considérer que les formes — [j] ou [i] — de cette sous variable partagent un même phonème? Si nous considérons qu'elles n'ont pas le même phonème, alors il faut envisager de segmenter cette sous variable en deux autres sous variables. Cette dernière solution nous paraît la plus raisonnable. D'autant plus que rien ne laisse à penser que dans la morphologie des unités significatives, où figure la forme [i], cette dernière pourrait être une forme de surface de /j/ ou /g/ comme c'est le cas pour la sous variable $\wedge G_{2\beta\varphi}$. En conclusion, $\wedge G_4$ doit être divisée en deux autres sous variables $\wedge G_{4a}$ et $\wedge G_{4\beta}$. La première correspond à la forme [j] et la seconde à [i] et leurs phonèmes sont respectivement /j/ et /i/.

Cependant, tout ce que nous venons de dire reste de l'ordre de l'hypothèse au regard du nombre réduit de données. Cette analyse phonologique portant sur les sous variables $\wedge G_2$, $\wedge G_4$ et $\wedge K_2$ a tenté de démontrer que derrière cette distribution géographique uniforme, il était possible de rendre compte d'un phénomène de neutralisation. Et qu'il fallait, par conséquent, envisager de considérer certaines formes de ces sous variables comme des allophones de phonèmes. En outre, les formes de ces phonèmes correspondraient à d'autres sous variables. Cela n'a été possible que par la méthode comparative. En étudiant seulement le système phonético-phonologique interne de ces variétés berbères, il aurait été impossible de nous rendre compte que les unités phoniques de certains morphèmes étaient des allophones. Mais tout ce que nous venons de dire n'a de poids qu'au regard de la position théorique adoptée.

Cette dernière postule comme assertion l'unicité morphémique. Cette hypothèse établit qu'au niveau des représentations il n'y a qu'une seule unité morphémique, ce qui implique que toute forme différente au niveau du morphème est causée par d'autres facteurs — phonétique ou phonologique —. Ces allomorphes seraient donc tous à rattacher à une seule unité morphémique. Pour ceux qui ne seraient pas familiers avec la morphologie du berbère les mots @a[j]sum@ et @i[k]esman@ partagent un même morphème dénommé dans les études berbères : racine. Cette dernière au niveau de la surface à deux formes ![ksm]! et ![j]sm]!. Mais dans le cadre de l'hypothèse d'une unicité morphémique, il est impossible de considérer qu'au niveau sous-jacent, il existe une unité morphémique ![ksm]! et une unité morphémique ![j]sm]!. Il y a nécessité d'unifier les deux formes de surface en une représentation morphémique. C'est la démarche que

nous avons entreprise dans les analyses précédentes.

Toutes les études, à notre connaissance, sur les représentations phonologiques ou morphémiques partent du principe que la pluralité est au niveau de la surface et l'unicité est au niveau des représentations. C'est le constat que nous faisons de nos lectures des différents cadres théoriques qui nous ont été données de lire — depuis la Generative phonology (SPE) à la Théorie de l'Optimalité¹² —. Ainsi, en caricaturant, tout le jeu consiste à déterminer comment unifier ces formes de surface en une seule représentation sous-jacente. Toutefois, ce point de départ *a priori* est infondé et n'est qu'une hypothèse théorique. Les données que nous venons de discuter tendent à remettre en question ce principe d'unicité morphémique. Car il y a un dilemme qui s'installe.

En voulant unifier au niveau morphémique, nous multiplions les cas particuliers qui ne peuvent s'expliquer, tous, par une seule règle phonologique et encore moins par des phénomènes phonétiques. Dans ce cas, il ne serait pas vain d'envisager de questionner ce principe. D'après nous, il ne serait pas abscons d'avoir à l'esprit que l'unicité morphémique n'est pas inviolable. Car, comment expliquer qu'à Boukais, le pluriel de @a[j]sum@ ne se réalise pas @[k]esman@ — comme à Mazzer —, alors que la forme de l'AI de @[i]si@ se réalise bien @[k]essi@. S'il y avait une règle qui dit que [i]/[j] dans un certain contexte — disons après un schwa — doit se réaliser [k], alors pourquoi cette règle ne s'est pas étendue pour @a[j]sum@. Cela suppose inévitablement qu'il n'y a pas d'unicité morphémique^{13 14}, sinon il y aurait eu généralisation de cette particularité à toutes les unités morphologiques. Donc, si nous adoptons ce raisonnement cela peut expliquer l'évolution séparée et dissymétrique qu'a connu Mazzer et Boukais. Et qu'en outre, [j] n'est probablement pas un allophone de /k/ comme nous l'avions affirmé.

Cette précédente réflexion est fondée sur l'idée que toutes ces variétés partageaient dans un temps passé les mêmes représentations et que le temps faisant celles-ci ont évolué diversement pour aboutir à nos données actuelles. Ainsi, à T_{passé}, toutes les variétés disaient ![KSM]!, puis, à T_{présent}, certaines ont adopté : ![JSM]! ou ![KSM]!/![JSM]!. Donc, cette hypothèse peut être battue en brèche en supposant que les formes que nous

12. C'est la raison pour laquelle nous n'avons pas réalisé un état de l'art sur ces théories puisqu'elles partagent le même socle théorique. Nos réflexions n'allant pas au delà de ce questionnement initial, il était vain d'approfondir ce point.

13. L'unicité morphémique englobe le phonème et toute unité qui dépasse le segment du phonème : syllabe, unité morphémique, ...

14. Cette analyse peut s'étendre également au cas de @am[ʒ]er@ - @im[i]ran@/@ime[g]ran@. Cela impliquerait qu'il n'y aurait pas lieu de considérer [ʒ] comme un allophone.

avons sont le fait d'un processus dérivatif et innovateur ¹⁵ et non pas archaïque — cette nouvelle hypothèse permet toujours de conserver le principe phylogénétique —. Mais cette hypothèse apparaît comme très peu probable en raison du fait que ce phénomène n'est pas régulier. Un processus appliqué à seulement un cas est une supposition qui est difficile à approuver. Cette problématique ne peut s'expliquer par une innovation.

De plus, si nous élargissons notre objet d'étude à d'autres unités phoniques et notre terrain d'étude à d'autres variétés berbères, nous nous rendrions compte d'un niveau de complexité beaucoup plus élevé. Du reste, si nous jetons un œil plus attentif à ces données. Nous constaterons que l'opposition entre ces unités phoniques ne se fait pas au hasard. Elle apparaît quand le morphème appartient à des réalités grammaticales différentes. Quoi qu'il en soit, nous ne développerons pas davantage cette problématique. D'autres données sont nécessaires pour formuler un avis plus tranché sur la question. Ci-dessous un tableau qui récapitule nos propos précédents ¹⁶.

Variables	^G								^K							
	^G ₁		^G ₂		^G ₃		^G ₄		^K ₁	^K ₂	^K ₃	^K ₄				
Sous variables	^G _{1a}		^G _{1β}		^G _{2a}		^G _{2βφ}		^G _{2βγ}		^G ₃		^G _{4a}		^G _{4β}	
Nouvelles Sous variables	^G _{1a}		^G _{1β}		^G _{2a}		^G _{2βφ}		^G _{2βγ}		^G ₃		^G _{4a}		^G _{4β}	
Unités phoniques	[g], [j], [h]	[g], [j]	[h], [ʒ]	[ʒ]-[g], [ʒ]-[j]	[ʒ]	[g]	[j]	[i]	[k], [ʃ]	[j]-[k], [j]	[j]-[k], [j]	[k]	[ʃ]			
Phonèmes	/g/, /j/, /h/	/g/, /j/	/h/, /ʒ/	/g/, /j/, /h/ ou /g/, /j/	/ʒ/	[g]	/j/	/i/	/k/, /ʃ/	/k/, /j/	/k/, /j/	/k/	/ʃ/			

TABLE II.4.2 : Les représentations phonologiques des variables

Par ailleurs, cette variation linguistique que nous avons collectée met en évidence un autre aspect. Il n'y a pas de rapport structurel dépendant au niveau du mode d'articulation entre les consonnes voisées et non voisées. Lahmar, Boukais et Berrebi se démarquent en ayant une occlusive pour la consonne non voisée et une fricative pour la consonne voisée. Les données obtenues dans les autres localités pourraient faire croire que la direction de l'obstruction fonctionne de paire. Pour deux unités phoniques proches, soit un système linguistique les prononcera de manière fricative, soit elles auront une forme occlusive. Mais comme nous venons de le voir cette hypothèse est à écarter. Car, effectivement, il n'y a pas d'automatisme à la frication ou à l'occlusion. Un système linguistique peut se complaire des deux modes d'articulation.

15. Une hypothèse qui pourrait être formulée est de considérer que nous avons à faire à une dérivation au sens linguistique du terme. Ainsi, certains pluriels et certaines formes de l'AI se contruisent en ayant une alternance consonantique.

16. Dans le tableau II.4.2, le tiret entre les formes indique les unités phoniques possibles que peut adopter un phonème d'une unité morphémique.

S'il obstrue partiellement ou totalement certaines de ces unités phoniques, il n'y a pas de contrainte à ce qu'il choisisse d'obstruer totalement l'un et pas l'autre. Toutefois, il est évident également qu'il y a une tendance. Les consonnes perçues comme articulairement proche évoluent dans le même sens. Nous retrouvons cette tendance pour l'ensemble des unités phoniques de ces variétés. Cependant, il y a des exceptions. L'une d'elle est partagée par tous les nucléolectes, c'est celle correspondant au cas des labiales. Dans cette dernière zone d'articulation, la consonne voisée sera occlusive, alors que la consonne non voisée sera fricative. Ainsi, même en tenant en compte de ce dernier exemple et du contre exemple de Lahmar, Boukais et Berrebi pour la sous variable $\wedge K_1$, cela reste limité à ces seuls cas particuliers. Le reste de leur système phonético-phonologique fonctionne de paire. Cela ouvre des questions sur les raisons de cette disjonction au niveau seulement de ces variables. Mais aussi des interrogations sur ce tandem au niveau du mode d'articulation pour des consonnes proches sur un plan articulaire.

Cette structuration en paire n'est pas questionnable uniquement sur le plan du mode d'articulation, nous pouvons aussi pointer cette spécificité au niveau du point d'articulation. Les variétés qui utilisent des occlusives pour nos variables ont comme caractéristique de partager le même point d'articulation. Il n'y a qu'une différence de voisement qui les distingue. Alors que les variétés qui n'ont pas d'occlusives pour nos variables présentent des formes qui ont un point d'articulation différent. Les contreparties des consonnes [k] et [g] dans les autres localités sont [ʃ] et [j] et non pas [ʃ] et [ʒ]. Si nous devons tracer une ligne qui démarquerait les points d'articulation de ces consonnes, nous observerions que ces consonnes fricatives sont positionnées du même côté. Donc, il y a un mouvement articulaire qui est en avant de la zone vélaire. Ces deux consonnes sont positionnées dans la même zone coronale postérieure, bien qu'elles aient des points d'articulation différents. Ce cas est à souligner pour mettre en exergue tout de même le fait, même si elles n'ont pas le même point d'articulation, que l'idée de former une paire semble ne pas s'être évaporée totalement. Cette idée de paire reste également pertinente pour les variétés de Lahmar et Boukais, puisque les points d'articulation sont assez proches entre [k] et [j]. Seul Berrebi a fait le choix de contrevenir à cette tendance en ayant des réalisations totalement opposées [k] et [h].

Autre particularité qui peut s'avérer pertinente, c'est le fait que les consonnes non voisées ont une proportion à être plus marquées que les consonnes voisées. Encore une fois, nous constatons que la consonne non voisée se démarque de la consonne voisée en

étant celle qui marque une distance physique plus importante entre la consonne fricative [ʃ] et sa contrepartie occlusive [k]. Pour les consonnes voisées, il n'y a pour ainsi dire qu'un "cran articulatoire" qui sépare les variétés fricatives [j] des variétés occlusives [g]. Cette observation peut aussi s'étendre aux consonnes dento-alvéolaires (voir Partie II.4.9, page 107) qui auront une réalisation qui ne s'apparente pas à la consonne voisée [d].

Nous ne pouvons pas finir cette discussion sans parler de manière plus approfondie de la question de la distribution géographique des formes des sous variables. Pour les sous variables présentant des formes différentes, nous constatons que la distribution de ces formes ne se fait pas aléatoirement. Pour chaque réalisation, il y a un continuum géographique évident. À aucun moment, il existe une forme de ces sous variables qui serait usitée dans des zones géographiques éloignées l'une de l'autre et séparée par d'autres formes différentes. L'évolution se fait de proche en proche.

Nous pouvons aussi nous rendre compte à quel point les différentes localités berbérophones sont espacées l'une de l'autre. Comment la géographie et le climat ont façonné l'occupation de ce territoire par les berbérophones et en conséquence ont dessiné aussi la distribution linguistique. Car, bien évidemment, le lien entre l'espace et la langue n'est pas direct.

D'autre part, nous constatons une extension plus importante des variétés à tendance fricative. Mais avec une division territoriale qui n'est pas identique suivant les variables. Bien que la répartition des faits linguistiques se fasse entre plusieurs localités, nous observons tout de même des isolats. Certaines variétés linguistiques se distinguent de toutes les autres. Ainsi, des faits linguistiques ne sont retrouvés qu'au niveau d'un point d'enquête comme à Chellala et à Berrebi. Ce qui nous amène à formuler le constat que l'homogénéisation linguistique a été en quelque sorte endiguée. Donc, nous nous interrogeons sur les raisons de ce particularisme. Est-ce que cet isolat linguistique est la conséquence de leur isolat géographique? Si c'est le cas, nous pouvons en déduire que ces différentes localités ont conservé une certaine indépendance entre elles. C'est ce qui expliquerait qu'ils aient pu préserver ou développer leurs particularités linguistiques.

De manière générale, ces cartes linguistiques mettent en évidence que les variétés se répartissent en groupes de deux ou trois. Cette fluctuation dépend de la variable linguistique. Cette situation est vraie pour les variables que nous venons d'étudier, cependant, est-ce le cas avec d'autres variables linguistiques? C'est ce que nous allons voir dans la partie suivante. Donc, dans le prochain point que nous allons étudier, la question de la frication versus occlusion ne sera plus l'axe central de nos discussions par

la suite. Nous aborderons uniquement la question de la gémiation en ayant toujours la même problématique en ligne de mire, c'est-à-dire le rapport entre phonologie et phonétique. Cette fois-ci, l'articulation des idées se fera principalement entre affriquées et fricatives et la question du voisement sera aussi abordée.

II.4.1.4 Consonnes géménées

Cette partie sera dédiée à l'étude de la variation de variables qui ont comme forme des consonnes géménées¹⁷. Dans un premier temps, nous déroulerons les données qui seront soumises à analyse. Puis, nous les discuterons. Ainsi, nous avons déterminé six variables. Cinq d'entre elles comprennent des sous variables : $^{\wedge}CC_i$, $^{\wedge}GG_i$, $^{\wedge}SS_i$, $^{\wedge}ZZ_i$, $^{\wedge}QQ_i$.

Les quatre premières variables partagent un certain nombre de propriétés communes. Celles-ci sont relatives aux caractéristiques qui les différencient. Les oppositions qui les distinguent se situent par rapport à leur point d'articulation, mais surtout au niveau de leur mode d'articulation. Ces différences de constriction sont le marqueur de la variation linguistique. Quant aux deux autres variables, elles illustrent une opposition de voisement entre les variétés.

Les consonnes géménées étudiées dans le cadre de cette partie se rapportent uniquement à des consonnes tautomorphémiques. De plus, ces géménées peuvent se rapporter soit à des géménées lexicales ou des géménées morphologiques. Nous signalerons pour chaque variable étudiée à quel type de géménées doit être rattaché ces variables.

II.4.1.4.1 Le matériau d'analyse

Toutes les variables que nous allons voir présentent des formes différentes suivant les points d'enquête. Ces dernières sont censées définir des cognats phonologiques. De plus, comme dans la partie précédente, ces sous variables rentrent en contraste. De surcroit, la distribution géographique et les réalisations phonétiques varient d'une sous variable à une autre.

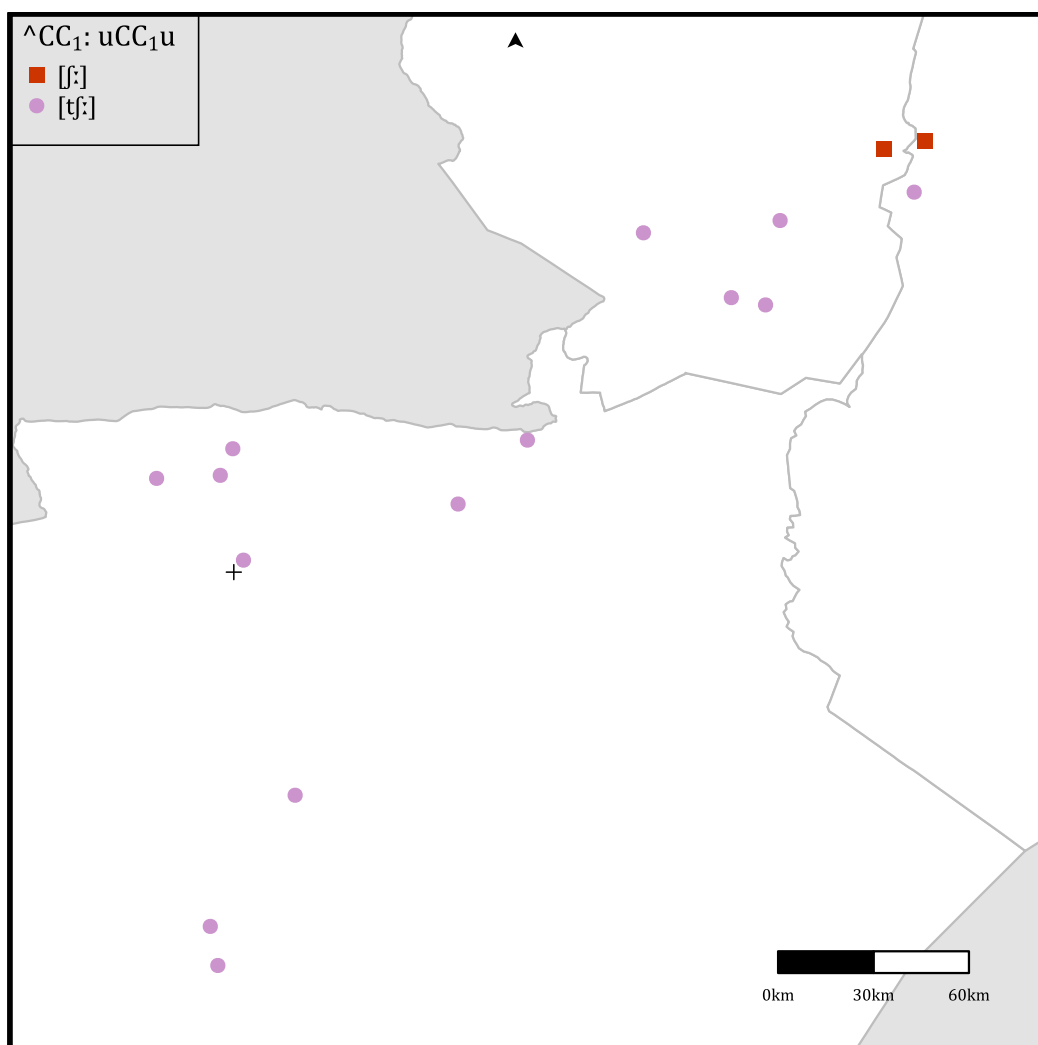
Ainsi, ces variétés du Sud-Oranais connaissent toutes une opposition entre consonnes affriquées et non affriquées. Puis, le mode d'articulation n'est pas le seul élément

17. Au sujet de cette terminologie, certains chercheurs ont débattu sur la dénomination adéquate à adopter pour caractériser cette unité sonore. Certains défendaient le terme de tension, alors que d'autres utilisaient le mot gémiation [Ridouane, 2003]. Pour notre part, nous nous rangeons à l'avis de Ridouane et d'autres. Donc, nous emploierons par la suite le terme de géminée pour dénommer cette unité phonique.

contrastif, il y a également le mode de voisement. Certaines variables représentent des consonnes voisées et d'autres non voisées.

II.4.1.4.1.1 La variable $^{\wedge}CC_i$: Frication et Affrication

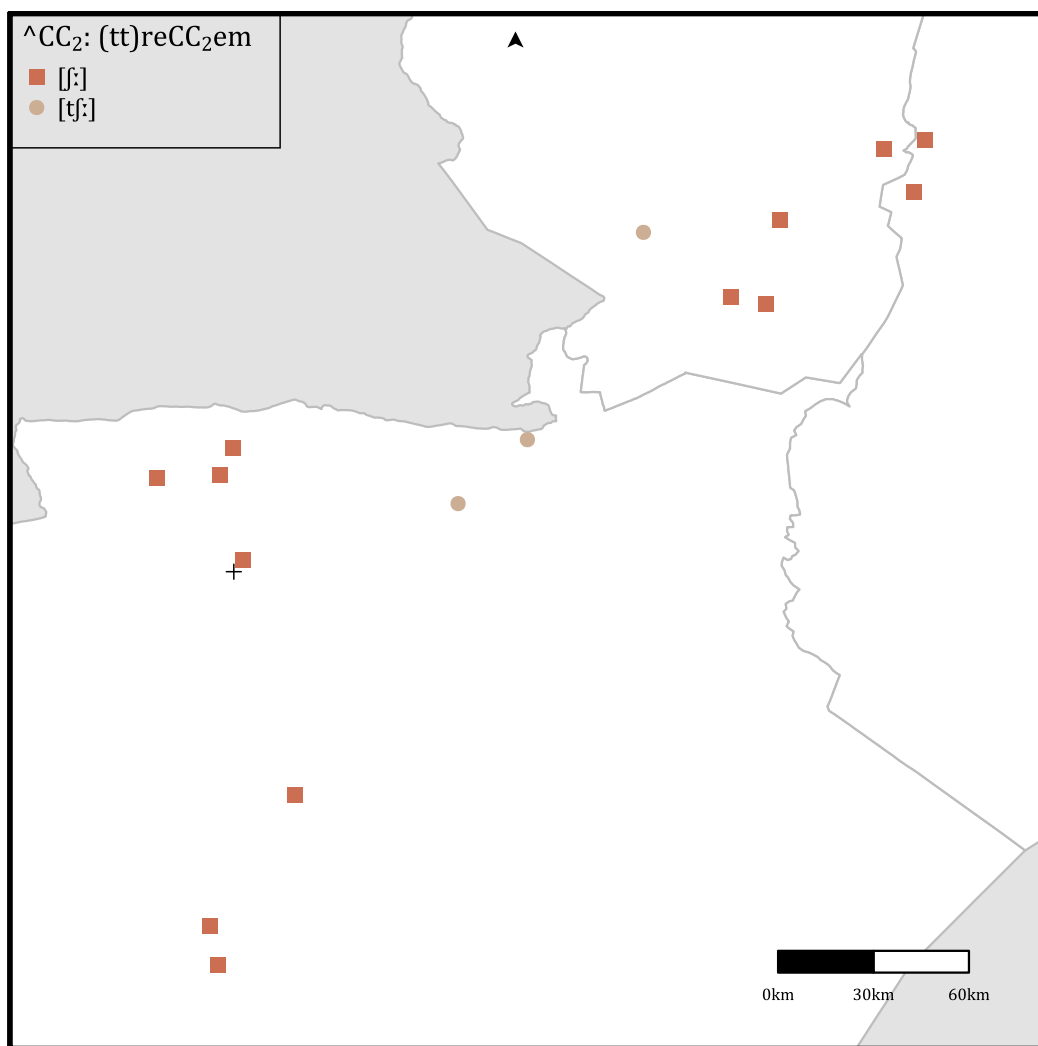
La première sous variable que nous allons voir a deux formes. Ces dernières correspondent à une consonne géminée affriquée alvéolaire et à une consonne géminée fricative alvéolaire. La forme affriquée est celle qui est la plus répandue.



Carte II.4.11 : La variation de $^{\wedge}CC_1$: uCC₁um \couscous\

Sur nos seize points d'enquête, ce sont les variétés d'Asla et de Chellala qui se démarquent. Ces derniers sont les seuls à avoir une géminée fricative pour cette sous variable. De plus, la géminée qui exemplifie cette sous variable est une géminée lexicale.

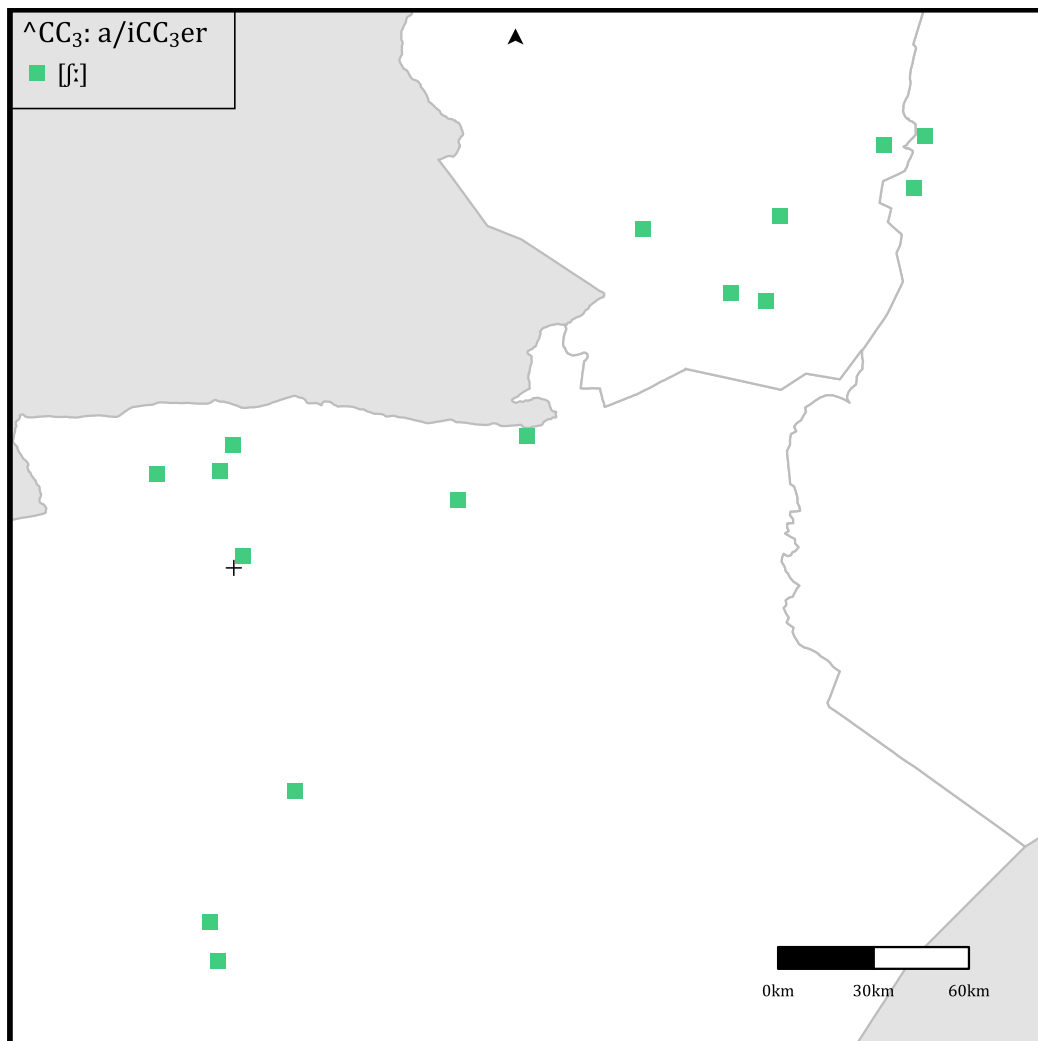
Mais cette dernière configuration est bouleversée lorsque nous avons à faire à une géminée morphologique. Ainsi, Asla et Chellala sont toujours dans la continuité et prononcent une fricative. Il en est de même pour Fendi, Beni Ounif et Sfisifa qui ont toujours pour forme une affriquée. Alors que les autres localités rejoignent la forme utilisée par Asla et Chellala.



Carte II.4.12 : La variation de $\text{^CC}_2:(tt)re\text{CC}_2em$ \ mouiller{2.SG.IMP_INT}

Cette fois-ci, la distribution est largement au profit des variétés fricatives et la forme affriquée est la forme minoritaire. Faisons remarquer que les variétés avec affrication tranchent en deux la distribution des variétés fricatives.

Avec la sous variable suivante, il y a une totale homogénéisation. Toutes les variétés utilisent la forme fricative. Même les variétés qui continuaient à utiliser une affriquée pour la sous variable $\wedge CC_2$ réalisent une fricative pour la sous variable $\wedge CC_3$.

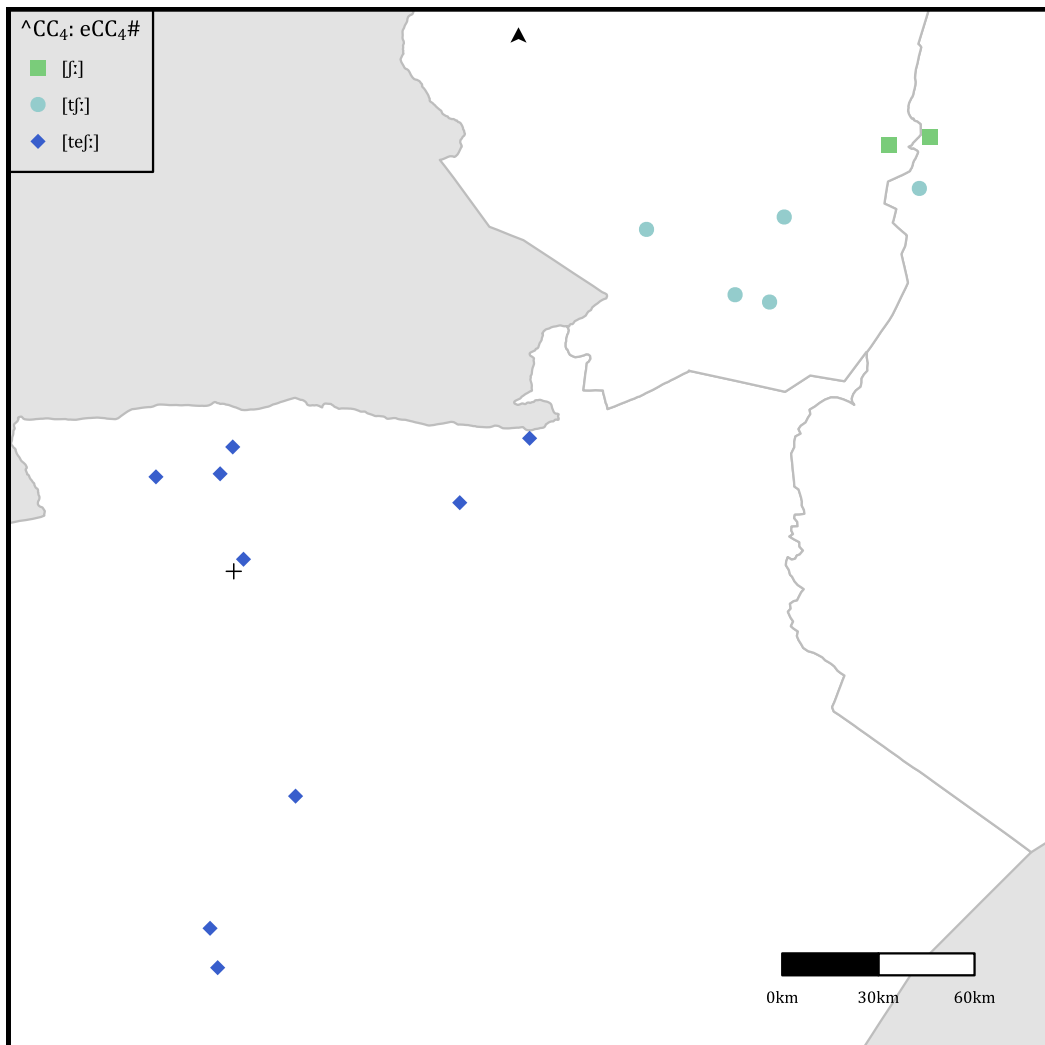


Carte II.4.13 : La variation de $\wedge CC_3 : i/aCC_3er$ \ongle\

Nous notons au passage également une variation à l'échelle de la voyelle initiale. Certains produiront ce morphème avec une voyelle fermée ||i|| et d'autres avec une voyelle ouverte ||a||.

D'autre part, cette sous variable est incarnée par une gémignée lexicale comme pour la sous variable $\wedge CC_1$. C'est aussi ce type de gémignée que nous retrouvons pour la sous variable suivante. Mais les formes que nous avons collectées se démarquent de $\wedge CC_3$.

Jusqu'ici les formes des sous variables de $\wedge CC_i$ présentaient deux formes possibles. Avec la sous variable suivante, nous pouvons constater que trois formes sont possibles. Nous retrouvons toujours une fricative et une affriquée, toutefois, la troisième forme se présente plus comme une chaîne d'unités phoniques. Ces formes ont lieu dans un contexte bien particulier. Cela se déroule uniquement lorsque cette sous variable est en final d'un morphème et lorsque nous avons à faire à une gémignée lexicale.



Carte II.4.14 : La variation de $\wedge CC_4 : eCC_4 \setminus \text{manger}\{2.SG.IMP\}$

Les formes fricatives et affriquées de cette sous variable sont dans la continuité avec les réalisations qui prévalaient pour la sous variable $^{\wedge}CC_1$. D'une certaine manière, c'est aussi le cas pour les autres variétés linguistiques. Car la forme $@[[t]]e[[ʃ]]@$ est phonétiquement proche d'une affriquée. La différence principale est qu'une unité sonore vocalique est incrustée au sein de la gémée affriquée. L'élément occlusif — avec probablement un point d'articulation différent $[t] / [t̚]$ ¹⁸ — et l'élément fricatif sont toujours présents.

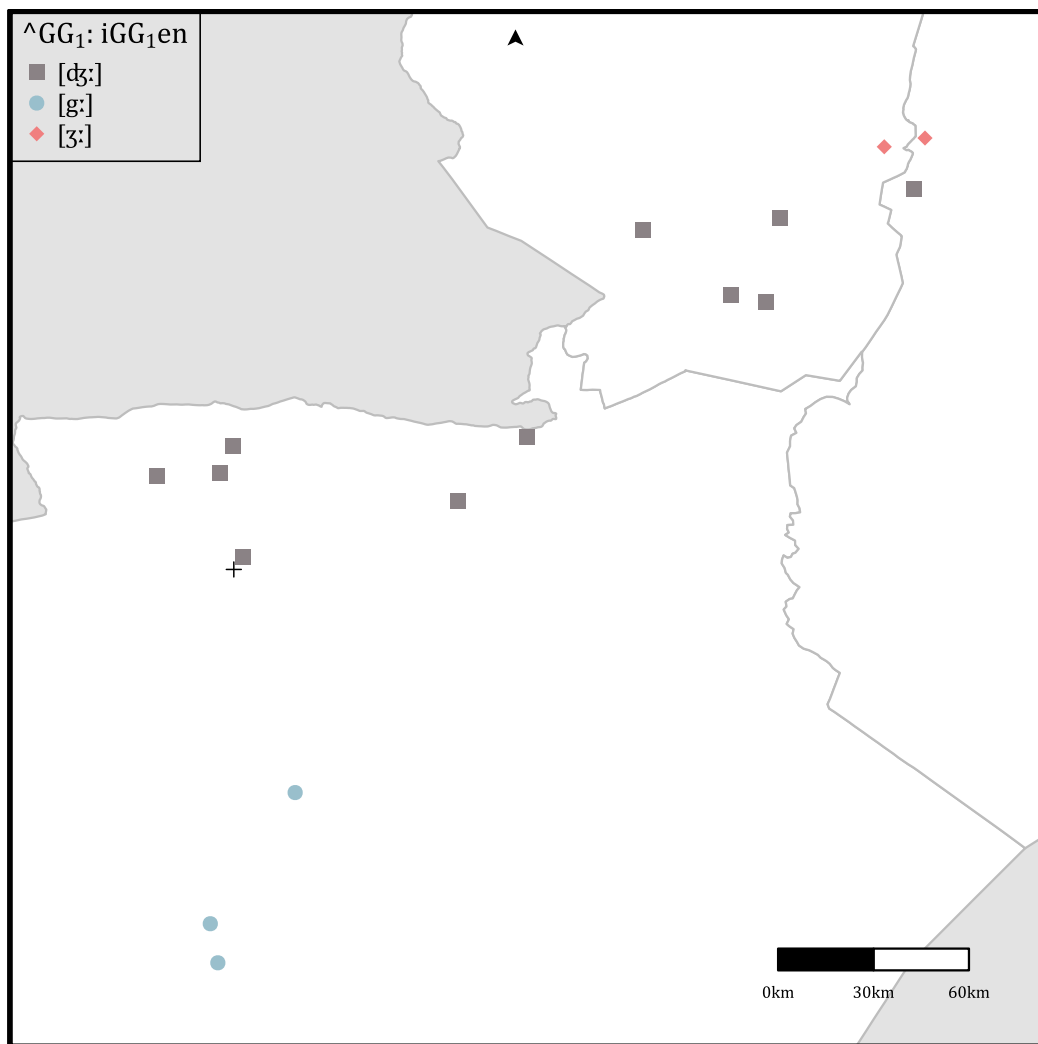
Quand cette unité morphémique n'est pas dans ce contexte, lorsqu'elle est, par exemple, suivie par une unité morphémique la réalisation de celle-ci adoptera les formes de la sous variable $^{\wedge}CC_1$. Ainsi, `\manger{1.SG.PRET}` se dira en fonction des variétés, soit $@[tʃ:]iɪ@$, soit $@[ʃ:]iɪ@$. Comme nous pouvons le constater l'unité sonore n'est plus présente quelle que soit la variante utilisée. Donc, il est évident que nous avons à faire à une seule unité phonique constituée de deux unités sonores.

Les formes que nous venons de voir sont toutes des consonnes sourdes, en revanche, pour les formes de la variable que nous allons étudier les choses sont différentes.

II.4.1.4.1.2 La variable $^{\wedge}GG_i$: Frication, Affrication et Occlusion

La première sous variable que nous présentons connaît plusieurs formes. D'autre part, dans le cas présent, nous avons à faire à une gémée lexicale. Ainsi, nous distinguons trois formes différentes. Nous rencontrons une opposition des variétés linguistiques entre forme occlusive et formes constrictives. Et au sein de ces dernières, une nouvelle opposition prévaut entre forme fricative et forme affriquée.

18. Cette rétractation est peut être à situer au niveau alvéolo-palatale, mais dans la suite de notre exposé, nous continuerons à utiliser le symbole t , pour le cas de cette affriquée, qui ne faudra pas interpréter comme étant une consonne alvéolo-dentale occlusive.



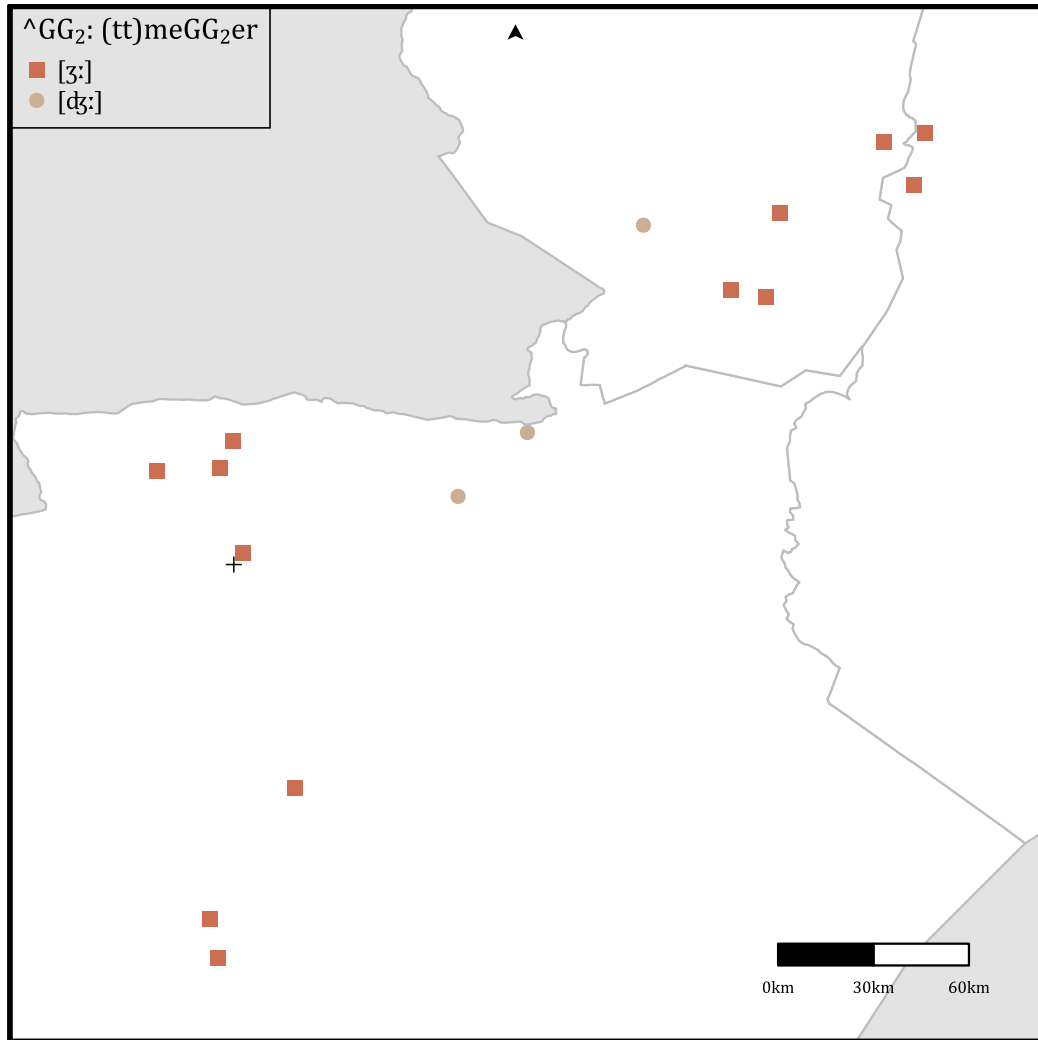
Carte II.4.15 : La variation de $\text{^GG}_1 : \text{iGG}_1\text{en}$ \un{NUM}

Une autre remarque concernant le point d'articulation peut être formulée. Les formes affriquées sont des consonnes coronales, alors que la forme occlusive est une consonne vélaire. Donc, au total, trois formes sont utilisées et l'affriquée [dʒ:] est la forme la plus répandue. Soulignons que les variétés septentrionales ne connaissent toujours pas de forme affriquée et que les variétés méridionales sont celles ayant une forme occlusive.

Toutefois, cette occlusion a laissé place à une fricative pour la sous variable suivante. De la sorte, les variétés d'Igli, de Mazzer et de Berrebi¹⁹ ont la même forme que celles du nord. Ce qui fait qu'Asla et Chellala sont toujours dans la continuité par rapport à la sous

19. Ces variétés du sud ont une autre particularité. La formation de l'*AI* se fait toujours avec le préfixe @tt@, même lorsqu'il y a une morphologie non concaténative à l'œuvre.

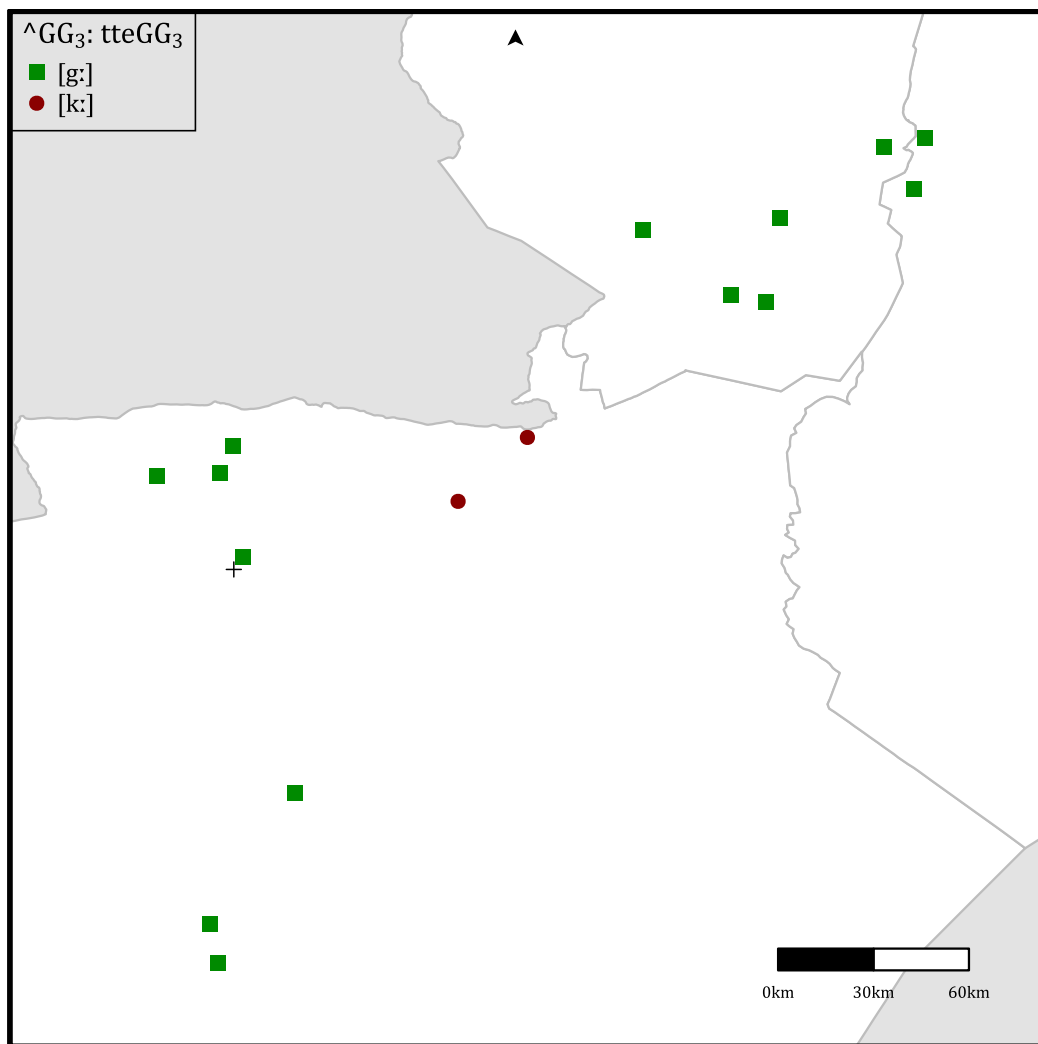
variable précédente.



Carte II.4.16 : La variation de $\wedge GG_2 : (tt)meGG_2er \setminus \text{moissonner}\{2.SG.IMP_INT\} \setminus$

Les variétés occidendo-centrales ont aussi laissé place à une fricative. Seules les variétés de Beni Ounif, Fendi et Sfissifa sont dans la continuité avec $\wedge GG_1$ en utilisant une affriquée. Évoquons aussi le fait que cette sous variable est définie par une gémignée morphologique. La forme simple de cette gémignée correspond à la sous variable $\wedge G_{2\beta\gamma}$. Mais le cas de figure inverse peut également exister comme le montre la sous variable suivante.

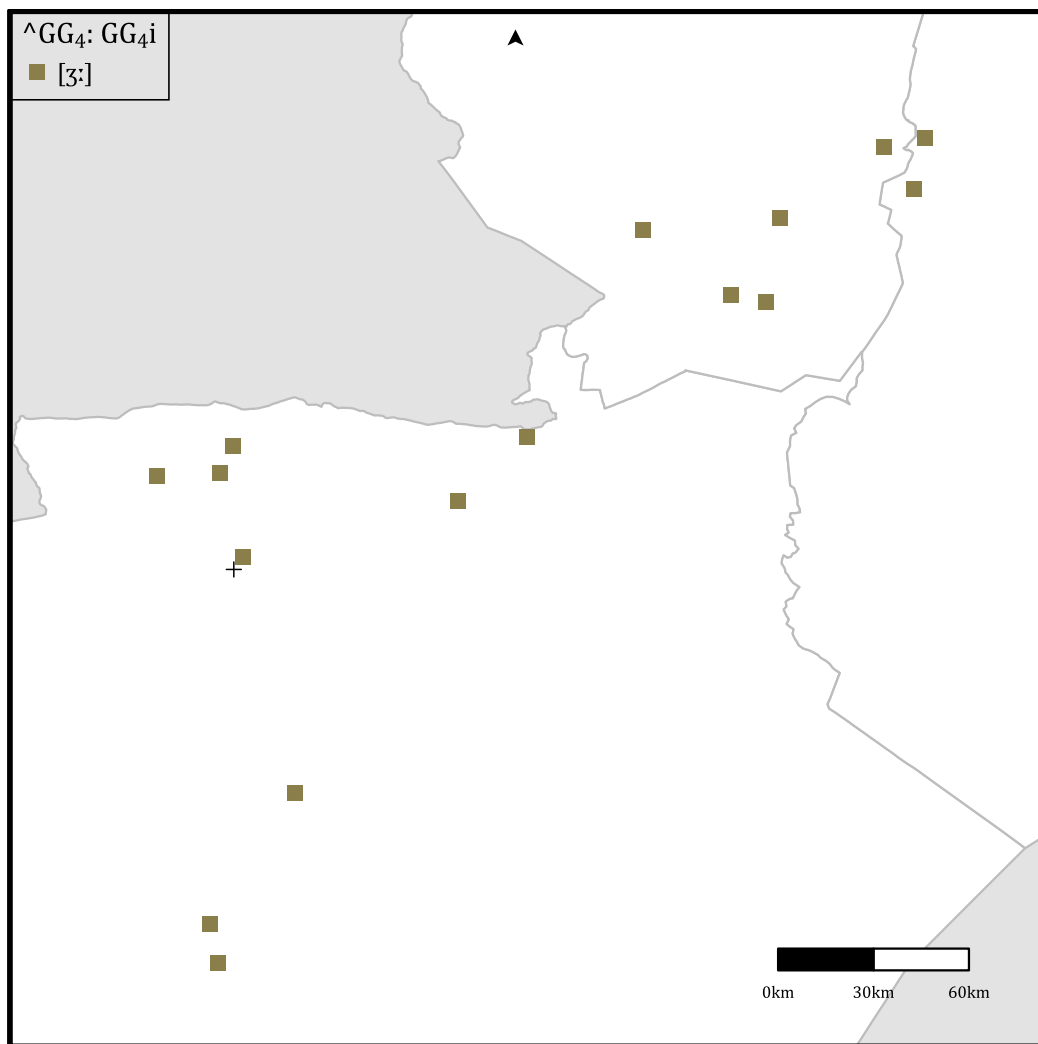
Cette fois, les variétés du sud sont dans la continuité avec la variable $\wedge GG_1$. Tandis que les autres variétés du Sud-Oranais épousent la forme occlusive. Cependant, cette répartition n'est pas, une fois encore, homogène.



Carte II.4.17 : La variation de ^GG_3 : tteGG_3 \faire{2.SG.IMP_INT}

Comme pour la sous variable précédente, les localités de Fendi et Beni Ounif se démarquent. Au lieu d'utiliser une consonne occlusive géminée sonore, leur réalisation sera sourde. Nous aborderons cette question de la différence de voisement plus en détails plus tard (voir Partie II.4.1.4.1.4, page 133). D'autre part, précisons que la géminée de cette sous variable est une géminée morphologique. Elle est le pendant de la sous variable $\text{^G}_{1\beta}$.

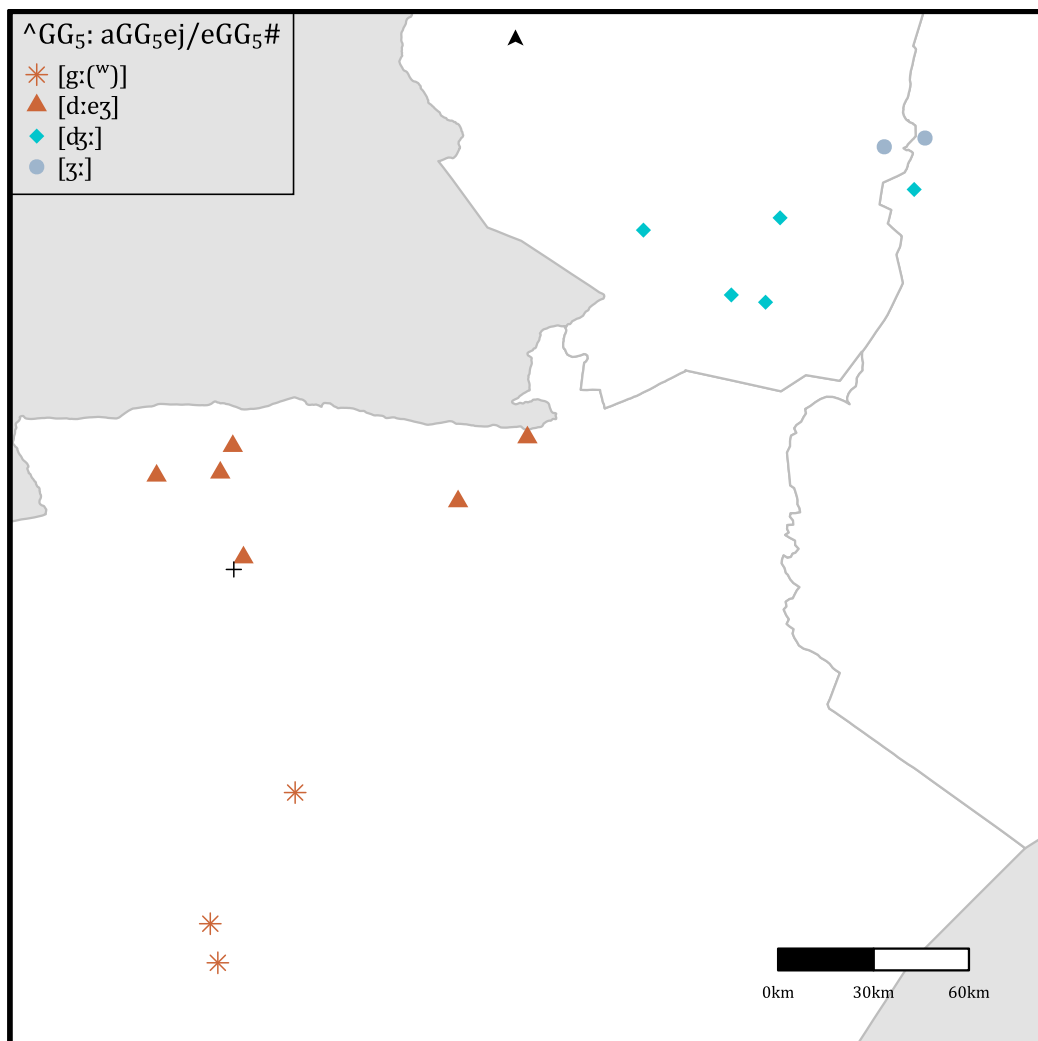
Mais tout ceci change avec la sous variable qui suit. Avec cette dernière, toutes les variétés utilisent la même forme. Toutes les localités vont rejoindre la prononciation d'Asla et de Chellala.



Carte II.4.18 : La variation de $\text{^GG}_4 : \text{GG}_4\text{i}$ \guérir{2.SG.IMP}

Et donc, dans cette situation, seules les variétés d'Asla et de Chellala sont dans la continuité systématiquement. Ces dernières vont invariablement utiliser la forme fricative. Alors que les autres vont soit utiliser la forme fricative, soit la forme affriquée.

Ce fait se poursuit avec la sous variable ^GG_5 , mais il y a deux formes de plus qui sont à l'œuvre. Ainsi, quatre formes sont utilisées dans le Sud-Oranais pour cette sous variable.



Carte II.4.19 : La variation de ^GG_5 : eGG_5 \laiser{2.M.SG.IMP}

Évoquons dans un premier temps le cas des variétés du sud qui se démarquent fortement des autres. Nous avons considéré que le morphème utilisé à Igli et ses environs formait un cognat morphémique avec les autres morphèmes utilisés dans les autres points d'enquête. Bien que l'ajustement entre les cognats phonologiques ne soit pas concrétisé par un alignement symétrique. Il nous apparaissait que la réalisation occlusive correspondait à la réalisation constrictive des variétés plus au nord. L'unité phonique [ɜ] dans @a[g:]e[ɜ]@ n'est pas géminée, ce qui n'est pas le cas pour l'autre consonne. C'est ce qui nous a poussé à considérer que [g:], [dʒ:] et [ɜ:] étaient les formes d'un cognat phonologique. Si tel est le cas, nous retrouvons donc la configuration qui prévalait pour la sous variable ^GG_1 . Mais cela n'explique pas pour quelle raison

les variétés du sud ont en plus [ʒ]. La cause est certainement d'ordre diachronique. Cependant, comme l'objet de notre étude est d'ordre synchronique, alors dans le doute, nous préférons écarter cette donnée pour l'instant.

En revanche, une différence notable existe entre la sous variable ^GG_1 et ^GG_4 . Cette dernière connaît une autre forme. Les variétés occidendo-centrales n'utilisent plus une affriquée. Cette dernière a fait place à une suite d'unités sonores qui semble se différencier de l'affriquée seulement par le fait qu'un vocoïde s'est intercalé entre l'unité sonore occlusive et constrictive. Une telle morphologie n'a lieu que lorsque cette sous variable est en final d'un morphème.

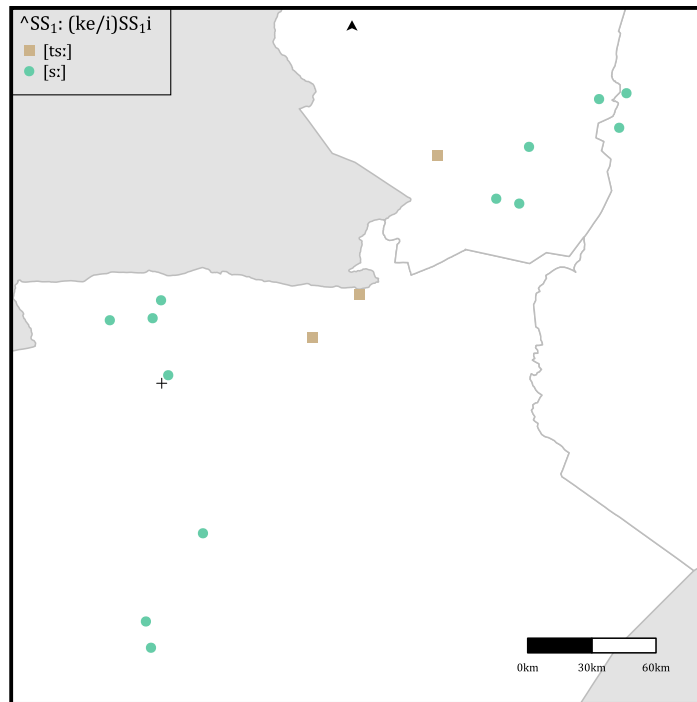
Par conséquent, cette unité morphémique lorsqu'elle n'est plus dans ce contexte a une réalisation différente. Celle-ci correspond aux formes de la sous variable ^GG_1 . Ainsi, si un morphème lui est suffixé, la forme sera soit une fricative ou une affriquée comme dans cet exemple : @i[dʒ:]u@ / @i[ʒ:]u@ \laiser{3.SG.PRET}\. D'autre part, finissons le commentaire de cette sous variable en indiquant que la géminée en question est une géminée lexicale.

Nous allons évoquer maintenant d'autres géminées dont le point d'articulation est bien plus antérieur que celui des variables que nous venons de voir.

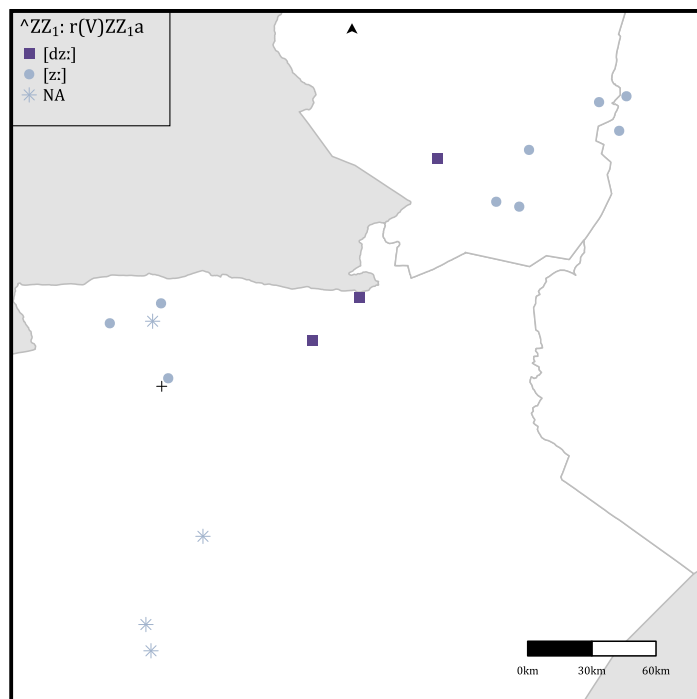
II.4.1.4.1.3 Les variables ^SS_i et ^ZZ_i : Frication et Affrication

Nous évoquerons ici même deux variables différentes, mais qui présentent une distribution identique. Les variétés du Sud-Oranais partagent les mêmes particularités pour ^SS_i et ^ZZ_i . Ainsi, il n'y a qu'une différence de voisement qui sépare ces deux variables.

Ces deux autres variables présentent aussi une variation entre plusieurs formes. Toutefois, le maximum de formes retrouvé pour cette variable est de deux. Ainsi, la première sous variable que nous allons voir est dans ce cas. Nous constatons dans un premier temps que les localités se partagent entre celles qui emploient une forme affriquée et celles qui utilisent une forme fricative. Cette dernière est la forme la plus répandue parmi les variétés du sud-oranais.



Carte II.4.20 : La variation de $\text{^SS}_1 : (\text{ke/i})\text{SS}_1\text{i}$ \ prendre{2.SG.IMP}



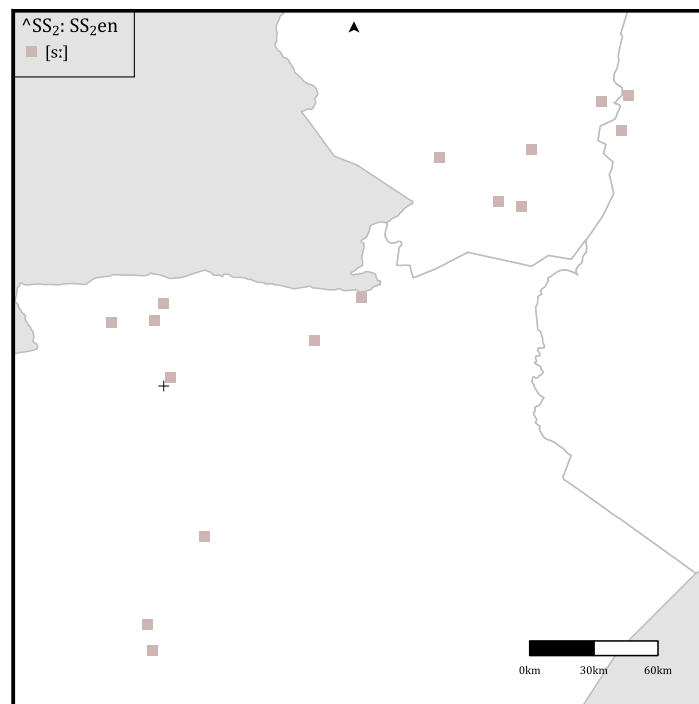
Carte II.4.21 : La variation de $\text{^ZZ}_1 : \text{r(V)ZZ}_1\text{a}$ \ chercher{2.SG.IMP_INT}

Les variétés de Beni Ounif, Fendi et Sfisifa sont celles qui utilisent la forme affriquée. La gémignée de ces sous variables correspond à une gémignée morphologique et provient de la gémignation de la consonne simple [z] ou [s].

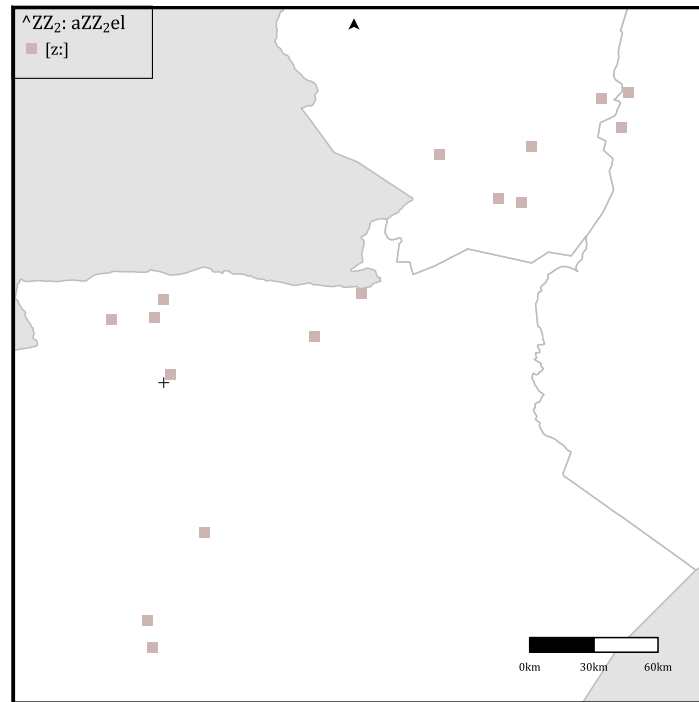
S'agissant de la sous variable $\wedge ZZ_1$, nous n'avons pas pu établir de cognats morphémiques pour l'ensemble du Sud-Oranais. Mais il n'y a pas de raison de penser que la réalisation des variétés non pourvues pourrait être différente de celle de $\wedge SS_1$.

Entre parenthèses sont indiquées les différentes réalisations de ces cognats morphémiques. S'agissant de l'exemple utilisé pour la sous variable $\wedge SS_1$, nous renvoyons à la partie précédente (II.4.1.3.1.2, page 102) où nous avons évoqué ce cas. S'agissant de la sous variable $\wedge ZZ_1$, "V" correspond à une voyelle qui peut être réalisée suivant les localités comme une voyelle ouverte ||a||, ou une voyelle fermée ||i|| ou encore comme une voyelle centrale ||e||.

Cette configuration entre variétés fricatives et variétés affriquées n'est pas respectée encore une fois. Puisque les variétés orientalo-centrales peuvent aussi utiliser des gémignées fricatives.



Carte II.4.22 : La variation de $\wedge SS_2 : SS_2en$ \savoir{2.SG.IMP}



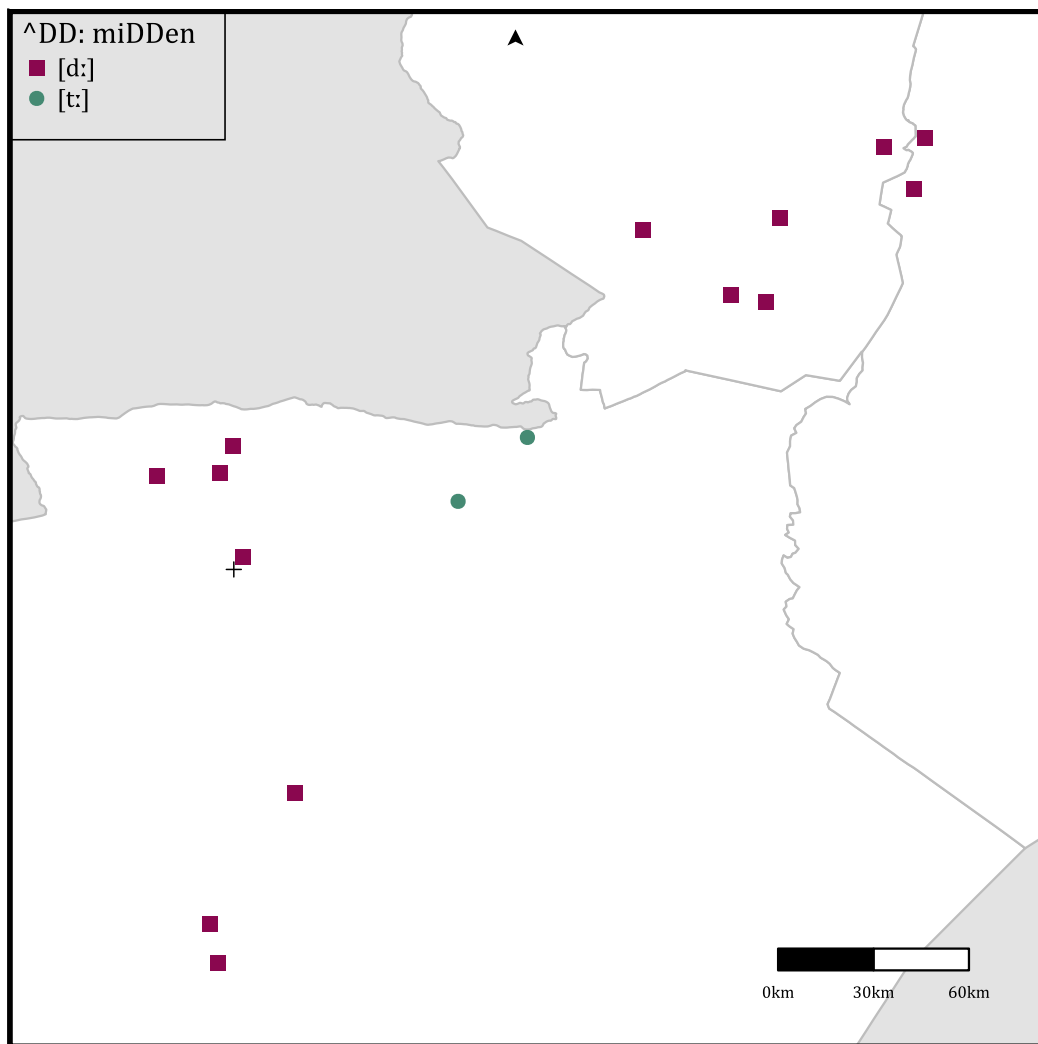
Carte II.4.23 : La variation de $^ZZ_2 : aZZ_2el \setminus \text{courir}\{2.SG.IMP\}$

Donc, Beni Ounif, Fendi et Sfisifa vont rejoindre dans leur réalisation les autres variétés. La gémignée de cette sous variable est une gémignée lexicale.

Les variables qui seront illustrées par la suite traitent d'un autre phénomène. Nous avons ainsi choisi d'évoquer une autre variation linguistique. Cette dernière met en lumière une différence de voisement qui touche certaines gémignées.

II.4.1.4.1.4 Les variables DD , BB et GG : Sonorisation et Assourdissement

Ce premier cas concerne en réalité plusieurs cognats phonologiques. Mais nous avons fait le choix de ne pas tous les évoquer. Nous avons préféré retenir une seule de ces variables pour illustrer cette variation. La variable DD servira donc de représentant à toutes ces variables. Comme nous pouvons le voir sur la carte linguistique, il existe des variétés qui pour cette variable vont utiliser une consonne sourde et d'autres une consonne sonore.



Carte II.4.24 : La variation de ^DD : miDDen \gens\

Les variétés qui utilisent des consonnes voisées pour ces variables sont les plus répandues. Celles qui utilisent des consonnes non voisées correspondent aux localités de Beni Ounif et Fendi. Les autres variables qui sont dans ce cas ont été répertoriées dans le tableau ci-dessous²⁰. Ces dernières ont la même distribution géographique que la variable ^DD.

20. À Berrebi, la forme utilisée est @[lbe]@ et non pas @[b:e]@. D'autre part, certaines variétés en fonction des morphèmes affixés n'auront pas une labiovélaire. Elles diront @[g:]ud@ au lieu de @[g:w]ed@

Variable	[^] BB		[^] GG	
Sous variable	[^] BB ₁	[^] BB ₂	[^] GG ₁	[^] GG ₂
Variétés sourdes	[p:]	[p: ^w]	[k:]	[k: ^w]
Variétés sonores	[b:]	[b: ^w]	[g:]	[g: ^w]
Exemples	[^] BB ₁ i \couper{2.SG.IMP}\	[^] BB ₂ a \père\	[^] GG ₁ ur \marcher{2.SG.IMP _I .NT}\	[^] GG ₂ ed \avoir_peur{2.SG.IMP}\

TABLE II.4.3 : Les formes des variables [^]BB et [^]GG

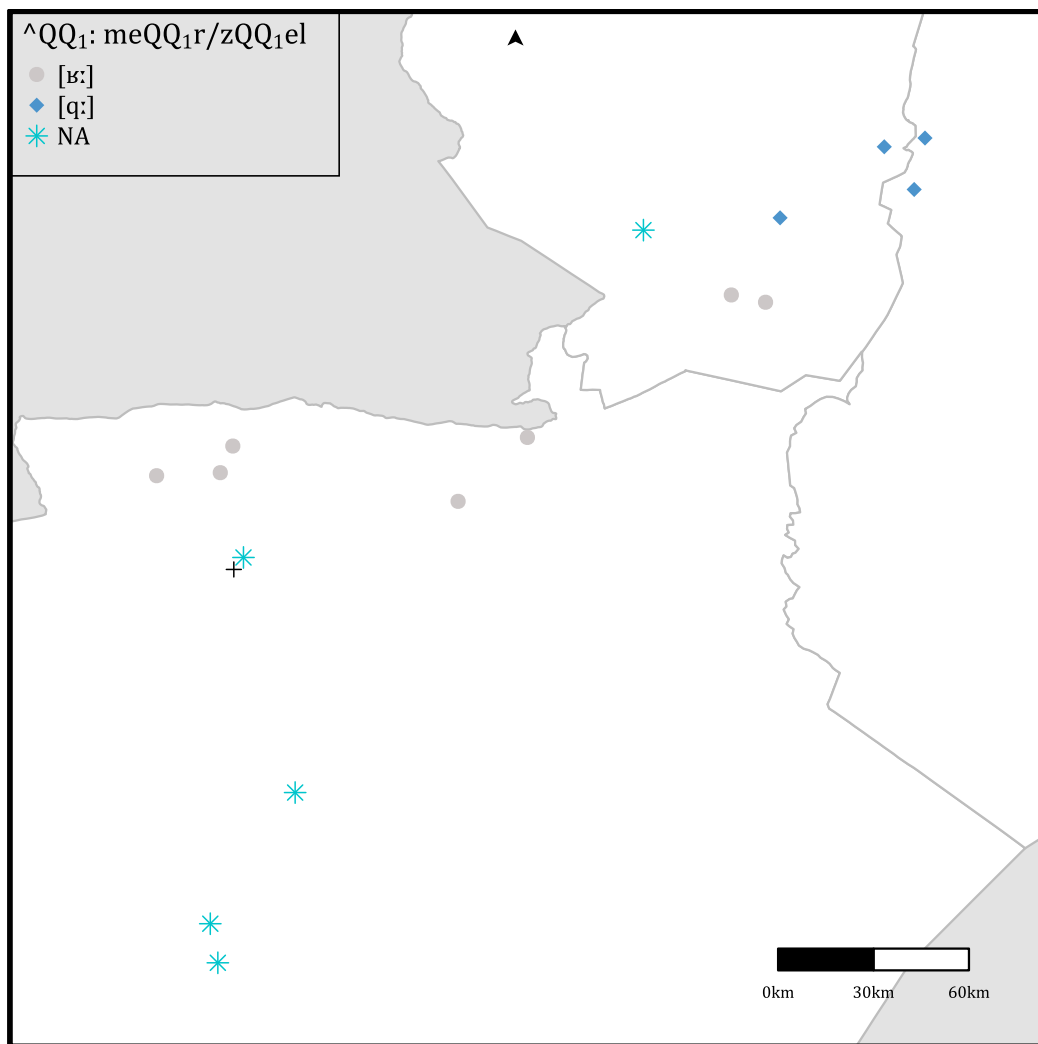
D'un point de vue phonétique, ces consonnes sourdes ne se distinguent pas de leurs contreparties sonores au niveau du point d'articulation. En revanche, il pourrait l'être au niveau du mode d'articulation. Nous percevons que ces consonnes ne sont pas de simples consonnes occlusives géminées. Cela reste à confirmer, néanmoins, il nous apparaît que ces consonnes sourdes sont des consonnes éjectives géminées. Dans le doute, nous avons préféré tout de même les considérer comme de simples occlusives.

Les exemples utilisés pour illustrer ces variables se rapportent pour la plupart à des géminées lexicales, mais des géminées morphologiques peuvent aussi présenter cette particularité. Ainsi, les localités de Fendi et Beni Ounif diront @he[k:^w]a@ \descendre{2.SG.IMP_INT}\, alors que les autres variétés utiliseront une forme sonore @he[g:^w]a@. Ces dernières sont le pendant géminé de la semi-voyelle [w].

Le point commun de toutes ces formes, c'est que totalement ou partiellement le mode d'articulation sera occlusif. Mais ce phénomène peut ne pas confronter seulement des consonnes occlusives. Effectivement, cette opposition entre sonorisation et assourdissement concerne également une autre variable qui ne présente pas les mêmes particularités que [^]DD. C'est cette dernière que nous allons évoquer par la suite.

II.4.1.4.1.5 La variable [^]QQ_i : Sonorisation et Assourdissement

Cette autre variable présente aussi une variation régionale qui s'articule autour d'une différence de voisement. Pour cette première sous variable, nous n'avons pas de données pour l'ensemble des variétés du Sud-Oranais. Les items dont nous nous sommes servis pour établir cette variation n'existent pas dans les variétés du sud. Donc, nous n'avons pas pu établir de cognat phonologique certain avec les variétés du sud.

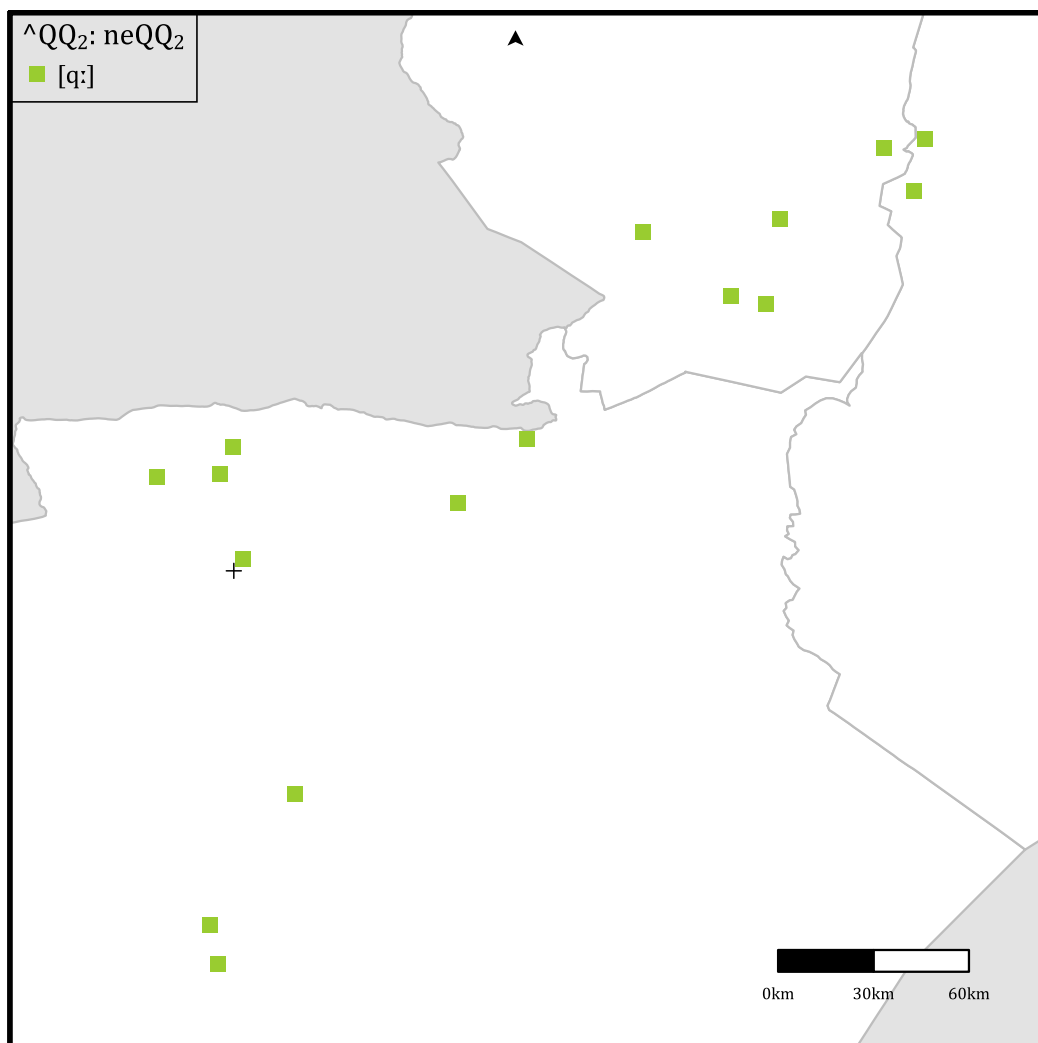


Carte II.4.25 : La variation de $^{\text{QQ}}_1$: meQQ_1er \grandir{2.SG.IMP_INT}\ / zeQQ_1el \se_réchauffer{2.SG.IMP_INT}\

En prenant en compte les données en notre possession, nous avons pu distinguer deux réalisations différentes pour cette sous variable. Il existe des variétés qui feront usage d'une forme sonore et d'autres d'une forme sourde. Cette dernière est retrouvée dans les variétés du nord et la forme sonore est employée par les variétés centrales. Bien que nous n'ayons pas trouvé de cognats morphémiques communs représentant cette variable pour l'ensemble du Sud-Oranais, il ne nous a jamais été donné d'entendre la consonne [ɮ:] dans les variétés du sud. Donc, nous en tirons la conclusion que ces variétés ont dans leur système phonético-phonologique uniquement la consonne [q:]. S'agissant de Sfissifa, cette variété est probablement à rattacher aux variétés centrales, mais tout ceci demande

confirmation.

D'autre part, les formes de cette variable se distinguent par leur mode d'articulation, le voisement n'est pas la seule particularité qui les différencie. La forme sourde correspond à une occlusive, alors que la forme sonore est une fricative. De plus, pour cette variable, nous avons à faire à une gémignée morphologique. La contrepartie simple de cette gémignée est la consonne [ɸ]. Toutefois, dans ces certains cas, cette sous variable sera prononcée [q:] par l'ensemble des variétés. Un exemple que nous pourrions donner est l'item @am[q:]ran@ \grand\ qui est usité dans tout le Sud-Oranais.



Carte II.4.26 : La variation de $^{\text{QQ}}_2 : \text{neQQ}_2$ \tuer{2.SG.IMP_INT}\

Cela étant, cette répartition ne constitue pas la seule possibilité. Effectivement, nous constatons que les variétés centrales ont également dans leur inventaire phonético-phonologique des consonnes occlusives uvulaires. Ainsi, les variétés de Beni Ounif et Fendi rejoignent les autres variétés, alors que ces dernières sont dans la continuité avec la sous variable précédente. La forme de cette sous variable correspond également à une gémignée morphologique, dont la contrepartie simple est toujours la consonne [ɣ].

Cependant, la sous variable $^{\wedge}QQ_2$ peut aussi représenter des gémignées lexicales comme par exemple @[q:]im@²¹ \être_assis{2.SG.IMP}\. Ce même constat peut être formulé pour des verbes qui ont [q:] à la forme non factitive et [ɣ] à la forme factitive comme dans le cas de @[q:]el@ \regarder{2.SG.IMP}\.

II.4.1.4.2 Discussion

Nous allons commencer par porter notre attention sur des questions de géographie linguistique. La distribution géographique de ces variables est très différente de ce que nous avons vu dans le cas des consonnes simples. Précédemment, une opposition nord sud était manifeste. Mais pour les consonnes gémignées, cette opposition s'articule principalement entre les variétés orientalo-centrales et les autres régions. Même si pour certaines variables la configuration avec trois blocs peut être visible, elle ne constitue pas la distribution principale.

D'autre part, nous ne constatons pas de particularisme. Chaque réalisation est partagée par au moins deux points d'enquête. Cependant, la distribution ne se fait pas de manière successive, il existe des ruptures. Dans le cas des gémignées, les variétés orientalo-centrales de par leurs réalisations phonétiques particulières coupent la distribution régionale entre le nord et le sud. Donc, il y a lieu de constater que la variation linguistique ne se fait pas systématiquement de proche en proche. Là encore, le continuum linguistique ne constitue pas une réalité intangible.

En outre, à l'analyse de ces matériaux, nous nous rendons compte par rapport à l'inventaire des gémignées que certaines d'entre elles sont usitées par toutes les variétés. La différence qui existe entre ces variables s'établit au niveau de la distribution de ces unités phoniques et de leur fréquence. Ainsi, une variété linguistique usitera davantage une gémignée, alors qu'une autre variété en emploiera une autre plus fréquemment.

21. Cela suppose pour que cet item soit considéré comme une gémignée lexicale qu'il y ait un phénomène de dégémination de la forme A à la forme AI

Toutefois, certaines géminées ne sont pas incluses dans toutes les variétés. Ainsi, les variétés d'Asla et de Chellala seront les seules à ignorer les affriquées. Et les affriquées alvéolaires seront beaucoup plus le fait des variétés de Beni Ounif, Fendi et Sfissifa.

D'autre part, nous constatons que certaines variétés pour une même variable n'ont qu'une forme, alors que d'autres alternent entre plusieurs formes. Par ailleurs, nous observons, pour ceux ayant plusieurs formes, que celles-ci ne sont pas conditionnées par le type de géminée. Si la géminée est lexicale, alors la forme pourra être soit une fricative, soit une affriquée et de même pour la géminée morphologique. Mais avec cette configuration, un problème phonologique se pose. Si nous considérons qu'au niveau des représentations, les géminées morphologiques sont identiques aux géminées lexicales, alors il devrait en surface avoir les mêmes formes. Cela est rendu inéluctable, s'il n'y a pas de contextes phonétiques ou phonologiques particuliers — comme c'est notre cas — qui pourraient interférer entre l'entrée et la sortie.

Si nous posons le problème de manière plus abstraite. Si [Y:] est le résultat d'un processus morphologique dérivant d'une consonne simple /X/, nous déduisons que [Y:] est la forme de surface de /X:/. Et que si d'autre part, nous retrouvons la même forme de surface [Y:] dans une géminée lexicale, nous déduisons automatiquement que sa forme sous-jacente est /X:/. Mais qu'en est-il si nous retrouvons aussi dans une variété linguistique une forme de surface qui serait [X:]? Si rien n'indique que cette forme a une distribution réduite, alors nous ne pouvons qu'admettre et définir que sa représentation phonologique correspond aussi à /X:/. Si tel est le cas, cela pose problème. Une représentation /X:/ peut-elle avoir deux formes [X:] et [Y:]? Avant de poursuivre et de répondre à cette question, nous allons récapituler les données²².

Géminées			
	Morphologique (Simple : [X])	Lexicale 1	Lexicale 2
Type I	[X:]	[X:]	[X:]
Type II	[X:]	[X:]	[Z:]
Type III	[X:]	[X:]	[Y:]
Type IV	[Y:]	[X:]	[Y:]

TABLE II.4.4 : Typologie des formes de surface des géminées

22. Le cas des assourdissements et des sonorisations ne sont pas pris en compte dans ce tableau.

Ainsi, le développement qui a été précédemment exprimé correspond au type IV de la table II.4.4. Ce type coïncide avec les variétés de Beni Ounif, Fendi et Sfissifa. Si nous continuons la discussion que nous avons interrompue. À notre sens, si nous souhaitons expressément confondre les réalisations phonétiques des géminées lexicales avec celles des géminées morphologiques sous une même représentation, alors nous allons au devant d'une complication inextricable. Cette quadrature du cercle ne peut être résolue qu'en stipulant que la représentation phonologique de la géminée n'est pas dépendante de son pendant simple. En d'autres mots, nous ne pouvons définir la représentation phonologique d'une géminée morphologique à partir de la forme simple dont elle dérive. Cela suppose une indépendance phonologique entre la consonne simple et la consonne géminée, comme il y a une indépendance entre deux consonnes simples, même si elles partagent des traits en commun.

Cela revient à dire que si une unité morphémique est l'objet d'une opération morphologique sur un segment [X] ayant comme représentation /X/, alors il n'y a aucune automaticité à ce que le segment dérivé ait en matière de qualité la même représentation que la forme simple. Cela sous-entend qu'entre la représentation de la consonne simple et celle de la géminée, cette dernière serait seulement agrémentée d'un temps de réalisation plus long par rapport à sa contrepartie simple. Donc, cela entraîne qu'il n'y a pas de raisons de se poser la question pourquoi la forme simple ne partage pas des propriétés phonétiques identiques avec la forme longue comme la spirantisation. Ce point est important et permet d'élucider bien des cas. Car, sans cela, nous ne pourrions comprendre les différentes formes de la variable [^]QQ.

Effectivement, ces observations précédentes sont appuyées par les données de cette dernière variable où la gémination de [Ɂ], selon les variétés, peut avoir deux formes différentes : [Ɂ:] et [q:]. Ces deux formes différentes, fricative et occlusive, peuvent même survenir au sein d'une même unité morphémique — dans ce cas la racine — : @me[Ɂ:]er@ et @ame[q:]ran@ comme nous l'avons vu. Cela renvoie au point discuté lors de la discussion sur les matériaux d'analyse des consonnes simples. À savoir, s'il était raisonnable d'acter que l'unicité morphémique est inviolable. Même si nous n'avons pas totalement tranché la question, nous constatons que cette nouvelle donnée contribue donc à remettre également en cause ce principe.

Par conséquent, il serait préférable de considérer les représentations du type IV comme n'étant fondamentalement pas très différentes de leur forme de surface. Même si nous cherchons à considérer que la représentation de la forme simple diverge de la surface — par un processus phonologique quelconque —, il sera toujours impossible de

réconcilier les formes des géminées lexicales²³. Donc, il serait plus sensé d'admettre que les variétés, qui sont dans ce cas, ont comme phonème en plus des géminées affriquées aussi des géminées fricatives. Ainsi, pour les variétés de type IV, les phonèmes de $^{\wedge}QQ_1$, $^{\wedge}QQ_2$, $^{\wedge}SS_1$, $^{\wedge}SS_2$, $^{\wedge}ZZ_1$, $^{\wedge}ZZ_2$, $^{\wedge}GG_1$, $^{\wedge}GG_2$, $^{\wedge}GG_4$, $^{\wedge}CC_1$, $^{\wedge}CC_2$, et $^{\wedge}CC_3$ sont respectivement /q:/, /k:/, /s:/, /ts:/, /z:/, /tz:/, /dʒ:/, /dʒ:/, /ʒ:/, /tʃ:/, /tʃ:/ et /ʃ:/ . Donc, il s'avère impossible de réunir toutes ces formes de surface en une représentation sous-jacente unique. Mais peut-on généraliser cette observation à tous les types ?

En effet, cette discussion a tourné essentiellement autour du type IV, qu'en est-il vraiment des deux autres types ? Il faut d'abord dire que les types II et III n'ont pas ce problème entre la forme simple et la forme géminée morphologique. Il n'y a plus de différences de qualité, mais uniquement de quantité. Toutes deux ont les mêmes modes d'articulation et les mêmes points d'articulation. Par conséquent, il ne serait pas inconcevable de poser comme hypothèse pour ces types qu'au niveau des représentations la géminée ne se distingue de la simple qu'au niveau quantitatif. Néanmoins, cela ne règle pas pour autant le problème des différentes formes de géminées — fricative et affriquée — pour les géminées lexicales. Y a-t-il lieu de considérer les formes affriquées comme les formes de surfaces d'un phonème fricatif géminé ? Si nous prenons en compte des considérations statistiques, il s'avère que les formes affriquées sont bien plus nombreuses que les formes fricatives. Par conséquent, il nous apparaît comme peu logique d'entrevoir le cas majoritaire sur la base du cas minoritaire. Il serait plus raisonnable d'envisager pour les types II et III que la représentation des sous variables $^{\wedge}CC_1$ et $^{\wedge}GG_1$ correspondent respectivement aux affriquées /tʃ:/ et /dʒ:/.

Au plan synchronique, l'analyse de ces données nous permet d'affirmer qu'entre la consonne simple et la consonne géminée les représentations ne sont pas nécessairement identiques en matière de qualité. Donc, la reconstruction de la représentation de la consonne simple — par exemple de forme fricative — à partir de la consonne géminée — par exemple de forme occlusive — n'est pas admissible. Cela vaut également pour le processus inverse, il serait inexact de déterminer la représentation d'une géminée — par exemple de forme affriquée — à partir de la consonne simple — par exemple de forme fricative —. Cette hypothèse semble être corroborée par une étude neurolinguistique [Ylinen et al., 2005]. Cette dernière fait la démonstration que les consonnes géminées et les consonnes simples ont des représentations différentes et sont surtout activées dans

23. Nous compliquerions encore plus les choses en introduisant un antagonisme avec le principe d'inaltérabilité — voir plus bas où nous évoquons cette question plus longuement —.

des régions distinctes du cerveau.

D'ailleurs, il faudrait nous interroger pour quelle raison, il n'y a pas d'homogénéisation comme à Asla entre les variables. D'après nous, le nœud du problème se situe au niveau des géminées lexicales qui ont des formes discordantes pour les autres variétés. Pour rappel, dans les autres variétés les géminées lexicales ne sont pas identiques, elles sont soit affriquées, soit fricatives. Mais comme avec les consonnes simples, les variables qui ont une distribution géographique homogène entre les variétés sont celles qu'il faut questionner. Celles qui ont cette distribution sont les sous variables $^{\wedge}GG_4$ et $^{\wedge}CC_3$. D'autre part, comme nous l'avons signalé, ces deux sous variables sont statistiquement très peu représentatives. Au point de vue synchronique, il semble que rien ne puisse expliquer ce non alignement avec les sous variables $^{\wedge}GG_1$ et $^{\wedge}CC_1$. Donc, la raison est certainement à chercher au niveau diachronique.

Il existe une autre dimension que l'observation des données permet de formuler. Ainsi, si nous comparons nos données avec ceux qui ont cours dans les études sur les géminées de ces dernières décennies. Il apparait que les conclusions qu'elles formulent ne peuvent s'appliquer à ces données. Rappelons que les géminées qui ont été étudiées précédemment sont toutes de vraies géminées. Cependant, certaines correspondent à des géminées lexicales et d'autres à des géminées morphologiques. De fait, une sorte de consensus s'est établie dans la communauté des phonologues pour ériger deux principes qui sont le principe d'intégrité et le principe d'inaltérabilité [Hayes, 1986].

Si nous tenons compte de ce premier principe et que nous le confrontons aux données des sous variables $^{\wedge}CC_4$ et $^{\wedge}GG_5$. Nous nous rendons compte de suite que l'intégrité a été violée. Puisque, une unité sonore s'est insérée au sein de la géminée. Cette unité sonore ne peut être considérée comme une voyelle intrusive causée par un phénomène de coarticulation provenant de l'unité sonore adjacente. Car les unités sonores environnant ce vocoïde sont des consonnes non voisées pour la sous variable $^{\wedge}CC_4$. D'autre part, soulignons que cette épenthèse n'a pas eu lieu pour toutes les variétés du Sud-Oranais. Il n'y a uniquement que celles qui présentent des formes affriquées qui connaissent ce phénomène — à l'exception du groupe septentrional —. Peut-être ce fait a-t-il son importance pour expliquer pourquoi il n'y a que ce type de géminées qui est affecté. Effectivement, les autres géminées (voir l'inventaire II.4.1.1, page 89), qui ne présentent pas deux segments en surface, ne connaissent pas l'insertion d'une épenthèse.

Par ailleurs, comme nous l'avons souligné, ces affriquées d'un point de vue articulatoire correspondent donc à deux segments phonétiques différents. Un segment

occlusif qui est suivi d'un segment fricatif. Ces derniers sont toujours retrouvés, même lorsque l'épenthèse survient. Mais un problème se pose, quel est le segment qui a la plus longue durée? Les deux segments ont-ils la même durée? Ou bien cela dépend du type d'affriquée? Car peut-être qu'une distinction est à faire entre les affriquées géminées occlusives et les affriquées géminées désocclusives²⁴. Ces deux dernières se distinguent par rapport au segment qui porte la longueur consonantique. De cette façon, pour les affriquées géminées occlusives l'allongement se fera sur le segment fricatif, alors que pour les affriquées géminées désocclusives se sera l'inverse. Parmi nos variables, nous avons ces deux types d'affriquées. L'affriquée géminée occlusive est représentée par les formes [tʃ:], [ts:] et [dz:]. L'affriquée géminée désocclusive est dépeinte par la forme [t:s] utilisée à Chellala (voir Partie II.4.1.3.1.3, page 107) et [d:ʒ]²⁵. Ainsi, pour répondre à nos questions, le segment fricatif serait celui qui a la plus longue durée pour le cas d'une nos affriquées et pour l'autre affriquée ce serait le segment occlusif²⁶. Et il semblerait que cela soit toujours le cas même quand le vocoïde s'insère au milieu de l'affriquée, donc en matière temporelle nous aurions par exemple [XeY:] et seule l'unité — fictive — Y serait géminée. Nous ne savons pas pour l'instant comment interpréter cette particularité, mais il nous apparaissait important d'évoquer ce fait.

En tout état de cause, le principe d'intégrité est une notion qui peut toujours s'appliquer si nous considérons qu'il agit au niveau de la surface, puisque l'intégrité de la longueur consonantique est toujours préservée sur l'élément fricatif ou occlusif. C'est une manière de conforter cette théorie et de l'accommoder avec les faits du Sud-Oranais. De surcroît, cette affirmation n'a de valeur que si nous tenons pour vrai l'hypothèse que nous avons formulée. Car, même si une des deux unités sonores a une durée plus longue

24. Ces dénominations sont de notre fait. Nous proposons ces étiquettes pour distinguer deux types d'affriquées qu'il est possible de retrouver en berbère. Les termes d'occlusion et de désocclusion sont fondés sur l'hypothèse d'une évolution diachronique qui aurait affecté ces consonnes. Cependant, cela n'implique pas inévitablement que si une affriquée est classée dans l'une ou l'autre de ces catégories qu'elle ait connu cette évolution.

25. Ce classement est provisoire, il est seulement basé sur notre perception. Une étude acoustique serait nécessaire pour déterminer avec certitude sur quel segment porte la longueur consonantique. De plus, il est envisageable qu'il puisse exister des affriquées avec une articulation quasi identique, mais qui se distinguerait par rapport à leur longueur consonantique par exemple comme dans [t:ʃ] et [tʃ:]. Rajoutons que si ces consonnes sont dans un contexte hétérosyllabique, nous formulons l'hypothèse que la coupure syllabique aura lieu sur le segment le plus long.

26. En outre, comme nous l'avons déjà mentionné, mais rappelons-le, ces unités phoniques représentent une seule unité phonologique. Il ne faudrait pas, par conséquent, voir dans nos transcriptions approximatives une indication contraire. Il serait d'ailleurs opportun à l'avenir d'envisager d'améliorer l'API pour prendre en compte des spécificités de la sorte.

que l'autre, il n'est pas certain qu'il ne faille pas considérer l'unité la plus courte comme une géminée. Pour s'en assurer, il faudrait confronter une affriquée simple avec une affriquée géminée et constater s'il y a une différence de durée entre le segment court de la géminée avec celui de la simple — sans accroissement de durée [XY(:)] / [XY(:):] ou avec [XY(:)] / [X:Y::] —. Si cela est le cas, alors notre hypothèse d'accommodation ne tient plus et avec le principe d'intégrité²⁷. En revanche, si ce principe a vocation à opérer au niveau des représentations, alors quoi qu'il en soit, il est difficile de continuer à soutenir cette idée. D'autant plus, si nous considérons que ces formes affriquées de surface sont la manifestation d'une forme sous-jacente d'une représentation qui serait non affriquée. À cet égard, il serait manifeste dans cette situation que le principe d'intégrité est un concept à remettre en cause. Enfin, mentionnons encore une fois que les faits décrits ne sont trouvés que dans une partie du Sud-Oranais, les autres variétés sont exemptes de ce phénomène, donc le principe d'intégrité est toujours une réalité possible pour ces dernières comme pour la majorité des langues berbères.

Poursuivons notre discussion en évoquant le cas du principe d'inaltérabilité. Ce dernier est fondé sur l'idée que la forme d'une unité phonique simple est la conséquence d'un *processus phonologique*. Et donc, le phonème de cette unité phonique renvoie à des traits différents de ce qu'il a en surface. Bien que ce postulat soit non fondé²⁸, il est de fait la position adoptée par bon nombre de chercheurs — car sur quoi se basent-ils pour affirmer que [θ] par exemple serait la forme de surface de /t/?²⁹ —. Ensuite, en se basant sur ce postulat, ils en déduiront du fait que ce *processus phonologique* n'apparaît pas pour la contrepartie géminée qu'il y a un phénomène d'inaltérabilité qui toucherait les géminées. Donc, si nous tentons de suivre le même cheminement que nos confrères en reprenant pour notre analyse la sous variable $\wedge G_{1\beta}$ et la sous variable $\wedge GG_3$. Alors, il nous a été permis de constater que le premier est le pendant simple de la deuxième. De plus, nous avons vu que la sous variable $\wedge G_{1\beta}$ était réalisée comme une constrictive [j] dans certaines variétés, mais pour la sous variable $\wedge GG_3$ la réalisation était occlusive. Donc, nous pourrions postuler que le phonème de $\wedge G_{1\beta}$ est /g/ — ce que nous n'avons pas avancé —. Et donc cela démontre que le phonème de la géminée est inaltérable,

27. Nous ne pouvons pas mener cette étude comparative, car les variétés du Sud-Oranais concernées ne connaissent pas dans leur inventaire phonético-phonologique d'affriquée simple.

28. S'il est fondé, il l'est sur des considérations diachroniques. Selon nous, un phénomène phonologique ne peut être déterminé que s'il est contextuel. Dans le cas contraire, si le chercheur persiste dans sa démarche, il est à ce moment-là en train de réaliser une reconstruction diachronique de son unité phonique.

29. Ridouane [Ridouane, 2008] fournit un certain nombre d'arguments pour justifier son postulat, mais c'est loin d'être le cas de tous les chercheurs.

puisqu'en surface elle ne se réalise pas [j:]. Mais encore une fois, un problème s'impose. L'unité phonique [j:] existe dans ces variétés, exemple @(a)[j:]ur@ \lune\. Donc, la forme simple [j] est-elle le pendant de [j:] ou de [g:]? Ou alors encore, elle est le pendant d'aucune géminée comme nous l'avons affirmé précédemment. Tout ceci montre que le procédé utilisé pour établir la réalité phonologique d'une consonne simple en relation avec une consonne géminée morphologique est épistémologiquement très contestable.

Pour conclure, faisons remarquer que nous n'avons trouvé qu'un cas d'opposition entre les formes [k:] et [ʃ:]. Ce cas correspond au verbe \passer_la_journée{2.SG.IMP_INT}\ qui peut soit être réalisé @(tt)[k:]al@ ou @[ʃ:]al@. Cette géminée est le fait d'un **processus morphologique**. Le pendant simple de ce dernier correspond à la sous variable $\wedge K_1$. Mais à Sfissifa, le rapport entre la forme simple et la forme géminée n'est pas linéaire comme dans les autres variétés. Ce **nucléolecte** associe la forme géminée occlusive à la forme simple fricative, ainsi ils diront à la forme de l'A @[ʃ]el@ et à l'AI @[k:]al@. Cette situation apporte de l'eau à notre moulin au sujet de la scission au niveau phonologique entre simple et géminée. Reconstruire la représentation de la géminée à partir de la forme simple n'est pas une méthode robuste. Donc, c'est le seul cas, en la matière, d'opposition entre occlusive et fricative s'agissant des non voisées. Ce qui est différent de ce que nous avons constaté avec la sous variable $\wedge K_1$. D'autre part, cela implique que pour certains cognats phonologiques, soit toutes les variétés prononceront [k:] comme dans cet exemple @e[k:]@ \ passer{2.SG.IMP}\, soit elles énonceront toutes [ʃ:] comme avec la sous variable $\wedge CC_3$. Il en ressort qu'il y a deux cognats phonologiques à distinguer : $\wedge KK_{1a}$ et $\wedge KK_{1\beta}$. La première sous variable renvoie à une forme occlusive qui est réalisée par toutes les variétés du Sud-Oranais, alors que la deuxième correspond à une opposition entre variétés occlusives et variétés fricatives.

II.4.1.5 Synthèse

Au cours de cette étude de géolinguistique, nous avons vu trois aspects différents : l'inventaire phonético-phonologique, la distribution régionale des unités phoniques et la représentation phonologique de ces résultats.

L'inventaire des unités phoniques que nous avons étudié est à peu de chose près identique pour toutes les variétés. Cela revient à dire que le tableau II.4.1.1 (page 89) que nous avons présenté peut dans l'état s'appliquer aussi pour les autres variétés. Donc,

la différence entre ces variétés a lieu sur un autre plan. Elle se trouve au niveau de la distribution. Ainsi, il n'y a pas de variétés fricatives et de variétés occlusives, elles le sont toutes, mais a des degrés divers.

L'autre type de distribution que nous avons vu est la distribution régionale. Les formes des variables n'occupent pas l'espace de la même manière. Certaines formes sont retrouvées uniquement dans une seule localité. Alors que d'autres sont utilisées par l'ensemble des variétés. Donc, les frontières changent perpétuellement.

D'autre part, les différentes exceptions que nous avons mentionnées mettent à mal la notion d'isoglosse. Il est toujours possible de multiplier les variables pour se conformer à ce concept. Cependant, quand la variation est interne au système linguistique, il apparaît impossible de postuler une notion indéfectible de frontière linguistique. La seule manière de résoudre ce problème est de se projeter au niveau des représentations. Une confrontation au niveau phonologique permet de mettre au clair des isoglosses. Ce manque d'ailleurs de frontière linguistique pour une variable, alors qu'elle est présente dans d'autres aires linguistiques berbères, peut être considérée comme un indice de neutralisation diachronique.

Donc, ces questions phonologiques sont fondamentales en dialectologie. Et nous en avons abordé d'autres. À savoir le rapport entre forme simple et forme géminée, mais également entre forme morphologique et forme lexicale. Donc, nous avons tenté de démontrer qu'au niveau des représentations, il n'y avait pas lieu de rattacher systématiquement la forme géminée à la forme simple ou inversement. Cette reconstruction des représentations n'est pas viable et elle est même contradictoire. Une fois c'est la géminée qui sert de prototype et la fois suivante c'est la simple. De plus, elles ne peuvent rendre compte de la complexité des formes en surface entre géminée fricative et géminée affriquée.

Enfin, nous nous sommes interrogé sur la question de l'unicité morphémique qui est le nœud gordien des d'études phonologiques. Cette théorie a ses limites comme nous avons pu le voir. Les allomorphes ou les allophones ne peuvent être définis systématiquement par une unité morphémique. Une approche plus souple de cette question pourrait peut-être débloquent des situations complexes.

Cette complexité a pu être mise en évidence en ayant recours au comparatisme. Ce dernier d'un point de vue épistémologique est une méthode solide pour rendre compte de phénomènes que l'analyse interne ne pourrait mettre en lumière. Ces derniers sont de l'ordre de la diachronie pour tous, mais peuvent questionner sur la réalité synchronique.

Encore faut-il que ce comparativisme soit basé sur une série de cognat suffisant. Ce qui n'est pas le cas systématiquement. Nous avons conscience du manque de robustesse de nos résultats qui sont seulement fondés sur peu de données. Toutefois, si nous prenons en compte l'ensemble des variétés berbères³⁰ où nous retrouvons les mêmes phénomènes, alors cela donne du poids à notre démonstration.

Ces questions phonologiques seront poursuivies dans la deuxième grande partie de notre thèse (). En revanche, cette fois-ci, il est impossible de mettre totalement sur le compte de l'entrée la réalisation phonétique de la sortie. Puisque, la sortie est indépendante de l'entrée dans la mesure où un phénomène physique est à l'œuvre. Ce cas de figure correspond aux trills. Et comme nous le verrons, il sera beaucoup plus compliqué de déterminer la réalité phonologique de certaines rhotiques alvéolaires. Car, dans les faits précédents, il ne rentrait en ligne de compte que des questions cognitives. Alors que dans le prochain point, des questions phonologiques et phonétiques se mêlent ce qui ne peut qu'embrouiller l'analyse linguistique. Mais avant d'évoquer cette situation, nous allons réaliser une étude dialectométrique.

30. Nous n'en ferons pas la démonstration ici, tout berbérissant a en tête ces problèmes qui sont rencontrés dans le monde berbère.

II.4.2 Dialectométrie

Nous avons vu dans notre étude de géolinguistique que les différents points d'enquête présentaient des similitudes. Ces dernières pouvaient s'étendre à tout le territoire du Sud-Oranais où se restreindre à quelques localités. D'autre part, un certain nombre de réalisations phonétiques était systématiquement partagé par les mêmes variétés. Alors que les autres variétés présentaient d'autres particularités phonétiques.

C'est ce dernier fait qui va nous intéresser dans cette étude. À savoir qu'est-ce qui les réunit? Précédemment, nous cherchions ce qui distinguait les variétés dialectales. Cette fois-ci, la réflexion est inversée. Nous allons tenter de mettre en évidence les variétés linguistiques qui se ressemblent au plan linguistique. Mais avant d'aborder la partition des variétés linguistiques, il nous faut redéfinir notre objet d'étude. À partir de maintenant, le terrain d'étude n'est plus la variable indépendante.

II.4.2.1 Les variables dialectales

Dorénavant, nos variables linguistiques sont dépendantes d'une autre variable. Précédemment, nos variables indépendantes étaient nos points d'enquête, mais pour une étude dialectométrique ce ne peut plus être le cas. Les variables indépendantes seront à présent nos variétés linguistiques. Cependant, avant de parvenir à corrélérer des faits linguistiques avec ces dimensions, il faut bien cerner son objet d'étude et se poser la question : quel est l'organisme vivant du linguiste ?

Cette interpellation qui fait une analogie avec la biologie n'est pas négligeable. Car l'objet d'étude du linguiste est difficile à saisir. Lorsque nous parlons de langue, de communauté linguistique ou de dialecte, tout ceci renvoie à une réalité construite. Parler de variété linguistique est un concept vaste. Ce qui fait que l'objet d'étude du linguiste est défini de manière *ad hoc*. Cela saute aux yeux, notamment dans la terminologie utilisée par les linguistes pour classer les différentes variétés linguistiques. Les chercheurs ont du mal à accorder leurs violons au niveau de la taxonomie. Avant de pouvoir définir des espèces, il faut d'abord déterminer l'unité primitive de son objet d'étude.

C'est pour cette raison que nous proposons un modèle théorique de l'objet d'étude. Ce dernier s'accompagne aussi d'une terminologie qui se veut à usage scientifique. Car la terminologie qui a cours s'avère problématique et cause des mésententes entre chercheurs.

II.4.2.1.1 Problème terminologique

La classification dialectale se trouve confrontée à différents obstacles. L'un d'entre eux se situe au niveau des dénominations utilisées pour désigner les différentes variantes linguistiques. Dans la dialectologie française, comme dans la dialectologie berbère³¹, les métalangages souvent employés sont : parler, dialecte et langue. Il existe d'autres appellations — langue régionale, idiome, patois, variante, ... etc —, mais elles ne contribuent pas à éclaircir la situation [Tillinger, 2013]. Ce problème n'est pas propre aux travaux dialectologiques en français, il se pose un peu partout dans le monde occidental, par exemple en littérature anglo-saxonne [Chambers and Trudgill, 1998, p.4]³².

Ces termes en question renvoient normalement à une organisation hiérarchique de la variation linguistique — une langue comprend des dialectes et un dialecte comprend des parlers —, mais il s'avère que d'un auteur à l'autre ces termes ont des sens qui se chevauchent ou bien sont corrélés à des statuts politiques ou de prestige. La conséquence de cet imbroglio sémantique est la controverse qui naît de l'utilisation de ces termes. En fonction des prises de position, une même variante linguistique pourra être considérée comme étant un parler, un dialecte ou une langue. Cette situation est principalement causée par la mésentente sur la définition à donner aux mots dialecte et langue [Léonard, 2012]. Bien que les dialectologues aient conscience de cette problématique, ils continuent tout de même à faire usage de ces dénominations. De ce fait, ils sont confrontés chaque fois à ce point épineux, qui peut se révéler être une tâche ardue pour le chercheur, comme le stipule G. Aurrekoetxea [Aurrekoetxea, 2013, p.133] dans l'un des titres de son article "time to be courageous".

Donc, avant d'aborder la question de notre nouvelle terminologie, il conviendrait d'expliquer plus longuement à quel niveau la confusion apparaît parmi les spécialistes en SDD. Nos lectures des différentes littératures portant sur la dialectologie ou de manière plus générale sur la variation linguistique fait apparaître deux positions bien distinctes. Il existe parmi les linguistes une composante qui perçoit la variation linguistique que sous l'angle de la distance linguistique et une autre qui aborde la question d'un point de vue statutaire. Ces deux écoles, si nous pouvons nous exprimer ainsi, emploient toutes les deux les mêmes termes pour qualifier leurs objets d'études. Ces mots en question que nous avons déjà évoqués correspondent à : parler, dialecte et langue.

31. Cette terminologie dans les études berbères, nous la devons à André Basset qui est pour ainsi dire le précurseur des études dialectologiques en berbère [Chaker, 1995].

32. Dans les études en langue anglaise, les termes utilisés sont : language, dialect, accent [Ibid.]

Ainsi, l'emploi de ces derniers, portant sur des aspects de la variation linguistique qui touchent à des réalités différentes, rend leur utilisation délicate, voire impossible. Donc, la vision des chercheurs qui utilisent cette terminologie traditionnelle dans une démarche classificatoire est sujette à question. D'autant plus, qu'il existe à l'intérieur de ces deux écoles des divergences. Ce qui ne fait qu'accroître l'opacité qui entoure ces mots. Pour rendre compte de ce fait, nous allons examiner plus en profondeur les différentes approches qui sont en cours à l'intérieur de ces deux écoles.

Si nous prenons le cas des chercheurs qui élaborent des classements en se fondant sur le critère de la distance linguistique. Nous constatons que les tenants de cette approche, du fait de leur méthode et des limites intrinsèques à leur modèle, ne permettent pas de rendre compte avec finesse de la diversité de la variation linguistique. Le premier reproche que l'on pourrait faire est leur vision d'une classification par hiérarchisation descendante. Leur point de départ est la langue qui est définie à priori — un apriorisme fondé généralement sur une conception politico-idéologique —. Par exemple, un berbérisant définira le kabyle comme étant une langue et pour un autre chercheur cela correspondra au berbère. En procédant de cette manière, cela débouche inévitablement sur des divergences comme nous venons de l'illustrer avec notre exemple précédent. Puis à partir de leur postulat, ils passeront au niveau inférieur et détermineront, quels ensembles de variétés linguistiques correspondent à des dialectes et ceux qui correspondent à des parlers. Mais là encore un autre problème surgit, c'est l'incapacité de ce modèle à mettre en évidence des regroupements de variantes linguistiques plus fines, car l'analyse est limitée à trois niveaux. Cette astreinte peut amener certains chercheurs à des dénis de particularismes linguistiques pour telle ou telle variante sous le prétexte que rien ne peut s'intercaler entre le dialecte et le parler ou bien à un niveau inférieur au parler. La dernière critique que l'on peut formuler à ce modèle est que généralement sous la dénomination de langue se cache une vision diachronique de cette entité. La langue serait à identifier à la langue mère ou encore à la protolangue ou à la langue originelle. Cela revient à considérer que la langue ne peut être attitrée à une réalité synchronique.

S'agissant des chercheurs qui octroient un statut particulier — représentationnel ou politique — à ces termes, c'est-à-dire qu'ils les perçoivent comme ayant des valeurs sémantiques péjoratives — dialecte, parler — ou mélioratives — langue —³³. Il faut signaler à leur encontre que nous rencontrons également en leur sein des divergences

33. Un sens qui est largement véhiculé dans le langage courant.

quant à leur manière de classer les variétés linguistiques. D'après Leonard [Léonard, 2012, p.51], il y aurait quatre tendances parmi les partisans de cette école :

(1) Paradoxe 1 : les dialectes ou les variétés n'existent pas, il n'existe que des langues.

(2) Paradoxe 2 : les langues n'existent pas, il n'existe que des dialectes ou des variétés.

(3) Paradoxe 3 : plus deux langues sont semblables, plus l'une des deux est autorisée à prétendre être différente par essence, en dépit de l'équivalence de substance.

(4) Paradoxe 4 : plus une variété linguistique ressemble à une langue de référence de rang H (rang haut), moins elle est une langue. Moins elle ressemble à une langue de référence de rang H, moins elle ressemble à une langue.

Comme nous pouvons le constater, ces différentes conceptions ne peuvent malheureusement pas être réconciliées. Par conséquent, il serait plus opportun d'avoir recours à une autre terminologie. Mais avant d'évoquer ce point, il existe une autre dimension qu'il est nécessaire de clarifier. Cela est relatif au champ disciplinaire qu'est la dialectologie que nous pouvons percevoir comme étant malmené. Car il ne s'agit pas seulement de remplacer un terme par un autre terme pour que le problème se dissipe entièrement. Pour cela, il nous faut d'abord définir notre champ d'étude. Qu'est-ce que la variation linguistique ? Toute variation linguistique est-elle du ressort de la dialectologie ? Nous allons tenter de répondre à ces questions.

II.4.2.1.2 Modèle théorique

Par conséquent, seuls des critères linguistiques doivent rentrer en ligne de compte. Cela étant dit, il est important de faire un certain nombre de constats pour mieux comprendre la complexité de la classification dialectale. On sait depuis longtemps avec l'étude de différents idiomes à travers le monde que la variation est visible dès l'échelle individuelle. Mais entre ces idiolectes, nous constatons qu'il existe des conventions linguistiques identiques. Ce "dénominateur commun" de faits linguistiques varie également, ainsi de groupes d'individus à groupes d'individus les conventions

linguistiques qui les unissent peuvent être plus ou moins denses. Ainsi, les écarts qui peuvent les différencier présentent des gradations variables. Puis, il faut souligner que, lorsque ce dénominateur commun atteint un seuil trop bas, l'intercompréhension entre ces groupes d'individus n'est plus assurée. Par conséquent, ce fait doit être caractérisé. Ainsi, la notion de communauté linguistique s'insère au sein de cette variation linguistique en y intégrant toutes les variétés linguistiques mutuellement compréhensibles.

Un autre point sur lequel nous souhaiterions discuter concerne les facteurs qui induisent la variation. On peut classer la variation linguistique suivant si les facteurs sont externes ou internes. Chaque type de variation peut être le fait de ces deux facteurs. Mais suivant quel type de variation linguistique auquel nous avons à faire, seuls les facteurs externes pourront faire l'objet d'une attention particulière. Ainsi, toute variation linguistique n'est pas nécessairement du ressort de la dialectologie et celles qui ne le sont pas ont un rapport avec les facteurs internes. Cette remarque concerne tout particulièrement les variations phonétiques. Ce choix s'explique par le fait, qu'avec ce type de variations linguistiques il y en a autant que d'individus, donc c'est une tâche impossible à quantifier et à mesurer. Cela implique que les seules caractéristiques linguistiques issues de facteurs externes ou internes qui seront prises en compte sont celles qui présentent une certaine régularité. En d'autres mots qui font partie d'un système conventionnel. À titre d'exemple, la variation acoustique entre hommes et femmes due à la différence anatomique faisant que les fréquences des unités phoniques des femmes sont plus élevées que celles des hommes, donc une variation d'ordre interne, ne doit pas être incluse dans une étude dialectologique. Mais le fait que des hommes prennent des voix aiguës, et que ces hommes appartiennent à une catégorie sociologique bien déterminée, est un phénomène qui doit être caractérisé. Car cet aspect est la conséquence d'un facteur externe.

Nous déduisons de ces quelques observations empiriques que le dialectologue doit s'intéresser à la variation linguistique correspondant à des faits linguistiques variant d'un groupe d'individus à un autre. Ainsi, seuls les faits linguistiques conventionnellement partagés par un groupe d'individus, mais ne l'étant pas par un autre, doivent susciter l'attention du dialectologue et être inventoriés. Au sein de cette variation, il existe tout un éventail de paliers qui doit être mis en évidence en commençant d'abord par caractériser les groupes humains partageant des caractéristiques linguistiques totalement identiques, ces derniers seront définis comme étant des "nucléoctes". Puis

l'ensemble des nucléolectes devront faire l'objet d'un classement en regroupant de proche en proche les variétés linguistiques qui partagent des conventions linguistiques communes. Ce groupement de nucléolectes sera dénommé "haplolecte".

Au cœur de cette segmentation, il serait à propos de discerner les haplolectes où est assurée l'intercompréhension. C'est à ce niveau que s'insérerait la notion de communauté linguistique. Mais cela n'exclut pas que parmi cette entité puisse figurer une hiérarchisation de différents types de parenté linguistique. Ainsi, selon notre approche, à l'intérieur d'une communauté linguistique, nous retrouverions donc une multitude de nucléolectes faisant office de vecteur de communication sans que chacun des membres de ces nucléolectes aient nécessairement le besoin d'adopter la variété linguistique de l'autre pour les besoins de la communication. Chaque membre d'un nucléolecte peut converser avec un membre d'un autre nucléolecte sans qu'il y ait d'interférence significative dans la réception du message. C'est ce constat entre variation linguistique mais conservation de l'intercompréhension qui nous impose d'identifier et de catégoriser cet ensemble de nucléolectes. Nous proposons de nommer ce groupement de nucléolectes par le terme de "communolecte". Cette réalité doit faire l'objet d'une identification particulière et systématique dans la description dialectologique, ce qui est moins le cas pour les autres niveaux formants des haplolectes qui n'ont pas besoin d'avoir une étiquette particulière. Nous proposons les catégorisations suivantes qui serviront à classer les différentes variétés.

Nucléolecte	désigne une variété linguistique qui parle un dialecte complètement identique
Allolecte	désigne une variante d'un nucléolecte
Haplolecte	désigne un groupe d'allolectes
Communolecte	désigne un haplolecte ou un groupe d'haplolectes formant une communauté linguistique
Famille linguistique	désigne un groupe de communautés linguistiques apparentées

TABLE II.4.5 : La nouvelle taxonomie des variétés linguistiques

II.4.2.1.3 Application au berbère du Sud-Oranais

Si nous considérons que le **nucléolecte** est notre unité primitive, alors nous avons défini que notre terrain d'étude comprenait au total dix-sept **nucléolectes**. À quoi est dû cette différence puisque ce chiffre ne correspond pas à notre nombre de points d'enquête? Cela s'explique par le fait que certaines localités peuvent comprendre plusieurs variétés. Cela est le cas à Igli. Dans cette ville, nous avons constaté que deux

variétés étaient en usage.

Ce fait s'explique par l'histoire de cette région. Il existait auparavant une localité du nom de Tuzdit qui était située entre Igli et Mazzer. Cette dernière a été détruite lors de la guerre d'Algérie. Tous ces habitants originaires de Tuzdit vivent actuellement à Igli. Donc, dans cette ville coexiste deux nucléolectes qui sont très proches, mais qui ont quand même leurs propres particularités. Les différences en question ne concernent pas le système consonantique, donc dans la suite de notre travail nous ne distinguerons pas ces deux variétés.

Par ailleurs, le chiffre que nous fournissons est provisoire, car dans les autres localités nous avons pu aussi constater qu'il existait des différences entre nos informateurs. Cependant, nous ne savons pas si cela est de l'ordre du conventionnel — donc mériterait d'être défini comme des nucléolectes différents — ou c'est un phénomène idiolectal.

Ici même, nous ne ferons pas la démonstration de nos conclusions, ce point sera donc à considérer comme étant un axiome de notre étude dialectale. Nous avons vu de manière partielle les différences qui existent entre nos nucléolectes dans notre étude géolinguistique. Mais il n'y a pas assez de variables linguistiques exposées pour rendre compte avec finesse des différences entre ces variétés.

Toutefois, ce sont ces variables linguistiques que nous allons reprendre pour mener notre étude dialectométrique. Ce travail consistera à déterminer les parentés linguistiques qui existent entre nos variables dialectales. Nous proposons de réaliser cela par deux approches différentes qui sont l'Apprentissage Non Supervisé et l'Apprentissage Supervisé.

II.4.2.2 Apprentissage Non Supervisé

Nous allons donc dans un premier temps explorer la méthode d'Apprentissage Non Supervisé. Nous allons échelonner ce travail pour rendre compte étape par étape des différentes méthodologies. À chacune de ces étapes, notre intention est d'utiliser plusieurs méthodes. Le but étant de comparer les résultats obtenus par ces derniers.

II.4.2.2.1 Le matériau d'analyse

Les méthodes en question correspondent à deux grands types de méthode de partition que sont la PHA et la Partition non hiérarchisée. Au préalable, il nous faut définir un nombre de partitions. Car la partition qu'elle soit hiérarchique ou non doit

définir une grandeur indiquant le nombre de partition "optimum". Cela est rendu nécessaire pour les méthodes k-mode et k-moyenne. Mais pour la méthode à PHA, nous en avons besoin pour réaliser par la suite une représentation des partitions sur les cartes linguistiques.

Afin de déceler ce chiffre adéquat, nous avons besoin d'une méthode. Nous pourrions choisir au hasard un nombre, néanmoins il est préférable de s'appuyer sur une méthode éprouvée en SDD qui est en partie conçue pour ce type de recherche préliminaire. Nous faisons référence à la méthode MDS. Nous avons appliqué cette méthode à nos données et le résultat sous forme de graphique est ci-dessous (II.4.1).

L'interprétation de ce scatterplot montrant la partition réalisée par cette méthode nous indique que trois groupes semblent se détacher. Les deux groupes du haut s'opposent au groupe du bas. Et les deux de gauche s'opposent à celui de droite sur l'axe de la composante principale.

Cela semble indiquer que par la suite, il serait préférable de choisir trois partitions pour regrouper nos variables dialectales. Ce choix étant fait, nous allons pouvoir entamer avec la première méthode de partition que nous avons choisie.

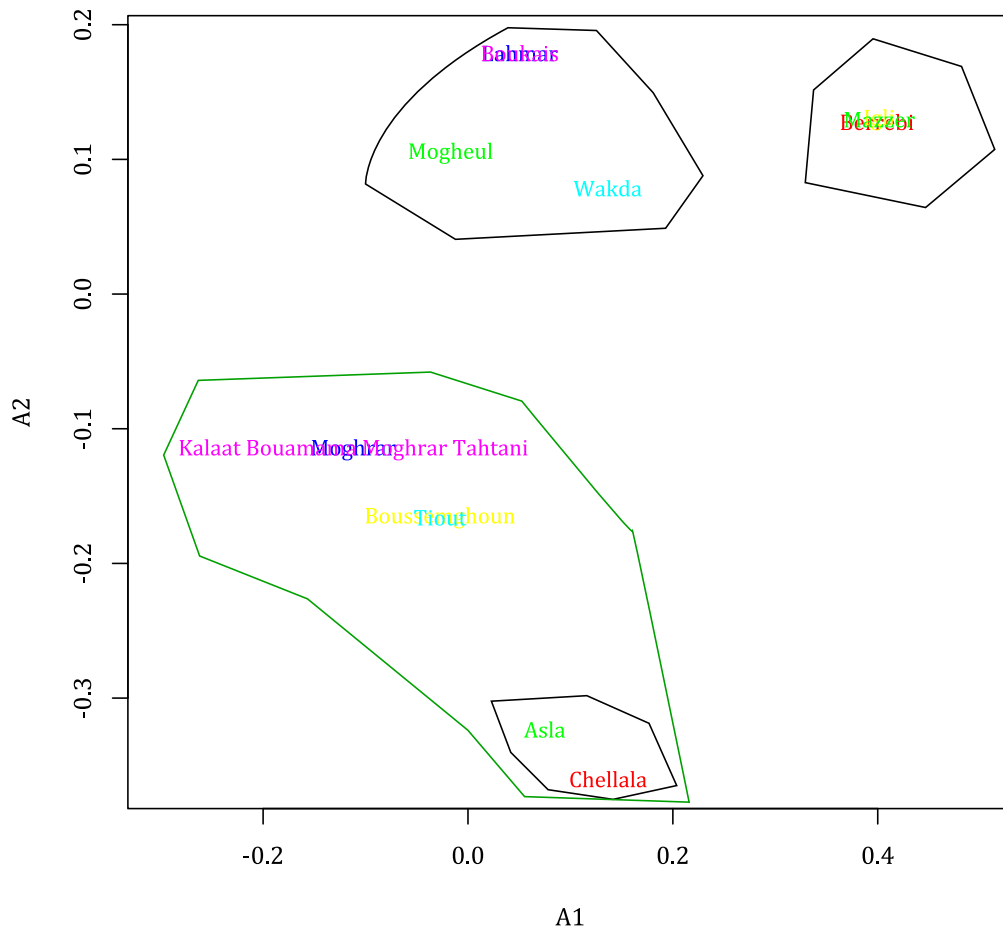
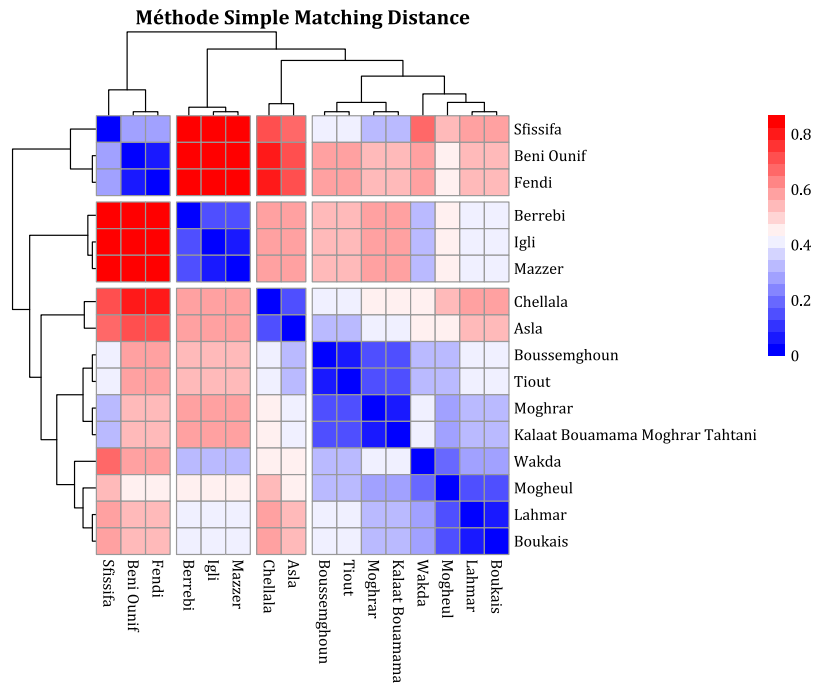


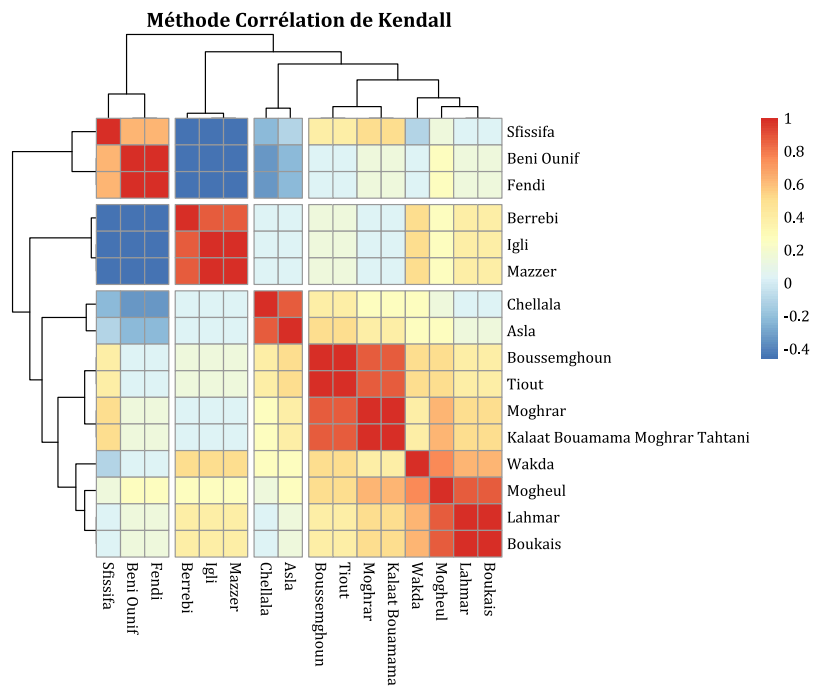
FIGURE II.4.1 : Partition des variables dialectales — méthode MDS

II.4.2.2.1.1 Partition hiérarchisée ascendante

Pour réaliser cette méthode de PHA, nous avons opté pour deux techniques différentes. Ces dernières nous permettent de créer nos matrices des distances. Ainsi, nous avons choisi d'employer la méthode Simple Matching Distance et la méthode Corrélation de Kendall. Les résultats de ces deux calculs différents sont présentés ci-dessous sous forme de heatmap. Cette dernière représente la matrice de similarité. Pour la première heatmap, plus les couleurs sont froides et plus les distances linguistiques sont faibles. Pour la deuxième heatmap, c'est l'inverse.



(a) Matrice de similarité et dendrogramme de la méthode Simple Matching Distance



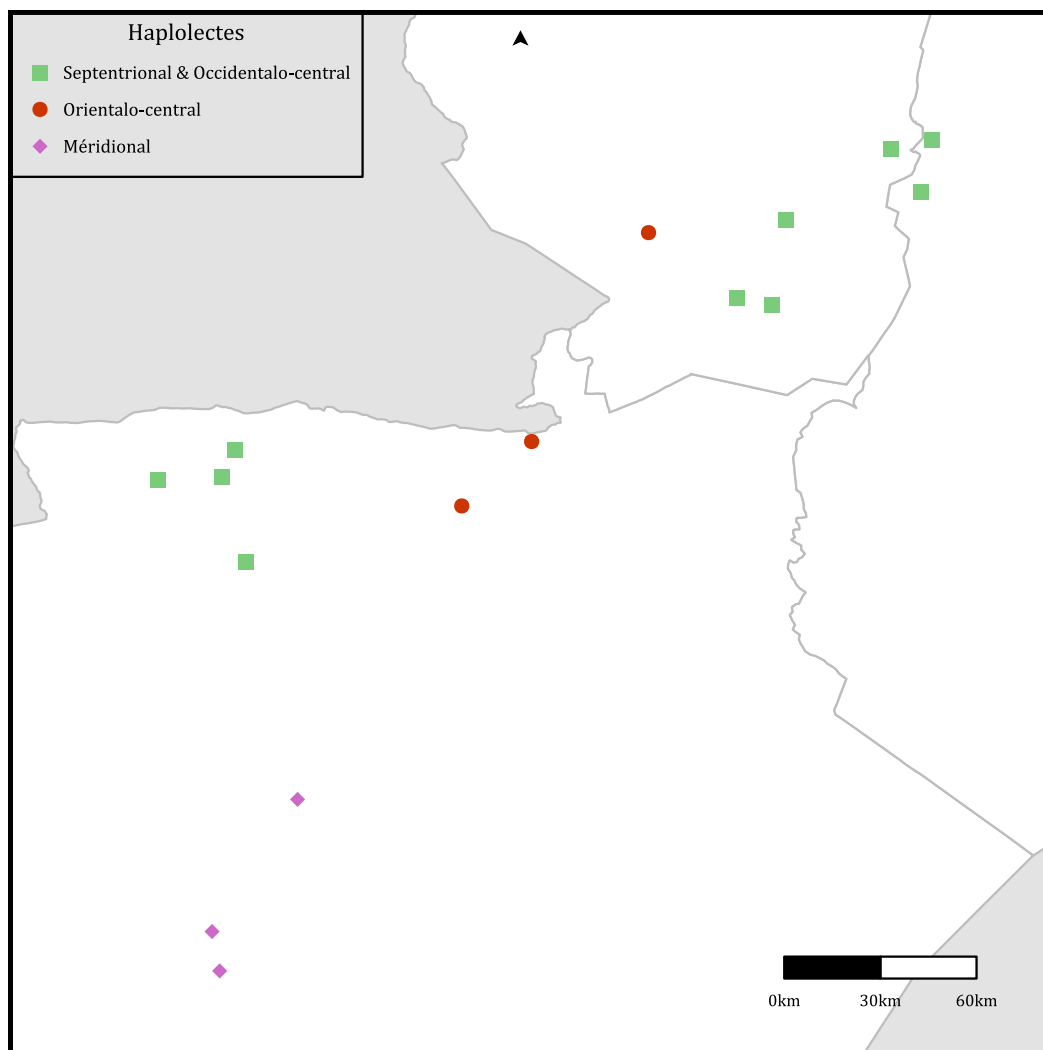
(b) Matrice de similarité et dendrogramme de la méthode Corrélation de Kendall

FIGURE II.4.2 : Comparaison des méthodes de partition hiérarchisée ascendante

Nous pouvons constater que les résultats sont strictement identiques. La méthode Simple Matching Distance et la méthode Corrélation de Kendall proposent les mêmes partitions. Bien que les valeurs, qui correspondent aux distances linguistiques entre les variables dialectales, soient différentes entre ces deux méthodes. Ils aboutissent toutes pourtant à classer les variétés linguistiques de la même façon.

Il est possible de visualiser d'un coup d'œil ce que nous venons de dire en regardant et la heatmap et les dendogrammes qui se trouvent sur ses côtés. La heatmap a été divisée en plusieurs morceaux. Ce découpage a été réalisé pour mettre en relief les partitions. Comme ces heatmaps donnent la possibilité de partitionner indépendamment les côtés. Alors, nous avons réalisé deux partitions différentes. Le côté gauche est découpé en trois haplolectes, alors que pour celui du haut, nous avons quatre haplolectes. En passant de trois à quatre, les variétés d'Asla et de Chellala forment à elles seules un haplolecte. Alors que, quand nous considérons trois partitions, ces variétés linguistiques sont regroupées avec d'autres variétés linguistiques.

À partir des résultats des méthodes précédentes, nous avons projeté sur une carte linguistique les partitions auxquelles ont abouti ces calculs. Comme nous l'avions annoncé, nous avons choisi de retenir le nombre de trois partitions. Ainsi, nous avons reproduit ci-dessous les partitions hiérarchisées des principaux haplolectes de la région du Sud-Oranais. Nous avons intégré, dans une première carte linguistique, les résultats sous formes de points. Les points de même couleur et de même forme représentent un même haplolecte.

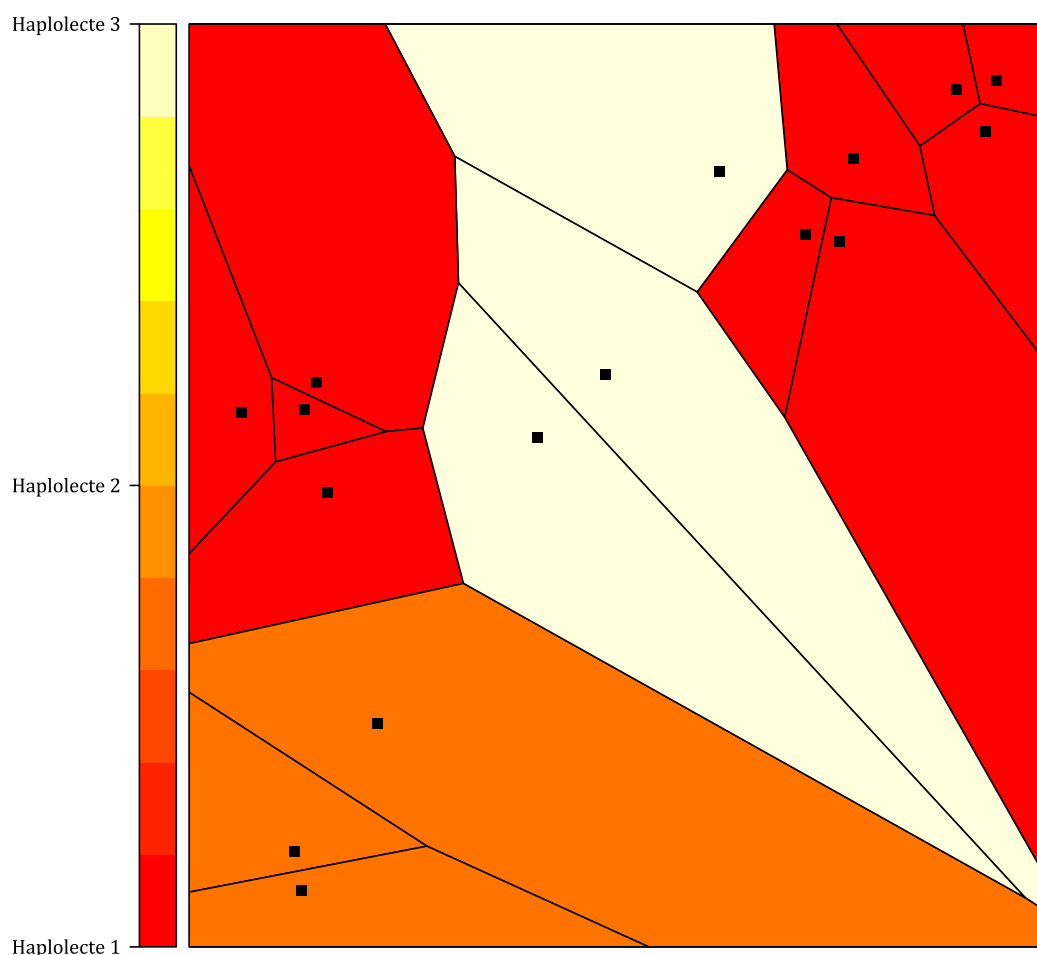


Carte II.4.27 : Répartition géographique des haplolectes de la PHA — Carte à point

L'analyse de la carte linguistique permet de mettre en évidence que l'haplolecte "septentrional et occidentalo-central" constitue la variété dialectale la plus importante. Les deux autres sont d'une dimension moindre. D'autre part, nous constatons une discontinuité géographique entre les variétés septentrionales et les variétés occidentalo-centrales. Ce fait est à corrélérer avec ce que nous avons déjà constaté sur certaines cartes linguistiques dans la partie Géolinguistique (Partie sur les Consonnes géminées II.4.1.4, page 119).

Toujours avec la même partition, nous avons représenté les haplolectes dans une carte choroplèthe. Les méthodes de PHA sont parfaitement adaptées à ce genre de représentation. En faisant varier la teinte, nous pouvons mieux décrire les distances

linguistiques. Comme, nous pouvons le voir sur la carte, les cellules sont très grandes. Cela s'explique par la faible densité de points d'enquête. Le diagramme de Voronoï établit le découpage d'un espace en se basant sur des points géométriques. Plus, il y a une densité de points importante dans un espace donné et plus les cellules seront petites.



Carte II.4.28 : Répartition géographique des haplolectes de la PHA — Carte choroplèthe

Cela étant, les résultats escomptés ne sont pas sans intérêts. Car ils renforcent visuellement davantage la rupture qui existe entre le nord et le sud. Nous pouvons constater que les variétés les plus méridionales sont plus proches, linguistiquement parlant, des variétés septentrionales que des variétés orientalo-centrales. Alors qu'elles sont bien plus proches géographiquement. Ce qui constitue un "paradoxe", car la distance géographique n'est pas corrélée à la distance linguistique.

Nous en avons fini avec la Partition hiérarchisée, nous allons voir ce qu'il en est avec les autres méthodes. Dans cette partie, nous verrons également plusieurs méthodes et

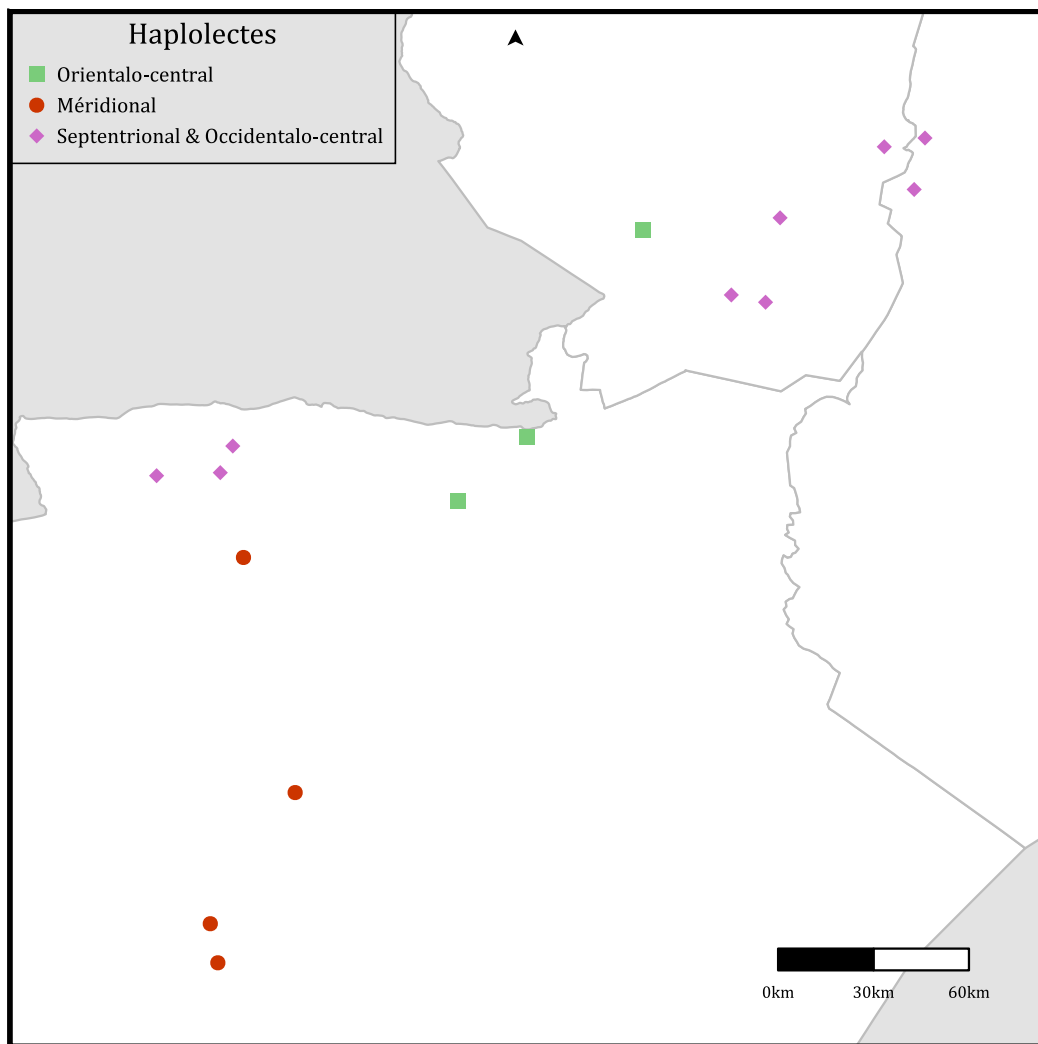
cela sera l'occasion de les comparer.

II.4.2.2.1.2 Partition non hiérarchisée

Après application de la méthode PHA, nous allons employer l'autre grande catégorie de méthodes qui font partie de la méthode par Apprentissage Non Supervisé. À cet effet, nous avons choisi de soumettre nos données à deux méthodes différentes que sont la méthode k-moyenne et la méthode k-mode.

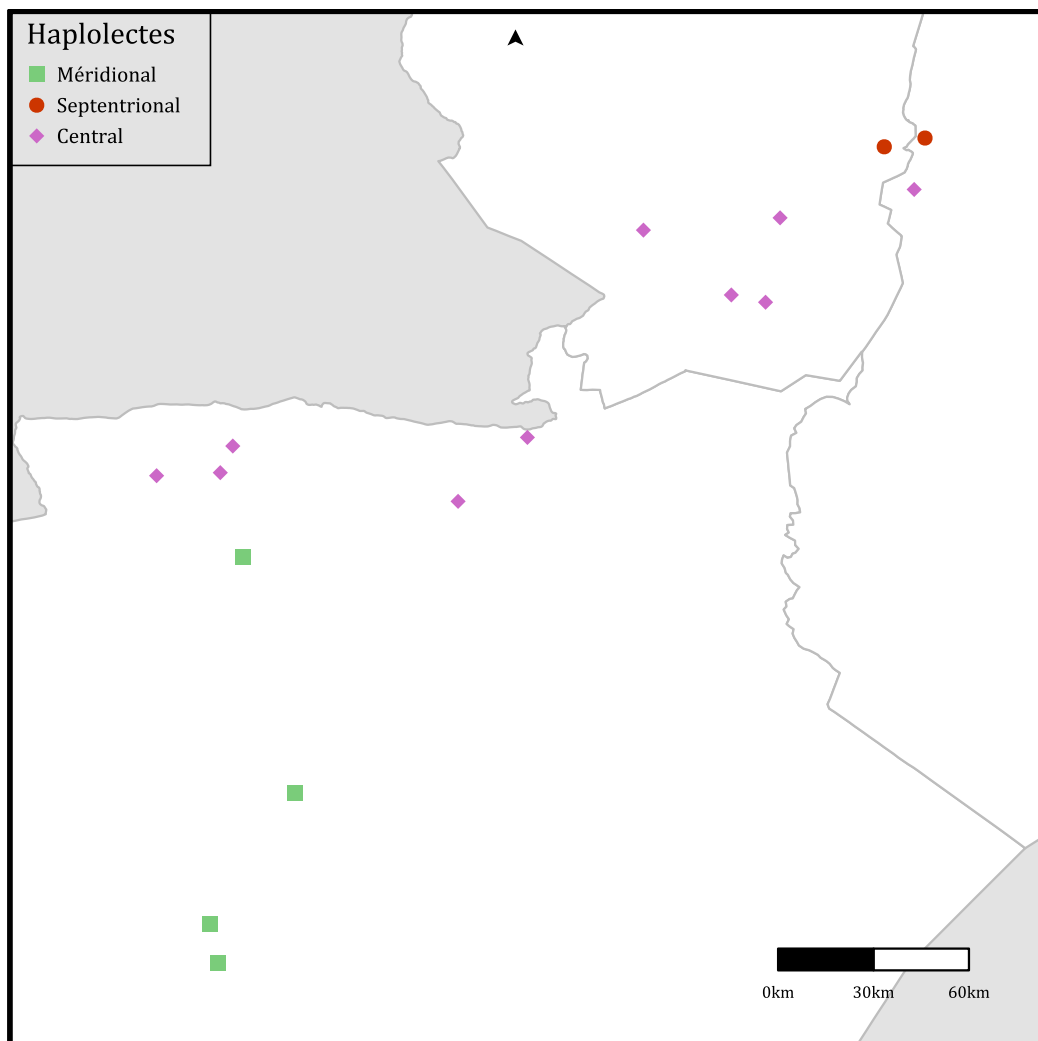
Ces deux méthodes partagent une aptitude commune. Leur principe est le même. En fonction du nombre de partitions adoptées, elles choisissent au hasard des variétés dialectales comme étant le "centre" de ces partitions. Puis, les autres variétés vont s'agréger autour de ces "centres". Et ce choix hasardeux est d'une importance capitale. Pour rendre compte de l'influence de ce procédé initial, nous allons successivement réaliser deux fois un calcul pour chacune de ces méthodes.

Nous allons commencer par aborder la dernière méthode que nous venons d'évoquer. Pour cette première phase, la méthode k-mode, à une variété près, a la même configuration que la partition que nous avons obtenue lorsque nous avons réalisé la PHA. Alors que la variété de Ouakda était rattachée aux variétés occidentalo-centrales précédemment, elle est cette fois-ci comprise dans les variétés méridionales. Cependant, l'haplolecte "septentrional et occidentalo-cenral" reste toujours le groupe linguistique le plus important.



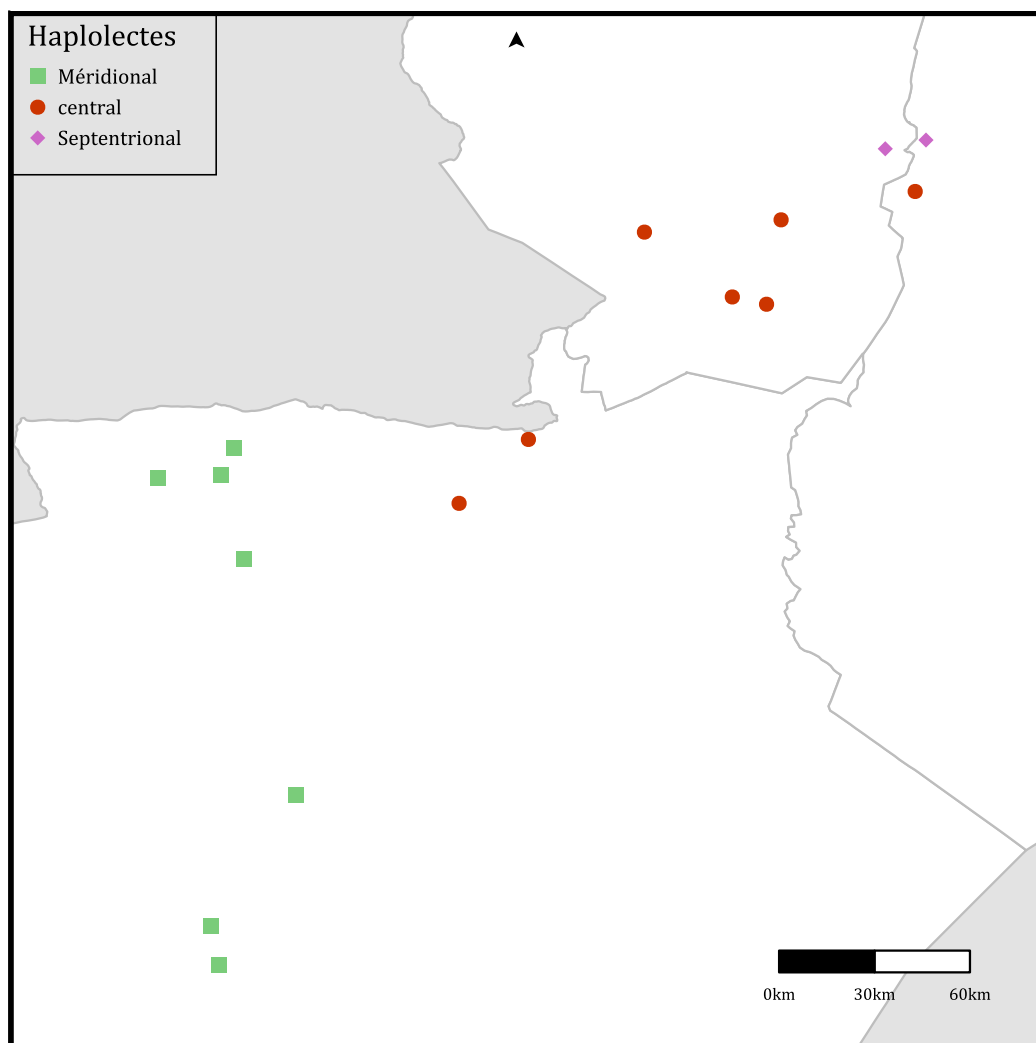
Carte II.4.29 : Répartition géographique des haplolectes de la méthode k-mode — Carte à point

Si nous réalisons une seconde partition de ces mêmes données, alors nous obtenons une nouvelle partition qui est différente de la première. Lors de cette nouvelle étape, de nouvelles variétés dialectales sont considérées comme étant les "centres" des partitions et donc cela aboutit à de nouveaux résultats. L'haplolecte "septentrional et occidentalo-central" disparaît au profit d'une nouvelle partition regroupant les ex-variétés centrales et une partie des ex-variétés septentrionales. Ce dernier constitue la plus grande partition qui est obtenue par cette deuxième phase. D'autre part, nous constatons que le continuum linguistique est désormais bien assuré.



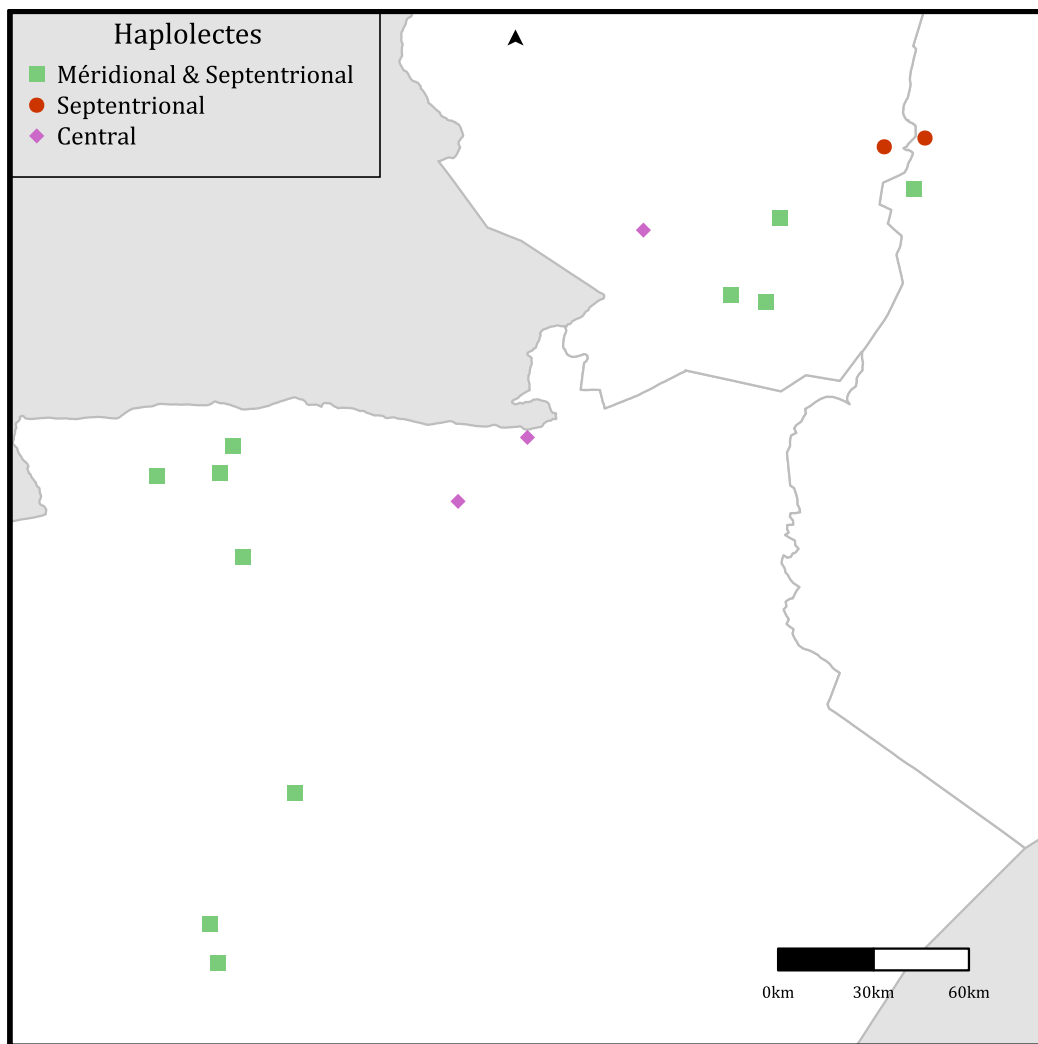
Carte II.4.30 : Nouvelle répartition géographique des haplolectes de la méthode k-mode — Carte à point

Regardons, maintenant, ce qu'il en est avec la méthode k-moyenne. Comme précédemment, nous allons réaliser deux phases de calcul. Au cours de cette première étape, contrairement à la méthode k-mode, avec la méthode k-moyenne, les choses changent presque radicalement dès le début. Nous voyons surgir deux grands haplolectes que nous avons désignés sous les termes de central et méridional.



Carte II.4.31 : Répartition géographique des haplolectes de la méthode k-moyenne — Carte à point

De plus, la rupture spatiale que formait l'haplolecte orientalo-central n'existe plus. Les haplolectes se suivent les uns après les autres. Mais si nous réalisons une nouvelle "fournée" de calcul, alors nous parvenons à l'observation inverse. Sur la carte linguistique ci-dessous, il n'y a plus de continuum. Le haplolecte "central" ressurgit et le méridional va même s'agrandir en regroupant les ex-variétés méridionales et une partie des ex-variétés septentrionales.



Carte II.4.32 : Nouvelle répartition géographique des haplolectes de la méthode k-moyenne —
Carte à point

II.4.2.2.2 Discussion

Il semble que les calculs qui établissent la distance linguistique entre deux variables dialectales a peu d'influence sur les résultats. Nous interprétons cette situation comme la conséquence d'une réalité extra-linguistique. Du fait que les variables linguistiques deviennent des variables numériques, elles vont être régies par des lois mathématiques. L'une d'elle est la loi commutative. C'est ce fait qui est à l'origine de ces similitudes. Tant

que le basculement de la variable qualitative à la variable numérique se fait en prenant en compte les autres observations linguistiques ou faits linguistiques. Et tout en faisant en sorte que les valeurs affectées soient fonction des autres variables afin de garder le même ordre de grandeur entre les variables. Alors cette loi s'appliquera et le cumul des ordres de grandeur sera le même quel que soit le calcul numérique utilisé. À la condition que les ordres de grandeur ne soient pas bouleversés. Donc, *in fine*, il se produira inévitablement une partition identique.

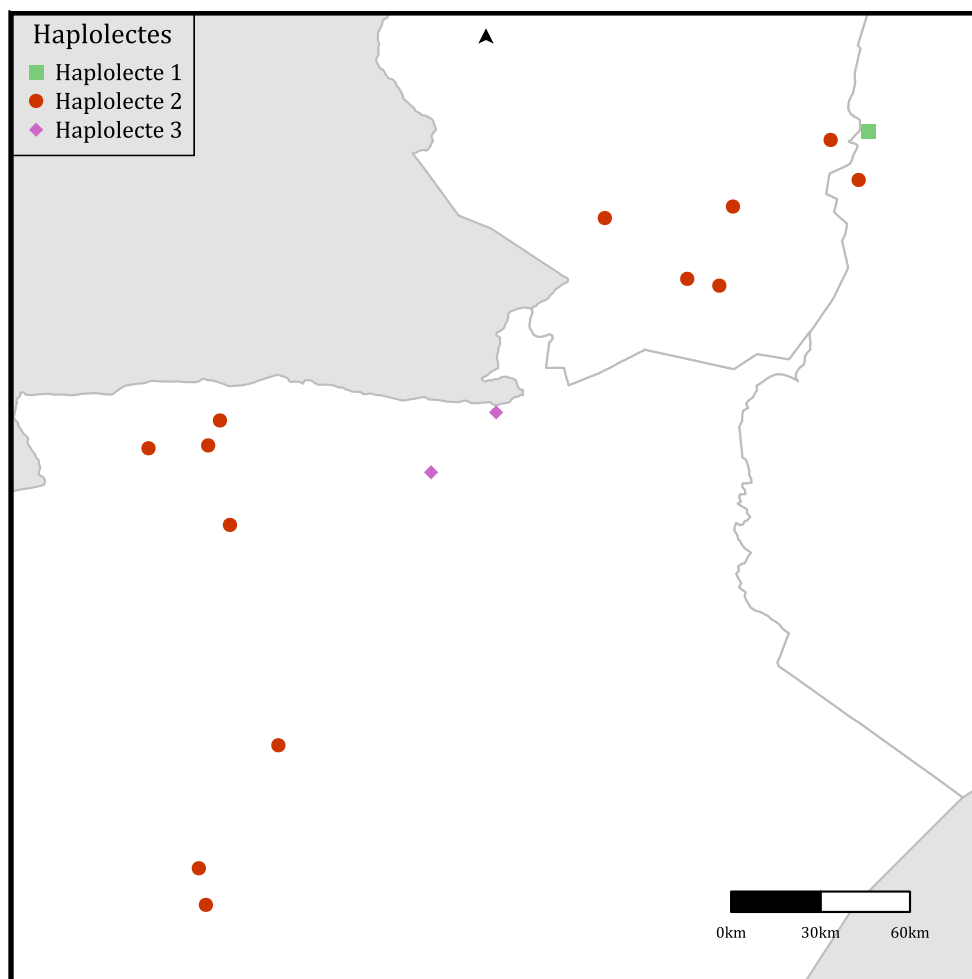
Un autre aspect à signaler, si nous comparons les partitions de ce dendrogramme avec ceux du **MDS**, nous nous apercevons que les résultats divergent. Sur le deuxième axe de la composante principale, nous voyons clairement que les variétés de Bousseghoun sont plus proches d'Asla que de Lahmar, mais pourtant sur la matrice de distance les choses sont différentes. C'est l'inverse qui se produit.

Par ailleurs, nous constatons que les variétés, qui dans le scatterplot — ou biplot — sont à proximité de la jonction des axes des deux composantes principales, sont celles qui basculent et alternent d'une partition à l'autre suivant les méthodes utilisées. Alors que les variétés bien déterminées par le **MDS** en ayant une projection dans le scatterplot totalement opposée sont celles qui ne se croisent jamais. Si cette corrélation est pertinente, alors en utilisant le **MDS**, il sera possible de prévoir quelle variable dialectale sera susceptible d'avoir une partition fluctuante. Cette fluctuation, que nous avons pu constater notamment pour les méthodes k-moyenne et k-mode, montre à quel point il faut prendre les résultats fournis avec circonspection.

Cela sous-entend également que les calculs mathématiques à la base de ces méthodes ne doivent pas différer fondamentalement. Puisqu'au final, les variétés définies comme étant totalement opposées par le **MDS**, le sont également dans les autres méthodes. Nous savons que le **MDS** a recours à la métrique du χ^2 . Et comme nous l'avons vu ce type de mesure ne donne pas de résultats plus différents que la méthode Simple Matching. Donc, cela expliquerait pourquoi nous retrouvons systématiquement bien identifié à part les variétés d'Asla et de Chellala, les variétés de Mazzer et d'Igli et parfois les variétés de Lahmar et Boukais — Le groupe d'Asla s'oppose au groupe d'Igli sur les deux axes, alors que le groupe de Lahmar s'oppose aux deux autres que sur un seul axe —.

Mais ce qui vient d'être affirmé n'est pas nécessairement exact pour toutes les méthodes. Car cela n'est pas le cas avec la méthode ROCK. Cette dernière a aussi été utilisée et les résultats qui en découlent sont très différents de ce que nous avons pu voir au cours des autres méthodes. Dans le cas de la méthode ROCK, nous constatons

que la partition se fait suivant une logique différente. Comme nous pouvons le voir sur la carte linguistique II.4.33. Il existe deux partitions qui correspondent uniquement à un nucléolecte. C'est-à-dire que ROCK considère que ces deux derniers ne présentent, d'une certaine manière, aucune parenté avec les autres variables. Ce fait nous est difficile à admettre et nous pourrions penser que le calcul mené s'est fait de manière aléatoire. Sauf que ce n'est nullement le cas. À tout bien réfléchir, il faut se remémorer les constats que nous avons énoncés dans la partie géolinguistique à savoir les particularismes linguistiques de certaines variétés linguistiques.



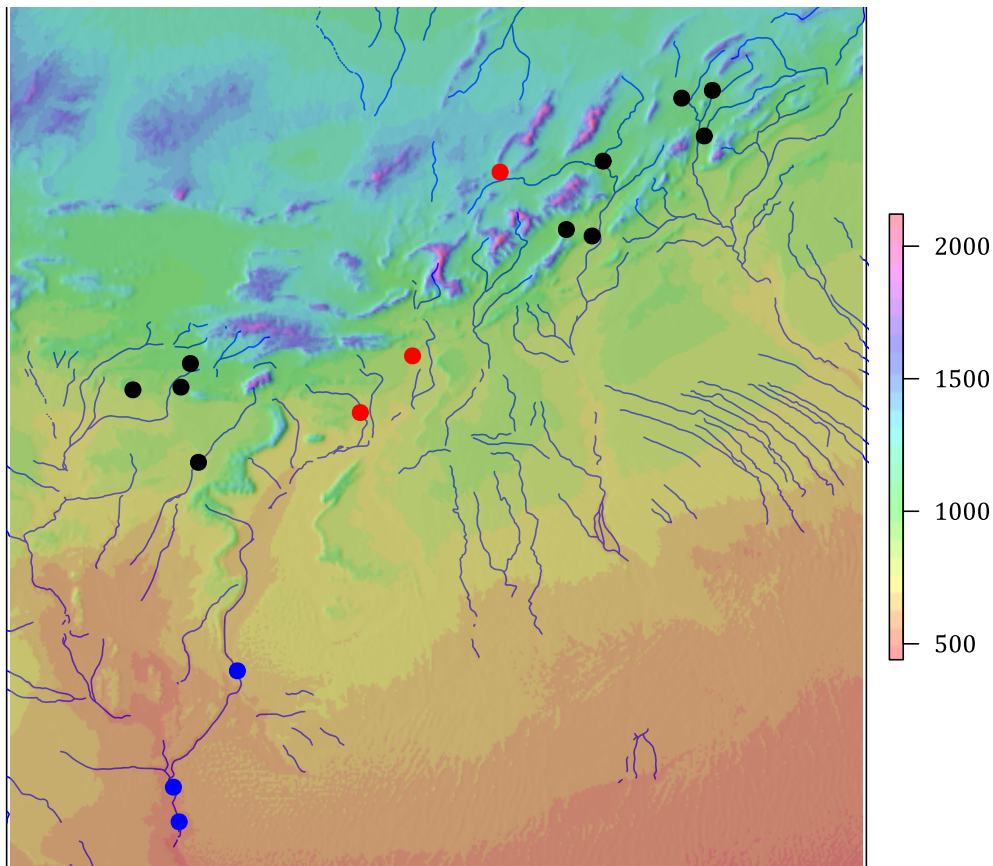
Carte II.4.33 : Répartition géographique des haplolectes de la méthode ROCK — Carte à point

C'est cette caractérisation que nous retrouvons lorsque nous observons le résultat fourni par ROCK. Les deux nucléolectes qui ont été séparés par cette méthode correspondent aux groupes de localités de Beni Ounif / Fendi et Chellala. Ces deux

derniers avaient des particularités phonétiques que nous ne retrouvions que dans ces seules oasis. Donc, il semble que ROCK mette en avant les particularismes plus que les correspondances ce qui est, comme nous l'avons dit, le contraire de la démarche en dialectométrie. Cependant, les propriétés de ROCK peuvent trouver des applications ailleurs en **SDL**, voire même en dialectologie si un axe de recherche nouveau apparaît et se développe dans le futur.

La question de la partition n'est pas le seul élément d'analyse que nous avons mené. Grâce aux cartes linguistiques, il nous a été possible aussi de corréliser ces partitions avec l'espace géographique. Cela nous a permis donc de constater que l'espace qu'occupait ces variétés dialectales n'était pas régulier, ni d'ailleurs équilibré. La partition, quelle que soit la méthode, tend à établir de grands groupes dialectaux à la défaveur d'autres qui restent de petites dimensions. Ce déséquilibre doit nous interroger. Nous n'avons pas de début de réponse à proposer pour expliquer ce fait. Probablement que c'est lié à la nature des calculs et non pas aux observations linguistiques. Car, suivant les méthodes, les variétés dialectales qui font parties de la plus grosse partition ne sont jamais les mêmes, alors que les observations linguistiques, elles, restent toujours identiques.

En outre, la réalité géographique peut être combinée de plein de façons différentes avec la variation linguistique. Au lieu d'utiliser des cartes linguistiques dont le fond de carte est gommé de toutes références topographiques, nous pourrions les utiliser. Car il est possible d'agencer en plus sur ces cartes linguistiques des éléments topographiques. Cet enrichissement peut peut-être donner des clés intéressantes pour comprendre la variation linguistique. La carte que nous avons affichée ci-dessous **II.4.34** entreprend de mêler les suggestions que nous venons de formuler.



Carte II.4.34 : Relief et variation

Cette carte reprend la partition réalisée par la méthode PHA. Nous pouvons faire un premier constat qui est que tous nos points d'enquête se trouvent à proximité d'un cours d'eau. Ensuite, sur cette carte est aussi représentée l'élévation. Donc, les zones violacées indiquent des chaînes de montagnes alors que les zones rougeâtres correspondent à des espaces peu élevés qui concordent avec le désert sableux. Maintenant que les choses sont posées, nous pouvons nous apercevoir que l'explication n'est pas à chercher du côté de la topographie. Cette dimension géographique n'aide pas à comprendre pourquoi le continuum linguistique n'est pas préservé. Sfissifa est séparé de Beni Ounif par des chaînes de montagne, donc les voies de communication sont plus compliquées entre les deux cités. Par ailleurs, la distance spatiale est plus grande entre Sfissifa et Beni Ounif qu'entre Moghrrar et Beni Ounif. Pourtant, Beni Ounif et Fendi partageront plus de spécificités — l'affrication notamment — avec Sfissifa. Donc, les enjeux doivent être cherchés davantage du côté de la sociologie.

Effectivement, il s'avère que la région septentrionale et la région orientalo-centrale —

tout particulièrement Figuig — ont une longue histoire d'unions maritales. Beaucoup des mères, des épouses, des grands-parents des habitants des variétés du nord proviennent de Figuig. Ce fait nous a été rapporté par nos informateurs sur place. Alors peut-être que ce point est à l'origine de cette proximité linguistique. Donc, il faut supposer que Sfisifa aurait davantage de liens matrimoniaux avec les variétés orientalo-centrales que les autres localités environnantes. Pour l'asserter, il faudrait mener une étude sociologique. Procéder à une telle entreprise peut rentrer dans le cadre d'un travail de dialectologie. Mais au lieu cette fois-ci de lier des variables du terrain d'étude avec des variables de l'objet d'étude, ce seront des variables de la population d'étude qui seront combinées avec des variables linguistiques.

II.4.2.3 Apprentissage Supervisé

L'autre méthode qui a cours en **SDD** est l'Apprentissage Supervisé. Les techniques s'inscrivant dans ce domaine sont nombreuses comme nous l'avons vu. Il ne s'agira pas pour nous de les voir en long et en large, nous allons axer notre recherche sur quelques-unes. Nous ne ferons pas comme dans la partie Apprentissage Non Supervisé une comparaison de nombreuses méthodes. Les techniques d'Apprentissage Supervisé ont leurs propres caractéristiques et surtout leur propre limite comme nous le verrons. De ce fait, nous utiliserons seulement deux algorithmes.

Ainsi, nous avons annoncé dans la partie méthodologie que nous allons faire usage des méthodes **CART** et Random Forest. Et c'est toujours notre intention, néanmoins nous avons quelque peu changé notre marche à suivre. Car, contrairement à ce qui a été possible de faire précédemment, les méthodes d'Apprentissage Supervisé demandent plus de données et des données à plusieurs niveaux et bien distribuées. Par conséquent, nous avons augmenté nos variables pour pouvoir utiliser cette méthode.

II.4.2.3.1 Le matériau d'analyse

La première fois que nous avons utilisé cette méthode les résultats que nous avons obtenus furent incomplets. Nous avons repris le même nombre de variables que pour la méthode d'Apprentissage Non Supervisé. Mais ce nombre est insuffisant. Comme nous pouvons le voir sur l'arbre de décision ci-dessous, l'algorithme utilisé ne parvient pas à faire émerger toutes les variables dialectales. Sur seize, seules onze peuvent être visualisées. Cela s'explique par le fait que les autres localités sont superposées

graphiquement avec d'autres nucléotectes, parce qu'ils partagent les mêmes faits linguistiques.

CART model tree

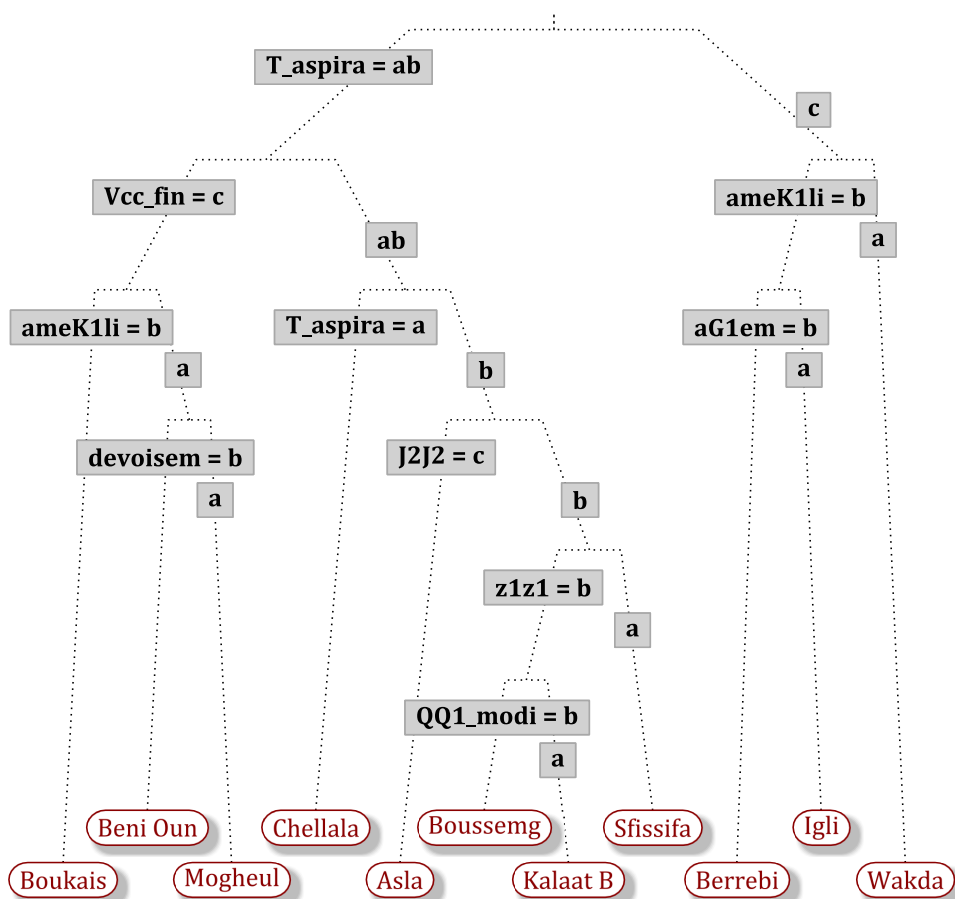


FIGURE II.4.3 : Échec de la partition des variables dialectales — méthode CART

Pour palier à cette difficulté, nous avons bien évidemment augmenté le nombre de nos variables linguistiques jusqu'à atteindre vingt-cinq variables. Ces derniers correspondent à des faits linguistiques ne se rapportant pas uniquement à des unités phoniques. Nous avons pris en compte également des unités lexicales et grammaticales. Nous ne précisons pas ici quelles furent les données que nous avons rajoutées. C'est un point subsidiaire qui n'est pas d'une importance fondamentale pour les questions qui nous intéressent. Les résultats avec cette nouvelle donne sont présentés ci-dessous.

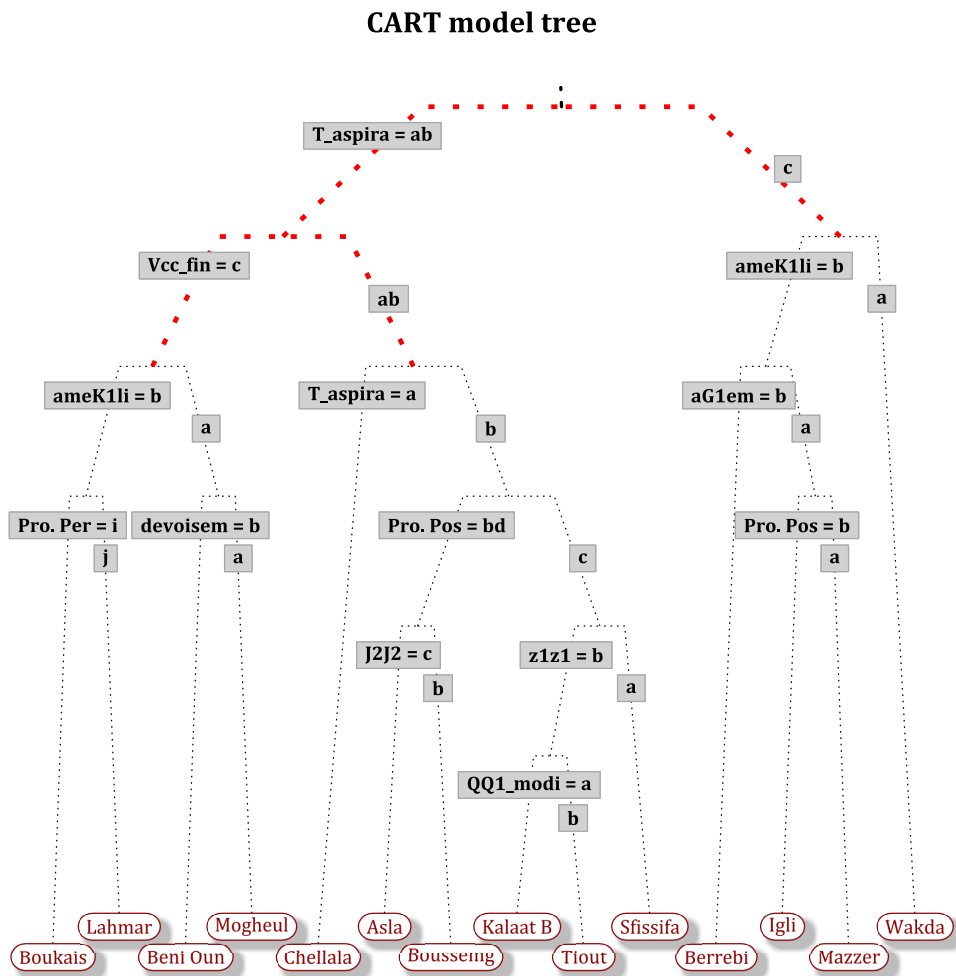


FIGURE II.4.4 : Partition des variables dialectales — méthode CART

Si nous observons le graphique, nous pouvons nous rendre compte qu'il y a plus de variables dialectales qui ont été classées, mais certaines localités manquent. Sur la figure II.4.4, il n'y a que quatorze nucléolectes de présent. Fendi et Igli ne sont pas visibles, car ils sont respectivement camouflés par Beni Ounif et Mazzer. En dehors de ce souci graphique toujours persistant, nous voyons apparaître trois grandes branches d'haplolectes illustrées par les lignes rouges pointillées.

Le produit de **CART** confirme le nombre de partitions trouvé lors de l'Apprentissage Non Supervisé avec la méthode **MDS**. Ainsi, la partition tout à gauche correspond aux variétés centrales, celle au centre renvoie aux variétés septentrionales, alors que la branche tout à droite renvoie aux variétés méridionales. Toutefois, les variables dialectales ne sont pas réparties exactement de la même manière. De plus, la qualité du

classement peut être remise en cause. Le seul cas qui peut sembler vraiment discutable est le regroupement de Ouakda avec Igli, Mazzer et Berrebi.

Pour vérifier si ces partitions sont conformes, il est possible de reprendre la méthode **CART** et de l'appliquer une nouvelle fois, en considérant cette fois-ci comme attribut cible non plus les variables dialectales en elles-mêmes, mais les haplolectes mis en évidence précédemment par **CART**. En parallèle, nous avons décidé d'utiliser la méthode Random Forest en reprenant le même classement précédent.

	CART			Random Forest		
	Central	Méridional	Septentrional	Central	Méridional	Septentrional
Central	5			5		
Méridional		4		1	3	
Septentrional			7			7

TABLE II.4.6 : Les matrices de confusion des méthodes CART et Random Forest

Comme nos partitions des variables dialectales sont fondées sur la méthode **CART**, il n'y a aucune surprise à ce que la matrice de confusion pour cette méthode ne présente aucun décalage. À l'inverse, les prédictions de Random Forest indiquent qu'il y a une variable dialectale qui serait mal classée. Selon cette dernière méthode, un des nucléolectes aurait plutôt sa place parmi les variétés centrales au lieu d'être classé dans les variétés méridionales. Ainsi, cela confirme notre intuition au sujet de Ouakda.

Si nous tenons compte des résultats de Random Forest, alors nous pouvons considérer que le classement de départ n'est pas totalement parfait. Nous pouvons donc envisager de réajuster nos partitions en intégrant Ouakda dans les variétés centrales. En réalisant ce réajustement, nous obtenons de nouvelles matrices de confusion qui fournissent une concordance parfaite entre nos partitions et les prédictions pour les deux méthodes. Autrement dit, aucun faux positif ou faux négatif n'est trouvé.

II.4.2.3.2 Discussion

Nous venons de voir la méthode d'Apprentissage Supervisé qui nous a permis de montrer que cette méthode pouvait être aisément utilisée pour traiter des variables linguistiques. Par ailleurs, l'autre aspect original de notre travail est que notre approche englobe l'ensemble du processus d'analyse de données. Nous soulignons ce point, car, généralement, cette méthode est employée pour évaluer la robustesse des partitions

obtenues lors de l'Apprentissage Non Supervisé. Dans notre cas, les regroupements de variables dialectales ont été trouvés en appliquant **CART**, donc nous n'avons pas tenu compte des résultats de l'Apprentissage Non Supervisé. Ainsi, ce n'est qu'ensuite que nous avons examiné si ces haplolectes étaient pertinents. Cette dernière analyse a été conduite en reprenant **CART** et en introduisant en même temps Random Forest.

En procédant de cette manière, dans un premier temps, nous nous sommes rendu compte que Random Forest prédisait des partitions différentes. Cela nous a permis d'envisager qu'il fallait réajuster les regroupements. C'est ce que nous avons fait. À notre surprise, avec ce nouveau partitionnement, nous nous attendions à ce que **CART** le rejette, il en fut tout autre. Cela nous a intrigué, cependant nous n'avons aucun élément de réponses pour expliquer cet accommodement. Il semble que **CART** permet des configurations plus souples qui peuvent être des solutions possibles de partitions, alors qu'avec Random Forest, le choix est bien plus restreint. Ce dernier fait n'est en rien surprenant, car cette méthode est plus adaptée pour les cas complexes comme le sont nos données. Dès lors, pour parvenir à expliquer ces divergences, une étude plus minutieuse des algorithmes serait nécessaire pour comprendre les orientations prises par **CART**. Du fait que ce domaine ne s'inscrive pas dans notre champ de compétences, nous sommes dans l'incapacité de pousser plus loin cet examen.

D'autre part, un des avantages de **CART** et de Random Forest, c'est qu'ils réalisent en même temps une cross-validation des données en essayant de vérifier si les résultats obtenus au départ sont reconfirmés au cours d'autres phases de calculs. Pour cela, **CART** et Random Forest emploient tous deux la méthode k-fold qui consiste à segmenter aléatoirement les variables linguistiques en part égale et à appliquer à chaque segment le même algorithme. Mais pour que les résultats prédictifs soient pertinents, il faut des données en nombre, sinon rien ne permet d'affirmer que cela n'est pas le fruit du hasard. Cette question de quantité ne concerne pas uniquement que la précision des prédictions, ce point a aussi son importance pour la partition des données. Puisque, certaines méthodes comme le Conditional Inference Trees n'arrivent pas à former des regroupements de variables dialectales en raison de ce problème quantitatif.

Ainsi, bon nombre des méthodes utilisées — et que nous avons testées — en Apprentissage Supervisé présentent cette particularité. Elles sont conçues pour traiter des données en grande profusion. Toutefois cela ne veut pas dire que ce fait constitue un obstacle infranchissable. Il est aussi possible en linguistique d'atteindre cette exigence. Après tout, la richesse d'un système linguistique est vaste. En revanche, cela suppose une

enquête extensible. Donc, un chercheur devra collecter des données en masse à divers endroits d'un terrain d'étude dans un temps raisonnable. À notre avis, cela pourrait décourager beaucoup de chercheurs d'entreprendre ce genre d'études sachant ce qu'ils doivent réunir comme condition.

Une autre limite à mettre sur le compte des méthodes d'Apprentissage Supervisé, c'est que certaines sont basées sur des mesures statistiques. Donc, cela suppose de faire l'hypothèse que nos données suivent un modèle distributionnel particulier — par exemple une loi normale —. Mais il est difficilement envisageable de soutenir cette hypothèse. Car nous ne pouvons pas dire qu'un système linguistique suive une quelconque forme de centralité — à part par exemple pour des mesures menées sur des variables phonétiques —.

L'autre aspect à dire à l'encontre de ces méthodes, c'est qu'elles ne proposent pas de rendus visuels de leur résultat. La méthode **CART** est l'une des seules à offrir la possibilité de créer un graphique. Avec cette dernière, la lecture des résultats peut se faire de manière visuelle. En revanche, avec d'autres, cela ne sera pas le cas. Par exemple, dans les méthodes Random Forest et Bagging que nous avons aussi utilisées, les résultats sont difficilement accessibles. Pour pouvoir les présenter dans une étude scientifique, il faut les mettre en forme manuellement par soi-même.

Pour l'instant, nous avons fourni que des critiques négatives. Malgré cela, nous voudrions finir par dire que tout de même, cette méthode doit faire partie de l'arsenal du dialectologue. L'Apprentissage Supervisé est très développé dans plein de domaines. D'ailleurs, ces derniers manipulent plus de variables linguistiques — des chaînes de caractère plus exactement — qu'autres choses. Alors, il n'y aurait pas de raisons qu'il en soit autrement en dialectométrie. Car l'avantage qu'offre cette méthode, outre de classer les variétés dialectales, c'est qu'elle indique quelles sont les variables vraiment distinctives — Sur les figures, cela correspond aux valeurs affichées au niveau des nœuds des arbres —. Celles qui expliquent plus la variation que d'autres. Ce fait peut être exploité par le dialectologue à d'autres fins.

II.4.2.4 Synthèse

Ce travail exploratoire démontre que les outils utilisés pour partitionner les différents nucléoctes n'aboutissent pas tous au même résultat. Mais que d'autre part, les partitions pouvaient être totalement différentes. Ce constat a été réalisé sur un petit échantillon

de méthodes et il en reste bien d'autres que nous n'avons pas présentées. Et ces autres méthodes donnent toutes aussi des résultats différents.

Donc, cela doit nous alerter sur le fait de se garder de conclusions définitives. Les partitions automatiques se basent sur des calculs qui n'ont rien avoir avec la réalité cognitive. Établir des distances entre variables linguistiques, puis entre variables dialectales est assurément un procédé très approximatif. Modéliser un système linguistique par des chiffres peut paraître un gage de rigueur, cependant cela est loin d'être le cas.

D'autant plus, si les variables linguistiques servant à caractériser les variétés dialectales sont relativement peu nombreuses. Comme c'est notre cas. D'ailleurs, cette question quantitative ne se pose pas de la même manière suivant les méthodes. Les méthodes d'Apprentissage Non Supervisé peuvent passer outre, toutefois ce n'est pas le cas pour les méthodes d'Apprentissage Supervisé. Donc, un dialectologue sera quelle méthode retenir en fonction du nombre de données qu'il a. Il est inutile de mener une étude dialectométrique en se basant sur la méthode d'Apprentissage Supervisé si les données sont limitées.

Le fait d'éprouver ces méthodes n'a pas été notre seul objectif. Pourvoir dégager une partition fiable des variables dialectales fut aussi l'un d'eux. Si nous recoupons les résultats obtenus lors de l'Apprentissage Supervisé et de l'Apprentissage Non Supervisé, alors il apparaît que les différents nucléolectes peuvent être classés dans trois grands haplolectes. Le contour de ces derniers n'est pas établi encore avec certitude, néanmoins nous pouvons affirmer que les variétés du Sud-Oranais s'insèrent au moins dans l'un des haplolectes que nous avons dénommés : septentrional, méridional et central.

En perspective, un autre point aussi à souligner est que l'haplolecte septentrional et l'haplolecte central sont proches l'un de l'autre par rapport à l'haplolecte méridional. Les variétés dialectales retrouvées au sein de ce dernier haplolecte furent considérées, depuis André Basset, comme une partie intégrante des variétés du Sud-Oranais. Nous n'avons pas remis en cause cette idée au cours de nos analyses, car nous sommes contraint à ne pas sortir des limites fixées à notre terrain d'étude. Cependant, nous avançons l'hypothèse que pour le berbère des oasis d'Igli et de Mazzer, notamment, il y a une forte présomption selon laquelle ces variétés pourraient être incluses dans une autre branche dialectale. Nous pensons aux variétés qui se situent dans la wilaya d'Adrar — Gourara & Touat—. Le berbère présent dans cette région semble très analogue à l'haplolecte méridional au point qu'ils pourraient former un même haplolecte. Cela n'est

pour l'instant que des suppositions, seule une autre étude dialectométrique pourrait affirmer ou infirmer cette idée.

CHAPITRE **II.5**

Appendice

— Remarques :

1. Le corpus est séparé en deux parties. Le premier tableau correspond aux items et le second tableau comporte les gloses des items. La glose et les items sont présentés dans le même ordre de présentation.
2. Les formes entre parenthèses correspondent à des variantes en usage dans les différentes localités :
 - (x/y) prononcé soit x, soit y en fonction des variétés ;
 - (x) cette unité phonique est facultative, elle est prononcée ou non en fonction des individus ou en fonction des variétés ;
 - (V) plusieurs voyelles peuvent être réalisées en fonction des variétés
3. La glose proposée est approximative. Le sens de ces items peut d'un point d'enquête à un autre être différent. Seules les particularités grammaticales seront identiques.
4. Tous les items présents dans ces tableaux ne sont pas nécessairement usités dans toutes les localités du Sud-Oranais. La transcription utilisée correspond à la transcription usuelle en usage à l'INALCO .
5. Nous avons placé a^{G_{1a}}ujil sous la variable ^{G_{1a}}, mais nous n'avons aucune certitude quant au cognat phonologique dont il dépend. Car ce cognat morphémique n'a pas été retrouvé à Berrebi.
6. Pour certaines variables — Exemples : ^{G_{1a}} , ^{K₁} —, nous n'avons pas indiqué toutes les données que nous avons en notre possession, alors que pour d'autres variables, ce sont les seules données que nous avons.

Variables	^G ₁		^G ₂		^G ₃		^G ₄		^G ₅		^G ₆	
Sous variables	textasciicircum G _{1a}	textasciicircum G _{1b}	textasciicircum G _{2a}	textasciicircum G _{2b}	textasciicircum G _{3a}	textasciicircum G _{3b}	textasciicircum G _{4a}	textasciicircum G _{4b}	textasciicircum G _{5a}	textasciicircum K _{2a}	textasciicircum K _{2b}	textasciicircum K ₄
Nouvelles Sous variables	textasciicircum G _{1a}	textasciicircum G _{1b}	textasciicircum G _{2a}	textasciicircum G _{2b}	textasciicircum G _{3a}	textasciicircum G _{3b}	textasciicircum G _{4a}	textasciicircum G _{4b}	textasciicircum G _{5a}	textasciicircum K _{2a}	textasciicircum K _{2b}	textasciicircum K ₄
Glose	orphanin marcher(2.SG.IMP) long grandir(2.SG.IMP) suspendre(2.SG.IMP) puiser(2.SG.IMP) s'envoler(2.SG.IMP)	(adv) coudre(2.SG.IMP)	faucille faucilles	oiseau pantalon invité orphanin être supérieur(2.SG.IMP) trembler(2.SG.IMP)	homme mâtis cabas houe mosquée	livre	vert	noir brûler(2.SG.IMP) dérober(2.SG.IMP) labouer(2.SG.IMP) dépêner(2.SG.IMP) dépêner	viande viandes	prendre(2.SG.IMP) prendre(2.SG.IMP_INT)	prendre(2.SG.IMP) prendre(2.SG.IMP_INT)	terre lait
Variables	^T & ^TT											
Sous variables	^T											
Glose	étoile pigeon entrer(2.SG.IMP) mourir(2.SG.IMP)			oublier(2.SG.IMP)								
Variables	^G (fricative / occlusion / affrication)											
Sous variables	^GG ₁											
Glose	un jurer(2.SG.IMP) jeter_un_oeil(2.SG.IMP) oreille		pétrir(2.SG.IMP_INT) moissonner(2.SG.IMP_INT)	grandir(2.SG.IMP_INT) marcher(2.SG.IMP_INT) jeter(2.SG.IMP_INT) pétrir(2.SG.IMP)		gouverner(2.SG.IMP) gale	laisser(2.SG.IMP)	compter(2.SG.IMP_INT) prendre(2.SG.IMP_INT)				maître connaître(2.SG.IMP) être_vieux(2.SG.IMP)
Variables	^ZZ											
Sous variables	^ZZ ₁											
Glose	chercher(2.SG.IMP)			courir(2.SG.IMP) chauffer(2.SG.IMP)								
Variables	^BB											
Sous variables	^BB ₁											
Glose	tiret(2.SG.IMP_INT) couper(2.SG.IMP) éduquer(2.SG.IMP)			pière								avoir_peur(2.SG.IMP) descendre(2.SG.IMP_INT) aller(2.SG.IMP_INT) genre devenir(2.SG.IMP_INT) rouge
Variables	^QQ											
Sous variables	^QQ ₁											
Glose	grandir(2.SG.IMP_INT) chauffer(2.SG.IMP_INT)			cruser(2.SG.IMP_INT) tourner(2.SG.IMP_INT) percer(2.SG.IMP_INT) attacher(2.SG.IMP) lire(2.SG.IMP_INT) grand être_sec(2.SG.IMP_INT)								

TABLE II.5.2 : Les gloses des items

Deuxième partie

Étude des Rhotiques alvéolaires : Phonétique expérimentale et Phonologie

CHAPITRE III.1

Introduction

Les études de phonétique et de phonologie sont les parents pauvres des études berbères. Notamment les études de phonétique comme le rappelle Kamal Naït-Zerad [Naït-Zerrad, 2004, p.15] :

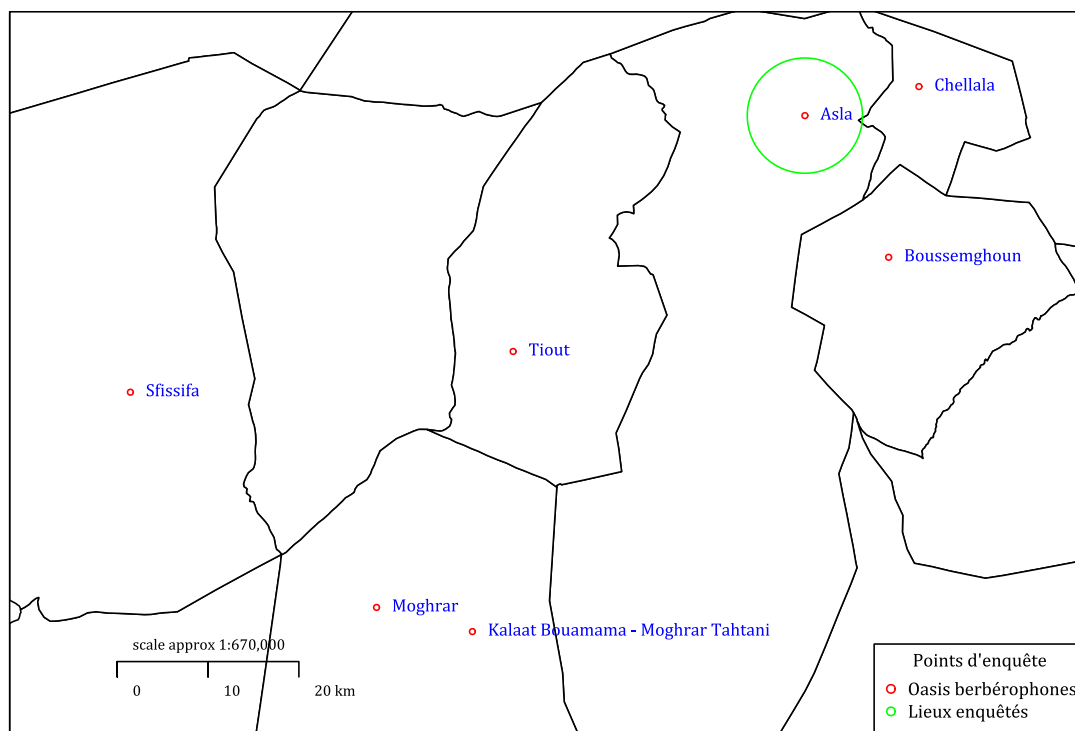
Il reste encore plusieurs parlers dont on n'a pas de système phonologique précis et les questions de phonétique sont loin d'avoir été étudiées sous tous leurs aspects.

Pour combler ces lacunes, nous avons entrepris de réaliser une étude phonétique, mais également une étude phonologique sur les rhotiques alvéolaires. Avant d'entamer cette étude, nous allons introduire dans quel cadre celle-ci se déroulera. Pour cela, nous allons redéfinir notre terrain d'étude et présenter plus longuement notre objet d'étude. La question de la population d'étude sera vue dans la partie méthode (Partie III.3.2, page 221).

III.1.1 Terrain d'étude

Le Sud-Oranais reste toujours notre terrain d'étude. Néanmoins, nous avons restreint la collecte de nos données à une variété de cette région. Nous avons choisi de porter notre attention sur la *Tmaziyt de l'Atlas Saharien (TmAS)*. Si nous nous sommes limité à cette branche du berbère, c'est toujours pour des questions de faisabilité. Il serait impossible avec les moyens que sont les nôtres de réaliser une étude expérimentale pertinente — donc forcément quantitative — sur l'ensemble de notre terrain d'étude.

Ce choix étant fait, il nous fallait d'autre part choisir un point d'enquête. Puisque, la *TmAS* est parlée à différents endroits qui sont espacés par plusieurs kilomètres. Donc, ce faisant, la *TmAS* est la langue de plusieurs communautés berbères. Comme nous l'avons vu, cette langue est parlée dans sept localités. Sur la carte ci-dessous, nous récapitulons où se situent géographiquement et administrativement ces oasis.



Carte III.1.1 : Les localités parlant le Tamazight de l'Atlas Saharien

Nous avons d'emblée choisi de faire d'Asla notre point d'enquête principal et cela pour des questions de "ressources humaines". Trouver des informateurs volontaires, bénévoles et disponibles, n'est pas une chose aisée dans cette région où le berbère est en voie de disparition. À Asla, nous avons la possibilité d'avoir plusieurs informateurs qui sont prêts à payer de leur personne. Cela étant, nous nous devons de bien cerner l'objet d'étude que nous allons étudier.

Donc, avant de procéder à cette étude expérimentale, nous nous sommes assuré de la représentativité de notre objet d'étude. C'est-à-dire si nos phonèmes étaient bien usités par l'ensemble des catégories sociologiques — âge & sexe principalement — et s'ils ne présentaient pas de variations au niveau géographique. Ce fut la première étape de notre recherche. Nous avons donc procédé à des enregistrements dans les localités de Tiout, de Sfissifa et de Boussemgoun. Dans chacun de ces points d'enquête, nous avons enregistré un informateur. Nos informateurs à Boussemgoun et à Tiout étaient des hommes âgés et à Boussemgoun, c'était une jeune femme. Ces enregistrements ont été complétés par ceux entrepris à Asla auprès d'une dizaine d'hommes d'âges différents

allant de l'adolescence à un âge mûr.

Ces enregistrements ont été réalisés lors de notre première enquête de terrain en 2015. Les variables que nous avons décidées d'étudier de plus près ne comprenaient pas seulement les rhotiques alvéolaires et étaient limitées à quelques items. Nous avons identifié plusieurs particularités phonétiques que nous trouvions problématique. C'est après analyse de ce corpus et en fonction des axes de recherche que nous avons définis que nous avons décidé de nous focaliser sur les rhotiques alvéolaires. L'année suivante, en 2016, de retour sur le terrain, nous avons ainsi réalisé notre étude expérimentale sur les rhotiques alvéolaires. Les données de cette première enquête ne seront pas présentées au cours de cette thèse.

III.1.2 Objets d'étude

D'abord, de manière générale, le système phonétique et phonologique de cette langue berbère n'est pas très différent de celui des autres langues lui étant apparentées comme nous l'avons vu précédemment. Ainsi, les processus morphologiques, les processus phonologiques et l'inventaire des phonèmes existant en **TmAS** sont retrouvés en partie ou totalement dans d'autres systèmes linguistiques berbères. Cependant, cette variante du Sud-Oranais présente tout de même un certain nombre de particularités phonétiques en comparaison, d'abord, sur un strict plan interne à la famille berbère et ensuite, plus largement, d'un point de vue pan-linguistique.

Ainsi, l'une de ces singularités concerne les rhotiques alvéolaires et le **TmAS** en connaît quatre types : [r], [r], [r^h] et [r^h]. La première de ces rhotiques est retrouvée dans toutes les langues berbères, mais ce n'est pas le cas de la vibrante alvéolaire roulée [r^h] qui semble, dans les langues berbères qui ont fait l'objet d'une description, avoir été peu inventoriée. L'autre aspect original de ces rhotiques est sur un plan pan-linguistique. La rhotique alvéolaire pharyngalisée n'est pas un son répandu dans les langues du monde. Dans la littérature, la mention de ce phonème n'est pas retrouvée abondamment, seuls quasiment les langues berbères et certaines langues sémitiques semblent avoir dans leur système phonétique ce son.

Donc, pouvoir faire une étude acoustique de ces unités phoniques constitue en bien des cas un travail original. D'autant plus, si par ailleurs, la langue étudiée n'a jamais fait l'objet d'une étude acoustique comme c'est le cas pour les langues berbères du sud-ouest algérien. D'autre part, les rhotiques en berbère, pour quel que langue berbère que ce

soit, n'ont quasiment jamais attisé la curiosité des phonéticiens. Par conséquent, cela constituera, là aussi, sur ce plan, une recherche innovatrice pour les études berbères.

Sur le plan des études phonologiques, notre travail est aussi à considérer comme novateur dans ce domaine. Puisque, les études intégrant la dimension psychique des rhotiques alvéolaires sont généralement laconiques. Ainsi, notre travail sera aux antipodes des études habituelles. Etant donné que nous avons souhaité nous attarder en détail sur les représentations sous-jacentes des rhotiques alvéolaires.

III.1.3 Plan

Pour toutes ces raisons, nous avons choisi d'étudier au cours de cette seconde partie les rhotiques alvéolaires. Pour ce faire, nous allons dans un premier temps faire l'état de l'art des études réalisées sur le Sud-Oranais de manière générale, mais aussi sur d'autres langues berbères. Cette première partie sera complétée par une synthèse des travaux sur les rhotiques et plus particulièrement les rhotiques alvéolaires rencontrées dans différentes langues. Une étude comme la nôtre qui intègre la phonétique doit passer par une revue de la littérature qui touche à d'autres systèmes linguistiques. Cela doit nous permettre d'avoir une vue d'ensemble des phénomènes qui gravitent dans l'environnement des $\|r\|$.

Puis dans un second temps, nous présenterons la méthodologie que nous avons utilisée et quels sont les points sur lesquels nous allons focaliser notre recherche. Cette partie aura pour but de mettre en exergue la manière et les étapes qui nous ont permis d'obtenir nos données. Puis, ce qui nous a amené des matériaux bruts aux matériaux traités. Ce processus méthodologique et technique est la pierre angulaire de toutes études expérimentales. Ne pas maîtriser le processus de bout en bout peut fausser les résultats que nous pourrions obtenir où nous serions dans l'incapacité de pouvoir interpréter les résultats comme il se doit. Donc, ce savoir en matière de méthodologie est capital.

Par la suite, nous mettrons en avant les résultats de nos recherches et nous les discuterons. Cette troisième partie comprendra plusieurs sous-parties qui étudieront des points très particuliers. Comme, il nous est impossible de pouvoir aborder toutes les facettes de ces rhotiques alvéolaires dans le cadre de cette thèse. Donc, nous sommes contraint de délimiter notre champ d'étude à certains aspects. Par conséquent, pour des questions de faisabilité, nous avons fait le choix de limiter notre étude à certains points

que nous allons détailler. Nous avons prévu de diviser notre étude en deux parties :

— Phonologie

Dans un premier temps, nous comptons faire une étude phonologique de ces rhotiques, voir quelles sont leurs distributions et pourvoir déterminer ensuite s'ils constituent des phonèmes ou des allophones. La méthode distributionnelle et des paires minimales seront utilisées pour mettre en évidence ce fait.

1. Rendre compte de la distribution des rhotiques alvéolaires.

Ce travail consistera à déterminer si les $\|r\|$ sont d'une manière quelconque en contraste.

2. Déterminer les représentations phonologiques.

Nous nous assignons comme tâche de caractériser les différents allophones des rhotiques alvéolaires et de déterminer quels sont les phonèmes de nos unités phoniques.

— Phonétique

Puis par la suite, nous comptons étudier certaines particularités de ces rhotiques alvéolaires afin de constater s'il existe certaines caractéristiques physiques qui les distinguent les unes des autres. Donc, nous allons prendre soin d'étudier les points suivants que nous allons énumérer :

1. Déterminer le mode d'articulation de la première articulation.

Cette étude cherchera à dégager si comme dans les autres langues, le TmAS a différents types de mode d'articulation.

2. Déterminer la nature du vocoïde.

Cette analyse a pour objectif de définir si les vocoïdes trouvés lors de la réalisation de $\|r\|$ sont inhérents à ces derniers où sont une unité sonore à part entière.

3. Mesurer la durée d'occlusion de $[r]$ et de $[r^{\text{h}}]$.

Avec cette mesure, nous souhaitons vérifier l'hypothèse suivante à savoir si la seconde articulation affecte la durée de l'occlusion.

4. Mesurer la durée des vocoïdes de $[r]$ et de $[r^{\text{h}}]$ qui apparaissent avant consonne /r/C.

En mesurant cette durée, nous souhaitons pouvoir nous rendre compte si la pharyngalisation affecte la durée de ces vocoïdes. De plus, l'autre hypothèse que nous souhaitons constater avec cette mesure est si la place occupée dans la syllabe par ces rhotiques joue également sur la durée des vocoïdes.

Enfin, nous concluons cette étude en résumant les principaux résultats auxquels nous sommes parvenu et nous indiquerons les perspectives de recherche future.

CHAPITRE III.2

État de l'art

Les rhotiques sont retrouvés dans la majorité des langues du monde. Il forme une classe de phonèmes qui sont différents sur bien des aspects. Il y aurait environ 75% des langues au monde qui auraient dans leur inventaire phonétique des rhotiques. Mais seulement 18% des langues auraient plus d'un type de rhotiques [Maddieson and Disner, 1984, p.73]. Les langues berbères, pour la plupart d'entre elles, font partie de ces 18%. Ainsi, les langues berbères ont au minimum des rhotiques alvéolaires et des rhotiques fricatives dans leur système phonétique [Galand, 1960, p.1216].

Bien que la langue berbère que nous étudions contienne, elle aussi, plusieurs types de rhotiques — voir l'inventaire phonémique dans le tableau II.4.1.1 page 89 —, nous comptons délimiter notre champ d'étude à l'analyse des rhotiques alvéolaires. Nous allons donc dans un premier temps présenter l'inventaire des travaux portant sur notre objet d'étude en prenant soin d'abord d'exposer ceux concernant la famille des langues berbères. Puis, nous évoquerons les recherches portant sur diverses langues du monde.

III.2.1 Les rhotiques alvéolaires en berbère

Comme nous l'avions indiqué précédemment en Introduction, le berbère du Sud-Oranais comprend quatre rhotiques alvéolaires ([r], [r], [r^h] et [r^h]). Toutes ces rhotiques ne sont pas forcément présentes dans toutes les langues berbères comme nous allons le voir maintenant. Dans toutes les descriptions des langues berbères que nous avons consultées [*rifain* : [Lafkioui, 2007]; *kabyle* : [Chaker, 1973], [Naït-Zerrad, 2001]; *chleuh* : [El Mountassir, 2003]; *touareg* : [Prasse, 2003]; *chaoui* : [Penchoen, 1973]], les auteurs font mention de la présence d'une rhotique alvéolaire battue dans ces langues. Cette consonne connaît aussi une version pharyngalisée, comme c'est le cas pour bon nombre de consonnes en berbère, et que nous retrouvons largement dans toutes les variétés berbères. Au sujet de cette consonne [r^h], en tachelhit, Ridouane [Ridouane, 2003, p.10] affirme que les /r/ pharyngalisées, que connaît cette langue, sont pour la plupart empruntées à l'arabe. Donc, jusqu'ici le berbère du Sud-Oranais est en osmose avec ses variantes sœurs, mais le parallèle s'arrête là. Au niveau des trills¹, il y a peu de langues berbères où il y est mentionné l'utilisation de ce phonème. Il n'y a en réalité que quatre autres langues berbères, comme celui de Zwara en Lybie, qui semblent avoir dans son système phonétique des trills :

1. Nous emploierons indifféremment les anglicismes trill et tap et les termes français correspondants : roulée et battue.

Alveolar /r^ɸ/ are tapped in the onset and briefly trilled in the coda, while geminate occurrences are long trills, as in /'r^ɸr^ɸuz^ɸ/ 'rice' [Gussenhoven, 2015, p.7].

Dans certaines variétés du rifain, il y aurait également des trills comme le rapporte Dell et Tangi [Dell and Tangi, 1992] [Dell and Tangi, 1993] :

In our phonetic transcriptions [r] represents a voiced alveolar flap, as in Spanish pero 'but', and [rr] represents a voiced alveolar trill, as in Spanish perro 'dog'. [...] Geminated /r/ is realized as [rr] [Dell and Tangi, 1993]

Nous pouvons citer une autre variété du rifain, le parler d'Ibquyen situé dans la province d'Al Hoceima, où nous retrouvons l'utilisation des trills. Comme en kabyle, par ailleurs, nous avons aussi constaté qu'il existait des trills — observation faite par nous-même sur le terrain —. En chleuh également, il existe des trills [Coleman, 1999], mais qui ne seraient que le pendant géminé d'une battue [Ridouane, 2014]². Enfin, à Siwi, Naumann [Naumann, 2012, p.264] signale que le berbère de ce lieu comprend aussi dans son système phonétique des trills qui seraient la forme géminée d'une tap.

Donc, cela va sans dire, bien que les trills ne soient pas systématiquement indiquées dans les autres langues berbères, cela ne signifie pas que le Sud-Oranais serait unique dans son cas et qu'il faille conclure qu'elles n'existent pas dans d'autres systèmes phonétiques berbères. Il ne serait pas étonnant, même il serait fort probable de découvrir qu'ailleurs en berbère, ces trills existent. Car il y a une autre remarque à faire à propos de ce que nous pouvons trouver dans la littérature berbérissante.

En nous basant sur les lectures de ces différents écrits, il en ressort que certains chercheurs berbérissants semblent ne pas appliquer les conventions de transcriptions phonétiques en usage. Cela se ressent particulièrement dans la terminologie utilisée. Dans la quasi-totalité des ouvrages consultés cités précédemment, leurs auteurs notent la battue en utilisant le symbole API pour la roulée [r] ou bien n'utilisent pas de transcription phonétique. De plus, certains chercheurs se contentent sans plus de précision d'indiquer que cette consonne, qu'ils répertorient, est une vibrante. Ce qui laisse planer le doute quant à la nature réelle de cette consonne, est-ce une battue ou une roulée? Nous pourrions bien croire qu'il s'agit de cette dernière, d'autant plus qu'ils utilisent comme symbole [r]. De plus, ces chercheurs ont très bien pu confondre ces deux consonnes et les noter de la même manière, d'où notre hypothèse que nous pouvons très

2. Nous remercions M. Ridouane de nous avoir communiqué cette information.

bien découvrir des roulées ailleurs par une attention plus accrue du système phonétique respectif de ces langues. Il existe également des chercheurs qui désignent les taps incorrectement en les nommant trills [Sadiqi, 1986, p.IV] [Kossmann, 1997, p.14].

Nous pourrions rajouter en guise de commentaire de conclusion que les différents chercheurs cités ne s'épanchent pas suffisamment sur ces phonèmes. Et donc, nous manquons d'informations sur les particularités de ces rhotiques alvéolaires, à savoir leurs distributions, les mutations phonétiques qui les touchent, leurs particularités physiques (articulatoire / acoustique)³. Nous ne trouvons que très peu de choses dans les études berbères, ce qui nous amène à jeter un regard sur les études menées sur d'autres langues du monde.

III.2.2 Les rhotiques alvéolaires dans le monde

Les rhotiques ont fait l'objet d'étude dans bien des langues et ont été analysées suivant bien des manières. Nous allons donc rapporter ici toutes les principales questions abordées par les chercheurs au sujet de ces rhotiques.

III.2.2.1 Problème de classement et de dénomination : Rhotique / Liquide / Tremulant

Les sons que nous étudions sont désignés par bien des noms. Les rhotiques alvéolaires sont d'abord traditionnellement considérées comme des sonantes et sont donc distinguées des obstruents. Ce premier classement connaît d'autres subdivisions que nous allons voir. Dans la littérature scientifique, nous retrouvons trois appellations utilisées qui sont : rhotique, liquide et tremulant⁴. La dernière appellation est essentiellement retrouvée dans la littérature anglo-saxonne et n'est pas très répandue. Par ailleurs, ce terme semble correspondre à des classements disparates selon les auteurs. Pour certains, tremulant est le synonyme du mot anglais "vibrant" — en français : vibrante — [Abondolo, 2015, p.522], alors que pour d'autres, il s'agit d'un

3. Dans les travaux de Coleman [Coleman, 1999] [Coleman, 2001], nous pouvons trouver des évocations sur certaines particularités acoustiques des rhotiques alvéolaires. Mais l'axe central de ses recherches n'était pas l'étude des rhotiques alvéolaires. Naumann [Naumann, 2012] est le seul qui ait étudié réellement le cas de ces rhotiques alvéolaires. Toutefois, sa principale motivation est de caractériser uniquement ces rhotiques alvéolaires en utilisant les formants.

4. Le mot tremulant est écrit dans son orthographe anglaise.

équivalent terminologique du mot rhotique [Hult, 1986]. Donc, le mot tremulant n'est pas celui retrouvé le plus couramment, ce sont plutôt les termes rhotique et liquide qui ont la préférence des chercheurs. Cependant, ces deux classes précédemment citées ne regroupent pas en leur sein exactement les mêmes sons. À noter également que certains chercheurs considèrent les rhotiques comme étant une sous-classe des liquides et donc cela correspond à ce qui est appelé par d'autres chercheurs les vibrantes. Nous n'allons pas ici faire l'inventaire en longueur des sons qui constituent ces classes, car cela est bien connu de tous et nous considérons qu'il ne s'agit pas d'un point crucial à développer. Nous renvoyons à l'ouvrage de Ladefoged et Maddieson [Ladefoged and Maddieson, 1996, p.215-244] pour de plus amples renseignements. Toutefois, nous allons tout de même nous attarder sur ce qui constitue et différencie la classe des rhotiques et des liquides. De manière générale, la classe des liquides regroupe des sons qui ont leur point d'articulation au niveau coronal, alors que la classe des rhotiques englobe des sons qui peuvent avoir un point d'articulation au niveau bilabial jusqu'au niveau pharyngal. Alors, il est à se demander, au vu de l'hétérogénéité de ces classements, qu'est-ce qui peut bien constituer l'unité de ces classes ?

S'agissant des rhotiques, il a été défini dans un premier temps que les sons de cette classe avaient comme caractéristique commune d'avoir le F3 faible, c'est ce qu'avaient affirmé Ladefoged et Lindau — cité par [Hall, 1997, p.107] —. Mais cette affirmation a été contestée par ces mêmes auteurs dans des travaux ultérieurs [Ladefoged and Maddieson, 1996, p.244], [Lindau, 1980, p.165]. Ils avaient constaté dans différentes langues que toutes les rhotiques n'avaient pas cette particularité, leur F3 allait au-delà de 2500 Hz. Toutefois, cela n'a pas désarçonné les tenants de la classe des rhotiques du bien-fondé de l'hypothèse de cette classe, perçue comme naturelle. Hall fournira des arguments à l'appui de l'existence des rhotiques notamment d'un point de vue phonologique [Hall, 1997, p.107-129] en cherchant à prouver que les rhotiques constituent un trait distinctif à part entière. Son travail s'inscrit dans le cadre de la "Distinctive Feature Theory". Pour Dickey [Dickey, 1997], c'est également au niveau phonologique qu'on retrouve ce qui caractérise les rhotiques. D'après ce chercheur, qui est un adepte de la théorie autosegmentale, ces derniers ont la particularité d'avoir un double nœud au niveau du lieu. Il y a d'autres auteurs qui définissent la classe des rhotiques sur d'autres bases. Pour Wiese [Wiese, 2003, p.350], le critère prosodique détermine la classe des rhotiques. D'après lui, leurs échelles de sonorité sont homogènes et se situent entre les latérales et les voyelles. Alors que d'autres auteurs arguent que ce qui fonde cette

classe est plutôt d'ordre diachronique et graphique [Ladefoged and Maddieson, 1996]. Ils s'avèrent effectivement que, dans nombre de langues, les rhotiques se transcrivent par le symbole "r" ou par un signe qui dérive de ce dernier. Et que certaines rhotiques sont des mutations d'autres rhotiques comme en français où le phonème /r/ est devenu /ʀ/.

À propos de la classe des liquides, nous trouvons également dans la littérature des tentatives pour démontrer l'unité de cette classe. Certains chercheurs trouvent qu'au niveau phonologique les liquides ont des caractéristiques communes. Celles-ci sont résumées dans le travail de Backley [Backley, 2011, p.165] où il synthétise les approches sur lesquelles se fondent certains phonéticiens pour définir la classe des liquides. Nous allons relever pour illustration certaines de ces positions. D'abord, cette classe est vue comme un classement par défaut, car ne rentrant pas dans des classements préétablis, alors par commodité ces consonnes furent regroupées ensemble. Ou bien, les liquides sont perçues comme une même classe pour des considérations diachroniques, car dans certaines langues, certaines rhotiques ont évolué en latéral — rhotacisme — ou inversement — lambdacisme —. La dernière vision qu'il avance et celle qui voit l'unité des liquides sous l'angle de la phonologie, et lui-même se réclame de cette approche.

Soulignons par ailleurs que certains auteurs contestent le bien-fondé de la classe des liquides et ne voient pas de raisons de mettre ensemble latéral et rhotique, puisque ces derniers ne partagent pas assez de points en commun. Colantoni et Steele [Colantoni and Steele, 2005] avancent comme arguments, en se basant sur le français et l'espagnol, que le comportement des rhotiques est proche de celui des fricatives en donnant comme exemple que les rhotiques comme les fricatives provoquent l'allongement vocalique des voyelles qui sont en leur contact. Ainsi, les rhotiques comme les fricatives ou les approximantes sont sensibles à leur environnement. Pour Hall [Hall, 1997, p.129], il en va de même, selon lui, la classe des liquides n'a pas lieu d'être et la classe des rhotiques doit prévaloir. Une étude expérimentale de Proctor [Proctor, 2009] compare les liquides sur la base de différents paramètres phonétiques et constate des parallèles — même articulation dorsale de la langue — et des divergences entre les latérales, les battues et les roulées, mais il ne conclue pas à l'existence ou non de la classe des liquides.

Donc, il existe des chercheurs qui souscrivent à la thèse qui consiste à nier l'intérêt d'établir une classe de liquides, cependant nous ne nous retrouvons pas d'auteurs qui auraient une opinion inverse, c'est-à-dire qui discuteraient du bien-fondé de l'existence de la classe des rhotiques. Ce classement semble faire l'unanimité auprès de la majorité

des phonéticiens, donc c'est la raison pour laquelle, nous reprenons à notre compte également cette approche et avons décidé de classer nos objets études dans la classe des rhotiques.

III.2.2.2 Caractéristiques phonétiques des rhotiques alvéolaires

Les études phonétiques sont à la jonction de deux dimensions physiques : la réalité biomécanique et la réalité acoustique. La première de ces réalités renvoie à l'appareil phonatoire. Ainsi, l'observation de ces rhotiques s'est faite de plusieurs manières. Les études articulatoires sur les rhotiques alvéolaires ont été réalisées soit à partir de palatogramme / électropalatogramme [[Heselwood et al., 2011], [Nicolaidis and Baltazani, 2011], [Kochetov and Colantoni, 2009]] ou par électromagnétisme [[Katz et al., 2015], [Howson et al., 2013]] ou soit à partir d'imageries numériques : X-ray [[Delattre and Freeman, 1968], [Lindau, 1980], [Bolla, 1981], IRM [Proctor, 2009], ultrason [Scobbie et al., 2013], [Lawson et al., 2013]]. Il existe également des études acoustiques qui peuvent comprendre en parallèle aussi dans leurs travaux des études articulatoires [[Recasens and Pallarès, 1999], [Howson et al., 2014]] ou bien qui ne prennent en compte que des paramètres acoustiques [[Dhananjaya et al., 2012], [Henriksen and Willis, 2010], [Baltazani, 2009], [Schmeiser, 2009]].

Effectivement, la réalité acoustique peut être mesurée et quantifiée. Les appareils qui captent cette énergie se rangent dans la catégorie des microphones. À l'intérieur de cette catégorie, il existe toute une variété de microphones qui se distinguent par leur fonctionnement technologique. Mais le principe de base est le même pour tous les microphones qui est celui de transformer une énergie mécanique en énergie électrique⁵. La captation de cette énergie est décomposable en trois paramètres physiques : le temps, la fréquence et l'amplitude. L'ensemble de ces paramètres a été pris en compte par les chercheurs, comme nous allons le voir.

Au cours de cette revue de la littérature, nous allons seulement centrer notre attention sur les types de rhotiques alvéolaires qui nous intéressent. Étant évident que les rhotiques alvéolaires connues et recensées ne se limitent pas à nos trois rhotiques.

L'une des premières caractéristiques que l'on peut évoquer d'ores et déjà concerne le mode d'articulation. Bien que la majorité des descriptions linguistiques présentent

5. Dans la partie Méthodologie III.3.3.1 page 224) nous présentons de manière schématique le fonctionnement d'un microphone qui est celui que nous avons utilisé pour notre étude expérimentale.

les taps et les trills comme ayant un seul mode d'articulation, il s'avère en réalité que celles-ci peuvent en avoir plusieurs. Bien que ces sons soient mis dans la classe des sonantes, donc considérés comme ne connaissant pas de phénomène d'obstruction. Il fût, toutefois, recensé trois principaux modes d'articulation qui sont : le mode occlusif, fricatif et approximant [Blecua, 2001]. Ces trois modes se distinguent par leur degré d'obstruction. Ci-dessous une figure où nous avons en vis-à-vis le spectrogramme d'une tap approximante et celui d'une tap occlusive.

À la comparaison des deux images, on se rend bien compte sur le spectrogramme de gauche qu'il y a de l'énergie au-dessus de 1 kHz, alors que sur celle de droite, cela est complètement absent. Ce qui implique pour ce dernier une occlusion totale et pour l'autre seulement une constriction. Ce point que nous venons de voir est partagé par l'ensemble des rhotiques alvéolaires, mais il en est tout autre pour d'autres critères. C'est la raison pour laquelle, nous allons présenter séparément les caractéristiques importantes de ces unités phoniques. Nous allons commencer par la tap.

III.2.2.2.1 Tap

Les taps ont été découvertes dans bien des contrées de par le monde. L'examen de la base de données de l'*UCLA Phonological Segment Inventory Database (UPSID)*⁶ indique que 127 langues sur 451 ont dans leur système phonétique ce son. Ce nombre correspond plus précisément à l'inventaire des flaps et des taps. Si nous avons pris en compte également les flaps, c'est en raison des confusions courantes faites entre ces deux dénominations qui pourtant correspondent à deux façons différentes d'articuler la langue [Ladefoged and Maddieson, 1996]. Donc, il se peut très bien que les chercheurs désignent par flap une tap. Par conséquent, il faut relativiser ce chiffre que nous donnons. D'autre part, comme les études phonétiques portant sur les taps pharyngalisés étant rares, nous avons fait le choix de ne pas aborder cette unité phonique ici même, mais plus loin dans une partie à part (Partie III.2.2.2.3, page 203).

III.2.2.2.1.1 Articulation

Les taps sont des phonèmes qui utilisent comme articulateur principal la langue. Ils se caractérisent par un mouvement de la pointe de la langue qui vient se placer au niveau

6. Cette base de données peut être consultable ici : <http://web.phonetik.uni-frankfurt.de/upsid.html> (Consulté le 15/08/2016).

alvéolaire. Quand le contact est établi, la pointe de la langue se retire. Et il n'y a qu'un seul contact effectué [Ladefoged and Maddieson, 1996]. Le point d'articulation le plus souvent mentionné dans la littérature est la zone alvéolaire, mais selon les langues, il peut occuper d'autres positions, notamment au niveau dental. Certaines langues peuvent avoir une opposition contrastive entre plusieurs taps qui se distinguent au niveau de leur point d'articulation (dental, alvéolaire ou rétroflexe) comme en Toda [Spajić et al., 1994] [Spajić et al., 1996]. Au sujet de la forme de la langue, celle-ci lors de la réalisation de la tap se trouve dans une position basse au niveau prédorsal avec une forme concave et connaît une rétractation au niveau postdorsal [Recasens and Pallarès, 1999, p.144].

III.2.2.2.1.2 Durée du contact

Un autre point qui a attiré l'attention des chercheurs concerne la durée d'obstruction des taps. L'étude typologique de Recasens et Spinoza [Recasens and Espinosa, 2006] rapporte que celle-ci peut varier entre 15 ms et 30 ms selon les langues. Cette durée est aussi fonction de son environnement phonétique et de la place qu'occupe la tap dans le mot (initial, final, médiane). Ainsi, d'après l'étude de Recasens et Pallares [Recasens and Pallarès, 1999, p.153], si la tap se trouve entre deux voyelles [a] la constriction maximale durera 18,4 ms et s'il se trouve entre deux voyelles [i] cela durera 39,6 ms. Cependant, Blecua [Blecua, 2001] ne fait pas le même constat. Selon elle, d'un point de vue statistique, les voyelles qui sont dans l'environnement phonétique des rhotiques alvéolaires — comme l'accentuation — n'ont pas d'influence significative sur la durée de la tap. Toutefois, elle observe que le contact des taps n'a pas la même durée lorsqu'ils sont dans un groupe consonantique. Suivant la consonne qui précède, la durée peut être plus ou moins longue. Ainsi, si la consonne est coronale ou labiale, la durée sera plus courte, mais si elle est vélaire, celle-ci sera plus longue.

III.2.2.2.2 Trill

Les trills sont retrouvées dans plusieurs langues à travers le monde. La base de données de l'UPSID recense 155 langues sur 451 qui font usage de ce phonème. En cela, c'est le son parmi les rhotiques qui est le plus courant dans le monde [Maddieson and Disner, 1984, p.79]. Nous ne mentionnerons au cours de ce passage en revue que les trills simples puisque notre étude bibliographique ne nous a pas permis de trouver des études phonétiques en abondance portant sur les trills pharyngalisées. Toutefois, la

question des trills pharyngalisées sera discutée par la suite de manière globale avec les taps pharyngalisées (Partie III.2.2.2.3, page 203).

III.2.2.2.1 Articulation

D'un point de vue articulatoire, la réalisation de ce son se fait par la vibration de la pointe de la langue, il existe plusieurs études qui en font la démonstration [[Bolla, 1981]; [Ladefoged and Maddieson, 1996]; [Recasens, 1991]; [Recasens and Pallarès, 1999]; [Halle et al., 2000]; [Solé, 2002]]. Ces constatations semblent faire l'unanimité chez les chercheurs et ont été faites pour des langues non apparentées entre elles. S'agissant du point d'articulation, selon les langues, il peut se situer soit au niveau alvéolaire, c'est le cas de la majorité des langues, soit se situer au niveau dental — UPSID —. Dans un article de Recasens et Spinoza [Recasens and Espinosa, 2006], ces derniers font un recensement de travaux et découvrent une variabilité plus conséquente au sujet du point d'articulation qui peut être plus ou moins en avant au niveau dental comme alvéolaire. Recasens [Recasens, 1991] constate par ailleurs que les trills connaissent une légère seconde articulation au niveau vélaire. Effectivement, il apparaît grâce à l'imagerie que le dorsum de la langue se rétracte vers la partie postérieure de la cavité buccale comme l'a aussi constaté Kavitskaya [Kavitskaya and Iskarous, 2009] et Bolla [Bolla, 1981] pour le russe. Une étude récente affirme que la racine de la langue est également impliquée dans l'articulation de la trill. Celata et Ricci [Celata and Ricci, 2017], grâce à l'endoscopie, ont pu observer que lors de la production de [r] le larynx remontait et que l'épiglotte avait une position plus en arrière et plus basse. Dans une autre étude sur la langue malayalam, il y est fait comme constat que la constriction est réalisée avec la zone laminaire de la langue [Punnoose, 2011]. Pour conclure, nous tenons à faire remarquer qu'aucune étude ne spécifie quelle partie de la langue réalise le mouvement mécanique vibratoire, est-ce tout le corps avant de la langue, muscle compris, ou c'est seulement la muqueuse qui se meut comme c'est le cas pour les plis vocaux⁷ ?

7. Les particularités physiologiques sur la vibration des plis vocaux sont relativement bien documentées [Giovanni et al., 2014]. Ces études permettent de mieux comprendre la réalité physique qui entoure cet organe. Alors que s'agissant des trills, la précision des études ne va pas aussi loin et la réalité biomécanique de la pointe de la langue est quasiment absente des études phonétiques.

III.2.2.2.2 Aérodynamique

L'aérodynamisme, mis en œuvre pour la production des vibrantes, a aussi fait l'objet d'études. La cause principale à l'origine des battements de la pointe de la langue est due à l'effet Bernoulli comme le mentionne entre autres Ladefoged et Maddieson [Ladefoged and Maddieson, 1996]. Mécaniquement, le processus se déroule de la manière suivante : Lorsque la pointe de la langue rentre en contact avec le palais, l'air va s'accumuler derrière le dos de la langue et une hausse de la pression va en découler qui va défaire l'occlusion. Cela du fait que cette pression intra-orale étant plus forte que la force musculaire et la pression atmosphérique qui s'opposent à elle. Au cours de la phase de désocclusion, la pression intra-orale va baisser et comme la force musculaire de la langue devenant plus forte que cette pression, alors l'occlusion va se reproduire. C'est cette mécanique qui va se répéter automatiquement et de manière cyclique. Il y a dans ce processus d'abaissement de la pointe de la langue aucun effet musculaire qui rentre en ligne de compte. Citons les propos de McGowan qui évoque le phénomène que nous venons de décrire, mais de manière plus imagée :

In the simplest terms, tongue-tip trills might be described in the following way. The tongue tip is viewed as a sprung trap door, with the oral cavity providing a container, or plenum, where air can have a greater pressure than atmospheric. The pressure difference between the plenum and the atmosphere causes the trap door to open as the torque provided by that pressure difference is sufficient to overcome the torsional tension of the trap door spring. The air rushes out, and the air pressure in the plenum drops so that the torsion spring closes the trap door over the mouth of the plenum. The plenum is refilled by air flowing from the lungs through the glottal resistance, and the cyclic process continues.[McGowan, 1992, p.2903]

Il faut également mentionner pour que cet effet mécanique puisse avoir lieu qu'il y a certains paramètres à respecter. Ces paramètres déterminants sont le flux d'air et l'aperture entre la langue et le palais. McGowan a réalisé des études expérimentales, il en résulte que pour que la vibration se déroule, il faut une quantité de débit d'air et une force musculaire adéquats. Pour obtenir un débit d'air adapté, cela suppose une maîtrise contrôlée des poumons et de l'ouverture de la glotte. Les études réalisées par Solé [Solé, 1998] [Solé, 2002] pour le catalan vont plus loin et indiquent que la pression intra-orale (P_o) nécessaire à l'activation de la vibration doit se situer entre 4 et 5,4

cmH₂O. De plus, ce dernier mentionne que, pour que le voisement se fasse correctement, la pression intra-orale ne doit pas être supérieure à ces valeurs sans quoi la vibration des plis vocaux devient inopérante. Et si la pression est en dessous de ces valeurs alors la vibration de la pointe de la langue se retrouve gênée. Les travaux de Gibson [Gibson, 2015] parviennent au même constat, il observe que le déclenchement de la trill opère lorsque le débit atteint une valeur inférieure à 600 dm³/s — qui après conversion et correspondance au niveau des unités de mesure revient au même résultat trouvé par Solé — comme nous pouvons le constater sur la figure ci-dessous.

III.2.2.2.3 Nombre et durée des contacts

Les trills peuvent avoir un nombre de contacts différents. En serbe, Zygis [Zygis, 2005] rapporte qu'on y a observé que le nombre de battements de la pointe de la langue varié entre deux et trois. En catalan, Solé [Solé, 2002] a remarqué que le nombre de contacts était de quatre et pouvait aller parfois jusqu'à cinq ou six. Pour la même langue, une autre étude [Recasens and Pallarès, 1999, p.51] indique que le cycle des contacts varie entre deux et cinq. En russe, Bolla [Bolla, 1981, p.99] a constaté une vibration de la langue qui pouvait comprendre trois ou quatre battements. La durée de chaque cycle n'est pas forcément égale, bien au contraire. Ainsi, il existe plusieurs études menées sur les durées des cycles des trills dont voici ci-dessous un aperçu des valeurs mesurées, données en ms, pour ces cycles — un cycle comprend deux phases : une phase d'occlusion⁸ et de désocclusion —.

	1er cycle		2ème cycle		3ème cycle		4ème cycle	
	occl.	déoccl.	occl.	déoccl.	occl.	déoccl.	occl.	déoccl.
Polonais [Zygis, 2005]	80 (initial)		20	30	20	30		
Polonais [Jaworski and Gillian, 2011]	17	20	17	20	17	20		
Espagnol [Quilis, 1981]	15	18	15	18	15	18		
Walpiri [Jaworski and Gillian, 2011]	17	20	17	20	x	25	x	25
Catalan [Recasens and Pallarès, 1999] : /ara/		71,6						
			105,2					
Catalan [Recasens and Pallarès, 1999] : /iri/		84						
			95					

TABLE III.2.1 : Durée en ms des cycles des trills dans différentes langues

Comme nous pouvons le constater sur le tableau III.2.1, toutes les études ne donnent

8. La phase d'occlusion à l'initial peut être difficilement quantifiable, si nous devons nous baser que sur des mesures acoustiques.

pas pour chaque cycle, ni pour chaque phase les valeurs mesurées. Néanmoins, avec ce petit aperçu, nous pouvons nous rendre compte d'une certaine homogénéité au niveau des valeurs. De plus, grâce à l'étude de Recasens et Pallares, il apparaît que la durée des cycles peut être influée par l'environnement phonétique. Ce dernier n'est pas le seul paramètre qui rentre en jeu au niveau de la durée, il y a également la place de la rhotique à l'intérieur du lexème qui peut jouer un rôle. Dans des ouvrages plus généraux, comme celui de Ladefoged et Maddieson [Ladefoged and Maddieson, 1996, p.218], nous trouvons indiqué une valeur d'environ 50 ms pour chaque cycle et 25 ms pour chaque phase. Il existe certains chercheurs qui expriment ce nombre de cycle en termes de fréquence (= 1/durée moyenne du nombre de cycles). Ainsi, Lindau [Lindau, 1980] évoque dans une étude trans-langagière une fréquence moyenne de 25. Dans une autre étude [Catford, 1977, p.130], c'est une fréquence moyenne de 30 qui est avancée. Par ailleurs, Kavitskaya [Kavitskaya and Iskarous, 2009] remarque que le contact avec la zone alvéolaire n'est pas systématiquement réalisé, la pointe de la langue effectue un mouvement, mais ne tend pas jusqu'à l'occlusion complète. Une telle observation fut également relevée par Recasens pour l'espagnol [Recasens, 1991].

III.2.2.2.2.4 Difficultés d'articulation et Rhotiques alvéolaires

Les études qui se penchent sur le processus d'acquisition du langage [[McGowan, 1992]; [Vihman, 1996]] indiquent que les trills sont parmi les sons qui sont maîtrisés le plus tardivement par les enfants. Cette constatation est assertée dans toutes les langues qui ont fait l'objet d'une étude. Les âges vers lesquels les enfants commencent à produire des rhotiques commencent vers l'âge de quatre ans en polonais par exemple [Żygis, 2005, p.148]. Par ailleurs, l'acquisition des rhotiques passent généralement par plusieurs étapes. En hébreu, il a été constaté que les rhotiques avant qu'elles ne soient maîtrisées étaient soit substituées par d'autres phonèmes (/j/, /x/, /l/, ...) ou soit n'étaient tout simplement pas prononcées [Ben-David and Bat-El, 2016, p.59], [Cohen and Ben-David, 2016] par les enfants. D'après Żygis, il en est de même en polonais, les enfants tendent à prononcer des /j/ ou des /l/ en lieu et en place de la rhotique [Żygis, 2005].

En outre, il fut observé chez les adultes natifs que les trills connaissaient beaucoup de variantes libres du fait de leur difficulté à être prononcées. En polonais, il a remarqué que les locuteurs pouvaient à la place des trills réaliser des taps, des flaps, de multiples taps/flaps. Le même constat a été fait par Simackova en serbe et en tchèque [Żygis,

2005]. Dans ces langues, les trills pouvaient par ailleurs devenir des approximantes ou des voyelles. Mais ces variations, même si elles se limitent à quelques phonèmes, ne sont pas pour autant conventionnelles. Car pour un même phonème prononcé dans un même morphème les locuteurs peuvent prononcer soit des fricatives, des taps ou des approximantes comme le mentionne Jaworski [Jaworski and Gillian, 2011] pour le polonais. Même si ces variations peuvent être conditionnées comme nous l'avons vu précédemment, le fait que les locuteurs ne produisent pas le même phonème est une indication de leur difficulté à prononcer les trills comme il se doit.

III.2.2.2.3 Pharyngalisation et Rhotiques alvéolaires

Les taps comme les trills peuvent être pharyngalisés. Ces sons sont retrouvés dans bien des langues, notamment les langues afroasiatiques. À notre connaissance, il n'existe pas d'études acoustiques spécifiques aux rhotiques alvéolaires pharyngalisées. Toutefois, il existe des études plus globales où la rhotique alvéolaire fait partie des unités phoniques étudiées parmi d'autres. Nous allons seulement dans cette partie évoquer les particularités des sons pharyngalisés au plan articulatoire. La pharyngalisation est une articulation secondaire qui touchent exclusivement les unités phoniques coronales et labiales. Concrètement, ces sons ont deux points d'articulation l'un antérieur et l'autre postérieur. Ce dernier est réalisé grâce à la racine de la langue qui vient marquer une constriction au niveau pharyngal. Cependant, certaines études indiquent que la seconde articulation pourrait avoir lieu non pas au niveau pharyngal, mais au niveau uvulaire :

Nos observations montrent que /S T χ q/ n'induisent pas de compression du sphincter aryépiglottique bien qu'elles aient une constriction dans la cavité pharyngale. Ces observations semblent être en accord avec les études relativement récentes qui semblent montrer que les emphatiques seraient des uvularisées. En effet, pour Herzallah, (1990) et Bin-Muqbil (2005), les emphatiques seraient produites avec le recul du dos de la langue sans l'implication de sa racine [Zeroual et al., 2014].

Plus loin dans le même paragraphe, Zeroual évoque une autre particularité articulatoire des consonnes pharyngalisées de l'arabe marocain. Il remarque au moyen de la naso-endoscopie que /t^s/ et /s^s/ présentent une rétractation au niveau dorsal de la langue. C'est ce qui expliquerait selon le même auteur pourquoi la langue est sujette à une postériorisation, car cela serait la conséquence de l'affaiblissement du dos

de la langue. Mentionnons d'autre part que certains chercheurs ont observé pour les consonnes pharyngales /ħ/ et /ʕ/ que le point d'articulation était plutôt réalisé au niveau aryépiglottal ou épiglottal [*arabe marocain* : [Zeroual and Clements, 2015, p.263]; *chleuh* : [Ridouane, 2003, p.162]]. Donc, il est possible d'envisager que la réalisation du point d'articulation de la rhotique alvéolaire pharyngalisée soit située au niveau des zones précédemment citées. Ainsi du fait qu'aucune étude n'a apporté de réponse sur ce point, nous pouvons que faire l'hypothèse que la seconde articulation se trouve soit au niveau uvulaire, pharyngale ou (ary)épiglottale. De même, nous pouvons supposer que cette articulation postérieure diffère entre la trill et la tap. Quant à la première articulation, nous postulons qu'elle ne doit pas être très différente de celle des rhotiques alvéolaires simples comme nous l'avons indiqué plus haut.

III.2.2.2.4 Différences phonétiques entre les taps et les trills

Dans une étude, Gibson [Gibson, 2015] présente les caractéristiques principales qui différencient les taps des trills. Selon lui, en espagnol, les taps présentent moins de chevauchement articulatoire que les trills. D'autre part, ces derniers auraient une attaque et une durée globale des voyelles précédentes plus longue. Il rajoute que le débit d'air est plus élevé pour les trills que pour les taps et que les trills atteignent plus rapidement leur maximum — au contact de la zone alvéolaire — que les taps. Cependant, toujours chez le même auteur, il constate que les consonnes non voisées n'ont pas d'effets sur les rhotiques.

Une autre étude portant sur le catalan [Recasens and Pallarès, 1999] a tenté également de distinguer les aspects démarquant les taps des trills. Selon leurs travaux, ils arrivent à la conclusion que les trills présentent une différence articulatoire nette par rapport aux taps. Ils arguent que les trills, en comparaison aux taps, ont une position prédorsale de la langue plus basse, une occlusion plus rétractée et exercent une influence sur les voyelles adjacentes, mais à l'inverse, ils sont peu sensibles à celles-ci⁹. De plus, ils constatent que les trills connaissent un "couplage" de la pointe de la langue et de la zone prédorsale plus importante. Une autre différence constatée concerne la durée d'occlusion des taps qui peut être plus ou moins longue suivant l'environnement phonétique, alors qu'avec les trills, il y a moins d'emprise possible avec les voyelles adjacentes. Ainsi, si la tap se trouve entre deux voyelles [a] l'occlusion durera 18,4 ms

9. Proctor [Proctor, 2009] fait la même observation grâce à l'EMA.

et s'il se trouve entre deux voyelles [i] cela durera 39,6 ms. Toujours chez les mêmes auteurs, ils constatent que la coarticulation dans un environnement intervocalique CVC ne se produit pas de manière identique entre la battue et la roulée. Enfin, Recasens et Spinoza [Recasens and Espinosa, 2006] mentionnent également que le premier contact de la trill est plus long que celui de la tap. Plusieurs chercheurs notent que parfois la frontière entre tap et trill n'est pas assurée, une trill peut être réalisée avec un seul contact et une tap avec plusieurs contacts :

Acoustic and articulatory data examined in these experiments indicate that there is not always a clear phonetic distinction between taps and trill. [...] intervocalic 'trills' can be produced with a single contact, while word-final 'taps' are often realized with multiple contacts [Proctor, 2009].

III.2.2.2.5 Accentuation et Rhotiques alvéolaires

Certaines études ont également constaté que certains phénomènes acoustiques touchant les rhotiques pouvaient être corrélés à l'accentuation. Des travaux menés [Jaworski and Gillian, 2011] sur cette question en polonais ont montré que les locuteurs avaient plus tendance à prononcer tel allophone au lieu d'un d'autre suivant que le segment soit accentué ou non. Si le segment est accentué, statistiquement parlant, les Polonais préfèrent plutôt prononcer la fricative au lieu de la tap, alors que s'il est non accentué, c'est l'inverse qui se déroule. En walpiri, une langue aborigène, l'accentuation affecte les durées des cycles des trills. De plus, comme en polonais, cela semble contraindre le choix d'une réalisation phonétique plus qu'une autre [Gillian, E. and Jaworski, S., 2014]. Ainsi, si la rhotique se trouve dans une position post-tonique, les locuteurs auront plutôt tendance à prononcer des trills au lieu de taps. Un cas identique existe également en anglais américain où les consonnes dentales [d] et [t] deviennent un flap en position post-tonique. Dans une langue amérindienne, l'arop-lokep, les trills semblent avoir plus de contacts si elles sont accentuées que si elles ne le sont pas [Raymond and Parker, 2005].

III.2.2.2.6 Articulation et Contexte

La coarticulation est un autre aspect sur lequel se sont penchés les chercheurs également. Leurs résultats montrent que l'environnement phonétique dans lequel se

trouve la rhotique affecte leur point d'articulation ou leur mode. Ainsi, en italien [Celata et al., 2014], une étude expose ce phénomène en étudiant l'influence des voyelles /a/, /i/ et /u/ sur la consonne /r/. D'autres études indiquent que la rhotique alvéolaire peut aussi affecter son environnement comme Zeroual [Zeroual et al., 2008] a pu le constater en arabe marocain où la tap bloque la spirantisation de [t] lorsque toutes deux forment un groupe consonantique [tr]. En anglais, les voyelles en contact avec la rhotique alvéolaire approximante lui empruntent ses particularités articulatoires après élision de ce dernier [Clark and Hillenbrand, 2007].

Recasens [Recasens and Espinosa, 2006, p.82] affirme qu'en catalan la voyelle [i] au contact des trills accuse une baisse du contact de la langue au niveau palatal. Et réciproquement, les trills au contact de [i] vont avoir le dos de la langue au niveau prédorsal plus bas, mais au contact de [a] la mâchoire est plus ouverte [Recasens and Pallarès, 1999, p.163]. Cependant, ces mouvements articulatoires précédemment cités sont plus prononcés pour les taps que pour les trills, toujours selon Recasens. Ce dernier en déduit à partir de ces constatations que les trills sont plus contraints de manière générale que les taps. Une autre différence, constatée par Recasens et Pellares [Recasens and Pallarès, 1999, p.153], concerne la durée de constriction des taps qui peut être plus ou moins longue suivant l'environnement phonétique. Il est en de même pour les trills.

Une étude sur l'espagnol menée par Gibson [Gibson, 2015], sur les trills en final d'un mot, indique également que les voyelles précédant une rhotique influent sur la position de la langue et sur le temps de réalisation de la rhotique. Ainsi, il fait la constatation que les voyelles postérieures et mi-fermées au contact de la rhotique ont une durée de phonation plus courte que les autres voyelles. En conséquence, l'attaque de la rhotique sera à l'inverse plus longue et sera plus courte avec les autres voyelles. Toutefois, la durée totale, c'est-à-dire entre le début de la phonation des voyelles et la fin de l'attaque de la rhotique, reste identique quelle que soit la voyelle. Toujours selon Gibson, les voyelles ont un impact sur les caractéristiques aérodynamiques des rhotiques. D'autre part, il affirme que les trills en espagnol sont le fait d'une contrainte mécanique provoquée par la voyelle qui la précède. Ainsi, les voyelles [a] et [o] sont particulièrement propices à l'émergence des trills, en cela, du fait que la posture prédorsale de la langue est en position basse comme c'est le cas pour les trills. Cela implique qu'au cours de cette phase de chevauchement si le débit atteint un seuil critique, à partir de $0,181 \text{ dm}^3/\text{s}$, alors la vibration se met en place. Alors que Solé [Solé, 2002] fait le constat que P_0 est plus élevée au voisinage de [i] que de [a], ce qui va à l'encontre des résultats de Gibson.

III.2.2.2.7 Formants et Rhotiques alvéolaires

Sur le plan acoustique, les études qui se sont penchées sur cette question utilisent comme mesure les **formants**. Dans la plupart des études, les chercheurs ne prennent en compte que F1, F2 et F3. En se basant sur ces valeurs, lorsque nous consultons les études portant sur différentes langues [*russe* : [Fant, 1971]; *toda* : [Spajić et al., 1996]; *tamil* : [Narayanan et al., 1999]; *catalan* : [Recasens and Pallarès, 1999] nous relevons quelques convergences. Les valeurs de F1 sont aux alentours de 350-650 Hz, celles de F2 varient entre 1250-2050 Hz et quant à F3 les valeurs peuvent atteindre 2300-2700 Hz. Les chiffres cités ne reflètent nullement le cas de figure d'un type particulier de phonèmes, puisqu'ils correspondent à des mesures réalisées sur des rhotiques alvéolaires différentes. Ces valeurs sont recoltées à différents endroits du signal acoustique. Les emplacements en question se situent au moment du déclenchement articulaire — onset —, mais aussi lorsque le mouvement articulaire atteint sa cible, donc lors de la phase stable du signal acoustique, et enfin à l'interruption articulaire — offset — de la rhotique alvéolaire. Les résultats montrent effectivement que ces formants varient en fonction du contexte articulaire précédant le geste articulaire de la rhotique alvéolaire. C'est ainsi qu'une attention plus particulière est accordée à ces formants [Recasens and Pallarès, 1999] pour mettre en évidence la coarticulation entre les voyelles et les rhotiques alvéolaires en mettant en lumière les différences de trajectoire formantique. Par ailleurs, certains chercheurs tentent de reconstituer les mouvements et les positions articulaires à partir des valeurs des formants qu'ils mesurent. Cela est rendu possible par le fait que les fréquences de résonances de la cavité buccale sont proches de ceux des formants :

Conceptually these should be held apart but in most instances resonance frequency and formant frequency may be used synonymously [Fant, 1971, p.20]

Toutefois, cette approximation ne vaut que pour les sons voisés. Et l'autre remarque que nous pouvons dire est que cette première approche est peut-être acceptable pour des voyelles, en revanche, il n'y a aucune certitude que les formants rendent compte efficacement de la réalité articulaire de consonnes voisées.

Enfin, rapportons un dernier fait au sujet de ces formants qui concernent ceux des consonnes pharyngalisées. Il a été observé que ces derniers phonèmes se distinguaient de leurs contreparties non pharyngalisées au niveau de leurs formants. F2 est le formant

qui se démarque le plus significativement par rapport aux autres. Ainsi, les consonnes pharyngalisées auraient un F2 plus bas que celui des consonnes non pharyngalisées [Yeou, 1997]. Ce fait a aussi été constaté spécifiquement pour les rhotiques alvéolaires de la langue berbère de Siwa [Naumann, 2012, p.269].

III.2.2.2.8 Perception et Rhotiques alvéolaires

Ce champ d'étude englobe la réalité physique et psychique. Les recherches qui étudient comment la réception du son est réalisée sont abondantes, mais celles qui portent sur les rhotiques alvéolaires sont peu nombreuses. Cependant, il en existe quelques-unes que nous allons rapporter. Dans une étude publiée par les psychologues Cutting et Day [Cutting and Day, 1975], ces derniers mentionnent la proportion qu'ont les rhotiques alvéolaires et plus généralement les liquides d'être sujettes à une fusion phonologique si ces derniers sont dans un groupe consonantique — consonne occlusive + liquide —. Cette notion de fusion phonologique est utilisée dans le cas où un auditeur considère deux items lexicaux acoustiquement différents, mais proche, comme étant un même lexème au niveau de sa représentation¹⁰. Une autre confusion qui est constatée par ces mêmes auteurs, c'est la substitution de /r/ par /l/ qui est faite par les sujets lorsque ceux-ci sont placés dans un groupe consonantique. En outre, ce qui est observé dans une certaine position n'est pas constaté dans toutes les positions, ainsi un tel phénomène n'a pas cours si par exemple la rhotique alvéolaire est placée en final d'un mot. Donc, la position du phonème à l'intérieur d'un morphème a un impact au niveau de la perception.

Des résultats identiques ont pu être constatés en allemand [Mitterer and Ernestus, 2008], [Mitterer and Scharenborg, 2013]. Ces études cherchent à déterminer si les différentes rhotiques — synthétisées ou non — sont perçues comme étant des phonèmes différents ou des allophones. Les résultats observés dépendent de la place de la rhotique dans le mot, cependant, il en ressort globalement que les sujets ont des représentations plus abstraites et pas en adéquation avec la réalité physique. Cette perception induit que les sujets considèrent comme une même unité phonologique des réalisations acoustiques différentes.

Comme, nous venons subrepticement de le voir, la réalité physique n'est pas la réalité psychique. Donc, pour approfondir ce point, nous allons passer en revue les questions

10. Par exemple, lors d'un test expérimental, on fera écouter au sujet "PAY" et "LAY", celui-ci entendra pour les deux cas "PLAY".

phonologiques entourant les rhotiques alvéolaires.

III.2.2.3 Caractéristiques phonologiques des rhotiques alvéolaires

Les rhotiques alvéolaires ont aussi été approchées sous l'angle de la phonologie. Différents courants théoriques ont tenté de représenter la réalité cognitive de ces sons. La thèse de Bradley [Bradley, 2001] et la thèse de Dickey [Dickey, 1997] recensent la plupart des tentatives de représentations phonologiques menées sur les rhotiques alvéolaires. Donc, nous renvoyons au travail de ces chercheurs pour de plus amples informations. Au cours de cette sous-partie, nous évoquerons plus les processus phonologiques et morphologiques qui affectent notre objet d'étude.

III.2.2.3.1 Mutation et Rhotiques alvéolaires

III.2.2.3.1.1 Diachronie

Au cours de l'évolution des langues, les rhotiques ont été amenées à muter pour devenir d'autres rhotiques ou bien des non rhotiques, voire à disparaître. Nous allons citer sans aucune exhaustivité quelques exemples d'évolution des rhotiques à travers le monde :

	phonème original	phonème altéré
Français [Demolin, 2001]	/r/	/ʁ/
Allemand [Schiller, 1999]	/r/	/ʁ/
Rifain [Lafkioui, 2007]	/r/	∅
Anglais [Labov, 1976]	/r/	/ɹ/

TABLE III.2.2 : Exemple d'évolution phonétique de la rhotique

Au sujet des langues qui connaissent une élision de leur rhotique alvéolaire, il a été constaté que ces dernières étaient compensées par un allongement vocalique des voyelles adjacentes. De telle évolution ont été observée pour le rifain [Lafkioui, 2007] et la langue grecque de Samothrace [Katsika and Kavitskaya, 2015]. La rhotique peut aussi être un son épenthétique comme c'est le cas en anglais [Uffmann, 2007]. Ou bien encore, citons également comme évolution phonétique possible la métathèse que nous retrouvons dans les langues romanes [Webb and Bradley, 2009].

III.2.2.3.1.2 Dialectologie

Si au niveau historique les sons ont évolué cela implique possiblement une variation également en synchronie. Car toutes les variantes d'une langue ou d'une famille de langues n'évoluent pas de la même manière. Par conséquent, ces variations des ||r|| peuvent de nos jours se recouper avec des dimensions extra-linguistiques. Que cela soit sur le plan diatopique ou diastratique, on rencontre des productions de ||r|| qui peuvent être en adéquation avec des aspects sociologiques. Ainsi, de telles caractéristiques ont été observées pour nombre de langues dans le monde. Nous allons ici même présenter quelques cas de figure.

Dans un article d'Alvarez Lopez [[Alvarez López, 2010](#)], ce dernier expose le cas de variations phonétiques qui touchent les populations afro-latines, appartenant à des classes sociales basses et parlant des variantes de l'espagnol ou du portugais. Ces groupes de population ont comme caractéristique de présenter une alternance entre les consonnes /d/, /r/ et /l/. Un autre exemple bien connu, en rapport toujours avec des populations d'origine africaine, est celui des afro-américains qui ont tendance à ne pas prononcer la rhotique approximante [ɹ] dans certains contextes et qui a été mis au jour par plusieurs études dont celles de Labov [[Labov, 1976](#)]. Par ailleurs, cette prise de conscience de l'écart qui existe entre leur propre variété et la variante considérée par eux comme plus prestigieuse entraîne généralement des hypercorrections. En somme, les locuteurs voulant à tout prix se rapprocher de cette variété haute vont changer leur manière de parler, mais ils vont commettre des erreurs en surgénéralisant une règle linguistique.

Sur le plan purement de la variation géolinguistique, le domaine des langues romanes est un terrain qui offre beaucoup de cas de variations qui touchent les rhotiques [[Müller, 2011](#)]. On peut citer par exemple le cas de la variation qui existe dans les dialectes romans de l'est de l'Espagne, le catalan, où les rhotiques en coda et avant consonne sont soit des taps à Majorque ou à Valence et des trills à l'est de la Catalogne [[Recasens and Espinosa, 2006](#)].

III.2.2.3.1.3 Allophonie et Neutralisation

Au niveau des variations, il y en a une autre, appelée variante contextuelle ou allophone. Ces derniers sont des variantes d'un phonème original qui a cours que dans un contexte distributionnel très précis. En fonction des langues, les rhotiques peuvent

être sujettes à ce genre de variations. Les variations qu'elles présentent diffèrent d'une langue à une autre. Dans la littérature, nous retrouvons évoqué des exemples que nous allons reporter ici même pour illustration. En russe, d'après Zygis [Zygis, 2005] la rhotique palatalisée [r^j] n'est présente qu'au contact de voyelles antérieures et fermées ou de la semi-voyelle [j], dans d'autres contextes il y a la présence d'une trill.

La variation contextuelle peut ne pas déboucher forcément vers la création d'une nouvelle unité phonique, cela peut tout simplement évoluer vers une unité phonique déjà présente dans le système et donc on parle dans ce cas de neutralisation. Une telle situation a bien évidemment eu lieu aussi pour les rhotiques. On rencontre à travers les langues du monde des cas de figure qui correspondent à ce phénomène et que nous allons évoquer de ce pas. Dans les dialectes romans d'origine ibérique, la rhotique /r/ a connu un processus phonologique de lambdacisation dans certains contextes phonétiques [Müller, 2011, p.140]. Dans une langue berbère située dans la région de Timimoun — sud de l'Algérie —, les taps connaissent différentes évolutions, selon le contexte, elles sont devenues soit des /ʕ/, /ħ/, /h/, soit il y a eu élision complète [Boudot-Lamotte, 1964].

III.2.2.3.2 Contrastivité entre les trills et les taps

Bradley [Bradley, 2001] dans une étude typologique prenant en compte sept langues d'horizons différents montrent que ces deux sons peuvent avoir une distribution complémentaire ou distinctive. Cet hispanisant met en lumière trois positions qui sont révélatrices de ce phénomène, mais cela n'induit pas que les autres positions ne partagent pas ces particularités. Ces positions en question que peuvent occuper ces phonèmes sont : la position initiale, intervocalique et dans une attaque complexe — groupe consonantique hétéroorganique —. Selon les langues, les taps et les trills peuvent occuper les mêmes positions ou bien il y a une répartition complémentaire de ces positions entre les taps et les trills. Par exemple, en kurde, les taps et les trills peuvent être présentes en positions initiales et en attaques complexes, cependant en position intervocalique, seules les trills occupent cette position. Cette distribution se révèle être particulière à cette langue, alors que dans d'autres langues les données se présentent différemment. À partir de ces constats, il en déduit que les trills et les taps, lorsqu'elles co-existent dans une langue, ont une tendance à se neutraliser.

III.2.2.3.3 Sonorité et Rhotiques alvéolaires

La sonorité est un des phénomènes phonétiques connue depuis le XIX^{ème} siècle. Il fut constaté que les phonèmes ont des niveaux de sonorité propre. Cette dernière correspond physiquement à l'amplitude acoustique d'un phonème. Ainsi, il existe une hiérarchie de sonorité, en conséquence, il y a des phonèmes qui sont considérés comme ayant une sonorité plus ou moins élevée par rapport à d'autres. Mais cela n'est pas tout, car l'autre observation qui a été faite, et qui est d'un intérêt capital, est que la question de la sonorité a son importance dans la syllabation et la phonotactique [Clements, 1990]. De ce fait, l'ordre des unités phoniques dans une syllabe serait déterminé par cette sonorité. Les sons les moins sonores ne peuvent devancer les sons les plus sonores dans le cas de groupes consonantiques — ou suites consonantiques — et le noyau d'une syllabe doit correspondre au son le plus sonore. Le phénomène que nous venons de décrire est appelé le principe de sonorité séquentielle. De telles constatations ont pu être faites en chleuh [Dell and Elmedlaoui, 1997], alors que dans les autres langues berbères — sauf erreur de notre part — cette question n'a jamais été abordée. Dans cette hiérarchie de sonorité, le rang des rhotiques alvéolaires est juste avant celui des semi-voyelles et des voyelles qui, elles, se trouvent en première position [Clements, 1990]. D'autre part, le principe susmentionné bien qu'il soit un fait universel et reconnu par l'ensemble des phonologues, il existe pourtant des divergences concernant certains points touchant à ce principe de sonorité [Parker, 2012]. Certains chercheurs arguent que dans un groupe consonantique la première consonne et la deuxième doivent être éloignées d'un point de vue sonore de 2 rangs sur l'échelle de sonorité. Les tenants de cette hypothèse nomment ce phénomène la distance minimum de sonorité — Minimum Sonority Distance —. Par contre, une autre partie des phonologues affirme que dans cette suite de consonnes la distance de sonorité doit être répartie équitablement. Les partisans de ce courant phonologique appellent ce phénomène le principe de dispersion de sonorité — Sonority Dispersion Principle —. La question, qui est en jeu dans ce débat et qui divise les phonologues, est est-ce la classe des liquides ou la classe des glides qui occupe naturellement la deuxième position dans les groupes consonantiques. Cependant, certaines études remettent en question ce concept en mettant en évidence d'une part qu'en terme de sonorité, la tap et la trill ne peuvent être sur le même plan. La tap étant considérée comme plus sonore que la trill [Pons Moll, 2008]. D'autre part, certains chercheurs expliquent que si en espagnol ou en catalan il n'existe pas de trills dans un groupe consonantique, mais uniquement des taps, c'est pour

respecter le principe de sonorité séquentielle. Toutefois, comme certains chercheurs l'ont noté [Bradley, 2001, p.101], il existe des langues comme le basque qui n'ont pas cette contrainte et ont des trills dans des séquences avec des fricatives ou des occlusives.

III.2.2.3.4 Homorganicité et Rhotiques alvéolaires

Les rhotiques alvéolaires sont des consonnes coronales antérieures et ces dernières ont du mal à coexister avec d'autres consonnes qui ont presque le même point d'articulation : /n/, /l/, /s/, /z/, /d/, /t/. Les langues ont recours à plusieurs stratégies pour éviter que des consonnes alvéolaires soient dans le même environnement. Pour cela, soit il y a un phénomène de dissimilation ou bien d'assimilation — ou harmonie consonantique — qui est appliqué. Par ailleurs, ce phénomène a cours soit à l'intérieur d'un même morphème ou bien entre morphèmes. D'autre part, ce phénomène ne s'embarrasse pas de la localité comme définie dans le principe du contour obligatoire, ni du croisement des lignes d'association cher à la théorie autosegmentale [Alderete, 1997]. En latin, par exemple, il existe deux formes pour un même morphème : aris/alis. L'un ou l'autre apparaît en fonction du mot qu'il préfixe. Si ce mot contient un /r/ alors l'allomorphe utilisé sera "alis" et si c'est un /l/ l'autre allomorphe sera employé — des exceptions existent — [Bennett, 2013, p.405]. Parfois, l'ordre de ces sons homorganiques influe comme en hongrois ou en italien où l'assimilation entre /r/ et /l/ n'a pas lieu si /r/ précède /l/, mais si l'ordre est inversé, alors la séquence /r/ + /l/ devient une trill [Pons Moll, 2008]. Il arrive de même que l'homorganicité touche un même et unique phonème qui ne peut se suivre à lui-même dans un lexème ce qui provoquera une dissimilation comme en yindjibarndi et en warlpiri [Bennett, 2013, p.614].

III.2.2.3.5 Phonotactique et Rhotiques alvéolaires

Selon les langues, les trills ou les taps sont susceptibles d'occuper toutes les positions dans le morphème, inclus toutes les positions syllabiques, et être dans tous les environnements possibles ou au contraire ils ont une distribution limitée. Dans les langues slaves, Zygis [Żygis, 2005] indique que les trills peuvent être en position d'attaque, de coda et être adjacentes à une consonne. Ce qui n'est pas le cas de la trill palatalisée, qui pour les langues slaves qui en possèdent, peut avoir une distribution plus limitée. Par exemple, seules les langues russe et bas-sorabe ont des trills palatalisées en coda. En espagnol, les possibilités sont plus resserrées, les trills apparaissent

seulement en position intervocalique ou initiale, ou bien encore si elles sont précédées de consonnes nasales, latérales ou sibilantes [Bradley and Schmeiser, 2003]. Toujours d'après le même auteur, les taps ont aussi une distribution qui leur sont propres et qui est complémentaire avec la trill sauf en position en intervocalique où elles sont distinctives. Ce type de distribution est également le même en catalan [Recasens and Espinosa, 2006]. Certaines langues ont la particularité de pouvoir avoir en noyau syllabique une trill ou une tap comme c'est le cas dans certaines langues slaves [Šimáčková, 2001]. Cela fut également constaté dans une langue berbère, en l'occurrence le chleuh [Ridouane, 2015]. La phonotactique de ces rhotiques alvéolaires est aussi conditionnée par la sonorité comme nous l'avons déjà évoqué.

III.2.2.3.6 Gémination et Taps

Dans certaines langues [Castillan/Espagnol : [Harris, 1984]; Catalan : [Wheeler, 1979]; Chleuh : [Ridouane, 2014]; Siwi : [Naumann, 2012]], les trills sont considérées comme étant les formes géménées des taps. Dans cette hypothèse, il est considéré que les trills ne sont qu'une réalisation de surface d'une forme sous-jacente qui serait la géminée /r:/. Cette hypothèse ne fait toutefois pas l'unanimité chez les linguistes espagnols, ainsi Recasens et Pallarès [Recasens and Pallarès, 1999] s'y opposent en évoquant des considérations d'ordre articulatoire. Bien que certains chercheurs s'opposent à cette opinion, il existe nombre de descriptions qui montre que la gémination de la tap se réalise comme une trill comme c'est le cas en haoussa et en hindi [IPA, 1999, p.93 & p.102]. Une autre particularité qui est nécessaire de mettre en évidence est celle de la distinction entre les trills lexicales et les trills morphologiques. Dans le cas des langues hispaniques, il n'y a que des trills lexicales et pourtant cela pousse certains chercheurs à considérer ces trills comme des taps géménées. Alors que dans les langues comme le chleuh, une bonne partie des trills sont des trills morphologiques qui correspondent à la gémination de la tap. Donc, dans ce dernier cas, il est plus aisé d'analyser la trill comme étant la gémination d'une tap.

D'autre part, il existe également des langues comme le finnois où il y a une distinction entre trill simple et trill géminée. Dans cette langue, la trill géminée est la contrepartie de la trill simple. Ainsi, la trill simple se réalise avec deux contacts et la trill géminée connaît trois ou quatre contacts [Ladefoged and Maddieson, 1996]. Donc, la trill n'est pas forcément une forme de surface d'une tap géminée comme a pu le démontrer également l'étude typologique de Inouye [Inouye, 1995].

Par ailleurs, dans les études espagnoles [Bradley, 2001, p.55], certains des chercheurs qui défendent l'idée que la trill est une tap géminée affirment que ce dernier est une géminée hétérosyllabique¹¹. Soulignons par ailleurs que nous retrouvons dans la littérature que très peu de mentions concernant des trills qui seraient le fait de fausses géminées¹² obtenues par concaténation ou assimilation comme celles qui ont cours entre morphèmes (géminée post-lexicale / hétéromorphémique)¹³. Pour compléter cette sous-partie, nous souhaitons faire ressortir que les auteurs considérant les trills comme des taps géminées n'évoquent pas le fait, lorsqu'il s'agit de ce type de trills provenant de taps géminées, s'ils respectent le principe d'inaltérabilité et d'intégrité.

III.2.2.4 Voisement et Rhotiques alvéolaires

Les rhotiques alvéolaires ne sont pas nécessairement des consonnes voisées comme le laisse à penser le tableau de l'API¹⁴. Certaines langues connaissent dans leur système phonologique des rhotiques alvéolaires non voisées. Il a été fait mention de ce type de rhotiques dans plusieurs langues [[Ladefoged and Maddieson, 1996, p.221], *italien* : [Celata et al., 2014]]. Cependant, ce cas de figure est très rare dans le monde, par exemple les trills non voisées ne concernent que 1,5% des langues du monde [Maddieson and Disner, 1984].

Solé [Solé, 2002] explique que si les trills sont dans la très grande majorité des cas voisées, c'est parce que d'un point de vue du moindre effort et d'un point de vue perceptuelle, la trill voisée obtient un meilleur rendement. Par ailleurs, toujours selon le même auteur, les trills non voisées seraient moins affectées par la baisse de la pression intra-orale et continuerait à réaliser des cycles de contact même après cette réduction

11. Ce terme a très cours chez les hispanisants, mais rien n'est dit sur les trills ambisyllabiques à savoir s'il existe ou non. Ce terme d'ambisyllabique est utilisé pour désigner la formation de syllabes qui, pour un phonème géminé, peut se réaliser entre deux syllabes différentes — hétérosyllabique — ou dans une même syllabe — tautosyllabique — [Louali, 2002].

12. Dans le jargon des phonologues, ce terme désigne une géminée qui a les mêmes propriétés acoustiques qu'une vraie géminée, mais qui au niveau des représentations constituent deux segments, alors qu'une vraie géminée représente un segment — pour plus de précisions, voir [Ridouane, 2010] —.

13. En l'occurrence, Bradley [Bradley, 2001, p.99] évoque une telle situation dans la langue kurde, mais les auteurs, chez qui il rapporte ce fait, considèrent que la consonne en question est un flap et non une tap. D'autre part, Lahiri [Hankamer and Aditi, 1988] rapporte qu'en malayalam le -rr- tautomorphémique devient -tt- alors que le -rr- ne connaît aucune modification phonologique. Là encore avec ce dernier exemple, il n'y a pas d'assurance que l'auteur parle réellement d'une tap ou d'une trill.

14. Consultable sur le site internet de l'API : <https://www.internationalphoneticassociation.org/content/full-ipa-chart> (Consulté le 16/07/2017).

de pression, ce qui n'est pas le cas pour la trill voisée.

Enfin, la présence de rhotiques non-voisées, qui sont retrouvées dans diverses langues, ne correspondent pas nécessairement à des phonèmes à part entière. Il s'agit généralement de réalisations idiomatiques, causées par un défaut d'élocution. Ce fait est généralement signalé pour le cas des trills en final d'un lexème qui sont le plus sujet à ce phénomène. Solé explique qu'une telle situation est provoquée par la diminution de la pression subglottale qui parvient des poumons. Cette pression n'étant plus suffisante, celle-ci ne parvient plus à faire vibrer les plis vocaux.

III.2.2.5 Vocoïdes et Rhotiques alvéolaires

Au cours du cycle de production des rhotiques que ce soit pour les taps ou les trills, il fut observé l'apparition de ce qui est appelé selon les écoles un vocoïde ou un svarabhakti. Ce constat a été réalisé dans bien des langues à travers le monde [*Catalan* : [Recasens and Espinosa, 2007]; *Hongrois* : [Vago and Gósy, 2007]; *Grec* : [Baltazani et al., 2011]; *Polonais* : [Savu, 2014]; *Roumain* : [Avram, 1993]; *Tchèque* : [Pavlík, 2014]; *Castillan/Espagnol* : [Bradley and Schmeiser, 2003], [Schmeiser, 2009]; *Chleuh* : [Coleman, 1999]; *Siwi* : [Naumann, 2012]]. Il apparaît que perceptuellement ce vocoïde peut être audible dans certains contextes [[Hall, 2006]; [Schmeiser, 2009]; [Warner and Tucker, 2011]]. Ce vocoïde n'apparaît pas systématiquement, cela dépend dans quel environnement est prononcé la rhotique. Ainsi, ce vocoïde tend à apparaître s'il est suivi ou précédé par une autre consonne. Toutefois, ce contexte ne garantit pas une présence systématique du vocoïde, car en espagnol, Blecua [Blecua, 2001] a montré que ce vocoïde n'était pas visible dans certains groupes consonantiques. Recasens et Espinosa [Recasens and Espinosa, 2007] indiquent que la présence ou non de ce vocoïde semble être conditionnée par des facteurs dialectaux et contextuels — si la consonne est hétéroorganique et voisée alors le vocoïde apparaît plus souvent que s'il n'est pas dans cet environnement —. En espagnol [Schmeiser, 2009], ce vocoïde est retrouvé également en final d'un mot, mais jamais à l'initial. Ce qui n'est pas le cas en grec, où il fut observé dans cette position-ci [Baltazani and Nicolaidis, 2013].

La nature de ce vocoïde a fait l'objet de débat. Certains chercheurs considèrent ce vocoïde comme une épenthèse qui serait causée par la coarticulation d'une voyelle qui suit ou précède la rhotique alvéolaire. Pour d'autres, ce vocoïde est une partie intégrante de la rhotique alvéolaire. Pour Proctor [Proctor, 2009, p.102], qui défend cette dernière

hypothèse, ce vocoïde est une manifestation acoustique causée par la zone dorsale de la langue qui se déclenche et se surélève avant le contact de la pointe de la langue et qui persiste même après la fin du contact. Rappelons que, comme nous l'avons vu aux points III.2.2.2.2.1 et III.2.2.2.1.1, les rhotiques alvéolaires ont une seconde articulation et c'est celle-ci qui s'active avant et après l'activation de la partie avant de la langue.

Par ailleurs, d'autres chercheurs ont rapproché ce phénomène à des faits identiques qui ont été observés dans les mêmes conditions, mais pour d'autres phonèmes. Hall [Hall, 2003] liste un ensemble d'unités phoniques qui partagent ce même particularisme. La particularité qu'ont tous ces phonèmes est d'être voisée et d'avoir un point d'articulation qui peut s'étendre de la zone uvulaire à la zone glottale. Selon ce chercheur, ce qui est retrouvé dans le cas de ces phonèmes n'est pas à considérer comme étant une épenthèse, puisqu'il n'est pas de nature phonologique, mais phonétique. Ce qui l'amène à proposer un autre terme pour qualifier ces derniers qui est voyelle intrusive — intrusive vowel —.

Comme nous l'avons vu ci-dessus, ce vocoïde est systématiquement retrouvé lorsque la rhotique alvéolaire côtoie une autre consonne, donc cela implique que si la rhotique alvéolaire n'est pas dans les conditions pré-citées comme à l'initial d'un lexème, alors nous devrions ne pas constater de vocoïde. Les chercheurs Baltazani et Nicolaidis [Baltazani and Nicolaidis, 2013] ont pu démontrer en grec la présence de ce vocoïde, même à l'initial. Donc, cela a pour conséquence de remettre en cause la thèse du chevauchement qui a longtemps été défendue pour expliquer l'existence de ce vocoïde.

Les mêmes auteurs précédents, dans une autre étude, ont également mesuré la durée de ces vocoïdes. Ils ont pu montrer qu'en moyenne leur durée différée suivant la position syllabique de la tap. La durée du vocoïde de la tap en position /rC/ est plus longue que celle en position /Cr/. Une comparaison des résultats de Baltazani et al. et de Schmeiser montre que le vocoïde en position /rC/ a une durée qui est comprise entre 20 et 40 ms. Par ailleurs, Bradley et Schmeiser [Bradley and Schmeiser, 2003] mentionnent une ancienne étude de Gili Gaya [Gili Gaya, 1921] où il constate que pour une même position syllabique /Cr/ la durée de ce vocoïde pouvait être influée par plusieurs paramètres comme l'accentuation, le type de consonnes — dorsal ou labial — précédant la tap et enfin la position dans le mot — initial ou intervocalique —. Schmeiser [Schmeiser, 2009] dans la même perspective montre, avec la séquence /rC/, que la durée du vocoïde peut aussi varier selon si la consonne est hétéroorganique ou homorganique, voisée ou non, occlusive ou non. C'est dans chacun de ces premiers cas que la durée du vocoïde sera la plus longue.

III.2.2.6 Modélisation, Synthèse sonore, Simulation et Rhotique alvéolaire

Bien que cela ne soit pas familier aux linguistes, il existe également des recherches qui tentent de modéliser les sons de la parole. Plusieurs approches existent soit les chercheurs souhaitent reconstituer synthétiquement les sons naturels du langage, par addition d'ondes ou filtrage, sans prendre en compte la réalité articulatoire. Ou bien à l'inverse, ils partent d'un modèle articulatoire qui cherchera à produire des ondes acoustiques se rapprochant des sons de la parole. Cette approche est rendue possible par une modélisation des conduits vocal et nasal qui s'apparente plus ou moins à la réalité physique. Cette dernière approche peut être utilisée dans deux voies différentes, soit en simulant une configuration articulatoire particulière pour observer quel type de signal acoustique est produit. Ou soit en cherchant à rapprocher un son synthétique d'un son naturel pour déduire les différentes positions des articulateurs comme c'est le cas dans l'étude de Yeou et Maeda [Yeou and Honda, 2011] sur les rhotiques uvulaires.

S'agissant d'études adoptant la première méthode pour les rhotiques alvéolaires rien n'a été trouvé lors de notre recherche bibliographique. Mais le deuxième cas de figure a été exploré par McGowan [McGowan, 1992] où il a tâché de modéliser les trills en ayant recours à un modèle mécanique simplifié qui est ensuite modélisé par un circuit électronique.

CHAPITRE III.3

Méthodologie

Cette partie de la thèse sera consacrée à la présentation de la méthodologie employée pour analyser nos phonèmes. Cela ira de la constitution du corpus jusqu'aux matériels utilisés. Nous détaillerons ainsi tout le processus qui nous a permis d'aboutir à nos résultats. Pour rappel, notre travail a consisté à étudier les rhotiques alvéolaires d'une langue berbère parlée dans le sud-ouest algérien. Pour mener à bien notre étude, nous avons récolté un corpus que nous avons enregistré. L'enregistrement de ce corpus s'est fait grâce à un microphone à électret et un enregistreur numérique. Les données collectées ont ensuite fait l'objet d'un traitement pour pouvoir être exploitées. Pour les traiter nous avons eu recours à plusieurs logiciels qui nous ont facilité la tâche pour analyser nos données. C'est aussi au cours de ce chapitre que nous évoquerons quels sont les points que nous allons étudier et qui seront évoqués dans le chapitre Résultat III.4 (page 240).

III.3.1 Corpus

Pour pouvoir étudier nos phonèmes, nous avons bien évidemment établi un corpus. Cependant, ce dernier peut s'élaborer de plusieurs manières suivant la méthode choisie. Il nous a d'abord fallu déterminer la méthodologie à utiliser pour pouvoir dégager notre corpus. Dans la littérature, il n'y a pas vraiment d'élaboration type de corpus. Les chercheurs ne semblent pas se poser la question de la méthode la plus adéquate pour parvenir à des résultats représentatifs. Les méthodes utilisées sont toutes différentes et cela à tous les niveaux. Que nous sommes en droit de nous poser la question s'il est possible de comparer entre elles ces études tant elles emploient des méthodologies toutes différentes.

En SHS, il y a deux méthodes classiques qui sont mises en œuvre : la méthode quantitative et la méthode qualitative. Cette dernière n'est pas applicable dans le cas d'une étude s'intéressant à l'aspect physique de la parole, par conséquent seule la méthode quantitative est utilisable. Et en manière de recherche quantitative, surtout en sociophonétique, il semble tout de même qu'il se dégage deux méthodes que nous pouvons dénommer : méthode quantitative intrinsèque ou extrinsèque. La première méthode consiste à avoir un objet d'étude qui est collecté auprès d'un informateur ou un peu plus, mais dont la spécificité sera de recueillir l'objet d'étude auprès d'une large gamme d'items. Cela peut atteindre des centaines d'unité(s) prononcées par ce nombre d'informateurs réduits. Alors que la méthode quantitative extrinsèque met l'accent sur

la diversité des locuteurs, cela peut aboutir à des centaines de personnes [Labov William et al., 2008, p.27], avec un corpus réduit à quelques items seulement.

En somme, dans la première méthode, la quantité se trouve au niveau des items, alors que dans la deuxième, la quantité se situe au plan des locuteurs. Ainsi, les contraintes du terrain étant tels — moyens financiers réduits, informateurs en nombre limité, ... etc — que la méthodologie que nous avons adoptée était la seule à nous être accessible. Par conséquent, naturellement, nous avons fait le choix de la méthode quantitative intrinsèque et donc de réaliser un corpus comprenant plusieurs items. De ce fait, notre corpus se compose d'environ deux cent vingt items qui appartiennent à des catégories syntaxiques diverses : verbe, nom, préposition, adverbe — voir en Appendice (page 298) —. Rappelons par ailleurs que la quantité est le seul moyen de nous prévenir contre le particularisme, c'est-à-dire de réalisations phonétiques idiomatiques qui seraient propre à un individu. Donc, cet aspect est important pour la crédibilité des résultats d'autant plus si le chercheur n'est pas familier du terrain d'étude sur lequel il travaille, comme c'est notre cas.

Nous avons accordé une attention toute particulière à la distribution de nos rhotiques alvéolaires. Nous avons essayé, autant que faire se peut, d'avoir un corpus où ces unités phoniques pouvaient être dans toutes les positions possibles. Cela comprend les positions à l'intérieur du morphème — initial, médiane -intervocalique ou non- et final — comme les positions syllabiques — attaque -complexe ou non-, noyau et coda —. La combinaison de ces deux types de positions offrent des configurations très nombreuses qui sont potentiellement au nombre de 36. Notre corpus comprend également des items hétéromorphémiques intégrant la prononciation d'une rhotique alvéolaire en jonction avec un autre morphème. Donc, en somme, notre échantillon se veut le plus représentatif possible de la distribution de ces phonèmes et cela entraîne du même coup des réalisations phonétiques qui sont potentiellement très diverses du fait de ces différents environnements.

III.3.2 Population d'étude

Pour mener à bien notre étude, nous avons enregistré deux locuteurs natifs du TmAS résidant dans la localité d'Asla dans la wilaya de Naâma. Un deuxième locuteur s'est avéré nécessaire pour dégager sans ambiguïté les caractéristiques phonétiques de nos rhotiques et pour déterminer ce qui relevait du conventionnel

ou du particularisme. Signalons que ces deux personnes sont des frères, ce qui est d'un point vue sociolinguistique intéressant dans la mesure où la variabilité qui pourrait être constatée dans nos résultats ne pourrait être imputée principalement à des considérations extra-linguistiques. Donc, cela réduit les hypothèses possibles qui pourraient être envisagées. Nous avons réalisé un tableau présentant les principales caractéristiques de l'échantillon de notre population d'étude :

	Informateur 1	Informateur 2
Âge	47 ans	35 ans
Sexe	M	M
Métier	Maçon	Enseignant
Langues maîtrisées (autre que le TAS)	Arabe	Arabe
Niveau d'étude	Collège	Université
Fumeur	Oui	Non
Type de morphologie physique	Filiforme	Embonpoint
Type de phonation	Modal / Grave	Légèrement creaky

TABLE III.3.1 : Profils de nos informateurs.

III.3.3 Matériel expérimental

Dans cette présente partie, nous présenterons les caractéristiques des instruments utilisés dans le cadre de notre étude expérimentale sur les rhotiques alvéolaires. En phonétique expérimentale, il existe une flopée d'appareils permettant d'étudier comment les sons de la parole sont réalisés [Ridouane, 2006] [Stone, 2010]. Suivant que nous souhaitons mesurer ou visualiser tel ou tel phénomène, nous devons nous adapter en choisissant ce qui conviendra le mieux pour observer les caractéristiques des phonèmes. La phonétique expérimentale peut se focaliser soit sur ce qui produit les sons, c'est-à-dire les articulateurs, soit sur le résultat de ce que produisent les articulateurs, c'est-à-dire le mouvement mécanique du flux d'air.

Les instruments existant pour observer les articulateurs sont tous invasifs et si nous voulons une qualité d'images correctes, il faut, par ailleurs, d'imposants appareils. Ces deux points nous ont donc dissuadé de mener une étude articulatoire. Nous avons tout de même cherché une technologie qui serait non intrusive et peu encombrante. Nos recherches nous ont donc mené vers la piste de l'imagerie par transillumination, connue aussi sous la dénomination d'imagerie photonique ou optique. Nous nous sommes

déplacé au laboratoire de Physique des lasers à Paris 13¹ pour nous enquérir sur cette technologie et nous avons rencontré plus particulièrement l'équipe de recherche sur l'Optique des milieux aléatoires. Après l'échange avec Jean-Michel Tualle et Éric Tinet, membres de cette équipe, nous nous sommes rendu compte que cette technologie prometteuse n'est pas encore au point et donc nous ne pouvions en faire usage. Nous souhaitons ici remercier ces deux chercheurs de nous avoir accordé cet entretien et de nous avoir dispensé des conseils, même si au final notre quête a été infructueuse.

Ainsi, la seule alternative qu'il nous restait était d'étudier ces phonèmes en nous basant sur les caractéristiques mécaniques du flux d'air projeté par les articulateurs. Pour cela, nous pouvions soit étudier l'aspect aérodynamique de ce flux ou bien ses ondes acoustiques. Concernant les mesures aérodynamiques, il y a un certain nombre d'inconvénients qui ne nous ont pas permis de les utiliser, notamment l'encombrement des appareils qui rendent difficile leur emploi. Donc, au final, la seule option restant à notre disponibilité fut de nous rabattre sur l'analyse des ondes mesurées pour étudier les particularités de nos phonèmes. Mais là encore, ils existent une multitude d'appareils pour réaliser des mesures acoustiques.

Bien que cela soit le cas, en réalité les possibilités que nous avions à notre disposition étaient restreintes, le choix des appareils que nous avons utilisés s'est imposé presque à nous. Compte tenu que notre projet de thèse était une étude acoustique, alors le seul moyen existant pour enregistrer les sons est d'utiliser un microphone. Par ailleurs, la nature du terrain où nous avons mené notre enquête présente des conditions difficiles. C'est une zone désertique et nous étions perpétuellement en déplacement. Donc, il fallait du matériel portable, peu encombrant et pas trop onéreux. Parmi la gamme des microphones — microphone dynamique, micro électrostatique, micro à électret, micro à ruban, ... — et des systèmes d'enregistrement existant — carte son, carte d'acquisition, enregistreur numérique — pouvant répondre à nos exigences, seul le microphone à électret et l'enregistreur numérique nous a paru convenir. Ce sont ces instruments que nous allons vous présenter ci-dessous.

1. Le site du laboratoire : www-lpl.univ-paris13.fr/FR (Consulté le 20/06/2016).

III.3.3.1 Microphone à électret

III.3.3.1.1 Fonctionnement technique

Nous allons commencer d'abord par présenter le microphone à électret. C'est le microphone le plus couramment utilisé dans le monde et retrouvé dans toutes sortes d'applications. Le microphone à électret est un sous type de microphone électrostatique. Brièvement, nous allons présenter de quelle manière techniquement cet instrument arrive à enregistrer le son. Le microphone à électret, comme tout microphone de manière générale, sert à convertir la pression acoustique en un signal électrique. La variation de l'onde acoustique est répercutée en variation de tension électrique. Cette transformation d'une énergie mécanique en énergie électrique est possible grâce au champ électrostatique. Le dessin ci-dessus représente un microphone à électret type, dont les principaux éléments sont :

- l'électret qui est généralement composé d'un matériau isolant et léger (ex : téflon) est recouvert d'une fine pellicule métallique.
- une plaque arrière métallique trouée.

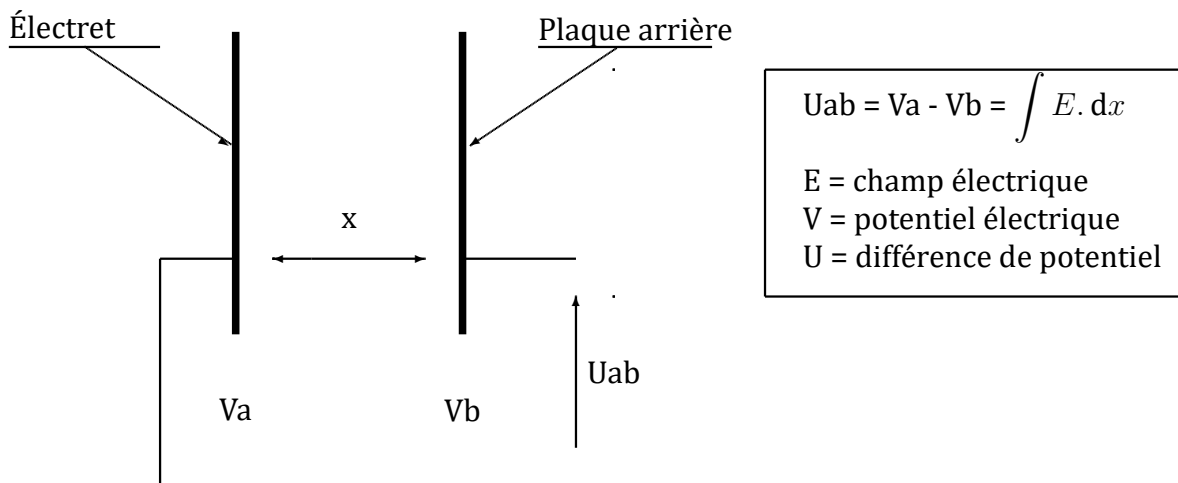


FIGURE III.3.1 : Schéma électrique simplifié du microphone à électret

Ces deux pièces doivent être éloignées et cette distance joue un rôle capital dans la mesure des ondes acoustiques. Ces dernières lorsqu'elles rentrent en contact avec l'électret vont le faire se déplacer et le rapprocher de la plaque arrière. Suivant la pression acoustique plus ou moins importante exercée sur l'électret, l'espace qui sépare ce dernier de la plaque arrière variera également. Cette variation est exploitée pour créer

une variation de la tension électrique — U_{AB} —. Cette dernière correspond à la différence de potentiel électrique entre les deux pièces et celui-ci est créé grâce à la polarisation électrique de l'électret et de la plaque arrière. C'est-à-dire qu'une des pièces a une charge électrique, donc un potentiel électrique, plus important que l'autre afin de créer une différence de potentiel. L'électret est celui qui contient les charges électrostatiques les plus élevées et qui ont été injectés en son sein après un traitement. Ce potentiel électrique est sujet à la distance, plus les deux pièces polarisées se rapprochent et plus le potentiel électrique sera fort.

Rajoutons également que la tension électrique dégagée aux bornes de ces deux pièces étant tellement faible qu'elle a besoin d'être préamplifiée, donc le microphone à électret est aussi composé d'un transistor et d'une résistance pour amplifier le signal. Pour faire fonctionner ces deux composants, il faut une source d'énergie, par conséquent, le microphone à électret est aussi pourvu d'une alimentation.

En résumé, la qualité d'un microphone à électret est essentiellement basée sur les propriétés des différents éléments que nous venons de citer précédemment.

III.3.3.1.2 **Avantage et inconvénient**

Les principaux avantages de ce type de microphone est qu'il est miniaturisable, qu'il est d'un prix abordable, qu'il a une bonne courbe de sensibilité et qu'il a une courbe réponse en fréquence large pouvant aller de 10 Hz à 30 kHz.

L'inconvénient principal de notre instrument est que ses performances ne sont pas optimales. Un niveau de bruit assez élevé — rapport signal / bruit — et une courbe de réponse en fréquence inégale. L'alimentation nécessaire est aussi un point négatif, car il contribue à rajouter du bruit appelé "tension phosphométrique".

III.3.3.1.3 **Le microphone à électret MM-701**

C'est le nom de la marque de notre microphone à électret. Celui-ci est du type micro-cravate destiné à un usage professionnel. Ses caractéristiques techniques sont les suivantes :

Courbe de directivité	unidirectionnelle (cardioïde)
Courbe de réponse en fréquence	100 Hz – 20 kHz
Sensibilité	-35dB à 1kHz
Alimentation	DC AA 1.5V battery
Impédance de sortie	600 Ω
Connecteur	prise jack

TABLE III.3.2 : Caractéristiques techniques du microphone à électret

Le point "fort" de notre microphone est sa courbe de directivité. Ce paramètre est d'une grande importance, car elle permet de réduire le champ de captation des sons aux ondes acoustiques provenant de devant le micro. Ce point, compte tenu de nos conditions d'enregistrement, est capital, puisque notre corpus n'a pas été recueilli dans une chambre sourde. De ce fait, cela permet de réduire sensiblement la captation de sons non désirés dans nos enregistrements.

III.3.3.2 Enregistreur numérique

Notre enregistreur numérique est un dictaphone numérique de la marque Olympus WS-813. Cet appareil est destiné à un usage professionnel. Le microphone à électret se branche sur le haut du dictaphone grâce à une prise jack. Ce type de connectique n'est pas généralement recommandé pour obtenir un signal de haute qualité dans la mesure où il ne permet pas de transmettre le signal dans les meilleures conditions.

Donc, ce signal "parasité", qui est la tension fournie par le microphone, est récupéré pour être traité par le dictaphone. Celui-ci se charge d'échantillonner le signal en signal numérique. Pour ce faire, il faut au préalable définir le format dans lequel nous souhaitons enregistrer nos données. Le dictaphone numérique propose plusieurs choix de format, nous avons choisi le format PCM (wav) qui a l'avantage ne pas compresser le signal et ne pas de le dénaturer. Sous ce format, la fréquence d'échantillonnage proposée est à 44,1 kHz / 16 b. Ces valeurs ne peuvent être modifiées, elles sont imposées, cependant elles sont largement suffisantes dans notre cas. Car le théorème de Shannon-Nyquist est respecté, puisque notre bande passante est à 20 kHz environ et que les fréquences qui nous intéressent se trouvent bien en deçà de 6000 Hz.

Après avoir réglé notre fréquence d'échantillonnage et avant de passer à l'enregistrement de notre corpus, il nous faut adapter le volume d'enregistrement avant chaque séance d'enregistrement avec nos informateurs pour avoir le bon volume : i.e. ni trop faible,

ni saturé. Le dictaphone permet de régler manuellement ce paramètre. Généralement, suivant l'informateur, le niveau se trouve entre 10 et 15.

III.3.4 Conditions d'enregistrement

Comme nous l'avons déjà mentionné, les enregistrements ne se faisaient pas dans une chambre sourde. Mais nous tenions tout de même à avoir d'assez bonnes conditions d'enregistrement. Donc, nous avons tenté d'atteindre cet objectif. La prise de sons s'est faite en maintenant le microphone à environ 20 cm de distance du locuteur bien que cela soit un micro-cravate, car nous ne voulions pas que les bruits de frictions du micro sur les vêtements puissent nuire à l'enregistrement. Donc, nous avons utilisé un dispositif pour maintenir le micro à l'écart du locuteur et pour qu'il reste statique. Rajoutons également que nos deux informateurs n'ont pas été enregistrés au même moment, par conséquent, nous allons présenter dans quelles conditions se font faites ces séances avec eux.

Pour le premier informateur, l'enregistrement a été réalisé dans une maison à l'écart de la ville et en début d'après-midi pour réduire au maximum les nuisances sonores. Ce moment de la journée est propice, car les habitants d'Asla ont tendance en cette période à cesser leurs activités pour faire la sieste comme il est habituel de le faire dans les régions méditerranéennes. Toutefois, cela n'a pas suffi à empêcher toute présence de bruits. D'une part, la pièce où nous nous trouvions était vide d'ameublement et assez grande, ce qui a eu pour effet d'accentuer la réverbération. Autre fait à mentionner toujours au sujet de l'enregistrement avec notre premier informateur est un autre bruit qui est intermittent et qui est celui d'un chant d'oiseaux. Mais ce tel désagrément a pu être contrôlé en cessant l'enregistrement lorsque l'oiseau piaillait et en reprenant ensuite la séance à l'arrêt de son chant. Donc, cela n'a pas nui outre mesure nos données. Le spectre fréquentiel du son de l'oiseau est très élevé et facilement reconnaissable, comme on peut le voir ci-dessous à l'image. Donc, il nous sera très aisé, dans l'éventualité de résultats non conformes de déterminer s'il est oui ou non responsable de cet écart. Le second informateur a été enregistré dans d'autres conditions. Nous avons conduit notre enquête le soir et à l'intérieur d'une habitation à l'écart de la ville afin d'éviter au maximum les nuisances sonores. Contrairement au cas de figure précédent, la pièce où nous étions était mieux insonorisée, car petite et meublée.

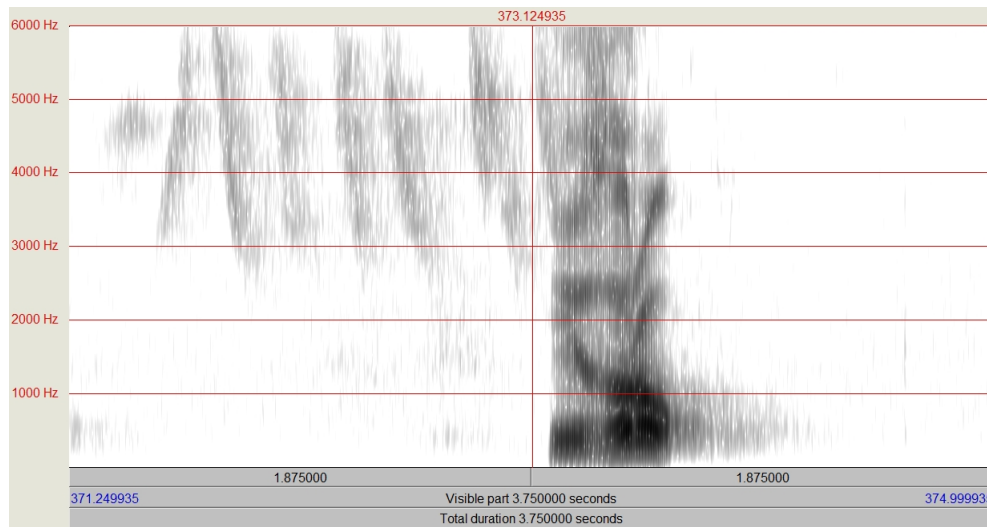


FIGURE III.3.2 : Extrait d'un passage du corpus présentant l'artefact acoustique du chant de l'oiseau.

III.3.4.1 La réverbération

Lorsqu'un enregistrement audio est réalisé en dehors d'une chambre sourde, nous courons le risque inévitablement d'être affecté par un phénomène de réverbération. Cela s'est présenté tout particulièrement lors des conditions d'enregistrement avec l'informateur 1 et elles ont aussi touché notre second enregistrement. En théorie, la réverbération a pour effet principalement de déformer l'intensité et la durée d'un signal, mais plus précisément :

it blurs temporal and spectral cues, flattens formant transitions, reduces amplitude modulations associated with the fundamental frequency of speech, and increases low-frequency energy, which in turn results in masking of higher speech frequencies [Hu and Kokkinakis, 2014]

En réalité, il serait plus juste de parler de réverbérations au pluriel, car il y a deux types de réverbérations appelés dans la littérature anglo-saxonne : early reflection et late reflection. C'est cette dernière réverbération qui peut nuire à la qualité sonore d'un enregistrement. La "late reflection" est le fruit d'une accumulation d'onde réfléchi et dont le temps de décalage avec le son original est plus grand que celui de la "early reflection" lorsqu'il arrive au récepteur. Le décalage pour la "early reflection" est de l'ordre de 50 à 100 ms, alors que pour la "late reflection" le temps est plus important

[Hu and Kokkinakis, 2014]². Cette distinction temporelle est importante au niveau des causes qu'elles provoquent sur les sons de la parole. Ainsi, d'un point de vue phonétique, des études montrent les dégâts causés par cette réverbération sur les phonèmes. Ces conséquences sont de deux ordres et sont désignées : overlap-masking et self-masking. Le premier est le fait de la "late reflection" et le dernier est le fait de la "self reflection" :

Overlap-masking is caused by the overlap of reverberant energy of a preceding phoneme on the following phoneme. This effect is particularly evident for low-energy consonants preceded by high-energy voiced segments (e.g., vowels). The additive reverberant energy fills in the gaps and silent intervals (e.g., stop closures) associated with vocal tract closures. Overlap-masking also occurs in vowels, but it is unlikely that the effect is significant, as the intensity of preceding consonants is much lower than the intensity of the following vowels. Self-masking is caused by the internal smearing of energy within each phoneme. This effect is particularly evident in reverberant sonorant sounds (e.g., vowels), where the formant transitions become flattened. This, in turn, produces confusion between monophthongs and diphthongs (e.g., see Nabelek and Dagenais, 1986; Nabelek et al., 1989). Self-masking is also evident in consonants in the initial position. However, the self-masking effect is substantially smaller when compared to the overlap-masking of consonants in the final position (Nabelek et al., 1989) [Kokkinakis and Loizou, 2011].

D'autre part, il faut prendre en compte la distance entre le récepteur — microphone à électret — et l'émetteur — l'informateur — pour juger de la qualité d'un signal. Plus l'émetteur est éloigné du récepteur et plus la réverbération prend le dessus sur le son original. Donc, à l'inverse si le récepteur est très proche de l'émetteur, comme c'est notre cas, les dégâts causés par la "late reflection" peuvent être négligés comme le mentionne Boothroyd [Boothroyd, 2004]. Dans l'article de ce précédent chercheur, nous y trouvons un exemple de signal dénaturé par la réverbération que nous reprenons ici pour illustration :

À la vue de cet exemple qui montre une détérioration prononcée du signal acoustique,

2. Aucun chiffre ne nous est proposé dans les littératures que nous avons consultées, mais on peut faire l'hypothèse que ce temps n'est pas supérieur au temps d'énonciation d'un morphème qui est pour notre corpus d'en moyenne environ 800 ms. Donc, il est possible qu'il y ait un chevauchement entre le son original et la réverbération type "late reflection".

nous pouvons affirmer qu'en comparaison, notre enregistrement est d'assez bonne qualité puisque comparable au spectrogramme supérieur de l'exemple ci-dessus. Dans la mesure où chaque phonème est clairement visible l'un de l'autre. Cela a été possible de masquer la réverbération grâce à la proximité de notre microphone avec notre informateur, mais aussi car nous enregistrons des items et non une parole continue, donc au final les sons qu'il prononce sont clairement audibles. Ce qui n'est pas le cas lorsque ces conditions ne sont pas respectées ordinairement. Cependant, cela ne signifie pas pour autant que cette réverbération a aucun impact, bien au contraire, bien que minime, il y a tout de même eu une influence. Donc, c'est pour cela que lorsque nous analyserons nos résultats nous ferons en sorte d'avoir en tête ce paramètre s'il y a des incohérences.

III.3.4.2 Le rapport signal sur bruit

Un autre paramètre à avoir à l'esprit est le rapport signal sur bruit qui est fondamental, notamment pour la lecture de spectrogramme. Si celui-là n'est pas satisfaisant, il rend difficile, sinon impossible, le déchiffrement des représentations visuelles du spectre. Ainsi, nous avons cherché à déterminer de quelle manière se manifestait cet aspect dans notre corpus audio. Nous avons fait le constat que le bruit de fond statique de nos lieux d'enregistrement étaient relativement faibles et qu'ils ne constituaient pas un possible fait majeur de dénaturation de notre son. Les valeurs recueillies du rapport signal bruit (**Signal Noise Rate (SNR)**)³ à différents endroits de nos enregistrements varient entre 35 et 45 dB. À noter que l'enregistrement avec l'informateur 2 présente un bruit de fond légèrement plus important par rapport au premier enregistrement avec l'informateur 1. Ces chiffres ne sont pas idéaux puisqu'on considère, communément, qu'il faut une valeur au dessus de 50 dB pour estimer que nous avons à faire à un bon enregistrement. Mais s'agissant d'enregistrements obtenus en dehors d'une salle insonorisée, de tels SNR sont tout à fait corrects.

3. La mesure du SNR s'obtient de la façon suivante : $SNR = 20 \log(\text{Amplitude maximum du signal} / \text{Amplitude maximum du bruit})$.

III.3.5 Le protocole

Pour obtenir nos données, nous avons fait éliciter notre corpus par nos informateurs. Pour ce faire, l'ensemble de notre corpus a été rédigé sur un calepin que nous avons fait lire à nos informateurs. Nous avons décidé de procéder ainsi pour ne pas influencer, par notre manière de prononcer, nos locuteurs du TmAS qui seraient tentés de nous mimer. La transcription utilisée pour rédiger nos items sur le calepin est la transcription usuelle employée et élaborée à l'INALCO⁴. Les signes usitées pour nos rhotiques sont les suivants : r - tap -, ɾ - tap pharyngalisée -, rr - trill - et r̥r̥ - trill pharyngalisée -. Néanmoins, la lecture des items n'a pas été sans difficulté, car le TmAS n'a pas de tradition écrite, c'est une langue orale, donc il n'a pas été aisé parfois pour nos informateurs de lire la transcription des items. Par conséquent, nous avons dû intervenir ponctuellement pour dire oralement à nos sujets quel était le mot que nous souhaitions qu'ils prononcent. Signalons également que les items qui avaient une morphologie assez semblable — à l'instar des paradigmes verbaux — ont été dispatchés dans le carnet, donc nous ne les avons pas faits suivre les uns derrière les autres, pour ne pas que l'informateur ait la tentation de les répéter de la même manière.

Ensuite, chaque item de notre corpus a été répété trois fois par nos informateurs. Nous leur avons demandé de bien espacer chaque répétition de l'item. Ces derniers ont respecté dans l'ensemble cette consigne. Nous n'avons pas exigé de nos informateurs qu'ils prononcent ces items insérés dans un énoncé cadre comme le recommandent certains chercheurs. Ils nous a semblé superflu de pratiquer une telle démarche. Nous justifions cette position en faisant remarquer qu'il n'existe pas d'énoncés naturels ou normatifs. Nos items appartenant à des catégories syntaxiques diverses, il nous aurait fallu élaborer plusieurs types d'énoncé cadre. Il s'ensuit que l'influence de ces derniers exercée sur nos items est un point qui nous échappe. De plus, tout acte de paroles est corrélé à une dimension socio-pragmatique. Et chacune de ces dimensions se distinguent par des faits linguistiques où l'aspect phonétique peut jouer. Le fait d'autre part que nous menions une étude expérimentale, où l'informateur se trouve dans une position où il est tout sauf dans son élément naturel, l'amènera inévitablement à adopter un comportement différent, car il est sujet au regard de l'observateur. Donc, additionner à l'item un énoncé n'aurait fait que biaiser le corpus et alourdir la séance

4. La notation usuelle à base latine du berbère : www.centrederechercheberbere.fr/tl_files/doc-pdf/notation.pdf (Consulté le 16/07/2017) & le tableau de Transcription usuelle, page xxiv.

d'enregistrements en provoquant de l'agacement chez nos informateurs. Cela aurait poussé l'informateur à avoir un débit de parole excessif et cela aurait nui à la qualité de l'enregistrement. Donc, nous aurions obtenu l'effet inverse de celui recherché.

Après l'application de ce qui vient d'être dit, nous finissons par obtenir un enregistrement prêt à l'emploi et dont les résultats nous apparaissent comme assez satisfaisants pour un enregistrement de terrain en terme notamment de rapport signal / bruit.

III.3.6 Traitement des données

Après que nos enregistrements soient enfin finis, il reste une dernière étape qui est le traitement de nos données. Ces derniers ont été effectués sur différents supports informatiques. Nous allons seulement vous présenter ici même, les principaux outils que nous avons utilisés pour traiter notre corpus.

III.3.6.1 L'étiquetage et les mesures : PRAAT

PRAAT est un logiciel spécialement conçu pour traiter les signaux acoustiques. Nous avons utilisé ce logiciel (version 5.4.04) pour réaliser différentes tâches. D'abord la première chose que nous avons faite, comme nos enregistrements sont enregistrés en stéréo, nous les avons transformés en mono en utilisant ce logiciel. Pour cela, nous avons choisi le canal gauche pour toutes nos données.

III.3.6.1.1 L'étape de segmentation

Puis, toujours avec ce logiciel, nous avons effectué la segmentation de nos données en étiquetant les morphèmes et les phonèmes qui nous intéressaient en l'occurrence ici principalement les rhotiques alvéolaires. La première étape fut d'effectuer un étiquetage sur chaque item. Pour cela, nous avons utilisé une segmentation automatique basée sur la détection du silence et du son. Sur PRAAT, il est possible de réaliser cette opération à condition d'utiliser un script à cet effet⁵. D'autres logiciels, avec une manipulation plus intuitive, peuvent aussi être utilisés comme Audacity. Cependant, la segmentation obtenue ne peut être réintégrée directement par la suite dans PRAAT pour continuer le

5. Script conçu par Mietta Lennes : http://www.helsinki.fi/~lennes/praat-scripts/public/mark_pauses.praat (Consulté le 20/07/2016).

traitement. Il faut passer par un script qui permet de réencoder au format "TextGrid". Que cela soit l'un ou l'autre moyen utilisé, cela nous a permis d'obtenir un résultat très satisfaisant, nous avons que très peu retouché les démarcations établies par le logiciel. Une telle opération nous a été facilitée par la méthode que nous avons utilisée pour recueillir notre corpus en espaçant convenablement chaque prononciation de nos items.

Ainsi, nous avons poursuivi notre travail sur PRAAT où chaque item de notre corpus a été étiqueté avec différents paliers. Nous allons seulement mentionner ici la segmentation qui concerne l'analyse statistique⁶. Ainsi, le premier palier correspondait à une transcription de l'item basé seulement sur l'encodage ANSI et prenant en compte des caractères qui étaient autorisés sous Windows pour la dénomination d'un fichier. Ce point est important, car l'encodage Unicode n'est pas encore totalement en usage sur tous les programmes informatiques. Par conséquent, cela rend illisible parfois l'usage du fichier, qui contiendrait des caractères Unicode, dans certains logiciels. Donc, nous avons utilisé un ensemble de signes de notre cru qui est toutefois en majorité inspiré des caractères X-SAMPA. Ci-dessous, les symboles ne correspondant pas à l'alphabet X-SAMPA :

Caractères non X-SAMPA	API
,	ɿ
P	ɸ
h'	ħ
E	ə
&	ː
R_	ʀ

TABLE III.3.3 : Caractères non X-SAMPA utilisés pour la transcription des items

Puis, le second palier est réservé à une transcription API, quant aux deux derniers paliers, ils sont affectés aux unités phoniques. Ainsi, le troisième niveau définit le début et la fin de nos rhotiques et les phonèmes lui étant adjacents. Si nous n'avons pas segmenté les autres phonèmes, c'est que nous avons considéré que dans le cadre de ce travail ils ne nous apportaient pas de le faire, si ce n'est nous rendre la tâche plus fastidieuse. Notre but était de rendre compte de l'environnement phonétique immédiat

6. Nous mentionnerons dans la partie Résultat III.4 (page 240) ce qu'il en est de la segmentation pour les autres parties.

de la rhotique susceptible de l'influencer ou elle-même influençant d'autres phonèmes. C'est-à-dire dans le cadre d'un traitement statistique pouvoir observer un phénomène de coarticulation. Enfin, le dernier palier se focalise sur la rhotique en elle-même et les différentes parties la constituant et que nous avons décidé d'étudier comme nous l'avons précédemment défini. Ci-dessous un exemple d'un item qui a été étiqueté :

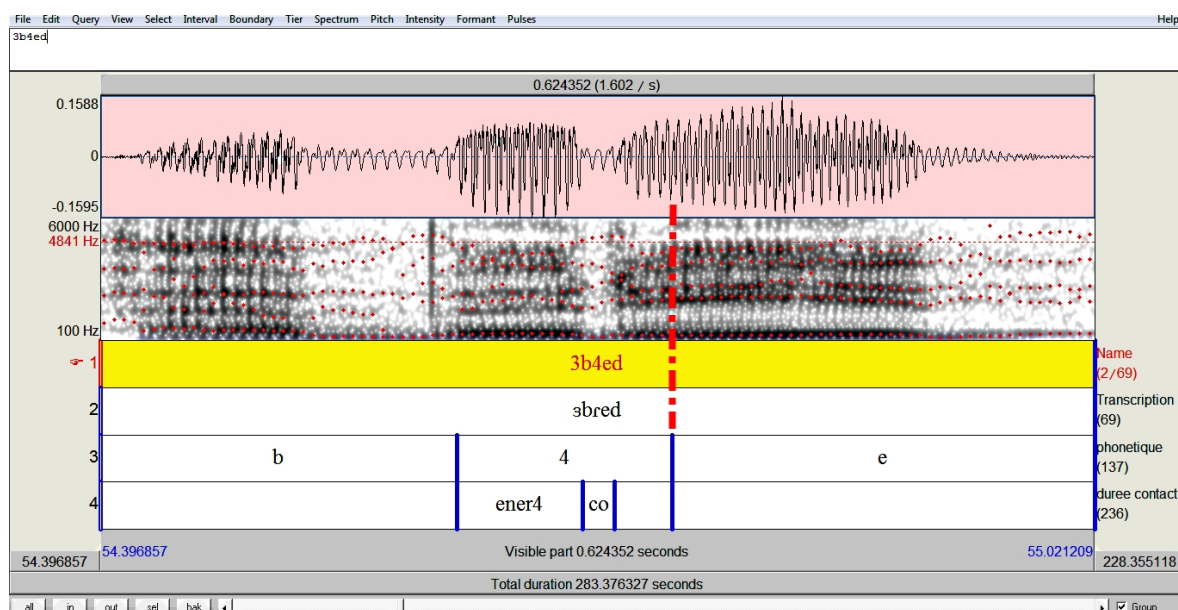


FIGURE III.3.3 : Exemple de la segmentation de l'item : @[ɜbred]@ \route\ — Imprimé écran de l'interface du logiciel PRAAT —)

III.3.6.1.1.1 Les frontières de la rhotique alvéolaire

Une question épineuse est celle de la segmentation de la rhotique alvéolaire. Il n'a pas été toujours facile de marquer où débutait et finissait les rhotiques. Ces dernières tout particulièrement lorsqu'elles sont comprises entre deux voyelles ne se laissent pas aisément délimiter. Par contre, lorsqu'elles sont précédées ou suivies d'une consonne ou si elles sont en final ou à l'initial, il n'y a aucun problème pour déterminer le début ou la fin de ces phonèmes. La figure ci-dessus illustre parfaitement ce problème, nous pouvons voir clairement où commence la rhotique, car elle est précédée d'une consonne, il est par contre difficile de dire où cela finit. Nous pouvons très bien arguer que la rhotique se termine à la fin de la tap, mais l'observation de la rhotique dans d'autres environnements, surtout si elle est suivie d'une consonne, montrent qu'il y a

systématiquement la présence d'un vocoïde qui suit la fin de la constriction ⁷. Donc, il est fort possible que même après une voyelle, il existe un vocoïde qui suit après la fin de la tap et qui n'apparaît pas distinctement sur le signal. Ce constat peut en être de même lorsqu'une voyelle précède la rhotique. De ce fait, nous avons préféré débiter et finir quelques peu après ou avant la constriction, même s'il n'y avait pas clairement d'évidence une frontière acoustique nette.

Toutefois, nous n'avons pas arbitrairement et au hasard décidé de placer le marquage à tel endroit ou à tel endroit, nous nous sommes basé, pour cela, du spectrogramme. Si nous nous référons toujours à la figure précédente III.3.3, le trait rouge pointillé — qui a été rajouté manuellement par nous-même — constitue la prolongation de la fin de notre rhotique tel qu'il a été réalisé sur le tier. Si nous observons attentivement l'image, nous pouvons constater qu'entre la fin de la tap et la réalisation pleine de la voyelle, une zone intermédiaire apparaît. Cet endroit contient des formants, notamment F3 et F4, différents des segments précédents et suivants et nous remarquons également une distribution de l'énergie sur le spectrogramme en haute fréquence inexistante. Ces particularités nous ont poussé à considérer cette partie comme étant intégrée à la rhotique.

III.3.6.1.2 Paramétrages pour le traitement du signal

À l'ouverture de PRAAT, celui-ci est prédéfini et propose de traiter automatiquement les données à partir de paramétrages standards. Nous avons décidé pour obtenir une meilleure visualisation de nos données de toucher à ces valeurs. Les nouvelles valeurs que nous avons utilisées sont celles-ci :

Intervalle en ordonnée (Hz)	0 à 5000
Fenêtre	Gaussien
Longueur de la fenêtre (s)	0,005
Intervalle dynamique (dB)	40
Nombre d'intervalle temporel (s)	750
Nombre d'intervalle fréquentiel (Hz)	250
Pré-emphase (dB/Oct)	8
Maximum (dB/Hz)	100
Compression dynamique	0

TABLE III.3.4 : Les paramétrages du spectrogramme utilisés sur PRAAT.

7. Ce point sera vu plus en détail par la suite (Partie refvocoide, page 283).

Nous avons principalement abaissé les valeurs de l'intervalle dynamique — Dynamic range — afin de réduire du champ du spectrogramme les fréquences ayant une faible énergie. Nous avons également changé la valeur de la pré-emphase — Pre-emphasis — pour accentuer les fréquences avec une forte énergie, donc correspondant aux formants, en faisant en sorte de choisir une valeur assez haute pour qu'il n'y ait pas que les basses fréquences qui soient mises en avant. L'autre paramètre que nous avons modifié est l'intervalle temporel que nous avons diminué. Pour choisir quelle valeur mettre, nous nous sommes basé pour cela sur la durée de nos items. La durée de nos items ne dépasse pas les 1,5 s, donc nous avons divisé cette durée par rapport à l'intervalle temporel qui est de 0,002 s. Après calcul, nous obtenons comme résultat 750 qui est la valeur que nous avons choisi pour le nombre d'intervalle temporel⁸. Nous n'avons pas touché aux valeurs des autres paramètres qui sont ceux proposés par défaut. Si nous n'avons pas trouvé nécessaire de changer ces derniers paramètres, c'est parce qu'ils n'amélioreraient en rien la qualité visuelle du spectrogramme. Pour le fenêtrage, les meilleurs rendus sont fournis par la fenêtre de Gauss et de Hanning. Comme, il n'y a pas de différence significative entre les deux, nous avons donc conservé la fenêtre de Gauss qui est celle proposée par défaut par PRAAT. Pour rappel, le fenêtrage est nécessaire du fait que nous avons à faire à un signal qui est fonction du temps dont les fréquences et les amplitudes varient continuellement, par conséquent la Transformée de Fourier ne peut s'appliquer correctement. Donc, il est nécessaire de réduire le champs du calcul à des portions du signal. C'est à cette intention que le fenêtrage a été mis en place pour permettre de faciliter le traitement du signal dans des intervalles. En conclusion, ce paramètre est important pour avoir une bonne résolution de notre représentation spectrale. Un autre point qui contribue à donner une certaine lisibilité du spectre est la longueur du fenêtrage. Si cette dernière est courte, nous obtenons un spectre à bande étroite et à l'inverse si la fenêtre est longue, la représentation spectrale est à bande large. La quasi totalité des spectrogrammes qui seront présentés dans cette thèse sont à bande large, donc nous ne l'indiquerons pas dans la suite de notre travail. Cependant, lorsque l'occasion s'y prêtera, nous mentionnerons l'utilisation du spectre à bande étroite (voir Partie III.4.1.2.2, page 267).

8. Cette méthode de calcul est fourni par le manuel de PRAAT. Le chiffre de 0,002 s est une valeur qui a été choisi en fonction de la pulsation glottale qui est d'environ 5 à 10 ms pour un homme. Donc, l'intervalle temporel à choisir doit être inférieur à la durée de cette pulsation glottale : http://www.fon.hum.uva.nl/praat/manual/Sound__To_Spectrogram___ .html (Consulté le 13/07/2016).

III.3.6.1.3 Mesure des propriétés physiques

Nous avons également utilisé PRAAT pour recueillir nos éléments de mesure tels que nous les avons définis précédemment. Pour ce faire, ce logiciel nous facilite la tâche en pouvant automatiser la collecte des propriétés physiques de nos données. Pour réaliser cette démarche, il faut recourir à un script, ce que nous avons fait. Celui que nous avons utilisé correspond à une adaptation du script qui a été développé par Cédric Gendrot. Ce script dénommé "Analyse tout" est disponible sur le site institutionnel personnel du chercheur⁹. Enfin, le script après qu'il soit compilé permet en sortie d'obtenir un fichier "txt" qui peut être converti par la suite en fichier "csv" et qui sera utilisé par la suite pour faire un traitement statistique de nos données.

III.3.6.1.4 Présentation des résultats

Les matériaux d'analyse seront tous illustrés par des spectrogrammes accompagnés de leur signal acoustique. Ces images se présentent de la manière suivante. L'axe horizontal pour le spectrogramme comme pour le signal acoustique correspond à l'axe temporel (unité : s). Cet axe est fragmenté tous les 0,1 s. L'axe vertical correspond pour le signal acoustique à l'amplitude (unité : Pa) et pour le spectrogramme à la fréquence (unité : Hz). Seul l'axe vertical du spectrogramme présente une graduation qui est de 1000 Hz.

Ces éléments physiques sont flanqués de descriptions linguistiques. Le signal sonore a été segmenté pour faire ressortir d'une part l'unité lexicale — le premier palier — et d'autre part l'unité phonique — le deuxième palier —. Le système de transcription utilisé est celui de l'API.

Cette transcription sera seulement perfectionnée pour les ||r|| afin de ne pas alourdir notre présentation. Donc, nous ne nous attarderons pas sur les autres unités phoniques. Par conséquent, toutes les particularités phonétiques qui pourraient exister dans les environs de nos rhotiques alvéolaires, comme les coups de glotte, les phénomènes d'aspiration ou d'affrication ne seront pas pris en compte. Ainsi, notre transcription sera plus orientée phonétique, mais sans atteindre un haut degré de précision.

Enfin, chaque illustration sera précédée par un court descriptif présentant la position lexicale et syllabique, ainsi que la syllabe. Donc, chaque item qui sera exemplifié sera

9. Le site internet : http://lpp.in2p3.fr/Equipe/cedric_gendrot/scripts.htm (Consulté le 13/07/2016).

décortiqué de cette manière. De plus, le tout sera légendé par un titre où figurera la traduction approximative et la glose de l'item.

III.3.6.2 Le traitement statistique : RStudio

Ce logiciel est, comme PRAAT, free et open source et il a été conçu pour le traitement statistique. RStudio utilise pour traiter les données le langage de programmation "R". Pour nous faciliter le traitement statistique de nos données, nous avons eu recours à différents "packages". Au sein de ces derniers sont inclus des fonctions — fonction au sens informatique — qui permettent de réaliser les calculs dont nous avons besoin (moyenne, écart-type, ... etc). D'autre part, RStudio comprend un module graphique permettant d'afficher les résultats obtenus. Donc, les résultats sous forme de graphique qui seront affichés par la suite proviennent de ce logiciel. Là également, Rstudio permet d'automatiser le calcul des données en permettant d'écrire un script.

III.3.6.2.1 Pré-traitement statistique des données

Les données que nous obtenons nous arrivent dans un état brut, donc elles ne sont pas directement utilisables de la sorte. Il faut, par conséquent, au préalable procéder à un ajustement. Du fait que nos données sont d'une taille relativement raisonnable et que c'est nous-même qui avons en partie recouru à sa description, alors cette étape ne fut pas la tâche la plus ardue. Puisque, cela a été facilité par l'étape précédente de segmentation. Par conséquent, en premier lieu, il nous a fallu choisir les variables que nous voulions analyser et ensuite faire le choix de ne segmenter que ces variables qui nous intéressaient. Donc, seules ces variables parmi les données collectées ont fait l'objet d'une exploitation. Par la suite, pour réaliser les calculs statistiques qui nous intéressaient, nous nous devions de regrouper les variables. Cette étape a été réalisée sur Rstudio qui offre la possibilité de manipuler les données pour les réorganiser et les assembler selon notre volonté — format long ou large, filtrage des données, ... —. L'étape suivante fut de chercher s'il y avait des données manquantes, dupliquées ou aberrantes. Lorsque ce fut le cas, nous avons rectifié cela sur PRAAT. À la fin de ce processus, les données peuvent être traitées statistiquement.

III.3.6.2.2 Traitement statistique des données

Les statistiques peuvent être cataloguées en deux écoles : la statistique descriptive et la statistique inférentielle. Nous avons pris le parti de ne pas faire de statistique inférentielle, donc ne pas prendre en compte les tests quels qu'ils soient — tests paramétriques ou tests non paramétriques —. Nous traiterons nos données en nous basant uniquement sur la statistique descriptive. Car le but de notre recherche est à des fins descriptives et non prescriptives. Nous envisageons seulement d'interroger les données en notre possession sans chercher à vouloir élaborer des prévisions.

La statistique descriptive peut s'apprécier aussi de différentes façons. Nous menerons uniquement des études univariées où la seule variable dépendante sera le temps. S'agissant des variables indépendantes, nous évoquerons leur cas plus précisément dans les parties concernées. Le fait de réduire le champ de l'analyse statistique à une étude univariée contraint de fait les techniques statistiques utilisables. Ainsi, notre étude statistique se résumera à une comparaison de la tendance centrale. Par ailleurs, nous avons fait le choix, pour des questions pratiques, d'utiliser comme méthode statistique l'exploration visuelle de données. Donc, nos résultats statistiques seront présentés sous forme de graphique et nous analyserons nos données à partir de ces figures.

CHAPITRE III.4

Résultats

Ici même dans ce chapitre apparaîtront nos résultats. Pour mémoire, ces derniers portent sur les rhotiques alvéolaires : [r], [r], [r^s] et [r^s]. Ainsi, nous allons étudier au cours de cette partie les points suivants. D'abord, nous allons mener une analyse phonologique. Nous y évoquerons la question de la distribution des rhotiques alvéolaires, puis nous étudierons les variantes allophoniques et la contrastivité de ces unités phoniques.

Par la suite, nous poursuivrons notre travail avec une analyse phonétique. Celle-ci portera sur deux domaines différents. En premier lieu, nous rendrons compte du mode d'articulation de la première articulation des ||r|| et puis nous continuerons avec une étude statistique et comparative sur la durée d'occlusion des taps simple et pharyngalisée. Cette partie sera suivie par une étude sur les vocoïdes, où nous tâcherons d'indiquer les contextes d'apparition de cette unité sonore. Enfin, nous finirons par une étude statistique sur la durée de ces vocoïdes.

III.4.1 Analyse phonologique

Cette analyse phonologique cherchera à déterminer quelles sont les représentations phonologiques de ces rhotiques alvéolaires. Si certaines ne sont pas des allophones d'une autre $\|r\|$. Pour parvenir à ce but, nous débuterons notre analyse par une étude de la distribution des $\|r\|$. Ensuite, nous observerons attentivement les environnements phonétiques dans lesquels apparaissent les rhotiques alvéolaires et quels types de réalisations phonétiques peuvent accomplir les $\|r\|$ phonologiques.

III.4.1.1 Distribution des Rhotiques Alvéolaires

Les rhotiques alvéolaires, comme nous l'avons précédemment évoqué, sont au nombre de quatre. Pour chacune de ces rhotiques alvéolaires, nous allons analyser leurs distributions. Pour ce faire, nous allons rendre compte de leur distributivité au niveau syllabique et lexical. Les positions lexicales qui nous intéressent sont la position initiale, finale et médiane. Au sein de cette dernière, nous porterons notre attention sur les rhotiques alvéolaires qui sont en position intervocalique et forment un groupe consonantique. Quant à la position syllabique, toutes les positions possibles seront prises en compte : attaque, noyau et coda.

La présente syllabification qui sera proposée n'a fait l'objet d'aucune enquête. Nous ne nous sommes pas enquis auprès des informateurs. C'est une syllabification intuitive basée sur les connaissances personnelles de l'auteur. Lorsque nous avons un doute quant à la segmentation syllabique, nous mettrons entre parenthèse le segment qui est incriminé. Cela concerne principalement le cas des géminées et des suites de consonnes. Par exemple, la segmentation pour une géminée¹ se présente-t-elle de cette manière — $\|r\|VC\cdot CV$ — ou comme ceci — $\|r\|V\cdot CCV$ — ? De même pour les groupes de consonnes peuvent-ils former une attaque complexe — $\|r\|CV\cdot$ — ou non — $\|r\|\cdot CV$ — ?

Ce sont des questions que nous n'avons pas tranchées pour l'instant. D'après les données que nous avons en notre possession au moment où nous rédigeons cette thèse, il semble que la syllabification de cette variété du berbère présente un certain nombre de particularismes. Pour nous assurer que nos analyses et nos affirmations soient fondées, il faudrait entreprendre une vaste étude sur ce point. Donc, nous ne traiterons pas de cette question ici même, car cela sortirait du cadre que nous nous sommes fixé.

1. Pour une vue d'ensemble de ces questions d'ambiguïté, voir [Kenstowicz, 1994, p.400].

III.4.1.1.1 Le matériau d'analyse

Nous avons extrait quelques items de notre corpus pour illustrer nos propos précédents. Ils proviennent tous de nos données qui ont été enregistrées avec l'informateur 1.

III.4.1.1.1.1 Tap simple

Cette consonne est retrouvée dans toutes les positions distributionnelles, à part une seule qui est la position finale d'un morphème. Dans cette dernière position, nous ne retrouvons que des trills. Au niveau des positions syllabiques, il peut être en coda, en attaque et peut également être le noyau syllabique.

— Tap en position médiane et en position syllabique de coda (syllabe : [3r])

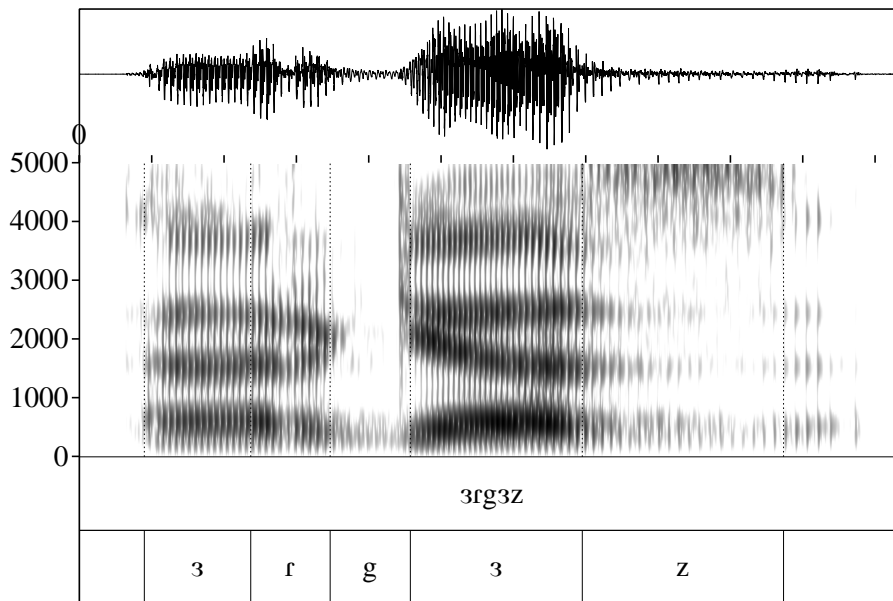


FIGURE III.4.1 : Spectrogramme de @[3rg3z]@ \homme\

- Tap en position médiane et en position syllabique d'attaque (complexe) (syllabe : [(3)res])

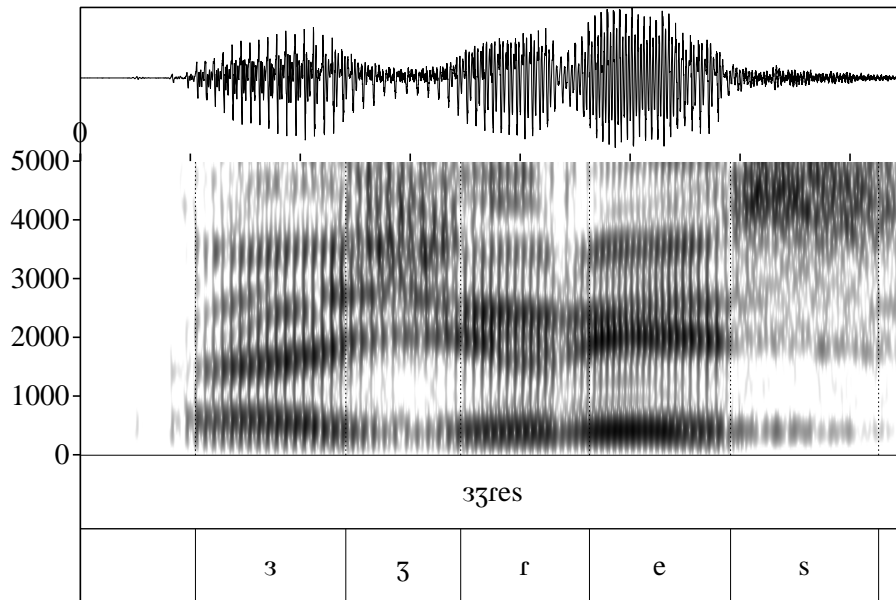


FIGURE III.4.2 : Spectrogramme de @[3̣res]@ \froid\

- Tap en position médiane (intervocalique) et en position syllabique d'attaque (syllabe : [re])

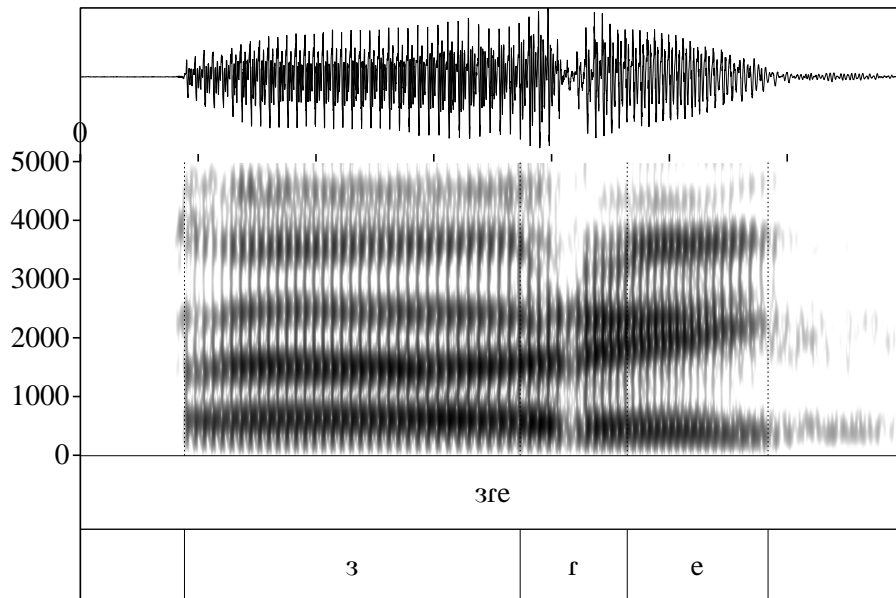


FIGURE III.4.3 : Spectrogramme de @[3̣re]@ \écrire{2.SG.IMP}\

— Tap en position syllabique de noyau (syllabe : [ʁf])

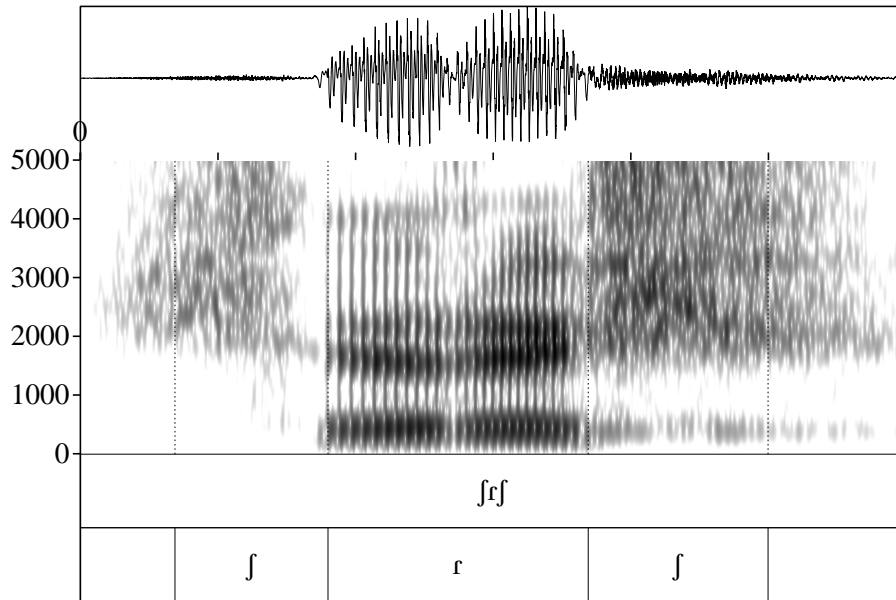


FIGURE III.4.4 : Spectrogramme de @[ʁf]@ \nouer{2.SG.IMP}\

— Tap en position initiale et en position syllabique d'attaque (syllabe : [rø])

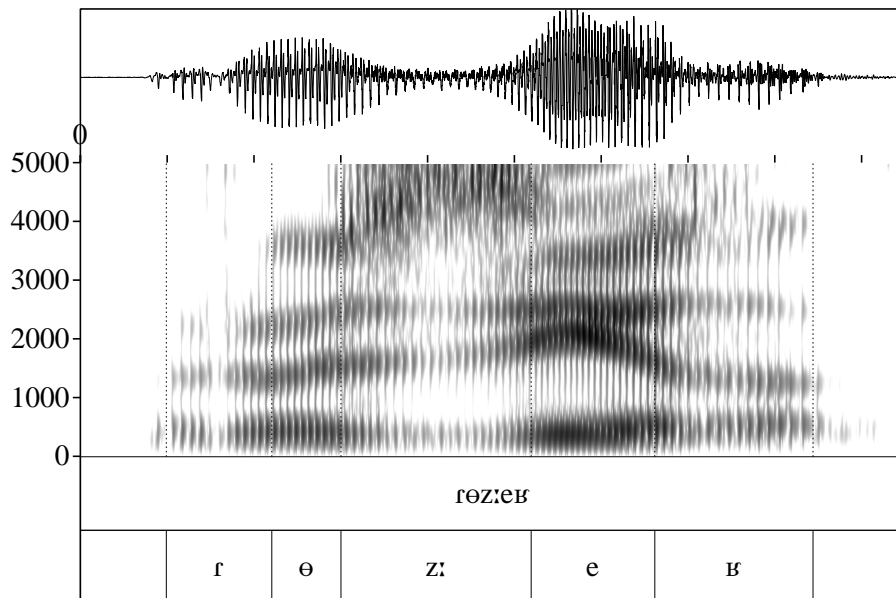


FIGURE III.4.5 : Spectrogramme de @[rø:ɛʁ]@ \chercher{1.SG.AI}\

III.4.1.1.1.2 Tap pharyngalisée

Les taps pharyngalisés ont exactement la même distribution que les taps simples.

- Tap pharyngalisée en position médiane et en position syllabique de coda (syllabe : [m^ɣɔr^ɣ])

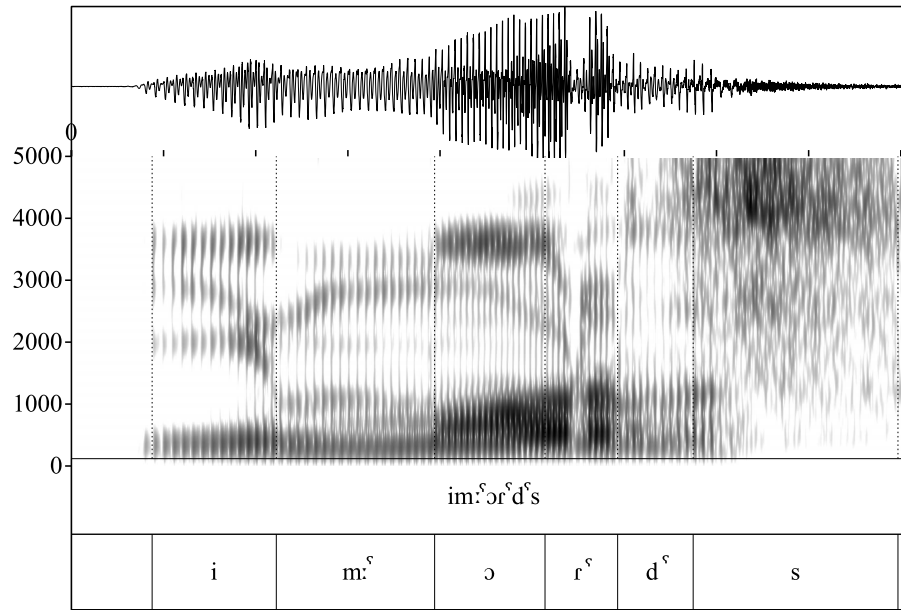


FIGURE III.4.6 : Spectrogramme de @[im:ɔrɔdɔs]@ \étrangler{3.M.SG.PRET}\

- Tap en position médiane et en position syllabique d'attaque (complexe) (syllabe : $[(\text{ʒ})r^{\text{ʃ}}\Lambda]$)

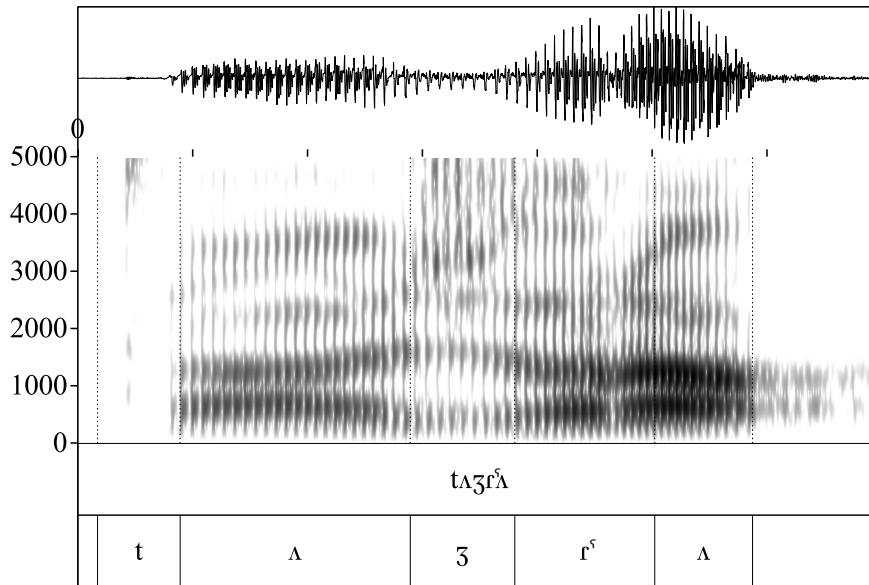


FIGURE III.4.7 : Spectrogramme de @[tʒrʃ]@ \récipient\

- Tap pharyngalisée en position médiane et en position syllabique de noyau (syllabe : $[(\text{ʁ})r^{\text{ʃ}}f]$)

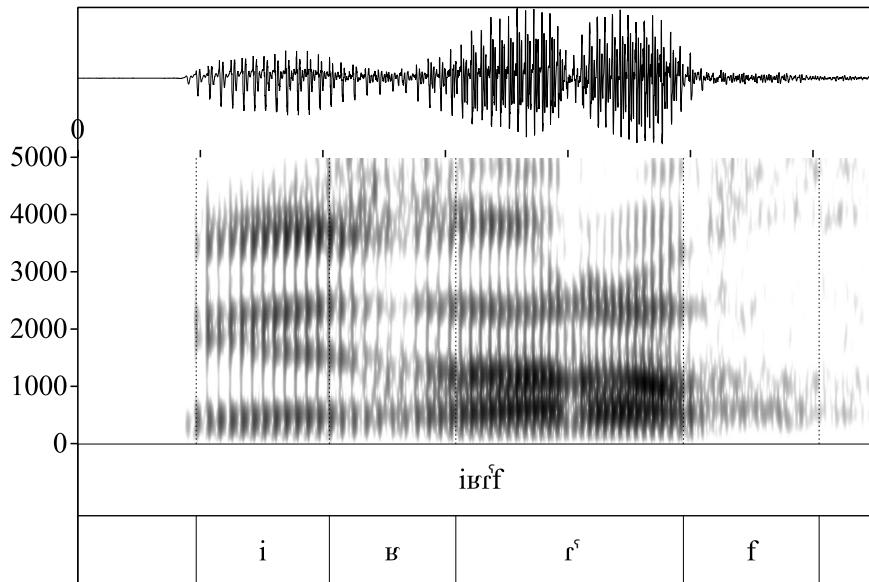


FIGURE III.4.8 : Spectrogramme de @[iʁrʃf]@ \geler{3.M.SG.PRET}\

- Tap pharyngalisée en position médiane (intervocalique) et en position syllabique d'attaque (syllabe : [rʕʌn])

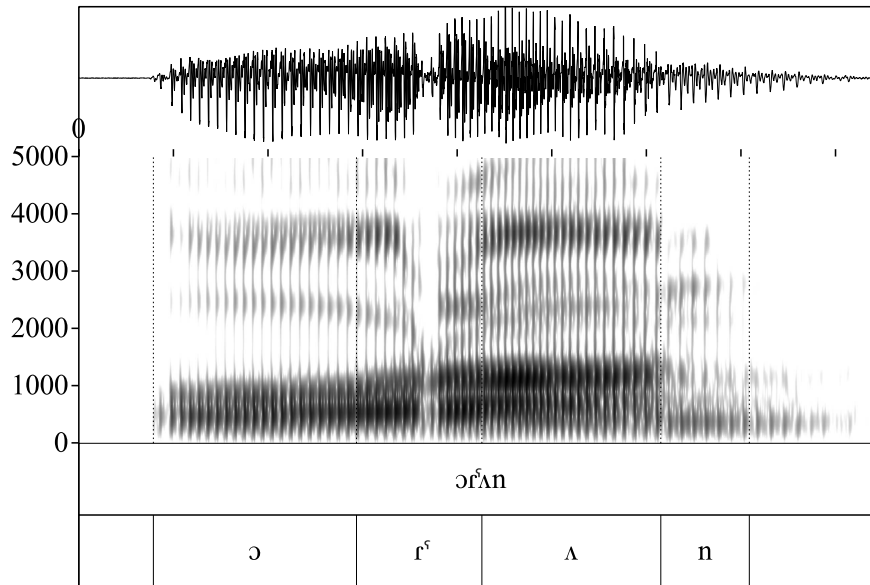


FIGURE III.4.9 : Spectrogramme de @[ʕrʕʌn]@ \poignée\

- Tap pharyngalisée en position initiale et en position syllabique de noyau (syllabe : [rʕ(b)])

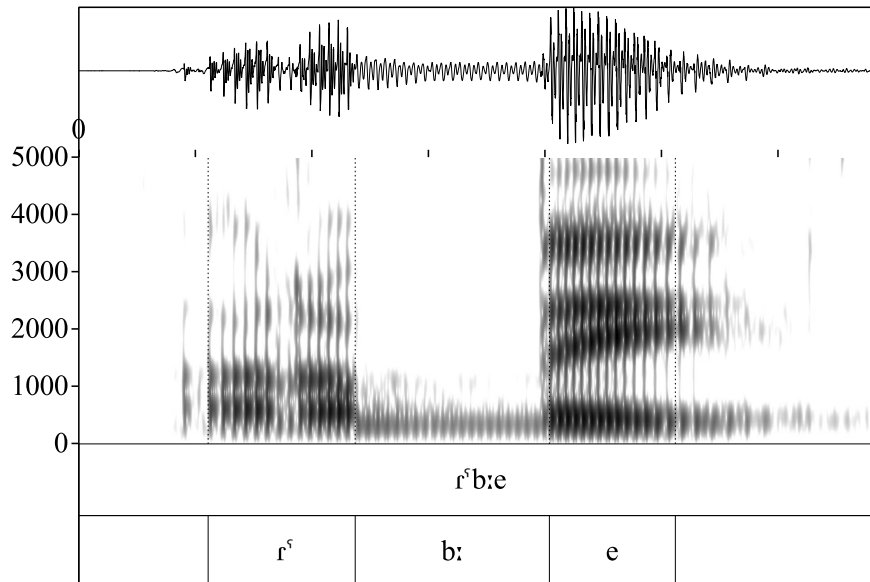


FIGURE III.4.10 : Spectrogramme de @[rʕb:e]@ \dieu\

III.4.1.1.1.3 Trill simple

La trill contrairement aux deux autres a une distribution plus complète. Nous le retrouvons dans toutes les positions à l'initiale, en médiane et en finale. De plus, elle peut être dans toutes les positions syllabiques en attaque, en noyau et en coda. Mentionnons également que le nombre de cycle, comprenant une phase d'occlusion et de désocclusion, auquel parviennent nos informateurs peut aller jusqu'à sept, mais dans la plupart des cas le nombre de cycles varie entre deux et trois. Nous allons illustrer tout ceci par des spectrogrammes d'items qui correspondent à ces distributions.

— Trill simple en final et position syllabique de coda (syllabe : [t:ər])

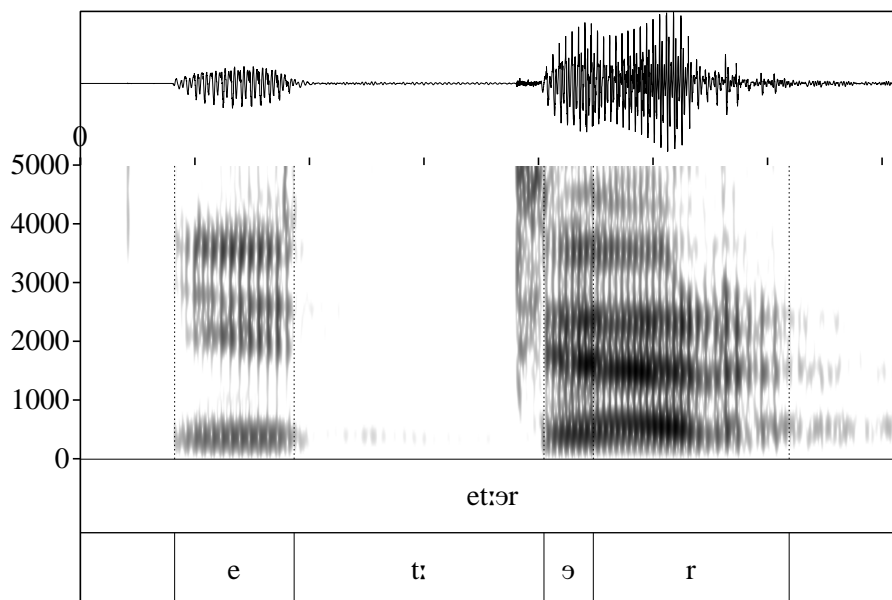


FIGURE III.4.11 : Spectrogramme de @[et:ər]@ \demander{3.M.SG.PRET}\

— Trill simple à l'initial et en position syllabique de noyau (syllabe : [r])

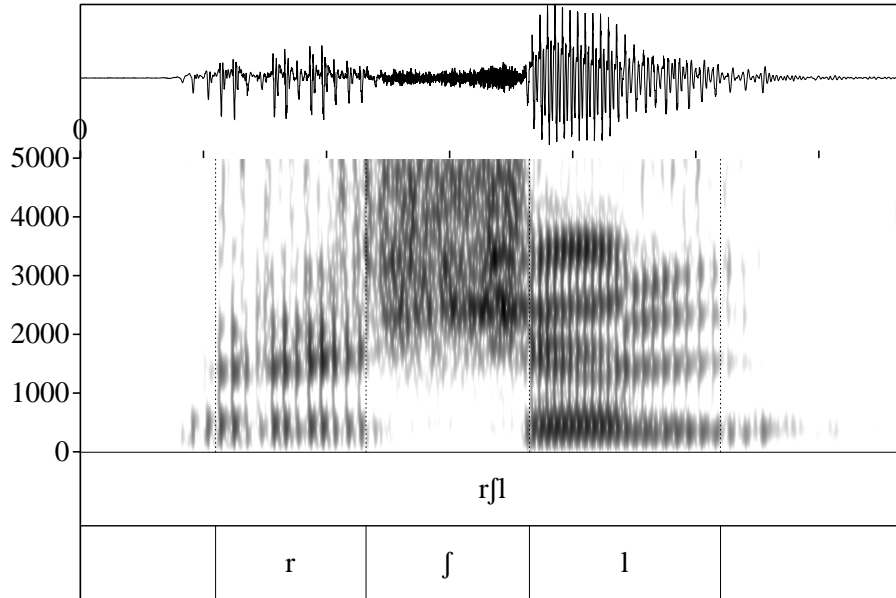


FIGURE III.4.12 : Spectrogramme de @[rʃl]@ \mariage\

— Tap en position médiane et en position syllabique d'attaque (complexe) (syllabe : [(d)re])

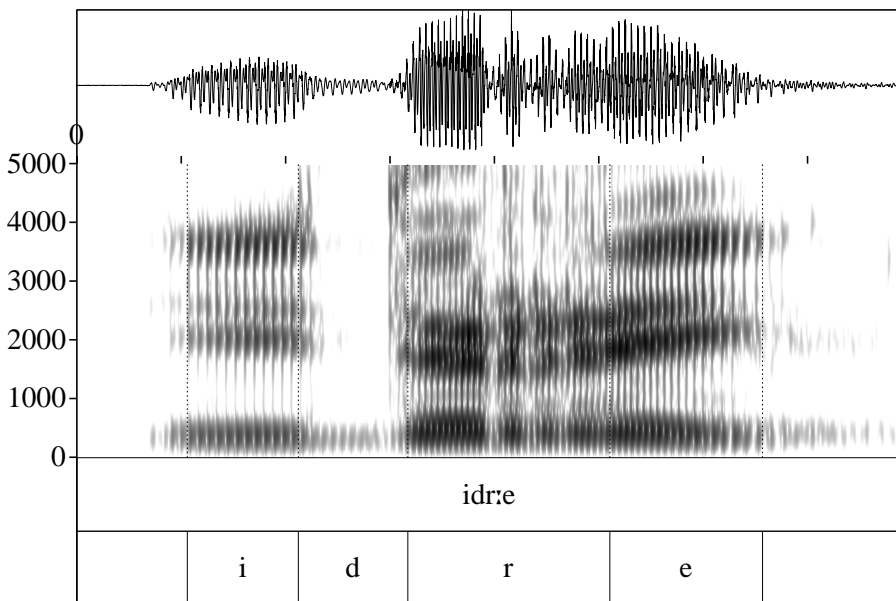


FIGURE III.4.13 : Spectrogramme de @[idre]@ \pousser_énergiquement{3.M.SG.IMP_INT}\

- Trill simple en médiane (intervocalique) et positon syllabique d'attaque (syllabe : [re]) -- Ici l'informateur 1 réalise sept cycles. C'est le nombre maximum de cycles que nous avons trouvé dans notre corpus —.

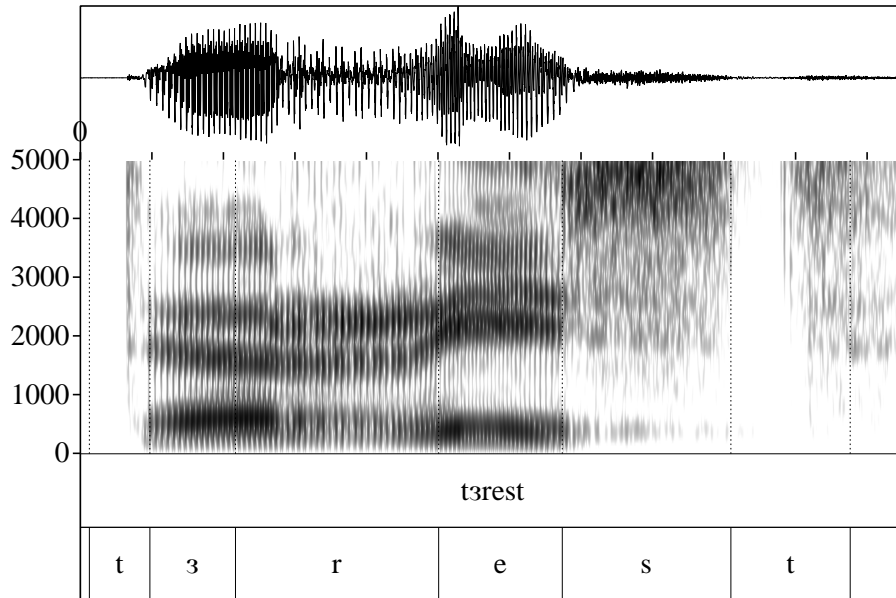


FIGURE III.4.14 : Spectrogramme de @[tʒrest]@ \argile_blanche\

- Trill simple en médiane et positon syllabique de noyau (syllabe : [ħr])

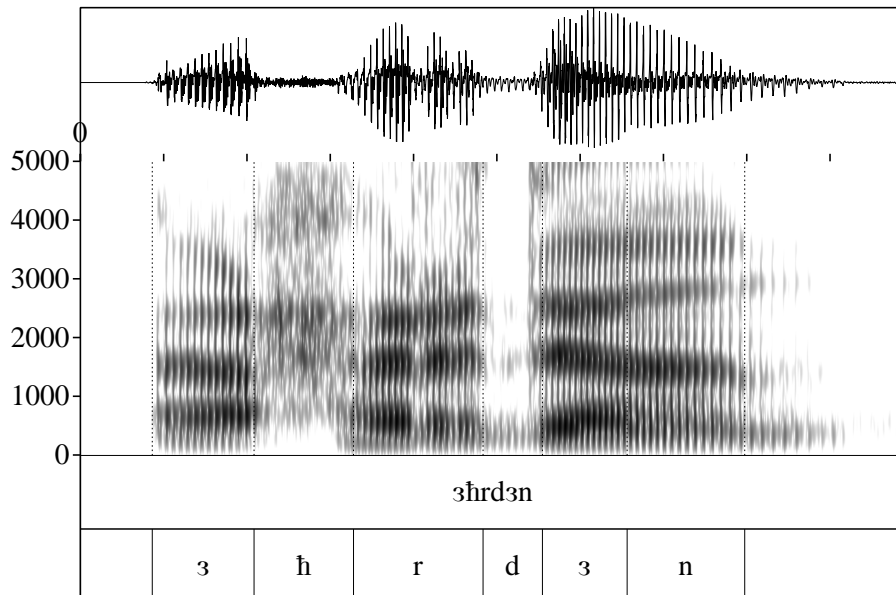


FIGURE III.4.15 : Spectrogramme de @[ʒħrdʒn]@ \uromastyx\

III.4.1.1.1.4 Trill pharyngalisée

— Trill pharyngalisée en final et position syllabique de coda (syllabe : [dʕΛrʕ])

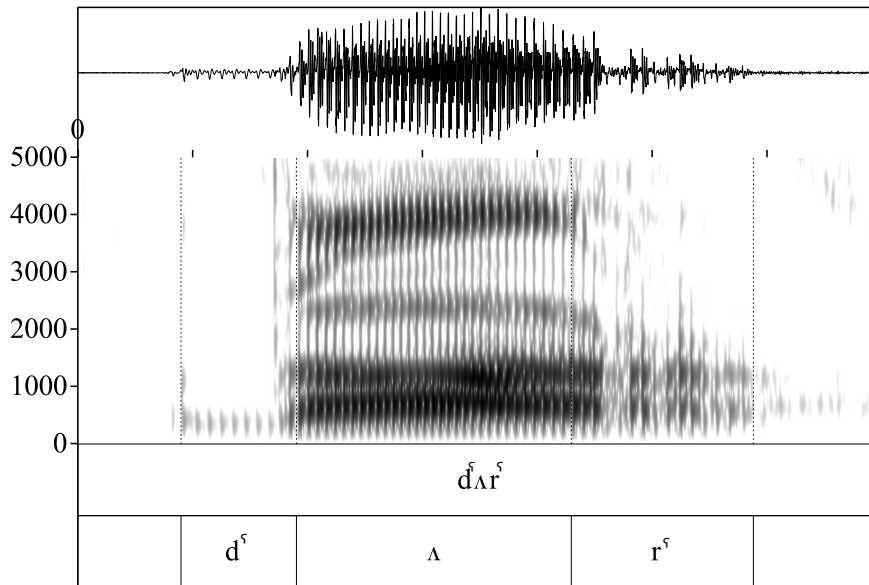


FIGURE III.4.16 : Spectrogramme de @[dʕΛrʕ]@ \pied\

— Trill pharyngalisée à l'initial et en position syllabique de noyau (syllabe : [rʕ])

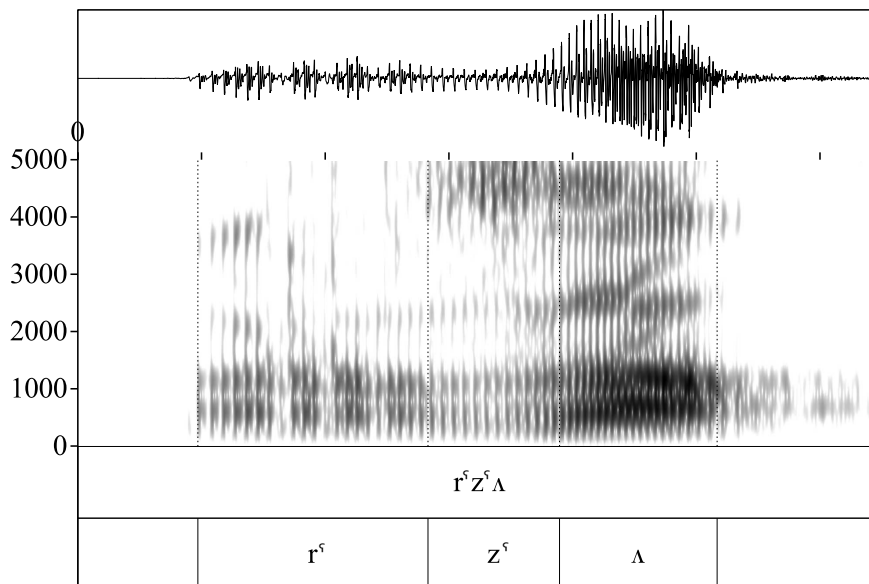


FIGURE III.4.17 : Spectrogramme de @[rʕʕΛ]@ \casser{1.SG.PRET}\

- Trill pharyngalisée en médiane (intervocalique) et positon syllabique d'attaque (syllabe : [r^sΛ])

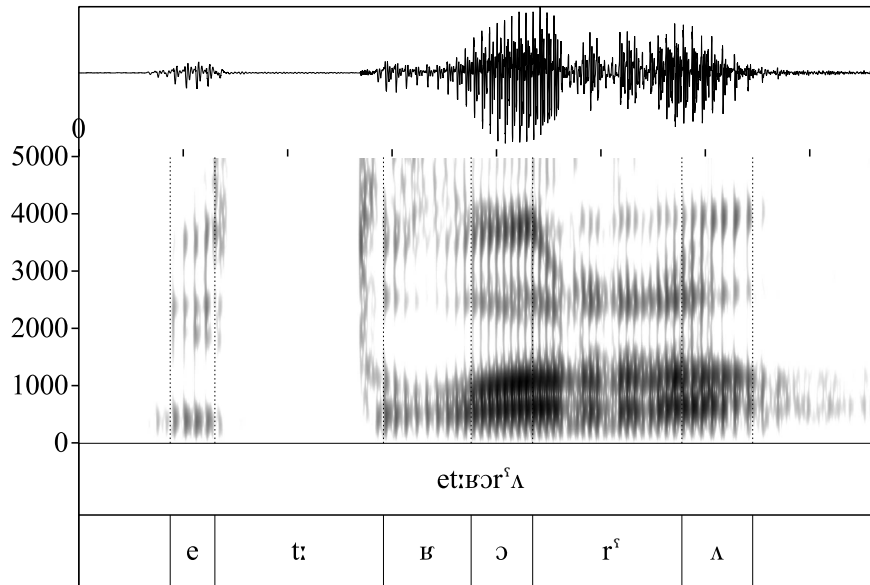


FIGURE III.4.18 : Spectrogramme de @[et:ʁɔr^sΛ]@\tromper{3.M.SG.AI}\

- Trill pharyngalisée en médiane et positon syllabique de noyau (syllabe : [ħr^s])

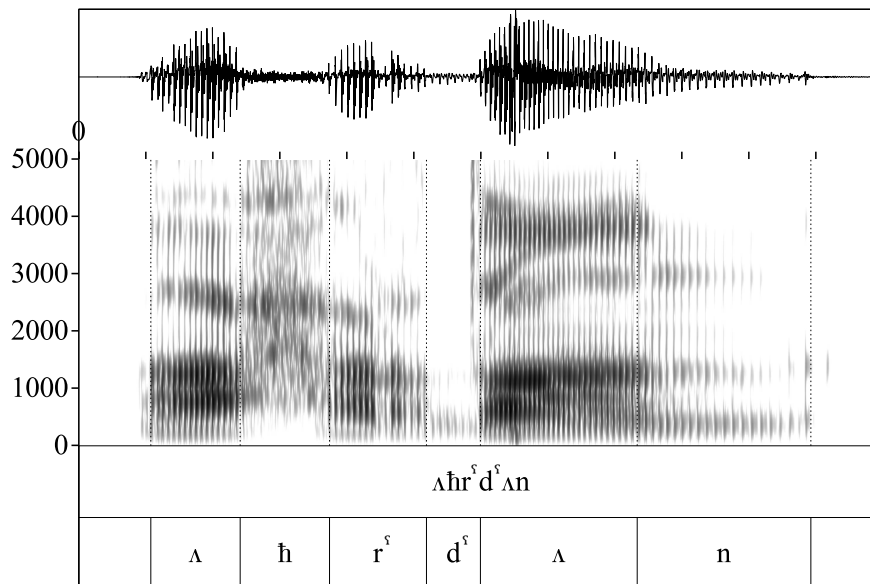


FIGURE III.4.19 : Spectrogramme de @[Λħr^sd^sΛn]@\esclave_affranchi\

III.4.1.1.2 Discussion

Les données qui viennent d’être montrées indiquent que les rhotiques alvéolaires en **TmAS** ont une distribution très large. Les taps simples et les taps pharyngalisés partagent la même distribution. Il en est de même entre les trills simples et les trills pharyngalisés. Cela sous-entend qu’entre la rhotique alvéolaire simple et sa contrepartie pharyngalisée, les positions syllabiques ou lexicales ne sont pas des critères de neutralisation. Mais si nous comparons les trills et les taps entre elles, nous constatons une différence.

Effectivement, la seule position qui n’est occupée que par un type de $\ll r \gg$ est la position finale. En cette position, seules des trills — simples ou pharyngalisés — peuvent y figurer. Cela implique que la tap est neutralisée en cette position. Nous reviendrons sur cette question dans la seconde partie de l’analyse phonologique.

L’autre constat que nous pouvons faire concerne les positions syllabiques, notamment celles correspondant aux groupes consonantiques. Toutes les rhotiques alvéolaires peuvent se trouver dans cette position. Nous avons choisi d’illustrer cette situation avec des consonnes (quasi-)homorganiques. Pour la tap simple et la tap pharyngalisée, la consonne qui précède est la consonne quasi-homorganique [ʒ]. Alors que pour la trill simple, celle-ci est illustrée avec la consonne homorganique [d]. Un premier commentaire s’impose, ces exemples n’induisent nullement qu’il faut considérer que les taps ne peuvent être présentes que dans ces environnements. Notre corpus indique que ces rhotiques ont des possibilités de distribution au regard des groupes consonantiques bien plus large que ce que nous venons de montrer. Les taps, simples ou pharyngalisés, peuvent aussi être précédées par une consonne homorganique :

Tap simple	Tap pharyn.
@[dre]@ \pousser{2.SG.IMP}	@[dr ^h ɔs]@ \peu{ADV}\
@[itre]@ \étoile \	@[ɬz ^h r ^h ɔ]@ \colline\
	@[s:r ^h emət]@ \bride_de_décor_du_cheval\

TABLE III.4.1 : Exemples de groupe consonantique de taps et de consonnes homorganiques

Toutefois, concernant la trill pharyngalisée, nous n’avons aucun exemple où celle-ci peut être considérée avec certitude comme étant en position d’attaque et comme étant précédée par une consonne. Mais il ne faudrait pas pour autant conclure que

ce manque est dû à des contraintes phonologiques. Cela s'explique par le fait, d'une part, que notre corpus n'est pas exhaustif. Et d'autre part, comme nous le verrons par la suite, que la trill pharyngalisée ne surgit couramment que dans l'environnement de consonnes pharyngalisées. Tout ceci limite inévitablement la probabilité d'avoir un groupe consonantique où y figure une trill pharyngalisée. Pourtant, nous arguons qu'une telle disposition est possible, car d'un côté nous retrouvons la trill simple dans cette position — la trill pharyngalisée étant pour partie un allophone de ce dernier —. Et d'autre côté, la trill pharyngalisée est retrouvée dans les mêmes environnements et positions que les autres $\llbracket r \rrbracket$, donc cela tend à indiquer que la trill pharyngalisée n'est pas dans une relation de complémentarité avec les autres $\llbracket r \rrbracket$, par conséquent, il n'y aurait aucune raison objective de ne pas rencontrer des trills pharyngalisées dans cette position.

De plus, nous avons considéré dans cette question de groupes consonantiques que le cas où $\llbracket r \rrbracket$ était assurément en position syllabique d'attaque, cependant, si nous prenons les cas où les $\llbracket r \rrbracket$ suivent une consonne tout en étant le noyau d'une syllabe. Alors nous trouvons des exemples où la trill pharyngalisée est devancée par une consonne, notamment des consonnes homorganiques — Les seules consonnes pharyngalisées et homorganiques avec $\llbracket r \rrbracket$ existant en TmAS sont $[z^{\text{f}}]$ et $[d^{\text{f}}]$ —. Ci-dessous des exemples illustrant cette situation et l'un de ces exemples sera par ailleurs illustré avec un spectrogramme :

Suite de consonnes avec des trills pharyngalisées

@[sdr^fd^fq]@ \explorer{2.SG.IMP}\

@[z:^fr^ft^f]@ \tempête\

@[z^fr^fd^f]@ \trier{2.SG.IMP}\

@[Λz^fr^feq]@ \pêt\

TABLE III.4.2 : Exemples de trills pharyngalisées précédées par une consonne homorganique

— Trill pharyngalisée précédée par une consonne quasi-homorganique

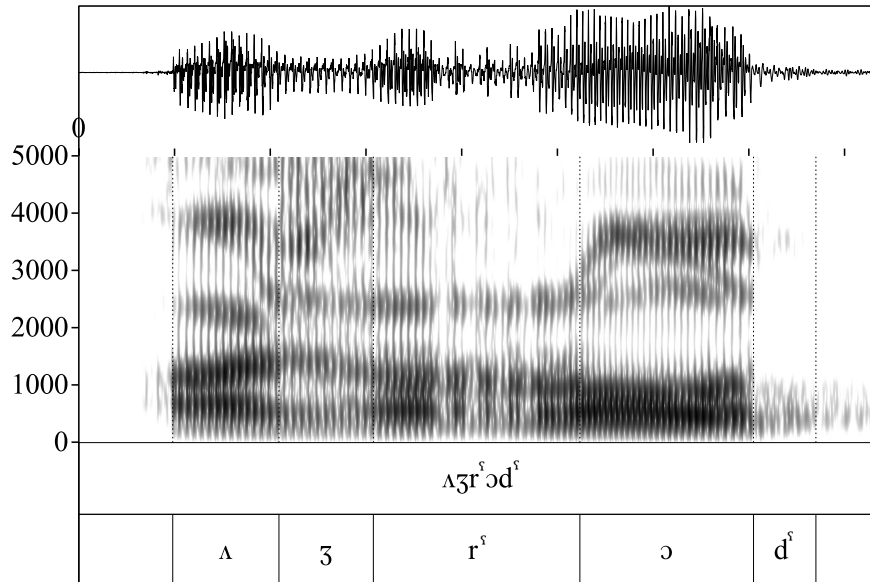


FIGURE III.4.20 : Spectrogramme de @[\u039b\u0303r\u0302\u0309d\u0302]\u00a0 \sorte_de_citrouille\

À la vue de ces exemples, notamment les deux derniers, il se pourrait très bien que nous puissions les considérer comme étant en réalité des groupes consonantiques formant une attaque complexe. Cependant, cela dépend de quelle manière, nous examinons les trills. Si nous partons du principe que les trills sont une réalisation de surface d'une tap géminée — ce point sera vu par la suite —, alors les règles de syllabification, celles qui sont usuellement appliquées, voudraient que nous apprécions ces trills comme étant ambisyllabiques. Donc, le découpage syllabique de ces morphèmes se présenteraient de cette manière : $/\Lambda \cdot \text{z}^{\text{h}} \cdot \text{r}^{\text{h}} \cdot \text{d}^{\text{h}}/$ ². Ainsi, ils contiendraient trois syllabes et non pas seulement deux si nous considérons l'autre hypothèse : $/\Lambda \cdot \text{z}^{\text{h}} \text{r}^{\text{h}} \text{d}^{\text{h}}/$. Mais rien n'est moins sûr au sujet de l'hypothèse à privilégier, d'autres investigations sont nécessaires pour trancher cette question.

Dans un tout autre domaine qui est d'ordre typologique, nos résultats remettent en question la hiérarchie que semblait avoir mis en évidence Bradley [bradley2001]. Selon lui, après analyse de plusieurs langues, au total sept, il s'était rendu compte que la contrastivité entre trill et tap semblait être régie par un ordre. Si nous reproduisons

2. Le découpage de la trill en deux suites de taps peut être vu comme étant la représentation de l'encodage de ce phonème au niveau squelettal [McCarthy, 1979]. Mais cette notation n'implique nullement que nous nous inscrivons dans cette théorie.

ici ses résultats, ils indiquent qu'une langue ne connaissant pas de contraste entre la trill et la tap dans une position donnée ne pouvait pas avoir de contraste dans une autre position. Ainsi, il avait établi l'ordre suivant :

Position 1	<	Position 2	<	Position 3	<	Position 4
Intervocalic		Word-initial		Heterorganic clusters, Word-final		Homorganic Clusters

Donc, une langue, comme l'espagnol, qui n'a pas de contraste entre la tap et la trill en position 2, ne pourrait selon cette hypothèse connaître de contraste en position 3. En revanche, en **TmAS** comme nous venons de le voir, les trills et les taps sont en contraste dans toutes les positions sauf une seule qui est la position finale — la position 3 —. Donc, si l'hypothèse de Bradley était correcte, le **TmAS** devrait connaître un contraste en cette position puisqu'elle n'ignore pas un contraste en position 4 — position homorganique —. Mais comme tel n'est pas le cas, cela remet en question l'universalité de cette hiérarchie qui ne peut s'appliquer sur le **TmAS**. Voire il est fort peu probable qu'en la matière une hiérarchie universelle puisse exister. Car, si nous devons élargir l'analyse à d'autres langues du monde — sept langues est un échantillon insuffisant pour avancer un fait universel —, nous constaterions des combinaisons variées de contraste qui seraient difficiles à réconcilier par une seule règle.

III.4.1.2 Variantes allophoniques

Au cours de cette partie, nous allons présenter les réalisations allophoniques possibles des rhotiques alvéolaires en **TmAS**. Les variantes que connaît un phonème peuvent être de deux différentes sortes qui sont dénommées couramment : variante contextuelle et variante libre. Donc, ces deux dimensions seront abordées séparément. Nous commencerons par l'étude des variantes contextuelles.

III.4.1.2.1 Variantes contextuelles

Les variantes contextuelles comme l'indique sa dénomination se réfèrent à une variante qu'on ne retrouve que dans une position ou des positions limitées. Cette distribution restreinte est un indice qu'une telle unité phonique est probablement la variante d'une autre unité phonique. Nous allons confronter nos quatre rhotiques alvéolaires entre elles pour déterminer lesquelles sont à considérer comme étant des phonèmes. Nous entreprendrons cette démarche en analysant dans un premier temps

les taps simples par rapport aux taps pharyngalisées. Puis le cas des taps vis-à-vis des trills. Et nous finirons par opposer les trills simples et les trills pharyngalisées.

III.4.1.2.1.1 Tap simple et Tap pharyngalisée

Les taps pharyngalisées ont un contexte phonétique moins fourni que les taps simples. Nous pouvons cataloguer les taps pharyngalisées suivant deux catégories. La première catégorie, qui est fonction du contexte phonétique, correspond aux taps pharyngalisées qui se trouvent au voisinage d'une consonne pharyngalisée. Quant à la seconde, elle est basée sur l'origine et correspond aux mots empruntés à l'arabe dialectal. Nous avons dans le tableau ci-dessous extrait de notre corpus quelques items qui relevaient de ces catégories :

Voisinage pharyn.	Emprunt à l'arabe
@[ʌzʳʰʌ]@ \colline \	@[əʳʰʌħ]@ \aller{3.M.SG.PRET}\
@[ʌbrʰdʰʌ]@ \oiseau\	@[rʰb:e]@ \dieu\
@[ʌqrʰdʰʌ]@ \grand\	@[rʰ]:d@] \arrêter{2.SG.IMP}\
@[zrʰwdʰ]@ \jeter{2.SG.IMP}\	@[frʰʌnsʰʌ]@ \France\

TABLE III.4.3 : Les contextes phonétiques de [rʰ]

En écartant les taps pharyngalisées du fait du contact avec d'autres consonnes pharyngalisées et celles issues d'emprunts à l'arabe, il existe toujours une multitude d'items qui sont composés de taps pharyngalisées et que nous ne pouvons d'une manière ou d'une autre faire remonter à une tap simple. Citons en exemple l'item : @[ʌrʰʌ]@ \accoucher{2.SG.IMP}\, où la rhotique alvéolaire est en position intervocalique, donc dans l'environnement d'aucune consonne, mais celle-ci est pourtant pharyngalisée. Cela soulève plusieurs questions, car si nous nous en tenons à cet exemple, nous pourrions trouver évident de considérer que cette tap pharyngalisée est étymologique. Dès lors, n'y aurait-il pas un moyen quelconque de rapprocher les taps pharyngalisées de ces items vers les taps simples? Nous ne voudrions pas conclure sans avoir exploré toutes les possibilités.

Une première remarque que nous pourrions avancer est que la tap pharyngalisée n'est jamais retrouvée dans un contexte intervocalique où il y aurait uniquement des voyelles hautes et antérieures. La tap pharyngalisée lorsqu'elle se trouve dans ce contexte est

toujours accompagnée par une voyelle basse ||a|| et/ou une voyelle haute et postérieure ||o||. Est-ce à dire que cette observation est un indice suffisant pour insinuer que ces dernières pourraient d'une quelconque manière influencer sur l'articulation secondaire de ||r|| ?

Selon nous, cette hypothèse est à écarter car nous trouvons dans notre corpus une tap simple qui se situe dans le même environnement précité. L'item en question correspond à @[ɜmzwɜrɔ]@ \premier\ que nous pouvons comparer également à [ɜng:ʷɹɔ] \dernier\. Il apparaît à partir de ces exemples que la tap pharyngalisée et la tap simple sont en contraste dans cet environnement vocalique. C'est pourquoi, il est peu probable que ces voyelles soient à l'origine de la pharyngalisation. Si nous devons persister à vouloir envisager la tap pharyngalisée comme dérivant de la tap simple, alors il nous faudra chercher l'explication au niveau diachronique. Au plan synchronique, l'analyse distributionnelle ne permet pas de rendre compte de cette hypothèse.

En conséquence, il nous faut conclure qu'il existe des taps pharyngalisées dans cette langue qui sont des phonèmes. Bien que nous n'ayons pas trouvé de paires minimales mettant en jeux ces deux unités phoniques. L'argument de la fréquence d'emplois de la tap pharyngalisée non réduite à un contexte distributionnel précis est suffisant en lui-même pour conclure que tel est le cas. Ainsi, il existe des taps pharyngalisées qui sont des phonèmes à part entière et d'autres qui sont des allophones de la tap simple. De ce fait, il nous faut aussi considérer que la tap simple devant des consonnes pharyngalisées est neutralisée en devenant une tap pharyngalisée. Mais rien ne garantit que la réalisation phonétique d'une tap pharyngalisée dans l'environnement d'une consonne pharyngalisée ne soit pas au niveau phonologique aussi pharyngalisée et que sa pharyngalisation ne soit en rien dû au contexte. Donc, il serait préférable lorsqu'une telle situation est trouvée de parler d'archiphonème |R|.

III.4.1.2.1.2 Tap et Trill

Les taps, comme nous l'avons déjà signalé, ne sont jamais retrouvées en final. Les seules rhotiques alvéolaires trouvées à cet emplacement sont des trills. D'autre part, les trills sont aussi présentes dans d'autres environnements phonétiques qui équivalent ceux de la tap simple. Donc, ceci ne laisse pas de place à l'idée que la trill pourrait être une variante contextuelle de la tap. Mais nous pouvons avancer l'hypothèse qu'une partie des trills sont des allophones de la tap. La seule possibilité qui pourrait tout de même accréditer cette idée est s'il y a eu neutralisation. Pour confirmer cette éventualité, il faut

d’abord déterminer si les taps et les trills constituent bien deux phonèmes différents. Il semble que cela soit le cas, car les paires minimales suivantes ont été trouvées :

Tap	Trill
@[dre]@ \pousser_violemment{2.SG.IMP}\	@[dre]@ \pousser_violemment{2.SG.IMP_INT}\
@[fr3]@ \labourer{2.SG.IMP}\	@[fr3]@ \labourer{2.SG.IMP_INT}\
@[qr ^s d ^s]@ \fracturer{2.SG.IMP}\	@[qr ^s d ^s]@ \fracturer{2.SG.IMP_INT}\

TABLE III.4.4 : Contraste phonologique entre tap et trill

Au vue de ces exemples, un autre constat apparait. La gémination des taps correspond à des trills. Les taps simples fournissent des trills simples et les taps pharyngalisées fournissent des trills pharyngalisées. Car, dans le système verbal du berbère, l’un des moyens pour obtenir l’aspect de l’AI ou *Inaccompli* est de partir du morphème qui est à l’aspect de l’A et de géminer soit la première consonne ou la deuxième consonne [Lahrouchi, 2005]. À partir de ces cas, nous formulons l’hypothèse que les trills simples retrouvées dans ces items sont en réalité les formes de surface d’une forme sous-jacente qui correspondrait aux taps géminées.

Ceci étant dit, nous revenons à l’étude des trills en final. Un autre aspect est à exprimer concernant ces derniers. Comparons ces deux réalisations de deux items prononcés par notre informateur 2 et qui ont le même radical @[ɤʒt:er||r]@ : @[ɤʒt:er]@ \épine_de_palmier\ et @[eɤʒt:ern]@ \épines_de_palmier\. Dans le premier item ||r|| est une trill et lorsque ce même ||r|| n’est plus en final puisqu’une autre unité morphologique le suit — la marque du pluriel —, celle-ci se réalise comme une tap.

Pour étudier ce cas au vu de ces données, deux hypothèses peuvent être formulées. Soit nous considérons tout bonnement que la trill a une distribution plus large que la tap simple et par conséquent, il n’y a pas de raisons de considérer la trill finale comme étant la réalisation phonétique d’un autre phonème³. Soit nous envisageons l’hypothèse que ce manque de contraste au niveau de cette position reflète un phénomène cognitif.

3. Si nous envisageons l’hypothèse qu’elles ne sont pas le produit d’un processus morphologique ou phonologique, alors cela ferait d’elles des géminées lexicales et si cette hypothèse est privilégiée, alors cela signifie qu’il existe au niveau des représentations des trills. Cette hypothèse ne va pas dans le sens du réductionnisme, donc elle peut être difficilement défendable.

Si nous considérons la deuxième hypothèse, alors il faut concevoir que la trill finale peut être le produit d'une gémiation. Cette idée est en soit probable pour les trills qui ne sont pas en final. Mais pour ceux qui le sont, il apparaît difficile d'appuyer cette thèse. Car, la prise en compte du système phonologique global du TmAS, voire de tout le berbère, n'apporte pas de l'eau au moulin de cette hypothèse. Puisque les trills en final sont observées même en final de racines supérieures au bilitère comme @[dəf:r]@\derrière{PREP}\. Alors qu'il n'existe pas en berbère de processus morphologique de gémiation qui porterait sur la troisième consonne d'une trilitère par exemple ou au-delà [Lahrouchi, 2009]. Les autres phonèmes avec ce type de racines ne présentent pas de la même façon de gémiation morphologique ou de géménées lexicales en final⁴. Donc, il n'y a pas de raisons de penser qu'il en serait autrement pour les taps. Les données indiquent qu'il serait plus raisonnable de considérer les trills retrouvées en final comme n'étant pas toutes, d'une part, des phonèmes à part entière et comme correspondant, d'autre part, pour certaines à des allophones de la tap.

Donc, il apparaît comme plus pertinent d'envisager la deuxième hypothèse, c'est-à-dire que cette trill finale n'est que la forme de surface d'une tap. Pour étayer davantage cette hypothèse, nous voudrions souligner qu'en TmAS les consonnes finales, pas uniquement les ||r||, ont une durée plus longue que celles qui seraient en médiane ou à l'initial. Mais cette durée ne doit en rien nous pousser à classer ces consonnes comme étant un phonème à part entière. La raison à cela est que ces phonèmes ne se comportent pas comme des géménées "classiques". Par exemple, la consonne géminée /d:^ɿ/ a pour réalisation de surface [t:^ɿ], cependant, dans le cas de ces consonnes finales la forme de surface restera [d:^ɿ], il n'y aura aucun dévoisement qui se produira. Nous en voulons pour preuve le spectrogramme ci-dessous qui illustre notre propos. Comme nous pouvons le voir, le signal acoustique enregistré pour ce [d:^ɿ] dépasse les 200 ms. Donc, c'est cette prolongation de durée qui aura pour effet d'enclencher la vibration de la pointe de langue comme nous l'avons décrit dans la revue de la littérature.

4. Dans le cas de géménées qui sont le fait d'une assimilation — /d/+t/ < [tt] — ou de fausses géménées, il est possible de trouver des géménées en final de morphèmes supérieurs aux bilitères.

— Allongement temporel de la consonne finale /d^s/ réalisée [d:^s]

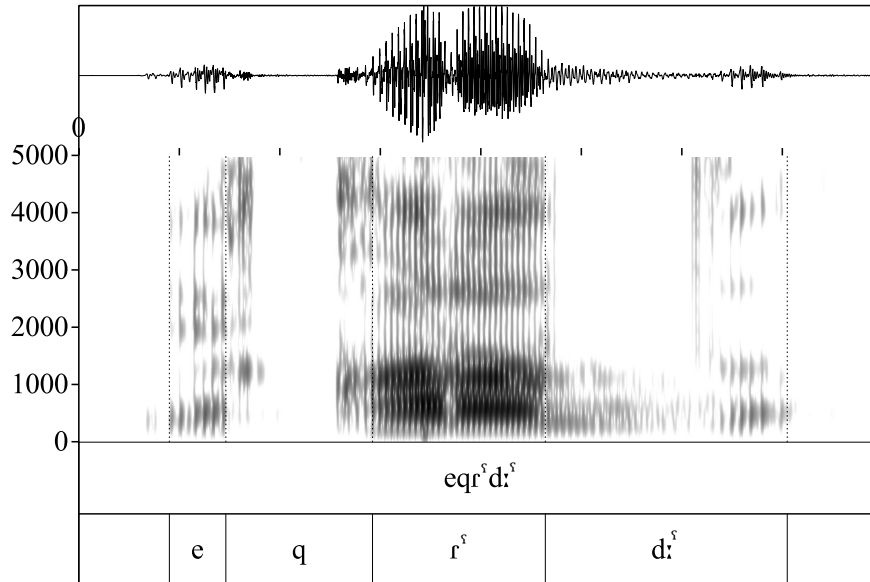


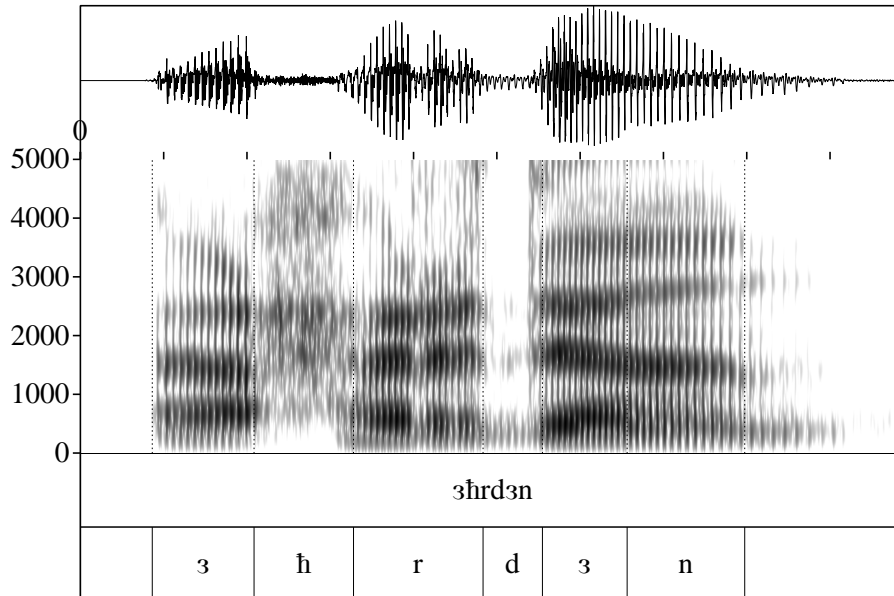
FIGURE III.4.21 : Spectrogramme de @[iqr^sd:^s]@ \fracturer{3.M.SG.PRET}\

D'aucuns nous diraient, mais pourquoi à ce moment-là la ||r|| finale a la même réalisation de surface que celles qui sont dans les autres positions. Nous répondrons que dans le cas présent, nous avons à faire à un phénomène phonétique et non pas phonologique. Si nous considérons que le dévoisement de /d:^s/ est d'ordre phonologique, donc au niveau des représentations nous aurions en contraste /d/ et /t:^s/ et au niveau de la réalité physique la continuité de ces représentations [d^s] et [t:^s]⁵, alors nous pouvons rendre compte de la situation des trills finals. Cette approche est celle que nous avons proposée dans la première partie de notre thèse (Partie II.4.1.4.2, page 138). Nous y avons formulé l'idée que la représentation des géminées n'est pas à reconstruire à partir de la simple. Ce nouveau cas de figure apporte de l'eau à notre hypothèse. Ainsi, la forme de surface des trills se rapporte à deux formes sous-jacentes⁶ qui sont la gémination d'une tap et une tap. Si la tap simple connaît une augmentation de durée, c'est en raison d'un phénomène annexe qui affecte la tap. Nous mettons ce phénomène sur le compte d'un processus phonologique. Cela ne correspond pas à un

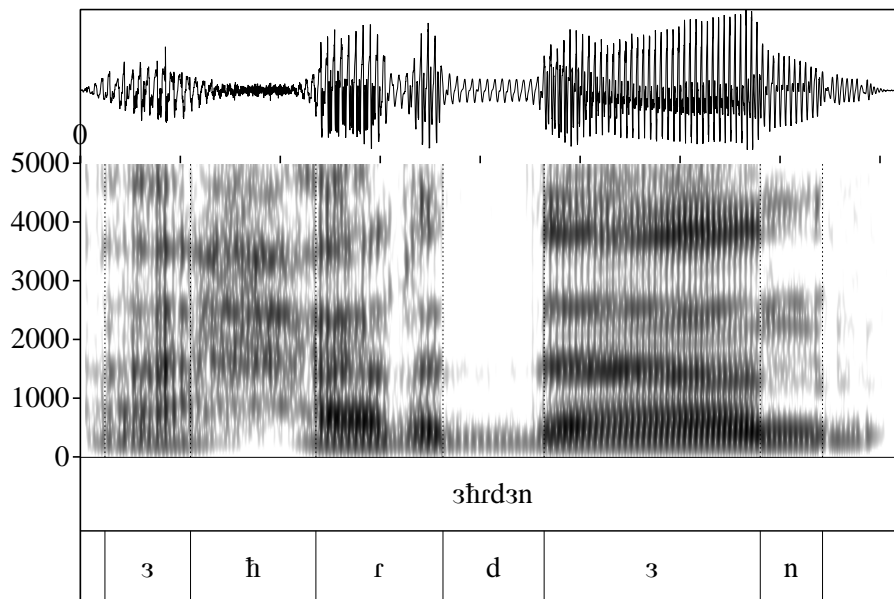
5. Certains chercheurs avancent l'hypothèse d'un phénomène physique. Leur explication est que la pression intra-orale étant plus élevée que la pression subglottale, au moment de l'occlusion, alors en conséquence les plis vocaux seraient empêchés de vibrer [Ridouane, 2003].

6. Voir une troisième si nous considérons l'hypothèse qu'au niveau des représentations la trill est présente comme cela est le cas dans d'autres langues du monde.

processus morphologique en l'occurrence la gémination.



(a) ||r|| est réalisée comme une trill par l'informateur 1



(b) ||r|| est réalisée comme une tap par l'informateur 2

FIGURE III.4.22 : Comparaison des spectrogrammes de @[ʒh||r||dɔʒn]@ \uromastyx\ réalisés différemment par nos informateurs

Cela est d'autant plus clair si nous comparons les données de l'informateur 1 à celles de l'informateur 2, nous constatons que ce qui est réalisé par l'un comme une tap est

produit par l'autre comme une trill. Nous avons relevé cela à plusieurs reprises dans notre corpus, exemples avec les items : @[ʒh||r||dɜn]@, @[ʒh||r||k:os]@, @[tɒq||r||^sneft]@, @[b||r||d:əfən]@ et @[ɛʌt:e||r||n]@. Ci-dessus (Figure III.4.22), une illustration de ce que nous venons d'évoquer à partir d'un de ces items.

Si nous portons une attention particulière au contexte d'apparition de cette variation, nous constaterons qu'elle survient en position médiane et lorsque le noyau de la syllabe est la rhotique alvéolaire. En cette position, très souvent notre informateur 1 réalisera une trill alors que notre informateur 2 produira plutôt une tap. Une autre contrainte est à relever, l'informateur 1 se gardera de réaliser une trill si la consonne qui suit est une fricative. Ainsi, avec les items comme @[ʒdr^sɛʌ]@ ou @[ʒbrʃɜn]@, les ||r|| sont réalisées comme des taps. Néanmoins, ces dernières observations ne sont en rien régulières, car des items comme @[ʒs||r||ʒl]@ ou @[ʒm||r||ez]@ y échappent. Donc, la trill chez notre informateur 1 apparaît si elle est dans des conditions très particulières.

En conséquence, tout donne lieu à penser que la tap dans certaines conditions peut se réaliser comme une trill et que cela est indépendant des représentations des unités phoniques. Nous n'avons pas identifié pour l'instant ce phénomène annexe, cela pourrait tout aussi bien être pour des raisons stylistiques ou être causé par l'accentuation. Il faudrait par ailleurs rapprocher ce phénomène d'une autre observation faite sur les voyelles en final qui connaissent un mouvement articulatoire plus intense en poursuivant leur trajectoire articulatoire au-delà de leur cible articulatoire, toutefois, la durée de ces voyelles n'est pas nécessairement très longue comme pour les consonnes. Cela se caractérise pour une voyelle ouverte ||a|| par une postériorisation et un degré d'aperture plus restreint. Pour les voyelles hautes, cela se caractérise par une diphtongaison⁷. S'il fallait intégrer ce dernier point dans une analyse plus globale, alors des considérations d'ordre syllabique pourraient être envisagées.

Pour conclure, nous laissons pour le moment cette question en suspens, dorénavant nous considérerons qu'une partie des trills finales comme étant la réalisation de surface des taps, ce qui implique une neutralisation de ce dernier en cette position. Et d'autre part, pour les trills réalisées en final d'une bilitère ou d'une monolitére, il est difficile de dire, sans recours au comparatisme intra-dialectal, si cette réalisation de surface est le fait

7. Pour les voyelles ouvertes ||a|| en final, nous n'excluons pas qu'il s'agisse également de diphtongues. Mais dès l'enclenchement de cette voyelle, celle-ci se présente déjà très en arrière par rapport à [ʒ] et suit le mouvement décrit ci-dessus. Nous n'avons pas transcrit ces particularités phonétiques dans nos transcriptions. Car, comme nous l'avions précédemment mentionné, notre attention se focalisera uniquement sur les rhotiques alvéolaires.

d'une tap ou d'une tap géminée. Donc, il faut envisager dans cette situation qu'il existe un archiphonème |R|.

Pour résumer, au niveau phonologique, il apparaît qu'il existe des taps et taps géminées qui, les uns comme les autres peuvent avoir une réalisation phonétique qui se manifeste sous la forme d'une trill. Peut-on affirmer pour autant que toutes les trills sont dans ce cas? N'y aurait-il pas au niveau des représentations une trill? Les trills simples lexicales sont aussi présentes dans notre corpus — voir la figure III.4.14 page 251 —, mais doit-on les considérer comme la manifestation au niveau physique d'un phonème qui serait une tap — géminée ou non — ou bien une trill? Cette question dépasse le simple cas de figure des trills, cela englobe la question de toutes les géminées qui ont une forme de surface différente de la forme de surface de la contrepartie simple. C'est ce que nous avons vu dans la Partie I de notre thèse (II.4.1.4.2, page 138).

Donc, nous ne nous lancerons pas ici même dans une étude théorique sur la question des représentations. Cette problématique reste ouverte, rien n'est moins sûr. Toutefois, si nous avons en ligne de mire une approche réductionniste et unificatrice, cela nous inciterait à considérer également les trills lexicales comme des taps géminées. Si nous privilégions cette hypothèse, alors la trill n'est pas à considérer comme étant un phonème. C'est cette idée qui sera développée dans la partie suivante⁸.

III.4.1.2.1.3 Trill simple et Trill pharyngalisée

Ces deux unités phoniques se distinguent aussi par leur distribution. La trill pharyngalisée est celle qui a une dispersion réduite. La trill pharyngalisée — lexicale ou morphologique — est retrouvée dans l'environnement d'une autre consonne avec une articulation postérieure comme les consonnes pharyngalisées. Nous retrouvons également des trills pharyngalisés qui ne sont pas dans ce contexte, ces derniers sont des emprunts à l'arabe. Ci-dessous des exemples qui illustrent nos précédents propos :

8. Pour ne pas compliquer davantage l'analyse des prochains matériaux, il est préférable de postuler cette hypothèse au risque sinon de nous embourber dans des abstractions rhétoriques. Pour rappel, notre but est seulement de mettre en avant des matériaux et de montrer la complexité qu'il y a à analyser ces derniers. Cette succincte analyse tendra à démontrer qu'il n'est pas aisé de déterminer la réalité phonologique des unités phoniques à partir des méthodes traditionnelles.

Voisinage pharyn.	Emprunt à l'arabe
@[iqr ^s d ^s]@\fracturer{3.M.SG.AI}\	@[fr ^s ʕ]@\verser{2.SG.IMP}\
@[ʌbr ^s d ^s ʌl]@\oiseau\	@[iʕr ^s]@\tromper{3.M.SG.PRET}\
@[sdr ^s d ^s q]@\exploser{2.SG.IMP}\	@[l ^s ʕ ^s ʌr ^s]@\voisin\

TABLE III.4.5 : Contextes d'apparition de la trill pharyngalisée

Le premier exemple présenté correspond, au niveau des représentations, à une tap pharyngalisée géminée qui est un allophone d'une tap simple géminée. Ainsi, le cas précédent se rapporte à une géminée morphologique. Les autres sont des géminées lexicales, et comme nous l'avons dit, elles sont aussi à considérer comme des taps simples géminées. Néanmoins, la question peut se poser : existe-t-il également des taps pharyngalisés géminés dans la représentation phonologique des locuteurs du TmAS? Il s'avère que les contextes que nous venons d'évoquer ne suffisent pas à rendre compte de tous les cas d'apparition de la trill pharyngalisée. Nous avons d'abord les cas des trills pharyngalisés qui se trouvent en final telle que @[ʌfr^s]@\feuille\ ou @[jemʕr^s]@\grandir{3.M.SG.PRET}\. Ces derniers cas sont comme nous l'avons vu plutôt à considérer comme étant la réalisation de surface de taps pharyngalisés.

Nous avons également dans notre corpus des items comme @[tʕr^sʕt]@\braise\ qui contiennent des trills pharyngalisés, mais qui ne sont ni des emprunts à l'arabe, ni dans l'environnement de consonnes pharyngalisés, ni en final. Toutes les trills en position intervocalique, à l'exclusion des emprunts à l'arabe, sont des trills simples comme par exemple @[ʕrʌ]@\enfant\. Donc, nous pourrions envisager que cette pharyngalisation est causée par la consonne uvulaire. Cependant, nous trouvons des items où des trills simples sont dans l'environnement d'une uvulaire comme @[ʕrd]@\pleurer{2.SG.IMP}\⁹. Cette uvulaire précède la trill et n'est pas après elle comme dans [tʕr^sʕt], dès lors peut-être est-ce la position après la rhotique alvéolaire qui est déterminante? Nous ne pouvons exclure totalement cette hypothèse, malgré tout, il est plus raisonnable de nous intéresser à l'étymologie de ce mot. Ce mot dérive du verbe @[r^sʕ]@\brûler{2.SG.IMP}\¹⁰ qui contient une tap pharyngalisée. Comme nous l'avons vu précédemment, la tap pharyngalisée est un phonème, pour autant dans l'item @[tʕr^sʕt]@\ a-t-on à faire à la réalisation phonétique d'une tap pharyngalisée géminée

9. Notre informateur 2 réalise cet item avec [r^s].

10. Cette unité lexicale ne fait pas partie de nos items.

ou bien seulement d'une tap pharyngalisée comme c'est le cas dans @[tʌqr̥neft]@ prononcé par l'informateur 1¹¹? Pour répondre à cette question, il nous semble plus rationnel de considérer la trill pharyngalisée dans @[təɾ̥ʌt]@ comme la caractérisation d'une tap pharyngalisée géminée puisque nos deux informateurs la prononcent de la même manière. Par ailleurs, nous avons dans nos items également des trills pharyngalisés lexicaux pour lesquelles, nous n'avons aucune idée de leur étymologie, exemple : @[ʌqr̥ɔ̃h]@\tête\. Donc, à moins de se lancer dans une étude diachronique, il n'est pas possible de pouvoir rattacher les taps pharyngalisés géminés à des taps simples géminés. Par conséquent, au niveau synchronique, il faut considérer les taps pharyngalisés géminés comme des phonèmes qui ont comme réalisation de surface des trills pharyngalisés.

III.4.1.2.2 Variantes libres

Ce type de variation affecte principalement les trills. La raison principale à l'origine de ces variations est la difficulté à laquelle se trouvent confrontés les locuteurs pour réaliser cette unité phonique. La production de ces trills, comme le rappellent différents chercheurs [Ladefoged and Maddieson, 1996], est difficile. Il s'en suit par conséquent des exécutions de la trill qui s'écartent de la réalisation canonique. Nous rencontrons principalement deux variantes libres dans nos données. La première est un phénomène de dévoisement qui opère dans la partie finale de la réalisation de la trill. La seconde correspond à une désocclusion inaudible de ||r||. Le contact de la pointe de la langue se prolonge au niveau alvéolaire, il n'est plus bref.

Nous rencontrons ce type de variations chez nos deux informateurs, mais ils ne s'appliqueront pas sur les mêmes items, ce qui renforce le côté aléatoire et libre de cette variation. Ci-dessous des exemples correspondant à ces variantes libres tirées des données de l'informateur 1. Les spectrogrammes présentés sont à bande étroite avec une fenêtre de longueur 0,05 s.

11. Notre informateur 2 réalise cet item comme une tap pharyngalisée.

— Dévoisement de la trill en final et en position de coda (syllabe : [z:ɜr])

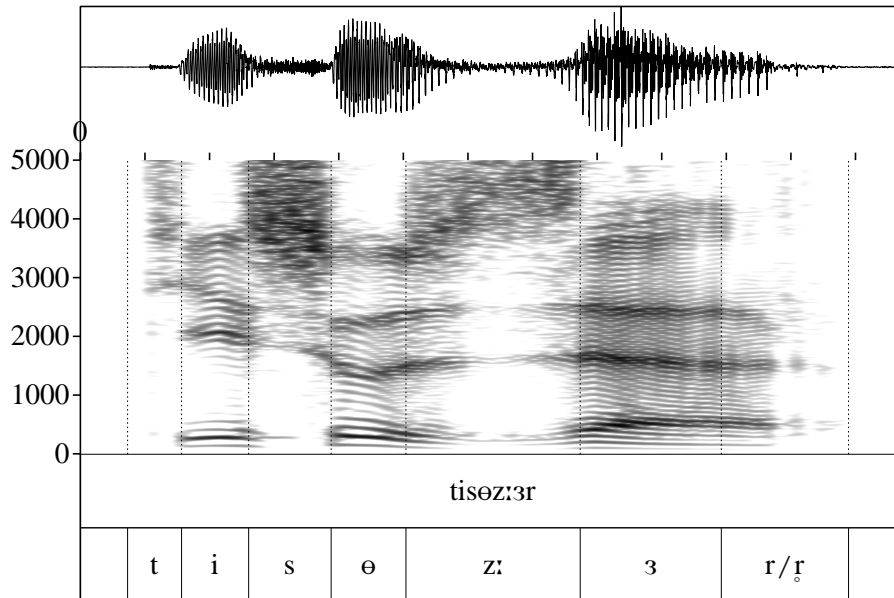


FIGURE III.4.23 : Spectrogramme de @[tisøz:ɜr/ɾ]@ \fourches\

Sur l'illustration, nous pouvons observer qu'au moment du contact de la pointe de la langue la présence de voisement disparaît. Nous pouvons affirmer cela dans la mesure où les premières harmoniques ne sont plus visibles. Ce qui révèle que les plis vocaux ne sont pas en train de vibrer. Cela est particulièrement évident lors du deuxième contact. Mentionnons par ailleurs que ce dévoisement n'opère pas pendant toute la production de la trill. À l'enclenchement de la trill, le voisement est toujours actif. Donc, c'est pour cette raison que nous avons opté pour cette transcription inusuelle : [[r/ɾ]]

— Désocclusion inaudible de ||r|| en final et en position de coda (syllabe : [dɔr])

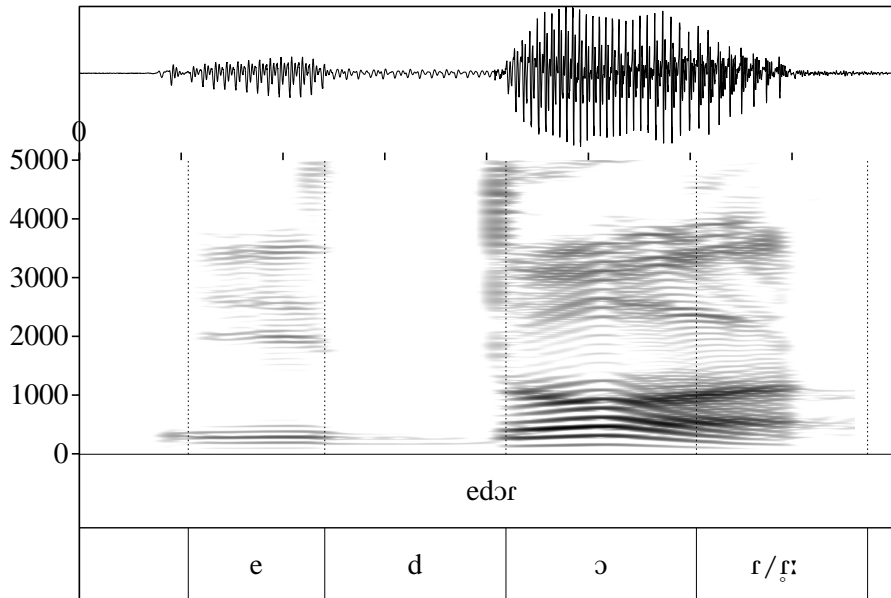


FIGURE III.4.24 : Spectrogramme de @[edɔr/ɣ:ˀ]@\tourner{3.M.SG.PRET}\

Sur cet autre spectrogramme, c'est le deuxième type de variante libre que nous avons illustré. Comme pour le précédent spectrogramme, nous avons choisi une présentation à bande étroite. Ce choix permet de mieux rendre compte du phénomène physique qui se déroule. Ainsi, nous pouvons observer d'une part un dévoisement qui se produit au contact de la pointe de la langue. De plus, ce contact ne s'estompe pas rapidement et ne réalise pas de cycles. Nous imputons ce dernier phénomène à une insuffisance de flux d'air qui ne permet pas d'entamer la vibration de la pointe de la langue. C'est en raison de cette caractéristique phonétique que nous dénommons cette unité phonique comme étant une désocclusion inaudible. Bien que conventionnellement cette dénomination ne renvoie pas précisément à ce phénomène, nous avons décidé toute de même de reprendre cette terminologie. Car il semble que le phénomène que nous venons de décrire ne porte pas d'étiquette particulière en phonétique. C'est ce qui nous a amené à choisir ce qui se rapprochait le plus de notre réalité. Et c'est pour cette même raison que nous avons opté pour cette transcription originale : [[r/ɣ:ˀ]].

D'autre part, grâce à ce défaut d'articulation, cela nous permet de constater qu'en final, il existe bien un allongement consonantique de ||r||. Car, comme nous le verrons par la suite, dans les cas normaux, le contact a une durée d'environ 20 ms. Mais sur le spectrogramme nous pouvons nous rendre compte que le contact avoisine les

100 ms. Cela appuie nos affirmations précédentes sur le fait que la trill finale est une réalisation de surface d'une tap. D'autre part, nous pouvons supposer, dans le cas qui nous intéresse, que le mouvement de la pointe de la langue ne s'apparente pas à une battue. Nous interprétons cette situation, par le fait que le contact n'est plus bref. Bien que le mouvement articulatoire n'est pas identique à une battue, nous avons tout de même repris la transcription de la tap [ɾ]. C'est un choix par défaut. Faute de mieux, nous avons fait usage de ce symbole.

III.4.1.3 Synthèse

Nous avons passé en revue les différentes rhotiques alvéolaires pour déterminer quelles pouvaient être leurs représentations. Nos analyses ont fait l'hypothèse que les trills étaient une réalisation de surface de taps géminées. Au niveau phonologique, les phonèmes présents sont : /ɾ/, /ɾ^s/, /ɾ:/ et /ɾ:^s/. Par ailleurs, nous avons dégagé le fait que les trills ne sont pas uniquement les formes de surface des taps géminées, elles le sont également pour les taps non géminées, mais seulement dans certains contextes. Nous ne souhaitons pas exclure totalement l'hypothèse d'une trill phonologique, car la prise en compte des particularités des géminées des variétés du berbère de cette région laisse à penser qu'en instance la situation est plus compliquée (voir Partie II.4.1.4, page 119).

D'autant plus que dans ce cas, bien que la réalisation finale de la trill soit le fait d'un phénomène physique, elle peut être une réalisation voulue, donc phonologique, ou bien une réalisation non désirée, donc phonétique. Cette indétermination a pour conséquence qu'il est difficile de savoir à partir de la forme de surface à quelle réalisation nous avons à faire. Cela vaut tout particulièrement pour le cas des trills lexicales qui ne sont pas le fait d'un processus morphologique. Le fait de reconstruire toutes les trills en prenant en considération ce dernier processus est une méthodologie discutable. La preuve de cela est le cas des trills finales. Donc, comme pour les géminées fricatives, il est probable que la trill, même morphologique, ne tient pas de la tap simple au niveau des représentations. Il faudrait postuler ainsi que la trill et la tap se distinguent aussi au niveau de la qualité et non pas qu'au plan de la quantité.

Nous avons d'autre part observé qu'il existait une variation d'ordre idiomatique. La trill pouvait être la forme de surface d'une tap dans des contextes bien particuliers autre qu'en final. Contrairement aux taps en final qui sont systématiquement des trills chez nos deux informateurs, cette réalisation n'est pas conventionnelle. Elle est propre à notre

informateur 1. Cette constatation va dans le sens que la tap simple peut aussi être une trill, cette dernière, selon l'hypothèse de départ, n'est pas une réalisation exclusivement de surface d'une tap géminée.

Par la suite, cette analyse a permis de faire ressortir les différents allophones de nos phonèmes. Il y a d'abord, ceux qui se caractérisent par leur contextualité et que nous venons à l'instant d'évoquer : [r] et [r^h]. Puis, il y a les allophones qui sont partiellement contextuels, mais qui apparaissent de manière aléatoire : [[r/r̥]] et [[r/r̥ː]].

Tous les aspects que nous venons d'analyser ne sont pas très différents de ce qui a pu être rapporté par différents chercheurs de par le monde. Donc, notre travail d'une certaine manière va dans leur sens, notamment s'agissant du rapport entre les trills et les taps géminées. Cependant, il apporte aussi de nouveaux éléments en invitant à ne pas considérer la trill comme étant inévitablement le pendant phonétique d'une tap géminée.

Cette variabilité que nous venons de brosser est loin d'être exhaustive. D'autres particularités phonétiques existent, qui toutefois ne sont pas de nature à remettre en question les phonèmes que nous avons mis en exergue. C'est ce que nous allons voir à présent.

III.4.2 Analyse phonétique

Dans cette partie, seules des considérations phonétiques seront prises en compte. Dans la partie précédente, la dimension phonétique n'était là que pour parvenir à remonter aux représentations phonologiques. Dans un premier temps, nous analyserons les particularités physiques de la première articulation. Puis, nous nous intéresserons à la nature des vocoïdes.

III.4.2.1 L'articulation primaire

Comme nous l'avons vu au cours de notre revue de la littérature, les rhotiques alvéolaires présentent au minimum deux articulations, c'est-à-dire deux constriction en des endroits différents de la langue. Durant cette partie, nous n'évoquerons que l'articulation primaire qui correspond à la constriction réalisée par la pointe de la langue. D'autre part, notre attention ne portera que sur le mode d'articulation de nos rhotiques alvéolaires. Nous chercherons à déterminer si ces $\|r\|$ sont occlusives, approximantes ou fricatives. Cette partie sera suivie par une étude quantitative sur la durée d'occlusion des taps.

III.4.2.1.1 Mode d'articulation de l'articulation primaire

Toute unité phonique peut être définie par au minimum deux dimensions qui sont le mode d'articulation et le point d'articulation. Pour rappel, le mode d'articulation définit entre autre le type d'obstruction réalisé par la langue. Mais, le mode d'articulation peut comprendre d'autres particularités articulatoires comme le type de cavité de résonance par où s'échappe le flux d'air. Ou encore le type d'écoulement du flux d'air produit par les poumons. Pour notre part, le mode d'articulation ne sera considéré, au cours de cette thèse, que sous l'angle de l'obstruction et ne prendrons pas en compte les autres particularités articulatoires.

Quant au point d'articulation, il précise à quel endroit au niveau de la cavité buccale l'obstruction est maximale. Nous n'avons pas consacré d'analyse à ce dernier aspect. Nous faisons l'hypothèse que le point d'articulation se situe dans la zone alvéolaire. C'est en partie grâce à ces deux dimensions qu'un classement des unités phoniques est effectué. Nous allons seulement nous intéresser à cette première dimension.

Les trois principaux modes d'articulation sont l'occlusion, la friction et l'approximante. Ces derniers sont utilisés pour classer les unités phoniques qui sont considérées comme étant des obstruents. Et comme nous l'avons vu, nos unités phoniques ne sont pas cataloguées dans cette catégorie. Par conséquent, les modes d'articulation que nous venons de citer ne leur sont traditionnellement jamais affectés. Elles sont caractérisées par un mode d'articulation propre dénommé battu et roulé. Mais, comme l'ont démontré plusieurs études, l'obstruction de la pointe de la langue ne se fait pas de manière identique. Il existe des degrés d'obstruction qui peuvent être identifiés à ce qui est trouvé pour les obstruents. Cette observation a amené certains chercheurs à vouloir reclasser les ||r|| dans la classe des obstruents [Baltazani, 2009].

Nous réduirons notre champ d'investigation à la tap simple et à la trill simple. Les contreparties pharyngalisées ne feront pas l'objet d'illustration, car elles se comportent à ce niveau de la même manière que les simples. Pour ce faire, nous analyserons successivement nos différentes rhotiques alvéolaires en commençant par la tap.

III.4.2.1.1.1 Le matériau d'analyse

À cet endroit, nous allons faire ressortir les différentes réalisations possibles du mode d'articulation de l'articulation primaire. Nous avons extrait quelques items de notre corpus pour exemplifier nos constatations. Nous utiliserons encore une fois les spectrogrammes pour certifier ce que nous alléguerons. Ces derniers proviennent toujours des données enregistrées avec l'informateur 1.

III.4.2.1.1.1.1 Mode d'articulation de la tap

Les taps qui figurent dans notre corpus ont en très grande majorité une obstruction de type approximant. Cela signifie que le contact avec la zone alvéolaire n'a pas abouti. Et qu'entre la pointe de la langue et le point d'articulation le passage de flux d'air est continu.

- Tap approximante en position médiane et en position syllabique de noyau (syllabe : [rs])

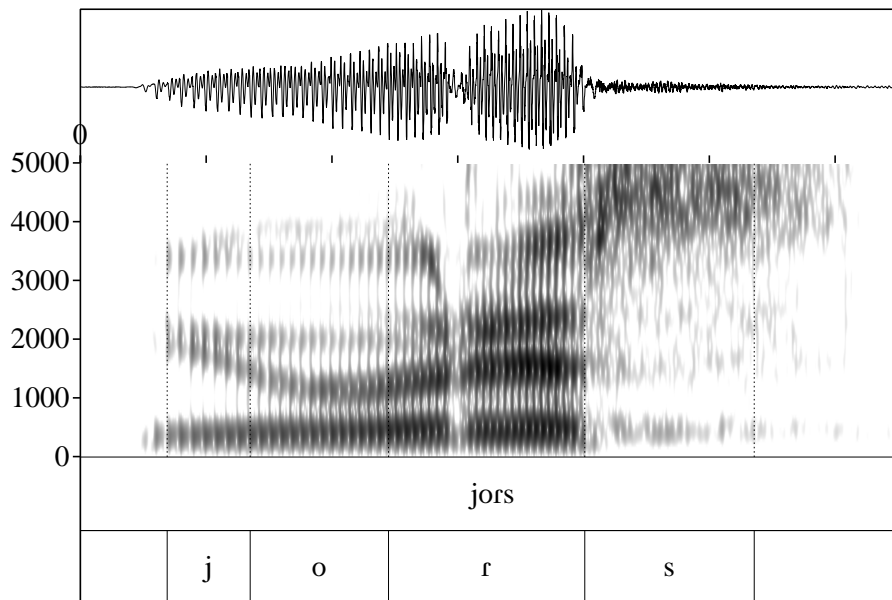


FIGURE III.4.25 : Spectrogramme de @[jors]@ \être_créateur_de{2.SG.IMP}\

Comme nous pouvons le visualiser sur le spectrogramme, lorsque la pointe de la langue atteint sa cible lors de la production de [r], le spectre n'est pas entièrement estompé. Il y a toujours de l'énergie dans les basses fréquences. Une partie de cette énergie correspond au voisement et se trouve dans les plus basses fréquences. Et l'autre partie du spectre, entre 500 Hz et 3000 Hz, indique que le flux d'air est continu lors de l'obstruction. Comme le spectre ne contient pas à cet endroit d'énergie dans les hautes fréquences, nous en déduisons que le mode d'articulation est de type approximant et non fricatif.

Toutefois, nous trouvons quelques items où le mode d'articulation est fricatif. Cette réalisation phonétique est contextuelle, nous l'observons que dans l'environnement de fricatives. Ci-dessous un exemple illustrant ce phénomène.

— Tap fricative en position médiane et en position syllabique de noyau (syllabe : [br])

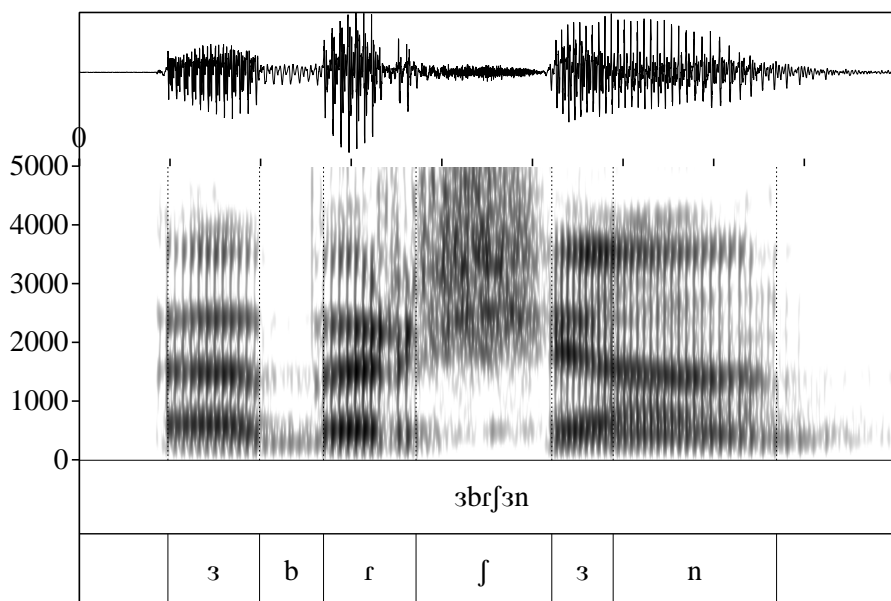


FIGURE III.4.26 : Spectrogramme de @3brf3n@ \ noir\

Contrairement au spectrogramme précédent, la zone correspondant à l'obstruction est totalement occupée par un spectre d'énergie. Des plus basses fréquences aux plus hautes fréquences, nous visualisons une trainée noire. Cela est le signe que le flux d'air est de nature turbulent, donc que le mode d'articulation à ce moment précis est fricatif. Nous attribuons cette frication à la proximité de [r] avec [ʃ], donc c'est un phénomène de coarticulation qui est à l'origine de cette réalisation phonétique.

Dans de rares cas, la réalisation de la tap peut être occlusive. L'item @[røz:ɛʁ]@, que nous avons déjà utilisé pour illustration (figure III.4.5 page 245), a ce type de mode d'articulation. Si nous observons attentivement le spectrogramme — que nous ne reprendrons pas ici même —, nous remarquerons que l'occlusion est réalisée que très brièvement. Si nous comparons cette réalisation avec celle donnée par Blecua (figure ?? page ??), il apparaît une différence de durée significative. En nous basant sur cette dissemblance, il est possible d'interpréter cette occlusion comme étant une réalisation secondaire, probablement idiomatique ou causée par l'environnement. Autrement, la durée d'occlusion aurait été plus étendue, c'est en tout cas l'assertion *ad hoc* que nous faisons.

En résumé, le mode d'articulation des taps est principalement approximant. Mais nous n'avons pas choisi de transcrire ces ||r|| par le symbole API : [ɹ], comme le font

certaines études. Parce que, bien que ce symbole soit utilisé pour décrire un son approximant, il ne décrit pas exactement la même unité phonique que nous retrouvons en **TmAS**. Il est certain que le mode d'articulation de la première articulation soit approximant, en revanche, il s'agit toujours d'une battue. Par conséquent, si nous devons choisir un symbole pour représenter fidèlement la réalité physique, nous utiliserions celui-ci : [[ɾ̩]].

La diacritique représentée sous la tap précise qu'il y a un abaissement. Cet abaissement concerne la pointe de langue et non pas une autre partie de la langue. Cette précision est importante, car cette même diacritique pourrait être utilisée pour décrire un fléchissement articuloire qui toucherait une autre partie des organes de la parole et qui pourrait être en jeu. D'autre part, l'**API** ne prévoit pas, malheureusement, de symbole pour décrire des degrés d'abaissement qui seraient utiles pour faire la différence entre une fricative et une approximante. Donc, ce symbole pourrait servir à dépeindre ces deux particularités phonétiques.

III.4.2.1.1.1.2 Mode d'articulation de la trill

Nous allons comme précédemment parcourir et présenter les différents modes d'articulation de la première articulation. Les trills se distinguent des taps par le fait que la pointe de langue réalise plusieurs battements — au minimum deux —. Leurs obstructions peuvent être approximantes au niveaux de tous les cycles réalisés par la trill comme nous le pouvons voir ci-dessous.

- Trill approximante en position médiane et en position syllabique de noyau (syllabe : [rn])

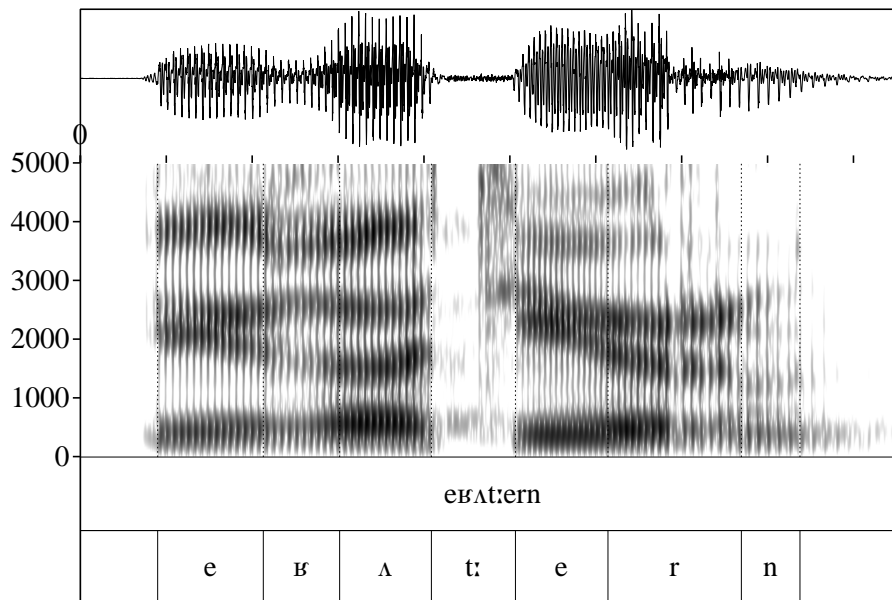


FIGURE III.4.27 : Spectrogramme de @[eβΛt:ern]@ \épines_de_palmier\

D'autre part, la particularité physique de ces battements est qu'ils sont totalement incontrôlés. Ce non contrôle de la vibration a aussi des implications quant à la nature du mode d'articulation. Les types d'obstructions rencontrés ne sont que très rarement identiques. Les contacts alternent entre approximant et fricatif. Ce constat peut être observé dans l'item @[idre]@ (figure ?? page ??) où le premier contact est fricatif, alors que les deux suivants sont approximants. Mais cette répartition est totalement aléatoire. Sur d'autres items, le mode approximant et le mode fricatif porteront sur d'autres cycles,

comme nous pouvons le constater sur le spectrogramme ci-dessous.

- Trill approximante et fricative en position médiane et en position syllabique d'attaque (syllabe : [ro])

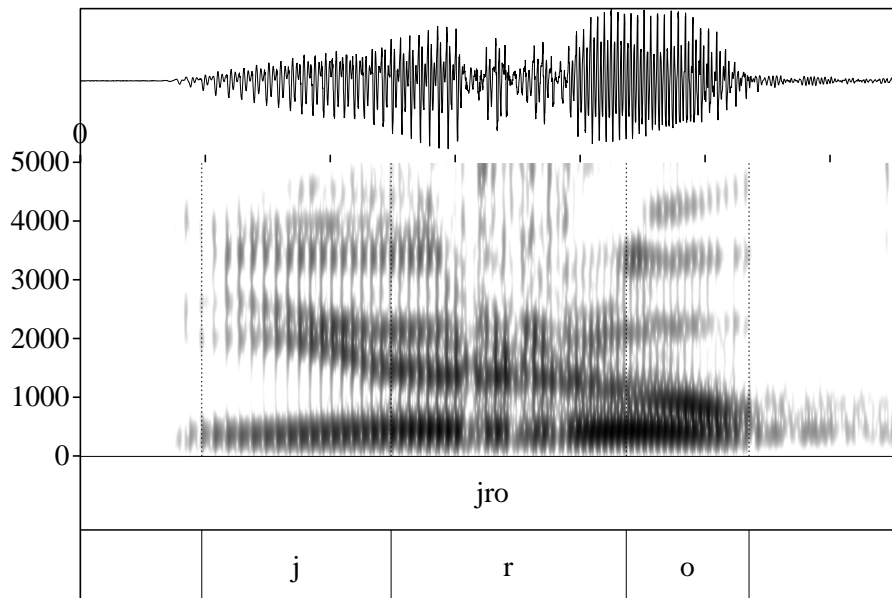


FIGURE III.4.28 : Spectrogramme de @[jro]@ \rendre{3.M.SG.PRET}\

Sur le spectrogramme, nous pouvons observer que contrairement à l'item @[idre]@, le premier et le deuxième contact sont approximants, alors que la dernière obstruction est fricative. Cela démontre le côté aléatoire et arbitraire de ces obstructions. Donc, le mode d'articulation dépend fortement des conditions aérodynamiques qui sont à l'œuvre dans la production de la trille. Cette aérodynamisme de toute évidence n'est pas en lui-même sous le contrôle total du locuteur. Chaque production de la trille semble relever du particularisme.

III.4.2.1.1.2 Discussion

Les matériaux que nous venons de visualiser et de commenter montrent sans conteste que le mode d'articulation de l'articulation primaire est multiple. Les ||r|| en TmAS peuvent être approximantes, fricatives et rarement occlusives. Cela induit que nous pouvons difficilement abonder dans le sens de certains chercheurs qui considèrent que les taps doivent être classées dans la classe des obstruantes. Car, dans le cas qui nous concerne, l'articulation primaire peut être à la fois et obstruante et sonante. Signalons

au passage que cette variation est aussi idiomatique, notre informateur 2 n'aura pas le même mode d'articulation que notre informateur 1 pour le même item. Celui-ci pourra produire des occlusives, là où figurent des approximantes chez l'informateur 1. Ou bien des fricatifs, là où l'informateur produira des approximantes. Ce sont pour toutes ces raisons que nous n'avons pas pris en compte cette dimension dans l'analyse phonologique, car cela ne relève que de la réalité physique. Cette variabilité ne peut pas être le fait d'un phénomène phonologique.

Nous avons précédemment évoqué quelques hypothèses concernant les raisons de cette disparité. Si nous considérons que le mode approximant est le mode normal de réalisation de $\|r\|$, alors il faut envisager que les autres réalisations sont causées par l'environnement. La coarticulation est la principale cause selon nous de cette variation. Par contre, les trills semblent être moins affectées par leur environnement, cela confirme les observations articulatoires faites dans d'autres langues. Il a été établi que les trills étaient moins influencées que les taps par le contexte phonétique. Si les trills ont des modes d'articulation différents, c'est en raison d'un autre phénomène qui est en lien avec les conditions aérodynamiques et biomécaniques. La force musculaire consacrée pour mettre la langue en position et pour la maintenir, comme les propriétés biomécaniques de la muqueuse, a inévitablement une influence sur le résultat final que sera la vibration. Ce dernier paramètre mériterait que nous nous y penchions de plus près avec une étude scientifique afin de cerner l'impact qu'il produit sur les trills.

Autre point à noter, en dehors du fait que les approximants et les fricatifs se distinguent par leur degré de constriction, une autre particularité les différencie. Les $\|r\|$ approximants connaissent une sorte de burst à la fin de l'obstruction, mais qui n'est pas systématiquement présente. Nous pouvons difficilement considérer cette impulsion comme étant l'équivalent des bursts réalisées par des consonnes occlusives puisque l'articulation de ces taps ne se fait pas en coupant le flux d'air — dans la majorité des cas —. D'autant plus que cette sorte de burst est extrêmement brève — moins de 5 ms — comparée aux occlusives voisées qui ont une durée bien plus longue. Cela implique que ce phénomène acoustique n'apparaît pas comme pouvant être relié à une concentration d'air qui s'accumulerait derrière la langue pendant la tenue de l'obstruction comme c'est le cas pour les occlusifs. Si c'est le cas, alors peut-être est-ce en raison de la rapidité de l'obstruction — environ 20 ms — qui créerait un différentiel de pression qui n'a pas eu le temps de s'échapper par le passage laissé par la pointe de la langue? Ou bien est-ce tout simplement la continuité du phénomène de turbulence qui s'accroît à la fin de la

réalisation de l'obstruction? Il est difficile de formuler une hypothèse fondée, car ce burst apparaît également au début du contact. Peut-être que l'un et l'autre de ces bursts ne représentent pas le même phénomène physique ou bien ils sont bien liés. Donc, si nous considérons la dernière hypothèse, il apparaît difficile d'arguer que nous sommes en présence d'un relâchement d'air compressé par une occlusion. D'autres études plus pointues sont nécessaires pour mettre au clair cette question. Nous ne pouvons dans notre situation qu'envisager des hypothèses.

III.4.2.1.2 Durée du contact de la tap

Dans cette étude, nous ne ferons pas de distinction entre les taps ayant l'une ou l'autre des modes d'articulation que nous avons observés dans la partie précédente. Ce qui nous intéressera au cours de ce travail, c'est de mener une étude comparative entre la durée de contact des taps simples et des taps pharyngalisés. Cette particularité des taps est régulièrement mentionnée dans la littérature scientifique comme nous l'avons vu. Mais à notre connaissance, une comparaison qui porte sur la durée d'occlusion des taps simples et des taps pharyngalisés n'a encore jamais été conduite. La raison en est simple, les taps pharyngalisés ne sont pas typologiquement très fréquentes dans les langues du monde.

III.4.2.1.2.1 L'analyse statistique

Notre corpus nous a permis de collecter 77 taps simples et 76 taps pharyngalisés ce qui fait un échantillon de même taille. Nous n'allons pas revenir sur la procédure que nous avons utilisée, car nous l'avons déjà évoquée. Pour pouvoir comparer l'ensemble des données de ces taps, il nous faut une présentation des données et des valeurs discrètes qui puissent rendre d'un coup d'œil de la situation. Le meilleur moyen pour cela est d'utiliser une boîte à moustache. Cette dernière permet de représenter la valeur médiane et de mettre en avant où se concentre la distribution d'une partie de nos mesures. Après calcul, nous obtenons les données chiffrées suivantes.

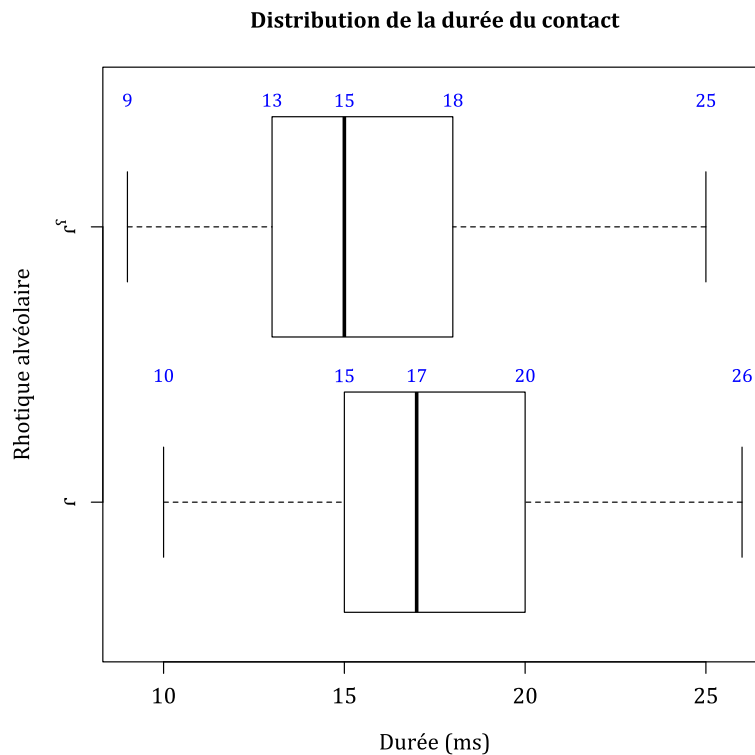


FIGURE III.4.29 : Boîte à moustache de la durée de contact des taps

D'après ces résultats, il semblerait que la durée d'obstruction des taps simples soit légèrement plus élevée que celle des taps pharyngalisées. La valeur médiane pour la tap simple est de 17 ms et pour la tap pharyngalisée de 15 ms. Donc, entre les deux, il existe seulement un écart de 2 ms. Le graphique à boîte à moustache nous permet également de constater une légère asymétrie pour les deux taps vers les valeurs plus basses. Au vue de la distribution, nos deux taps sont relativement identiques. Pour la tap simple l'étendue des valeurs est entre 9 et 25 ms et pour la tap pharyngalisée, elles sont comprises entre 10 et 26 ms.

Par ailleurs, nous avons mené un test de normalité pour savoir si nos échantillons suivent une loi normale, pour nous aider par un test d'hypothèse à trancher la question. Les résultats du test de Shapiro-Wilk et celui du test graphique Q-Q plot ne révèlent pas que nos données suivent une loi normale. La p-value pour la tap simple est de 0,014 et pour la tap pharyngalisée de $7,7 \cdot 10^{-6}$, donc elles sont inférieures à l'intervalle de confiance de 0,05 que nous nous sommes fixé.

Par conséquent, au vu de ces chiffres, nous ne pouvons réellement conclure à une différence significative entre leurs durées d'obstruction. En conclusion, il

n'y a pas d'éléments qui nous permettraient d'affirmer que le contact des taps pharyngalisés serait plus court que celui des taps simples. D'autant plus que les biais, qui sont inévitablement présents dans ce genre d'expérience, neutralisent les quelques différences que nous pouvons constater.

III.4.2.1.2.2 Discussion

Le battement de la pointe de la langue se réalise dans un espace de temps qui est grosso-modo le même à travers les langues. Si nous comparons les résultats de la tap du simple avec ceux des autres langues. Nous nous apercevons que nos valeurs s'inscrivent dans ce qui est retrouvé dans la littérature scientifique (voir Partie III.2.2.1.2, page 198). Nous pouvons faire ce rapprochement uniquement pour les taps simples. L'autre variable que notre étude comprend ne peut rentrer dans cette comparaison. Ainsi, l'originalité de notre étude est de mettre en vis-à-vis des taps simples et des taps pharyngalisés. Nos résultats montrent que presque en tout point la durée d'obstruction de nos deux $\|r\|$ sont identiques.

Cette similarité au niveau de l'obstruction de la première articulation peut à première vue sembler étrange dans la mesure où l'articulation générale d'une pharyngalisée par rapport à une simple n'est pas identique. Comme nous l'avons vu les pharyngalisés ont leurs propres caractéristiques articulatoires, notamment une position plus postérieure de l'arrière de la langue. En conséquence, nous devrions nous attendre à une différence significative. Mais en dépit de tout cela, les résultats indiquent clairement que ce n'est pas ce qui se produit. Ce fait n'est pas anodin et peut être interprété d'une autre manière.

Cette observation dit des choses sur l'articulation primaire par rapport à l'articulation secondaire. Si l'articulation secondaire ne semble pas avoir d'effet sur la durée d'obstruction, cela peut expliquer une indépendance articulatoire et phonologique entre ces deux parties de la langue. Nous pouvons avancer cette hypothèse par le fait que la tap simple ignore cette articulation secondaire ce qui n'est pas le cas pour la tap pharyngalisée. Mais pourtant leur articulation primaire se comporte de la même manière, ainsi nous pouvons déduire que le mouvement de la pointe de langue n'est pas conditionné par la racine de la langue.

Une dernière hypothèse peut être formulée à partir de ces résultats. La durée de la tenue de l'obstruction et le mode d'articulation étant les mêmes entre la tap simple et la tap pharyngalisée, alors nous pouvons supposer qu'il parcourt la même distance — à supposer que la vitesse de battement étant la même pour les deux —. Si tel est le cas, il y a

fort à supposer que le point d'articulation entre les deux est identique. Qu'une fois encore la postériorisation semble donc ne pas affecter fortement le mouvement articulaire de la pointe de la langue.

III.4.2.2 Les vocoïdes du tap

Comme c'est le cas pour bon nombre de langues, nous retrouvons également en **TmAS** ce phénomène. Nous allons brosser un descriptif rapide des particularités qui touchent à cette unité sonore. Nous nous attarderons que sur le cas des vocoïdes retrouvés pendant la réalisation des taps qu'elles soient simples ou pharyngalisées. Notre attention se portera sur l'essence de ces vocoïdes, sont-elles une manifestation intrinsèque des taps ou un phénomène annexe? Puis, nous prolongerons notre étude sur une particularité physique de ces taps, nous mesurons la durée de la réalisation de cette unité sonore.

III.4.2.2.1 L'environnement des vocoïdes

Nous avons fait le choix de présenter les données suivant leur place dans le lexème et au niveau de la syllabe. Cependant, notre corpus n'est pas assez exhaustif pour nous permettre d'avoir toutes les configurations syllabiques possibles pour chaque position lexicale et chaque type de rhotique alvéolaire. Mais cela ne signifiera pas pour autant qu'une telle combinaison soit impossible.

Le présentation des données sera structurée en mettant en avant la position lexicale, puis la position syllabique. Ce choix est motivé par l'importance que revêt la position lexicale pour déterminer la nature de ce vocoïde. En prenant en compte par ailleurs la position syllabique, cela pourra démontrer que la présence du vocoïde n'est pas conditionnée par sa place dans la syllabe, ce qui donnera du poids aux affirmations que nous formulerons plus tard.

III.4.2.2.1.1 Le matériau d'analyse

Nous allons donc pour chaque rhotique alvéolaire existant en **TmAS** présenter les données en y intégrant ces considérations syllabiques. Dans cette énumération des types syllabiques, des formations syllabiques tautomorphémiques comme hétéromorphémiques pourront être prises en compte. Nous illustrerons par la suite nos dires systématiquement par des spectrogrammes qui seront tout autant de preuves

concrètes des affirmations qui seront avancées par la suite.

Toutefois, nous ne présenterons pas pour chacun de ces cas de figure un spectrogramme. Tout ceci ne ferait que surcharger l'exposition de nos résultats et n'apporterait pas grand chose à la problématique que nous souhaitons traiter. Par conséquent, nous choisissons de focaliser notre attention sur les cas les plus intéressants de notre corpus. Ceux-ci sont tous les cas où la rhotique alvéolaire est intercalée entre des consonnes sourdes. Ainsi, parmi les catégories syllabiques que nous pourrions déterminer par la suite, nous n'exposerons que celles qui ont cette originalité. Donc, certaines configurations ne pourront pas remplir nos conditions, mais nous précisons tous les cas possibles que nous avons rencontrés dans notre corpus.

III.4.2.2.1.1.1 En position initiale

En cette position, le type de syllabe dans lequel peut s'insérer la rhotique alvéolaire se réduit à quatre possibilités : #r#, #r·C, #rC·C ou #rV(C)·. Dans les trois premières configurations, il faut considérer la rhotique alvéolaire comme étant le noyau de la syllabe.

III.4.2.2.1.1.1.a Unité phonique : Tap simple

— Tap en position initiale et en position syllabique de noyau (syllabe : [r])

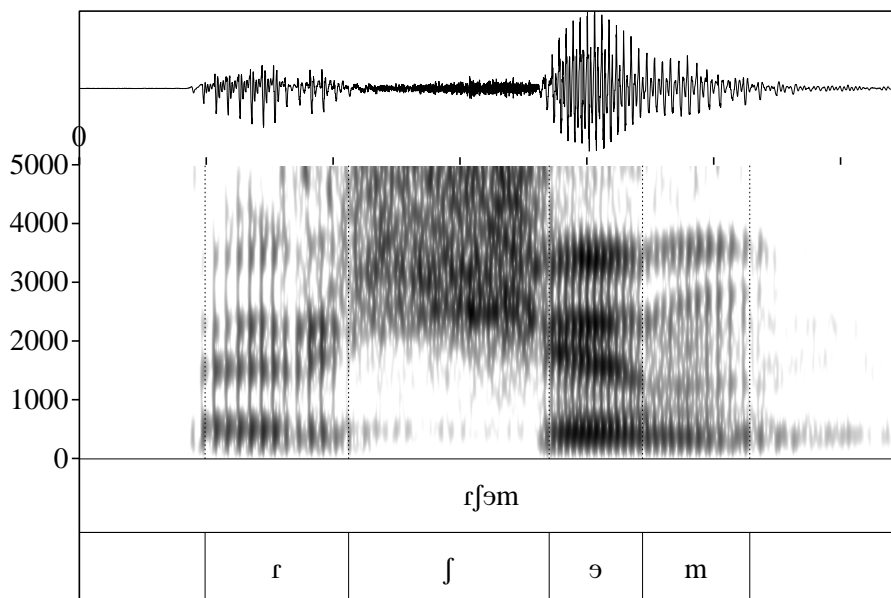


FIGURE III.4.30 : Spectrogramme de @[rʃəm]@ \ mouiller{2.SG.IMP}\

III.4.2.2.1.1.1.b Unité phonique : Tap pharyngalisée

- Tap pharyngalisée en position initiale et en position syllabique de noyau (syllabe : [r^ʕ])

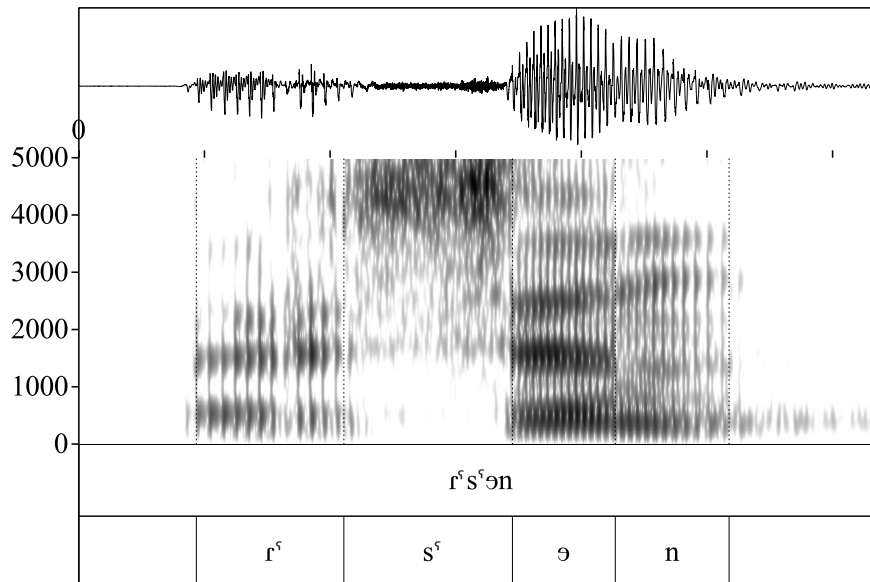


FIGURE III.4.31 : Spectrogramme de @[r^ʕs^ʕən]@ \être_calme{3.M.PL.PRET}\

III.4.2.2.1.1.2 Médiane

Les types de syllabe pouvant exister en cette position sont nombreux. Parmi eux, il faut distinguer, celles où la rhotique alvéolaire a comme position syllabique le noyau et celles où la rhotique alvéolaire se trouve soit en coda ou en attaque. Dans ce dernier cas, il faut subdiviser, encore, pour plus de clarté, et classer les types de syllabe entre celles qui forment un groupe consonantique et celles qui ont seulement des attaques et des codas simples. Pour une meilleure représentation de ces possibilités, nous avons réalisé un tableau résumant les types de syllabe rencontrés dans notre corpus en position médiane.

		Coda					
		Complexe		Simple		Sans	
Attaque	Complexe	∅	∅	C r VC	∅	C r V	∅
	Simple	∅	∅	r VC CV r r V r	C r C	∅	C r
	Sans	∅	V r C	V r	r C	r V	r

TABLE III.4.6 : Les types de syllabe en médiane

Ces différentes configurations ne feront pas l'objet d'une illustration. Certaines d'entre elles ne concordent pas avec les conditions que nous nous sommes fixées. Donc, par la suite, la présentation de nos matériaux se limitera aux cas les plus pertinents.

III.4.2.2.1.1.2.c Unité phonique : Tap simple

— Tap en position médiane et en position syllabique de noyau (syllabe : [fr])

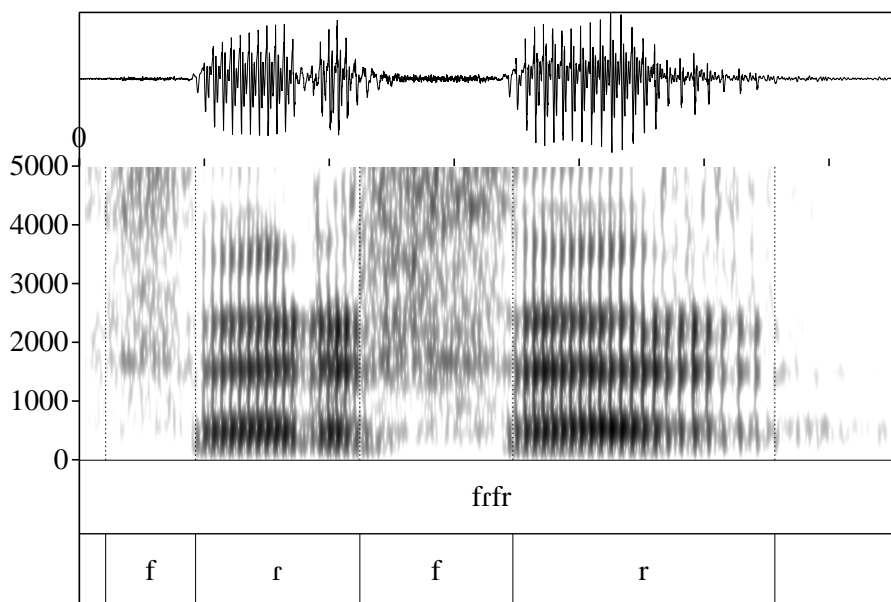


FIGURE III.4.32 : Spectrogramme de @[frfr]@ \s'envoler{2.SG.IMP}\

— Tap en position médiane et en position syllabique de noyau (syllabe : [sr])

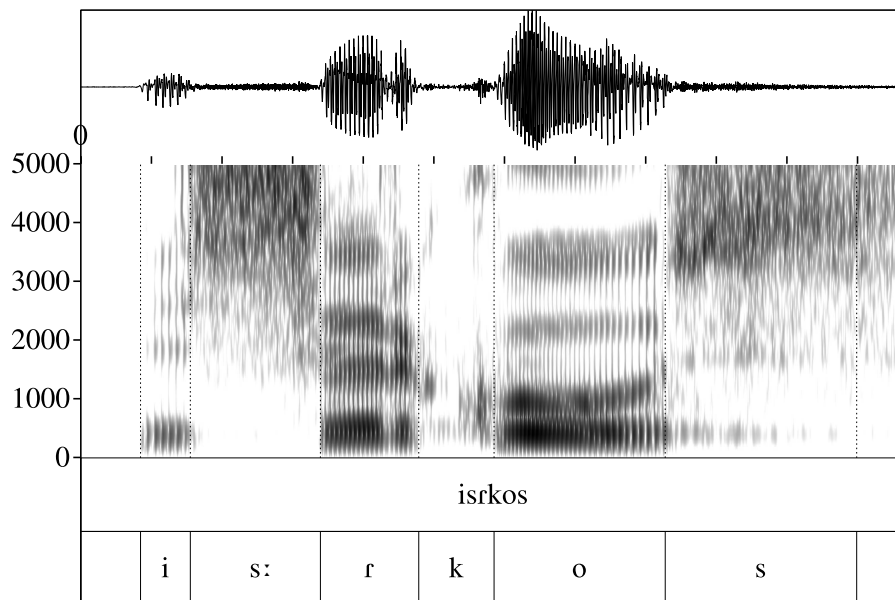


FIGURE III.4.33 : Spectrogramme de @[is:rkos]@ \mentir{3.M.SG.AI}\

— Tap en position médiane et en position syllabique de noyau (syllabe : [ʃrʃ])

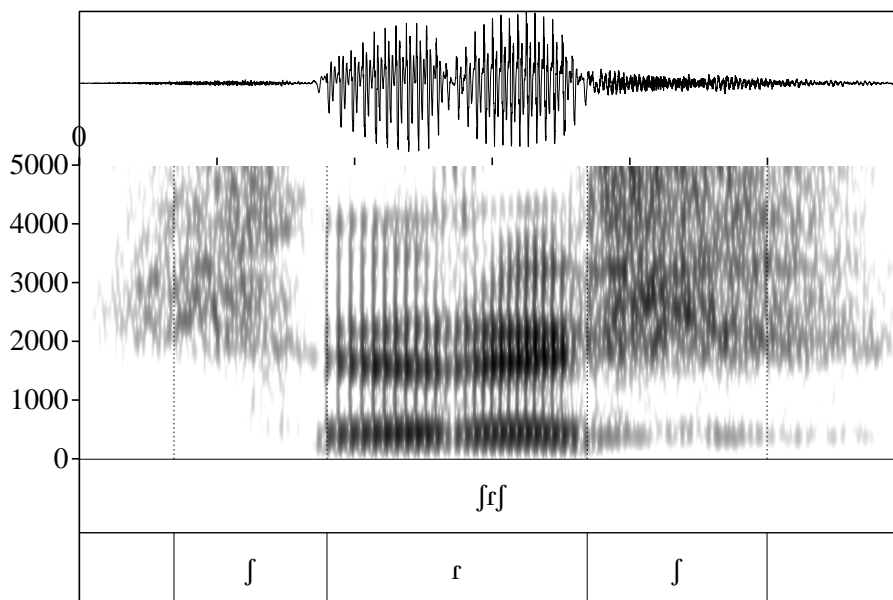


FIGURE III.4.34 : Spectrogramme de @[ʃrʃ]@ \attacher{2.SG.IMP}\

III.4.2.2.1.1.2.d Unité phonique : Tap pharyngalisée

- Tap pharyngalisée en position médiane et en position syllabique de noyau (syllabe : [r^ʰt^ʰ])

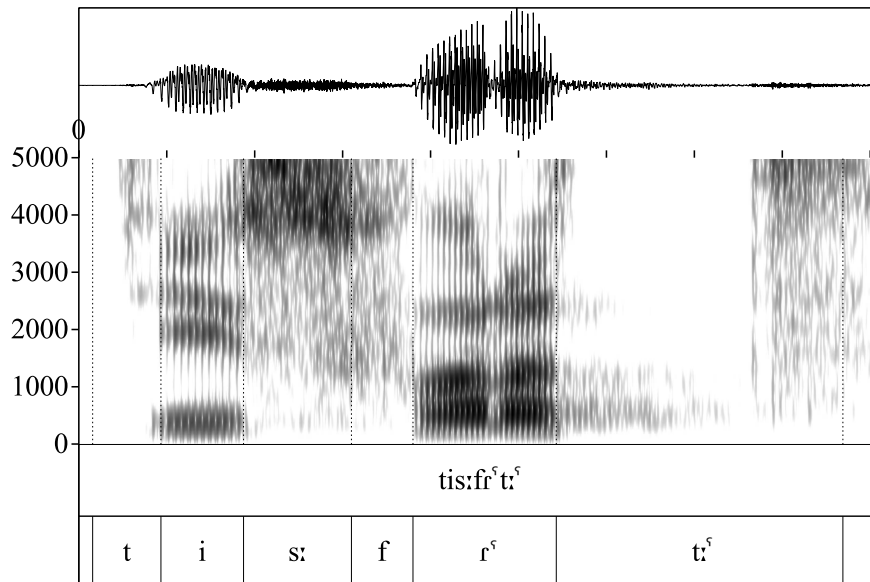


FIGURE III.4.35 : Spectrogramme de @[tis:fr^ʰt^ʰ]@\ balai\

III.4.2.2.1.2 Discussion

L'observation des données nous amène à formuler un certain nombre de constatations. D'un point de vue distributionnel, ces vocoïdes apparaissent en n'importe quelle position lexicale. De même pour les positions syllabiques, même si nous retrouvons dans les items illustrés une surreprésentation de la ||r|| en position noyau. Cela s'explique par le fait que nous avons volontairement choisi de réduire nos observations aux ||r|| n'étant pas adjacentes à des unités phoniques voisées. Cette limitation a pour conséquence d'exclure d'emblée les voyelles et donc c'est la ||r|| qui remplit le rôle de noyau en lieu et place de la voyelle. Mais des exemples, où la présence du vocoïde est attestée dans les autres positions syllabiques, ne manquent pas. Il suffit de revenir sur les spectrogrammes affichés dans les parties précédentes pour s'en rendre compte.

Cependant, il existera une différence notable entre les autres positions syllabiques et la position noyau concernant la manifestation du vocoïde. Si la ||r|| est en position coda ou attaque, nous n'observerons une nette apparition du vocoïde qu'après ou avant

la réalisation de l'obstruction. Tout dépendra si la $\|r\|$ est suivie ou précédée par une consonne. Si par exemple, $\|r\|$ est en position coda et qu'elle est suivie par une autre consonne, alors le vocoïde sera clairement visible sur le signal. Mais seulement entre la fin de l'obstruction et le début de la réalisation de la consonne adjacente. Alors que dans les spectrogrammes que nous venons d'afficher, nous apercevons clairement que l'obstruction est précédée et suivie par des vocoïdes. Il y a un double vocoïde en quelque sorte.

Cette paire de vocoïde apparaît systématiquement lorsque $\|r\|$ n'est pas à proximité de voyelles. Nous l'observons lorsqu'elle est insérée entre deux consonnes, mais également à l'initiale et suivie par une consonne. Et, chose encore plus importante, elle apparaît aussi quand ces consonnes sont sourdes. Ce point est capital. Car, si ces consonnes étaient uniquement voisées, l'apparition de ces vocoïdes serait expliquée par un phénomène de coarticulation. C'est d'ailleurs comme cela que la majorité des chercheurs l'expliquent. Pour les tenants de cette hypothèse, la vibration des plis vocaux ne se serait pas arrêtée ou aurait commencé en avance. Donc, c'est ce qui expliquerait cette présence de voisement et par la même occasion de ces vocoïdes. Même si cette explication est généralement avancée dans le cas où la $\|r\|$ est comprise entre une voyelle et une consonne, nous pouvons toujours reprendre cette même hypothèse pour les cas où $\|r\|$ est précédée et suivie seulement par des consonnes voisées.

Le fait par conséquent de relever des vocoïdes dans un environnement sans consonnes voisées permet de remettre en question d'une part cette hypothèse de chevauchement articulatoire. Car, de toute évidence, la vibration des plis vocaux se produit et se termine dans l'espace de production de $\|r\|$. Avant ou après, les données montrent incontestablement que les plis vocaux ne sont pas en activité, puisque nous n'observons pas de barre de voisement sur les spectrogrammes.

Cependant, il est toujours possible d'arguer qu'en réalité ces vocoïdes sont tout bonnement des voyelles qui seraient en l'occurrence des schwas. C'est d'ailleurs ainsi que la tradition berbérissante les considère. Puisque les items que nous avons reproduits en illustration sont systématiquement accompagnés d'un schwa transcrit soit comme ceci "e" ou comme cela "ə" comme dans cet exemple : @is:"ə/e"rkus@ \mentir{3.M.SG.AI}\. Si nous considérons cette hypothèse, alors au sein de cette paire de vocoïdes, l'un correspondrait à ce schwa et l'autre ne serait qu'une extension de ce schwa. Le deuxième vocoïde serait donc à interpréter comme étant là encore la cause d'un chevauchement articulatoire.

Cette interprétation pourrait être défendable pour les ||r|| en médiane, mais nous ne pouvons pas l'appliquer pour les ||r|| à l'initial. Si la ||r|| est à l'initial, il va de soi que nous ne pouvons pas attribuer la présence du vocoïde à un son voisé, puisqu'il n'y en a pas, ni avant ni après sur les exemples que nous avons fournis. Ni ne pourrait être attribué à la présence d'un schwa. Cette affirmation se base sur plusieurs observations. D'abord en nous basant sur des considérations non linguistiques, c'est-à-dire en prenant en compte les représentations que les berbères — linguistes ou non — ou des chercheurs de toutes origines qui ont travaillé sur le berbère ont sur cette unité phonique. Pour ainsi dire, il est très rare de trouver dans la littérature scientifique ou dans des ouvrages littéraires de transcription où la ||r|| est accompagnée par un schwa en position noyau et à l'initial. Pour le cas exclusivement du TmAS, André Basset transcrit ces ||r|| à l'initial sans schwa.

Si nous analysons cette question sous l'angle de la linguistique, le premier argument que nous fournirons est que si le vocoïde est en réalité un schwa — donc une unité sonore extérieure à ||r|| —, alors nous devrions également retrouver un schwa pour toutes unités phoniques présentes à l'initial et étant le noyau de la syllabe. Nous n'avons pour l'instant jamais constaté ce fait. Nous avons dans notre corpus le spectrogramme de @[lʕʕʕrʕ]@\voisin\ où [lʕʕ] forme une syllabe et est le noyau de cette syllabe, mais cette dernière n'est pas précédée ni suivie par un vocoïde. La rhotique alvéolaire est la seule unité phonique à être dans cette situation. Donc, il apparaît difficile de considérer ces vocoïdes comme étant une unité sonore à part entière.

Un autre argument qui irait dans le sens que nous défendons est de comparer les formants de ces vocoïdes avec ce qui est dénommé dans les études berbères des schwas¹². Si nous prenons l'exemple de @[rʕəm]@\mouiller{2.SG.IMP}\ (figure III.4.30 page 284) où figurent en même temps des vocoïdes et un schwa, nous pouvons constater, bien que les formants soient quasiment identiques, une différence quant à l'énergie dégagée par leur spectre respectif. Les formants du schwa, transcrit par le symbole ə, ont une énergie beaucoup plus intense comparés à ceux des vocoïdes. Cette différence d'énergie est un des critères qui distinguent les sonnantes des voyelles. Ce qui est, par conséquent, un indice quant à la nature différente de ces deux unités sonores et que nous ne pouvons considérer cette paire de vocoïde comme un schwa.

Ensuite, le deuxième enseignement que nous pouvons tirer est que si ces vocoïdes ne sont pas la conséquence d'une inertie articulatoire ni un schwa, alors ils sont inhérents aux ||r||. La vibration des plis vocaux commence bien avant la réalisation de l'obstruction

12. Nous reprendrons cette terminologie en dépit que cette unité phonique n'est en rien un schwa

et se poursuit après. Et les vocoïdes sont la manifestation de ce voisement qui a cours en dehors du mouvement articuloire de la première articulation. En conséquence, l'articulation qui probablement configure les propriétés acoustiques de ces vocoïdes est à chercher dans l'articulation secondaire réalisée par la zone dorsale de la langue. Le flux d'air modulé par la vibration des plis vocaux se trouve filtré par la configuration articuloire réalisée par l'articulation secondaire et c'est ce qui donne ces vocoïdes.

En résumé, il serait plus logique d'admettre que ces vocoïdes sont une partie intégrante de $\|r\|$. Tous les indices que nous fournit notre corpus penchent en cette direction. Maintenant, nous allons voir une autre particularité de ces vocoïdes, cette particularité en question est la durée des vocoïdes.

III.4.2.2.2 Durée des vocoïdes

Les vocoïdes peuvent être caractérisées par plusieurs propriétés physiques. Nous avons choisi de mener une étude sur l'une d'elles. Nous avons cherché à déterminer quel pouvait être le temps de réalisation de ces vocoïdes. Mais pour des questions de faisabilité, nous avons décidé de restreindre notre champ d'investigation. Cette restriction est aussi motivée par des questions d'effectif. Nous n'avons pas assez d'observations pour chaque variable. Donc, pour toutes ces raisons, nous ne prendrons pas en compte toutes les positions d'apparition des vocoïdes.

Les vocoïdes que nous avons choisi d'étudier correspondent à ceux que nous voyons apparaître avant une consonne. D'autre part, nous avons pris en compte aussi la position syllabique de $\|r\|$. Les positions syllabiques auxquelles nous nous sommes intéressées sont celles où la tap est en position noyau. Ainsi, nous avons distingué les configurations suivantes : $Cr^{\text{a}}\cdot C$, $r^{\text{a}}C$, $Cr^{\text{sa}}\cdot C$, $r^{\text{sa}}C$.

Ces quatre configurations mettent en jeu deux problématiques. De plus, nos configurations sont la combinaison de deux paramètres. Il y a une première problématique basée sur l'un de ces paramètres. Celui-ci concerne les types de syllabes. L'une est une syllabe ouverte $C\|r\|^{\text{a}}\cdot C$ et l'autre est une syllabe fermée $\|r\|^{\text{a}}C$. Donc, ceci constituera le premier point de notre étude. Cela nous permettra d'analyser l'influence de la syllabification sur le vocoïde. Puis, la deuxième problématique est basée sur le type de $\|r\|$. Nous avons discerné seulement les taps simples et les taps pharyngalisés. Le but étant de constater si ce paramètre a un impact sur la durée des vocoïdes.

III.4.2.2.1 L'analyse statistique

Les calculs menés sur les durées mesurées ont été appliqués sur nos observations. Les effectifs pour chacune de nos variables ne s'équivalent pas. Pour toutes nos variables, nous n'avons pas un nombre d'observations identiques et pour certaines variables les observations collectées sont assez réduites. Mais ces variables que nous avons choisies comparées aux autres variables laissées de côté sont relativement plus importantes. Les autres variables ne présentent qu'une à trois observations, donc c'est ce qui nous a poussé à ne pas les intégrer et à retenir seulement nos quatre configurations. Nous avons résumé sous forme de tableau les effectifs de nos variables : $Cr^{\text{a}}\cdot C$ ¹³, $r^{\text{a}}C$ ¹⁴, $Cr^{\text{sa}}\cdot C$ ¹⁵, $r^{\text{sa}}C$ ¹⁶.

Variable	Effectif
$Cr^{\text{a}}\cdot C$	17
$r^{\text{a}}C$	8
$Cr^{\text{sa}}\cdot C$	15
$r^{\text{sa}}C$	8

TABLE III.4.7 : Effectif des variables retenues pour la durée des vocoïdes

Pour pouvoir comparer ces variables, nous nous sommes fondé une nouvelle fois sur une exploration visuelle des données. Nous avons choisi comme représentation statistique la boîte à moustache. Cette dernière a l'avantage de pouvoir également représenter la dispersion en même temps que des valeurs discrètes issues de mesures statistiques.

13. Celui-ci correspond aux items : 3, 27, 61, 67, 72, 74, 77, 85, 88, 99, 107, 135, 173, 178, 163, 195, 202 (voir Appendice III.5, page 298).

14. Celui-ci correspond aux items : 79, 86, 89, 167, 175, 176, 177, 194 (voir Appendice III.5, page 298).

15. Celui-ci correspond aux items : 9, 11, 57, 82, 84, 101, 102, 110, 115, 124, 136, 179, 189, 198, 204 (voir Appendice III.5, page 298).

16. Celui-ci correspond aux items : 59, 64, 106, 112, 134, 148, 170, 190 (voir Appendice III.5, page 298).

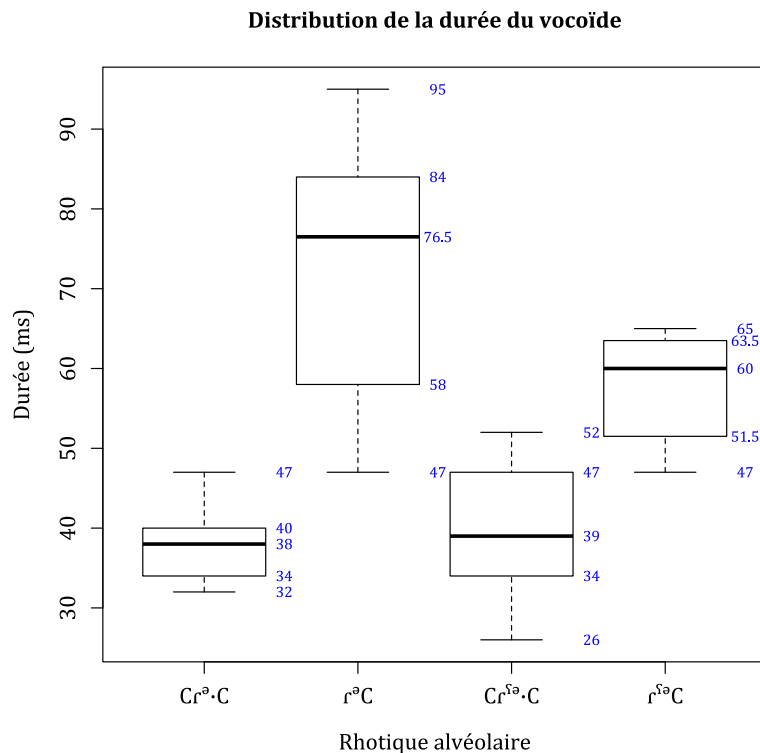


FIGURE III.4.36 : Boîte à moustache de la durée du Vocoïde

Un premier constat ressort est que les vocoïdes en syllabe ouverte ont une durée beaucoup plus courte que ceux en syllabe fermée que cela soit pour les taps simples ou pharyngalisés. Autre observation est que les vocoïdes des taps simples pour une syllabe ouverte ont une distribution plus courte que celles en syllabe fermée. Mais la situation est inverse lorsque nous comparons les taps pharyngalisés. Ce sont les vocoïdes en syllabe ouverte qui ont une distribution très large.

La comparaison des valeurs des médianes montrent qu'entre la tap simple et la tap pharyngalisée en syllabe ouverte les valeurs sont proches, environs 40 ms. Cependant, cela n'est pas le cas pour la syllabe fermée. La tap pharyngalisée présente une médiane moins élevée que la tap simple. Ce décalage est bien sûr le fait d'une trop large distribution.

III.4.2.2.2 Discussion

Les résultats que nous venons de présenter sont à relativiser. Car, ces échantillons n'ont pas la même taille entre eux, puisque le nombre des observations est environ deux fois moins élevé pour les syllabes fermées. D'autre part, un autre fait important à signaler

est que les données que nous avons collectées pour les syllabes fermées doivent être distinguées entre celles qui sont précédées par une voyelle et celles qui sont à l'initial. Dans le premier cas de figure, les vocoïdes ont une durée beaucoup plus longue. Ce cas de figure est principalement rencontré avec les taps simples, donc c'est ce qui explique cette large distributivité. Par conséquent, il aurait mieux fallu séparer à l'intérieur de cette configuration deux types de syllabes fermées : $V \cdot r^{\circ}C$ & $\#r^{\circ}C$. Cela permet de neutraliser le biais introduit par la coarticulation. Ainsi, nous n'avons pas retenu pour $V \cdot r^{\circ}C$ les vocoïdes dépassant les 90 ms¹⁷.

En réalisant cette distinction, l'écart constaté au niveau de la médiane entre les syllabes fermées n'est plus pertinent. Puisque, le cas de figure principalement rencontré pour les taps pharyngalisés dans le cas des syllabes fermées est celui-ci : $\#r^{\text{sa}}C$. Si nous reconsidérons cette nouvelle configuration, alors au niveau des chiffres, cela donne une médiane pour $r^{\circ}C$ de 58 s (effectif : 5) et pour $r^{\text{sa}}C$ de 60 s (effectif : 8). Si tel est le cas, nous pouvons plus facilement interpréter l'effet de la pharyngalisation sur les vocoïdes. Ce qui nous amènera à affirmer que cela soit pour la syllabe ouverte ou pour la syllabe fermée, la pharyngalisation n'a que peu d'influence sur la durée des vocoïdes. Puisque, nous constatons qu'en moyenne leurs durées sont assez semblables. Cette hypothèse, si elle est confirmée par d'autres études, nous aide à déduire quel est l'articulateur qui forge le flux d'air en sortie. Le vocoïde apparaît quand la pointe de la langue est rétractée et n'est plus en activité. Par conséquent, le point d'articulation à l'origine de ce vocoïde n'est pas à chercher dans la partie antérieure de la cavité buccale. Ni par ailleurs dans la partie postérieure, puisque comme nous venons de le voir, la racine de la langue, qui participe à la pharyngalisation, ne change rien à la réalité physique des vocoïdes. Donc, il reste la partie dorsale de la langue qui peut se mouvoir entre la zone vélaire et uvulaire. Ainsi, le point d'articulation possible du vocoïde doit se situer dans cet espace.

Ce point que nous venons d'évoquer traite l'une des problématiques que nous voulions soulever. Il reste donc une problématique qui concerne la syllabification. Cette fois-ci, les résultats montrent qu'il y a une véritable influence de ce paramètre. Les données montrent clairement que de manière générale le temps de réalisation pour une syllabe ouverte est plus court que dans le cas d'une syllabe fermée. Et cette différence est valable pour les taps simples comme pour les taps pharyngalisés. Selon nous, cette différence temporelle n'est pas due à un quelconque phénomène physique, la raison est certainement à chercher au niveau phonologique.

17. Nous avons écarté les items suivant : 175, 176, 177 (voir Appendice III.5, page 298).

Cette question phonologique est d'importance pour déterminer le choix de nos variables et c'est ce qui nous a poussé à exclure certaines configurations. Parmi les syllabes fermées, nous aurions pu aussi intégrer la configuration syllabique suivante : $C||r||^{\circ}C$. Mais nous avons préféré éviter de l'inclure, car la syllabification de ce type de syllabe n'est pas parfaitement établie à l'instant où nous rédigeons cette thèse. Comme nous l'avions précédemment évoqué dans la partie phonologie, il semble que la segmentation syllabique est particulière en **TmAS**. Il y a plusieurs phénomènes qui se superposent et qui sont difficiles à démêler. D'abord, il semble que suivant où la frontière syllabique se trouve le vocoïde peut être plus ou moins long. Par exemple, cette suite d'unités phoniques $C||r||C$ peut être syllabifiée de deux manières : $C||r||\cdot C$ ou $C||r||C$ ¹⁸. Si nous avons à faire au premier cas, il semble que le vocoïde sera plus long entre la première consonne et $||r||$, mais si c'est le deuxième, alors le vocoïde sera plus étendue entre la dernière consonne et $||r||$. Ce résultat laisse à penser que ces vocoïdes servent de noyau à la syllabe. Si tel est le cas, nous sommes en droit de nous poser la question si ce vocoïde n'est pas plutôt un schwa. Bien que nous ayons tranché la question à savoir que ces vocoïdes ne doivent pas être confondus à des schwas. Cela n'empêche pas que les schwas peuvent s'insérer entre une consonne et une $||r||$ comme c'est le cas avec les autres voyelles. L'inspection des spectrogrammes semble indiquer que cette hypothèse est probable. Ainsi, il faut aussi envisager qu'entre la consonne et $||r||$ deux types d'unités sonores voisées qui se ressemblent peuvent se suivre : "e"^o. Donc, nos cas précédents peuvent en réalité avoir quatre configurations au lieu de deux : $C||r||\cdot C$, $C||r||C$, $C^{\circ}e^{\circ}||r||\cdot C$ ou $C||r||^{\circ}e^{\circ}C$.

N'ayant pas pour l'instant tranché cette problématique en déterminant à quelle configuration syllabique se rattachait certains de nos items, nous avons donc remis à plus tard cette question. Cependant, l'enseignement que nous pouvons tirer de cette analyse statistique est l'importance que revêt la syllabification dans la durée de ces vocoïdes. Probablement qu'un des vocoïdes, formant la paire des vocoïdes de la $||r||$, joue un rôle primordial dans la détermination du "centre de gravité" de la syllabe. Que ce vocoïde soit en compagnie ou non d'un schwa. Pour cela seules des études statistiques multivariées pourront confirmer nos hypothèses.

18. La configuration $C\cdot||r||C$ fait partie de la configuration plus générale $||r||C$ qui a été prise en compte dans notre analyse statistique.

III.4.2.3 Synthèse

Cette deuxième partie de notre étude sur les $\|r\|$ arrive à son terme. L'analyse phonétique que nous avons réalisée a abordé nombre de questions. Nous avons dans un premier temps cherché à rendre compte des particularités de la première articulation. Ce travail était subdivisé en deux sous parties. La première sous partie fut de déterminer le mode d'articulation de l'articulation primaire. Cette dernière comme nous avons pu l'observer est dans la quasi majorité des cas une réalisation où l'obstruction n'est pas occlusive. Le flux d'air est plutôt continu et peut être obstrué de deux manières différentes. Soit le mode d'articulation est approximant, soit il est fricatif. Cette frication est principalement visible lorsque la $\|r\|$ est dans les environs d'une fricative. Cependant, les trills semblent moins affectées par cette coarticulation. Les constriction fricatives retrouvées sont plus le fait d'une réalisation aléatoire. Puis, nous avons tenté de déterminer la durée d'obstruction des taps. Nous avons pu constater que le contact se maintenait pour les deux types de tap dans un espace de temps équivalent. Ce qui nous a amené à conclure que l'articulation secondaire n'avait pas réellement d'emprise sur l'articulation primaire.

Puis dans un deuxième temps, nous nous sommes intéressés aux vocoïdes. Nous avons discuté de la nature de ces vocoïdes. Nous avons pu montrer que ces derniers étaient intrinsèques aux $\|r\|$. L'hypothèse de la coarticulation ne pouvait s'appliquer aux données que nous avons. Le chevauchement articulatoire n'a pas de sens dans un environnement où ne figure aucun son voisé à part la rhotique alvéolaire. Puis, ce travail sur les vocoïdes s'est poursuivi avec une analyse statistique sur la durée de réalisation des vocoïdes dans certains contextes. Les résultats ont pu démontrer que la durée des vocoïdes variait en fonction de la syllabation. Deux configurations étaient opposées, quand $\|r\|$ et la consonne suivante formaient une seule et même syllabe et quand ces mêmes unités étaient séparées par une frontière syllabique. Dans le dernier cas, la durée du vocoïde était plus courte, alors que dans l'autre cas la durée était plus longue. Deux autres configurations ont été confrontées qui sont la durée des vocoïdes des taps simples et celle des taps pharyngalisés. Il ressort après analyse qu'une nouvelle fois la pharyngalisation n'influe pas significativement et que cette articulation n'est pas à l'origine du vocoïde.

CHAPITRE III.5

Appendice

— Remarques :

1. La transcription des items a été menée avec l'API.
2. Les items retranscrivent les unités phoniques. Cela implique que toutes les unités sonores particulières ne seront pas retranscrites.
3. La glose proposée est approximative.
4. La forme nue des verbes correspond à la forme de l'impératif.
5. La forme nue des noms correspond à la forme du singulier.
6. L'ordre de présentation est l'ordre d'énonciation des items par les informateurs.
7. Les formes transcrites correspondent à celles élicitées par l'informateur 2.
8. Les données audios de ces items sont disponibles à cette adresse : https://sudoranais.shinyapps.io/Analysis_Processing_Rhotic_Alveolar/.

Item	Glose	Item	Glose		
1	ɜbred	route	29	ʌmor	part
2	tɜʁret	bâton	30	tɜsert	moulin_à_main
3	ɜbrʃɜn	noir	31	tisirʌ	moulins_à_main
4	ʌzʃrʃɔ	sorte_de_colline	32	erʃɔ	lire{3.M.SG.PRET}
5	ʌdʃrʃʌrʃ	montagne	33	eq:ʌrʃ	lire{3.M.SG.IMP_INT}
6	edorʃʌrʃ	montagnes	34	f:r	caler{2.SG.IMP}
7	tinzɜrt	nez	35	izʃo	voir{3.M.SG.PRET}
8	tenzɜr	nez	36	tʌʒrʃʌ	réipient
9	ʌbrʃdʃʌl	oiseau	37	ʌwrʃʌʁ	jaune
10	tmɜrt	barbe	38	ɜʒres	froid
11	ʌqrʃdʃʌl	grand	39	ʌzʃq:ɔrʃ	poutre
12	dziri	claire_de_lune	40	izʃʁʃʌn	poutres
13	jor	lune	41	dʃʌrʃ	ped
14	jerʃo	jeter{3.M.SG.PRET}	42	jg:or	marcher{3.M.SG.IMP_INT}
15	tɜrʃo	jeter{3.F.SG.PRET}	43	ʌrʃmʃʌs	sorte_d_arroche
16	ʌrʃɔ	accoucher	44	ɜʁzʃrʃ	rivière
17	ire	cou	45	ʌʁʃɔm	pain
18	tʌʁʃɔt:	épaule	46	k:ɜr	se_lever{2.SG.IMP}
19	ɜre	écrire{2.SG.IMP}	47	ʌrʌ	enfants
20	tɜrɜjt	écriture	48	tbrorɜ	grêlon
21	idmɜrɜn	poitrine	49	ʌzʃɜrʃʌrʃ	long
22	ij:ɜr	ongle	50	rʃl	marier{2.SG.IMP}
23	orʃʌn	poignée	51	tɜjmart	jument
24	erʃʌn	jardin_potager	52	def:r	derrière{PREP}
25	tʌrʃdʃɔnt	sorte_de_pain	53	ħɔnjr	avoir_froid{2.SG.IMP}
26	ʃ:wʌrɜ	sacoche	54	ʌmʒʌrʃ	le_fait_de_moissonner
27	brdɜʃ:n	aurore	55	ʌmʒrʃ	faucille
28	iʁɔrʃ	tromper{3.M.SG.PRET}	56	ʁʌt:er	épine_de_palmier

Item	Glose	Item	Glose
57	sdr ^s d ^s q	explorer{2.SG.IMP}	85 3srzəl serrure
58	ər ^s Λh	aller{3.M.SG.PRET}	86 3rjte pâte
59	r ^s z ^s m ^s Λ	ouvrir{1.SG.PRET}	87 t3rest argile_blanche
60	jor ^s Λr ^s	jouer{3.M.SG.PRET}	88 Λbrsew reste_de_fils
61	is:rkos	mentir{3.M.SG.IMP_INT}	89 tisfrst tibia_d'un_animal
62	et:ʀrq	affubler{3.M.SG.IMP_INT}	90 onqər sorte_de_datte
63	r ^s ɔlΛ	fuir{1.SG.PRET}	91 Λqrɔh tête
64	r ^s g:w ^s lΛ	fuir{1.SG.IMP_INT}	92 t3ft:ret fusaïole
65	jr ^s wΛl	fuir{3.M.SG.PRET}	93 orfem sorte_de_datte
66	jəmr ^s Λ	être_difficile{3.M.SG.PRET}	94 r ^s z ^s Λ casser{1.SG.PRET}
67	jr ^s Λ	mâchouiller{3.M.SG.PRET}	95 Λws3r vieux
68	əʀdΛr	choisir{3.M.SG.PRET}	96 rəz:εʁ chercher{1.SG.PRET}
69	əʀt:r ^s	choisir{3.M.SG.IMP_INT}	97 fr ^s d ^s balayer{2.SG.IMP}
70	iqr ^s d ^s	fracturer{3.M.SG.PRET}	98 tΛms ^s r ^s et salon
71	iqrd	s'accroupir{3.M.SG.PRET}	99 3hrd3n uromastyx
72	ΛbrdΛ	sable	100 3tbir pigeon
73	em: ^s or ^s d ^s s	étrangler{3.M.SG.PRET}	101 Λhr ^s d ^s Λn esclave_affranchie
74	jrkd ^s	danser{3.M.SG.PRET}	102 t3jr ^s z ^s ezt lièvre
75	ʁrs	égorger{2.SG.IMP}	103 Λs ^s :jr ^s sorte_de_plante
76	ʁr-s	chez{PREP}-lui{3.M.SG.OI}	104 3ʁr ^s 3s
77	ʁr-sən	chez{PREP}-eux{3.M.PL.OI}	105 ΛʁΛr3s sorte_de_datte
78	ʁr-e	chez{PREP}-moi{1.SG.OI}	106 r ^s b:e dieu
79	is:erd	laver{3.M.SG.PRET}	107 frfr s'envoler{2.SG.IMP}
80	edɔr	tourner{3.M.SG.PRET}	108 ifriwn ailes
81	ʃrəʒ	semmer{2.SG.IMP}	109 s:r ^s emət bride_de_décor_du_cheval
82	zr ^s wΛd ^s	jeter{2.SG.IMP}	110 Λqr ^s bɔs appareil_du_cheval
83	et:Rɔr ^s Λ	tromper{3.M.SG.AI}	111 3mrez bosse_crânienne
84	tΛqr ^s neft	pétiole_de_palmier	112 r ^s s ^s dΛ puer{1.SG.PRET}

Item	Glose	Item	Glose		
113	emer ^ʃ ʌn	faucilles	140	tɔmort	pays
114	tisøz:ɜr	fourches	141	jeq:ɔr	sécher{3.M.SG.PRET}
115	ʌsr ^ʃ wʌt	dépiquage	142	ɜzrer	grappe_de_raisin
116	ʌr ^ʃ nʌn	aire_de_piquage	143	jrɔ	rendre{3.M.SG.PRET}
117	tifokʌr	sacs	144	fr ^ʃ ʌns ^ʃ ʌ	France
118	ɜridɜl	boiteux	145	tɜrb:qt	chausette
119	dre	pousser_violemment {2.SG.IMP}	146	id:r	vivre{3.M.SG.PRET}
120	ʌd ^ʃ r ^ʃ	appuyer{2.SG.IMP}	147	tiz:ɜrnen	prière_musulmane_de_midi
121	ɜd:ɜr	baisser{2.SG.IMP}	148	jor ^ʃ f	torréfier{3.M.SG.PRET}
122	iqr ^ʃ d ^ʃ	fracturer{3.M.SG.AI}	149	ɜmzɜwɜrɔ	premier
123	ʌfr ^ʃ	feuille	150	ʌmg:ʌr ^ʃ ɔ	dernier
124	ɜdr ^ʃ ɜʌl	aveugle	151	ɔɜrefən	glace
125	r ^ʃ :d	arrêter{2.SG.IMP}	152	l-ɜɜɔr-en	là_bas{PREP-ADV-DEM_DIST}
126	ɜzøfre	célibataire	153	ʌwr ^ʃ -ɔ	ici{ADV-DEM_PROX}
127	erdən	blés	154	ɜtrɜs	homme_brave
128	təjdrt	épi_de_blé	155	ɜrfət	crible
129	ɜ ^ʃ ʌr ^ʃ	entre{PREP}	156	ʌfr ^ʃ ʌj	enclos
130	ɜdrɜz	tisserand	157	izmr	agneau
131	ʌjr ^ʃ ʌd	lion	158	ofref	mouton
132	frɜ	verser{2.SG.IMP}	159	ʌz ^ʃ wr ^ʃ	racine
133	tʌr ^ʃ t ^ʃ :ʌ	branche_de_palmier	160	tɜɜmrt	coin
134	tər ^ʃ dwen	branches_de_palmier	161	tɜrdəst	empan
135	ɜrɜbef	école_coranique	162	ɜj ^ʃ :er	enfant
136	ʌmr ^ʃ wʌs	dette	163	tɜsr ^ʃ ɔft	groupe_d'ovine
137	tʌnw:ʌr ^ʃ t	baratte_en_peau	164	ɜr ^ʃ d	pleurer{2.SG.IMP}
138	ɜrgɜz	homme	165	frerəs	être_irrité{2.SG.IMP}
139	ʌɜr ^ʃ m	ville_fortifiée	166	idre	pousser_violemment {3.M.SG.AI}

Item	Glose	Item	Glose		
167	ɾʃmɔ	mouiller{1.SG.PRET}	194	ɛʁɔt:ɛrn	épinés_de_palmier
168	ɛʒənʒɔɾʃ	bleu_foncé	195	ʒhɾk:os	chaussure
169	jemɔɾʃ	être_grand{3.M.SG.PRET}	196	itre	étoile
170	jɔɾʃf	geler{3.M.SG.PRET}	197	ɔʒʃɾʃəq	pêt
171	ʒʃɾʃ	serrer{2.SG.IMP}	198	ɔχɾʃbɛʃ	griffure
172	izwɛɾ	devancer{3.M.SG.PRET}	199	ɔʒɾʃɔdʃ	sorte_de_citrouille
173	wɾqt	feuille	200	egorʃ	rôter{3.M.SG.PRET}
174	idʃɾʃɔ	allumer{3.M.SG.PRET}	201	tɔɾʃɔt	accouchement
175	jors	être_créditeur_de{3.M.SG.PRET}	202	z:ɾdət	repas_copieux
176	ɛd-jɛɾs	POT-être_créditeur_de{3.M.SG.A}	203	z:ɾʃtʃ	tempête
177	it:ɛɾs	être_créditeur_de{3.M.SG.AI}	204	teqɾʃɔt	bouteille
178	jɾsɔ	être_posé{3.M.SG.PRET}	205	zʃɾʃdʃ	trier{2.SG.IMP}
179	jɾʃsʃɔ	être_calme{3.M.SG.PRET}	206	tɛɾʃɛn	braises
180	ɾʃsʃn	être_calme{3.M.PL.PRET}	207	tɛɾʃɛt	braise
181	id:əre	être_courvert{3.M.SG.PRET}	208	tɛɾgɔ	rigole
182	ɔfkɾʃɔn	tortue	209	ʃɾʃ	nouer{2.SG.IMP}
183	ɔʒɾʃɔn	crapeau			
184	tɔjʃɾɔɾʃt	poulie			
185	lɛʃɔɾʃ	voisin			
186	edre	pousser_violemment{3.M.SG.PRET}			
187	enrz	talon			
188	dɾʃɔs	peu{adv}			
189	ɔfrʃdʃɔs	chauve			
190	ɔɾʃn	farine			
191	ɔfrʃɔχ	oisillon			
192	tət:ɾɔ	mendicité			
193	et:ɾ	demander{3.M.SG.PRET}			

Conclusion générale

Nous arrivons au bout de notre étude. Notre description des variétés berbères du Sud-Oranais se conclut sur cette esquisse. Ce dernier mot est à propos compte tenu du cadre très limité de notre recherche. Peu de faits linguistiques ont été pris en compte, mais ils ont été analysés selon différentes approches. De plus, aucun travail académique d'envergure n'avait été mené sur ces variétés berbères. Que cela soit les travaux d'André Basset ou même ceux de René Basset les données qui ont été publiées étaient clairsemées. Ainsi, notre thèse vient combler un manque à ce niveau-là. De cette manière, notre étude a poursuivi deux directions de recherche complémentaires. Une première approche qui intègre la variation régionale et une autre qui intègre la phonétique expérimentale. Ces aspects ont été essentiels pour rendre compte de la réalité phonologique de certaines consonnes. Effectivement, ces dernières ont été les seuls faits linguistiques que nous avons retenus dans cette thèse. Toutefois, un regard plus attentif sur notre corpus peut aussi donner une idée sur d'autres particularités linguistiques que nous avons abordées de manière éparse. Cela étant dit, nous allons faire le bilan des résultats de notre recherche.

Notre thèse a visé dans un premier temps à dresser l'état de la situation du berbère dans cette région. Depuis fort longtemps, peu de choses nous parvenaient du Sud-Oranais. Par nous-même, nous avons pu nous rendre dans chacune des localités berbérophones connues comme telles. Les propos rapportés ici et là sur la disparition du berbère sont, hélas, exacts. Nous n'avons recensé que trois lieux, sur seize, où le berbère était encore présent auprès de toutes les générations. Ces villes correspondent à Asla, Boussemghoun et Ouakda. Cependant, dans ces deux dernières localités, le berbère est bien mieux implanté qu'à Asla. Quant aux autres oasis, il y a Igli qui est dans une situation intermédiaire. Chez eux, la non transmission généralisée du berbère est plus récente. Pour nos autres points d'enquête, les personnes parlant le berbère correspondent quasiment à la classe d'âge la plus âgée. Chez ces dernières, nous estimons que les berbérophones ne dépassent pas le nombre des dizaines. Donc, si rien n'est fait, dans les décennies qui viennent, il ne restera plus un seul(e) locuteur(rice) du berbère dans ces villes.

Il est certain que l'étude de la vitalité du berbère mériterait davantage d'attention. Les problématiques qui peuvent être étudiées sont nombreuses. Comme, pourquoi à Boussemghoun le berbère se maintient et non pas à Beni Ounif? Du fait que notre thèse s'inscrit dans le champ de la linguistique, alors nous ne pouvons nous y pencher plus.

Car ce travail est davantage en lien avec la sociologie du langage. Toujours est-il que nos observations sont un témoignage précieux dans une perspective de réhabilitation. Ce point n'est pas à l'ordre du jour, néanmoins, des signes favorisant cette possibilité s'entrevoient. Car, lorsque nous avons débuté notre doctorat fin 2014, le berbère en Algérie était considéré comme une langue nationale, mais les droits culturels sur le terrain étaient presque inexistantes. Ceci a quelque peu changé avec l'introduction de l'enseignement du "berbère". Donc, les conditions politiques sont plus favorables pour entreprendre une action dans ce sens.

Après cette vue d'ensemble sur la situation du berbère dans cette région, nous avons étudié la variation linguistique dans cette région. Cette étude a incorporé des méthodes utilisées dans d'autres disciplines. Les SIG et les SDD ont été des outils sur lesquels nous nous sommes appuyé. Avec la première de ces méthodes, nous avons réalisé une étude géolinguistique. Nous avons pu voir à quel point il existait des différences phonétiques entre ces variétés. La distribution géographique de ces particularités n'était guère plus stable. À chaque variable linguistique les isoglosses changent. Cela est tout ce qu'il y a de plus normal. C'est un phénomène universel et le berbère du Sud-Oranais n'y échappe pas. En revanche, il existe d'autres constats qui le sont moins.

Ainsi, des phénomènes de rupture du continuum linguistique ont été constatés et également des réalisations de faits linguistiques qui ne s'inscrivaient pas dans un continuum linguistique. Ils n'étaient que le propre fait d'un point d'enquête. Ces unités phoniques dont nous avons étudié la variation correspondent à des consonnes simples et à des consonnes géminées. Les particularités phonétiques à l'œuvre entre ces consonnes ne sont pas les mêmes. Les consonnes simples se caractérisent par une distinction entre les variétés occlusives et celles fricatives. Par contre, nous avons pu voir que certaines variétés pouvaient être occlusives pour certaines variables et fricatives pour d'autres. Alors que d'autres variétés étaient pour ces mêmes variables soit totalement fricatives, soit totalement occlusives. Par ailleurs, nous nous sommes rendu compte qu'ils partageaient le même inventaire phonétique. Rien ne les distinguait de ce point de vue-là. Mais la différence se trouvait au niveau de la distribution et de la fréquence d'emplois de ces unités phoniques.

Quant aux consonnes géminées, nous avons fait état de deux phénomènes différents. Le premier concerne principalement la distinction entre les variétés fricatives et les variétés affriquées. Il est apparu que certaines localités pouvaient utiliser les deux unités phoniques, alors que d'autres ne connaissaient que la fricative. Ensuite, nous

avons présenté la variation de variables dont l'assourdissement et la sonorisation constituaient le point de divergence entre certaines variétés.

Cette somme de faits nous a amené à nous interroger sur la réalité cognitive de ces unités phoniques. Alors que pour une variété, par exemple, le passage du singulier au pluriel se fait sans aucune modification phonétique — donc le morphème reste le même —, pour d'autres une évolution morphologique s'opère. Ce fait nous oblige à reconsidérer la représentation phonologique de ces unités phoniques. Toutefois, cela concerne des unités phoniques dans des contextes bien déterminés et qui ont été neutralisées. D'autre part, nous avons questionné le rapport entre les consonnes simples et les consonnes géminées. Dans certaines variétés, les simples, au niveau des surfaces, peuvent être rattachées à deux formes de géminées. Pour une même variété, la simple peut se distinguer de la géminée seulement au niveau quantitatif. De plus, elle peut aussi connaître une différence de qualité par rapport à leur contrepartie simple. Cependant, si nous examinons cette situation au niveau des représentations, il apparaît qu'il n'est pas évident de systématiquement aligner les représentations au niveau cognitif. Car, si aucun contexte particulier ne se dégage pour expliquer cette bifurcation, alors il est plus raisonnable de distinguer les représentations de ces deux géminées.

L'autre aspect ayant trait à ces géminées concerne le principe d'intégrité. Nos données ont pu mettre en évidence que ce dernier ne pouvait s'appliquer à certaines variétés du Sud-Oranais. Effectivement, ces géminées affriquées dans un contexte bien particulier voient s'insérer en leur sein un vocoïde. Par ailleurs, il en est de même du principe d'inaltérabilité qui semble dans ses fondements méthodologiques discutable. Car, comme nous l'avons vu en cherchant à appliquer ces hypothèses, il nous a été impossible de réconcilier des formes de surface avec des représentations sous-jacentes.

Par la suite, une étude dialectométrique a fait suite à notre étude géolinguistique. Nous avons entrepris d'explorer plusieurs méthodes pour jauger leur potentiel de partitions. Nous avons d'abord eu recours aux méthodes de la méthode d'Apprentissage Non Supervisé. Nous avons employé des méthodes qui avaient déjà été utilisées, mais également de nouvelles méthodes. Nous avons pu constater que les résultats fournis par ces méthodes ne concordaient pas. D'ailleurs, certaines d'entre elles étaient très influencées par leurs conditions initiales — nombre de partitions, choix des centres des partitions —. En fonction de ces derniers, les résultats obtenus n'étaient jamais identiques. Ils ne tendaient pas vers une solution unique.

Puis dans un second temps, nous avons utilisé la méthode d'Apprentissage Supervisé.

Ce travail exploratoire a cherché à évaluer l'utilisation de cette méthode dans le cadre de la dialectologie. Pour cela, nous avons utilisé la méthode CART et Random Forest. Les résultats auxquels nous avons abouti n'étaient pas probants au départ. Les conditions initiales — nombre de variables, nombre d'observations différentes — sont aussi très importantes et déterminent la qualité des résultats. Malgré cela, nous avons pu dégager trois haplolectes homogènes.

La deuxième étape de notre thèse a été de réaliser une étude de phonétique expérimentale sur les rhotiques alvéolaires. Cette étude a été conduite sur une variété du Sud-Oranais. En TmAS, elles sont au nombre de quatre : [r], [r̥], [r̥^h] et [r̥^h]. Nous avons entrepris d'étudier quelques particularités touchant à ces sons. Notre recherche a porté sur les deux facettes des ondes acoustiques que sont la phonologie et la phonétique. Nous avons entamé notre étude par une analyse phonologique et nous avons poursuivi avec une analyse phonétique.

Le premier volet de notre étude phonologique fut d'examiner la distribution des ||r||. Pour rendre compte de cette caractéristique, nous avons pris en considération les positions lexicales et syllabiques. L'objectif était de déterminer en quelles positions ces unités phoniques n'étaient pas distinctives entre elles. Ainsi, notre étude phonologique a permis de mettre en évidence que les ||r|| avaient une distribution large et que les différentes ||r|| étaient partout en contraste sauf dans une position, la position finale.

À la suite de ce travail, nous avons cherché à distinguer les formes sous-jacentes des formes de surfaces. Nous avons mis en évidence que les représentations phonologiques s'écartaient des réalisations phonétiques. Les formes sous-jacentes ont des traits phonologiques qui ne correspondent pas exactement aux traits physiques. L'analyse de notre corpus ne nous a pas permis de constater que les taps pharyngalisés étaient des allophones de taps simples. Bien qu'il existe des taps simples dans l'environnement de consonnes pharyngalisées, il s'est avéré impossible de considérer que tous les taps pharyngalisés sont le produit d'une distribution contextuelle. Certains de nos items ne comportent pas de consonnes pharyngalisées excepté des taps pharyngalisés. Ces données poussent à assigner aux taps pharyngalisés une représentation phonologique qui est indéterminée. Donc, nous avons considéré la forme sous-jacente de [r̥^h] comme un archiphonème. Car, il n'est pas certain que même dans l'environnement d'une consonne pharyngalisée, les taps pharyngalisés soient la représentation d'une tap simple /r/.

Puis, l'opposition de nos items entre eux nous a permis de dégager qu'une partie des trills était les formes de surface de taps géminés. Les taps comme les autres consonnes

du berbère connaissent un processus morphologique de gémination. Ce processus morphologique produit en final une trill. La réalisation articulatoire n'est plus un seul battement, mais un enchaînement de plusieurs battements. Bien que le mouvement articulatoire soit différent, nous ne pouvons interpréter la réalisation de cette trill que comme étant la forme de surface d'une représentation d'une tap géminée.

Ce dernier point est clairement établi, ainsi il ressort de nos analyses que les trills sont aussi une réalisation phonétique du phonème /r/. C'est en tout cas l'hypothèse que nous avançons. Ce fait survient en final d'un lexème où il n'y figure que des trills. Nous avons apporté un certain nombre d'arguments pour étayer notre hypothèse. Il y a deux points principaux qui concourent à ne pas considérer les trills finales comme des taps géminées. Le premier est qu'il est peu probable de retrouver un processus morphologique de gémination en final d'une trilitère et pourtant nous trouvons des trills en cette position. Le deuxième est que toutes les unités phoniques en final ont un temps de réalisation plus long que la moyenne. Ces deux aspects suffisent à faire la démonstration qu'il ne peut s'agir d'une tap géminée. Donc, par déduction, il est plus envisageable de considérer cette unité phonique comme une tap simple. Cependant, cette hypothèse demande à être mieux consolidée surtout d'un point de vue théorique en expliquant pour quelle raison les unités phoniques en final sont plus longues.

Cette description n'est qu'un premier pas, il reste encore bien à faire. Plusieurs de nos hypothèses ont besoin d'être consolidées par d'autres travaux. Ainsi, dans le domaine géolinguistique, nous pouvons envisager d'élargir notre terrain d'étude à d'autres variétés notamment celles du côté marocain. À très long terme, un Atlas linguistique de cette région peut être envisagé. Les variables prises en compte seraient bien évidemment plus larges.

La dialectométrie est un autre champ d'investigation que nous voudrions aussi développer. Actuellement, l'utilisation des mathématiques en linguistique se fait de manière intuitive. Nous réutilisons des outils mathématiques qui sont développés dans d'autres champs disciplinaires. Peut-être faudrait-il réfléchir à faire émerger des outils tirés à partir de données linguistiques? Mais dans un premier temps, notre souhait est de poursuivre notre travail sur les méthodes de partition, notamment les méthodes d'Apprentissage Supervisé. Ainsi, ce travail peut être un point de départ pour l'établissement d'une deep learning linguistique, c'est à cela que pourrait servir les méthodes d'Apprentissage Supervisé.

Notre ambition par ailleurs est de continuer à réaliser des travaux de phonétique

expérimentale. Nous pourrions recourir à d'autres matériels d'observation. Les études acoustiques ne peuvent pas être la seule voie pour comprendre la réalité physique du langage. Étant bien entendu que cela dépendra fortement des opportunités sur place en Algérie. De plus, les évolutions technologiques pourront faciliter cette tâche. Le développement d'appareils d'imagerie pour des examens non intrusifs, répondant aux contraintes vécues sur le terrain, aura un rôle prépondérant.

Nous comptons également poursuivre nos réflexions sur la réalité cognitive du langage. Nous avons vu qu'un tout petit aspect de cette réalité à travers les représentations phonologiques. Mais le langage ne s'arrête pas à cette dimension. Ce dernier est une entité très complexe. Pour rendre compte de son fonctionnement, cela implique inévitablement une analyse non linéaire ou modulaire des systèmes linguistiques — la notion de modularité renvoie à la structuration cognitive qui serait à la base du fonctionnement cérébral [Scheer, 2011, p.497] —.

Cependant, cela n'est pas suffisant, il faut aussi déterminer une méthodologie capable de décrire un système phonologique. La méthode déductive a ses limites comme nous avons pu le voir. Il y a lieu, par conséquent, de renforcer la phonologie expérimentale. Toutefois, celle-ci devrait également prendre pour objet d'observation le cerveau et ne plus se contenter des ondes acoustiques. Cette démarche serait mieux à même de rendre compte de la réalité cognitive. Ainsi, les données du Sud-Oranais peuvent aider à formaliser des hypothèses pour aboutir à une modélisation du langage plus proche de la réalité.

Alphabet Phonétique International L'Alphabet Phonétique International est une convention de transcription créé par l'Association du même nom. Cet alphabet est destiné aux linguistes. Son but est de donner à ces derniers le moyen de décrire la réalité sonore du langage avec précision et de faciliter la communication de ces données auprès de la communauté scientifique.

Site de l'association : www.internationalphoneticassociation.org/.. xxv

Aoriste Terme dénommant l'un des thèmes verbaux existant en berbère. Il correspond au thème aspectuel qui dénote un procès potentiel. 71

Aoriste Intensif Terme dénommant l'un des thèmes verbaux existant en berbère. Il correspond au thème aspectuel qui dénote un procès inaccompli. 71

Formant Le formant est une unité de mesure utilisée en phonétique acoustique. Il est généralement défini comme correspondant aux pics spectraux du spectre sonore. 206

Haplolecte Désigne un groupe d'allolectes. 36

Inaccompli Dans les études berbères, ce terme est l'équivalent de l'Aoriste Intensif. Il est utilisé par certains berbérisants. 260

Le principe d'inaltérabilité Ce principe stipule qu'il existe des phonèmes qui ne sont pas affectés par les processus phonologiques qui touchent d'autres phonèmes. 142, 214

Le principe d'intégrité Ce principe se rapporte au fait qu'il est impossible d'insérer un vocoïde dans une géminée.. 142, 214

Morphème Le plus petit segment significatif d'un système linguistique. Cette unité est le pendant physique de l'unité morphémique (Synonyme : Unité significative). 25

Nucléolecte Désigne une variété linguistique qui parle un dialecte complètement identique. 36, 92, 145, 153

Phonème Cette unité est le pendant cognitif de l'unité phonique. 25

Processus phonologique Les processus phonologiques sont à l'origine des allophones. Un phonème affecté par un processus phonologique aura une forme de surface différente. 140, 144

Processus morphologique Les processus morphologiques servent à former de nouvelles unités morphémiques à partir d'autres unités morphémiques. Cette dérivation est un procédé de morphologie non concaténative. 139, 145

Prétérit Terme dénommant l'un des thèmes verbaux existant en berbère. Il correspond au thème aspectuel qui dénote un procès accompli. 71

Signal Noise Rate Une unité de mesure définissant la qualité d'un enregistrement. Si ce rapport est élevé, l'enregistrement est de bonne qualité. 230

Tmaziyt de l'Atlas Saharien Le tmaziyt de l'Atlas Saharien est une langue berbère parlée dans les wilaya de Naama et d'El Bayedh. Les populations locales nomment leur langue : tmaziyt. 184

Unité phonique Unité sonore idéalisée fondée sur des considérations psychoacoustiques. Cette unité est le pendant physique du phonème. 25

Unité morphémique Cette unité est le pendant cognitif du morphème. 25

A

- Acoustique, 143, 152, 186, 193,
196, 201, 203, 205, 207, 208,
212, 215, 217, 218, 261, 279,
291
- Afrique
 Afrique du nord, 10, 13, 93
 Afrique subsaharienne, 53
- Al Bekri, 9
- Al Idrissi, 9
- Algérie, 4, 5, 7, 22, 23, 45, 80, 154,
211, 305, 309
- Almohades, 10
- Almoravides, 10
- Apprentissage Non Supervisé
- Corrélation de Kendall, 156
- Dendogramme, 59, 83, 157, 158, 166
- Distance linguistique, 156
- Heatmap, 59, 156
- Matrice des distances, 156
- PHA, 82, 83, 156
- ROCK, 166–168
- Simple Matching Distance, 156
- Apprentissage Supervisé, 56–58,
82, 154, 170, 173–176
- Bagging, 175
- CART, 83, 170, 172–175
- Conditional Inference Trees, 174
- Cross-validation, 59, 174
- Holdout Method, 59
- K-fold, 59, 174
- Leave-P-Out, 59
- Random Forest, 83, 170, 173–175
- Art rupestre, 8
- Atlas saharien, 5
- Autololes, 8
- Ayerm, 6

B

Basset, André, 21–23, 40–45
 Basset, René, 21, 23, 93
 Bavares, 8
 Bazinas, 9
 Berbère, 7, 40, 42, 63, 89, 290
 Chaoui, 22, 45, 191
 Chelha, 21
 Chleuh, 22, 41, 42, 95, 97, 191, 192,
 204, 212, 214, 216
 Civilisation berbère, 8
 Kabyle, 22, 44, 89, 90, 150, 191, 192
 Langue berbère, 93
 Mozabite, 22
 Ouarsenis, 22
 Rifain, 22, 43, 89, 97, 191, 192, 209
 Siwi, 214, 216
 Tabeldit, 22
 Tacelhit, 22
 Tmaziyt, 22, 184, 186, 188, 221, 231,
 254, 255, 257, 261, 266, 276,
 278, 283, 290, 295
 Touareg, 22, 89, 191
 Zénaga, 22
 Zénète, 22
 Berbérisant, 31, 40, 47, 52, 87, 89,
 93, 147, 150, 192
 Biomécanique, 196
 BULAC, 23, 30
 Bédouin, 6

C

Califats, 10
 Catalan, 214, 216
 Cheikh Bou 'Amâma, 10
 Chine, 53
 Chomsky, 25
 Chélif, 40
 Cognat, 92
 Cognat morphémique, 93, 111, 129,
 179
 Cognat phonologique, 92, 93, 129,
 135, 179
 Colonisation française, 10
 Concaténation, 215
 Consonne
 Affriquée, 119
 Alvéolaire, 90
 Approximante, 90, 195
 Battue, 195, 273
 Bilabial, 89, 92, 194
 Coronal, 23, 91, 108, 111, 117, 125,

- 194, 203, 213
- Flap, 215
- Fricatif, 92, 119, 145, 195
- Gutturale, 90
- Labiale, 91
- labiale, 203
- Labiovélaire, 91, 134
- Latéral, 195, 214
- Liquide, 194, 195, 208, 212
- Nasale, 91, 214
- Occlusif, 90, 91
- Palato-alvéolaire, 92
- Pharyngale, 91, 194, 204
- Rhotique, 193–196
- Rhotique alvéolaire, 186, 193, 221,
234, 254, 258, 264, 266, 283,
285, 290, 296
- Roulée, 195, 273
- Rétroflexe affriquée, 108
- Spirante, 90, 92
- Tap, 197, 198, 205, 212, 214, 217,
231, 235, 243, 254, 256–260,
262, 264, 265, 270, 273, 275,
281, 286, 291, 293, 296
- Tap pharyngalisée, 231, 246, 254,
258, 266, 281, 288, 293
- Trill, 191–193, 198, 199, 201–204,
215, 216, 218, 231, 243, 249,
254, 256, 267, 268, 270, 271,
276, 278, 279, 296
- Trill pharyngalisée, 231, 252, 254,
255, 265, 266
- Uvularisation, 91
- Vélaire, 92
- Vélarisation, 91
- Corpus, 23, 44, 63–65, 69, 81, 92,
110, 186, 220, 226, 229, 231,
233, 243, 254, 258, 264, 266,
273, 280, 283, 285, 290
- Corpus oral, 63, 69
- Corpus écrit, 63
- Liste Chaker, 64
- Liste Swadech, 64
- D**
- De Slane, 9
- Dialectométrie, 39, 51–53, 55–60
- Doutté, Edmond, 14
- E**
- Electromagnétisme, 196
- Erg occidental, 5
- Espagnol, 195, 201, 202, 204, 206,
210, 212–214, 216
- F**
- Fait linguistique, 48, 55, 64, 80
- Fenêtre, 267
- Formant, 194, 207, 290
- Français, 195

- G**
- Galand, Lionel, 42, 43
 Gilliéron, Jules, 39
 Goebel, Hans, 48, 51, 53, 59
 Gourara, 176
 Grec, 209, 216
 Groningen, 60
 Groupe consonantique, 198, 206, 208, 211, 212, 242, 254, 255, 285
 Guerrab, Said, 44, 45, 52
 Guir, 6
 Géminée, 86, 90, 91, 119, 138, 140, 144–146
 Dégémiation, 138
 Fricative, 142
 Gémiation, 90, 119, 132, 140, 214, 260, 262, 263
 Géminée affriquée, 119, 120, 123, 124, 126, 130, 139, 141–144, 146
 Géminée fricative, 120, 123, 124, 146
 Géminée lexicale, 141, 142, 261
 Géminée morphologique, 140, 145, 261, 266
 Géolinguistique, 40, 210
 Atlas linguistique, 41, 42
 Beam map, 59
 Carte choroplèthe, 60, 159
 Carte isoplèthe, 60
 Carte linguistique, 41, 43, 44, 46, 51, 59, 60, 96, 113, 118, 133, 155, 158, 159, 167, 168
 Honeycomb map, 59
 Géostatistique, 51, 60
 Autocorrélation Spatiale, 61
 Interpolation Spatiale, 61
 Gétules, 8
- H**
- Hadjra Mektouba, 9
 Haoussa, 214
 Haplolecte, 36, 70, 153, 158, 159, 163, 172, 174, 176
 Hindi, 214
 Homorganique, 213, 217, 254, 255, 257
 Hongrois, 216
- I**
- Ibn Khaldoun, 9
 Ifrenides, 9
 INALCO, 23, 45
 Indoeuropéen, 40, 93
 Inscription rupestre, 9
 Iran, 53
 IRM, 196
- J**
- Japon, 53
- K**
- Kabyle, 21
 Kossmann, Marteen, 23, 193
 Ksar, 6
 Kurde, 215

- L**
- Ladefoged, Peter Nielsen, 194, 200, 202
- Lafkioui, Mena, 43, 44, 49, 52
- Lexicostatistique, 51, 52
- Libyco-berbère, 9
- Léon l'Africain, 10
- M**
- Maddieson, Ian, 194, 200, 202
- Maghreb central, 22
- Marmol, 10
- Maroc, 23, 41, 43
 Ain Chair, 5
 Bouanane, 5
 Figuig, 5, 9, 10, 23
 Ich, 5
 Marocain, 6
- Mecque, 10
 Hajj, 10
- Microphone, 196
- Mode d'articulation, 89, 97, 106, 116, 119, 135, 137, 188, 197, 241, 272, 274, 276, 277, 282, 296
- Morphème, 25
 Hétéromorphémique, 221, 283
 Tautomorphémique, 119, 215, 283
- Moyen-atlas, 40
- Méridiens, 10
- N**
- Naït-Zerad, Kamal, 14, 52, 184
- Nucléolecte, 92, 98, 100, 108, 117, 152, 153, 167, 172, 175
- Néolithique, 8
- O**
- Oasis, 7
- Ottomans, 10
- P**
- Paire minimale, 188
- Palatogramme, 196
 Electropalatogramme, 196
- Palmier dattier, 7
- Paléolithique, 8
- Pays-bas, 60
- Phonotactique, 212
- Phonème, 25, 91, 186, 215, 266
 Allophone, 94, 111–115, 146, 188, 242, 255, 259, 261, 266, 271
 Variante libre, 269
- Phonétique, 23, 55, 64, 67, 86, 90, 95, 101, 114, 138, 147, 148, 175, 184, 186, 188, 191–193, 196, 198, 202, 205, 210, 217, 221, 222, 229, 237, 241, 258–260, 264, 266, 269–272, 275, 279, 296, 304, 307, 308
- Point d'articulation, 89, 91, 98, 99, 103, 108, 117, 119, 124, 130, 135, 198, 204, 206, 213, 217, 272, 283, 294
- Polonais, 201, 216
- Pop, Sever, 39
- Portugais, 210
- Première articulation, 272, 276,

- 282, 291, 296
- Principe
 Principe d'inaltérabilité, 142, 215
 Principe d'intégrité, 142, 215
- Processus
 Processus morphologique, 139, 145, 186, 270
 Processus phonologique, 140, 144, 186
 Assimilation, 94, 213, 215, 261
 Lambdacisme, 195, 211
 Rhotacisme, 195
- Q**
- Questionnaire, 44, 47, 65, 69, 72, 92
- R**
- Ridouane, Rachid, 119, 191, 192
 Rif, 43, 52
 Al Hoceima, 192
 Ibquyen, 192
 Rostémides, 9
 Roumain, 216
 Russe, 213
- S**
- Saa, Fouad, 23
 Sahara, 8
 Schwa, 115, 289, 290, 295
 Sciences de l'Information
 Géographique, 24, 29, 39
 Sciences des données, 24, 29, 36, 49, 50, 53, 59, 77, 149, 155, 170
 Deep learning, 49
 Dummy, 82
 Fouille de données, 49
 Fouille de textes, 59
 Intelligence artificielle, 49
 Java, 53
 Machine Learning, 49
 Python, 53
 R, 53
 Scala, 53
 Sciences du langage, 25, 29, 39, 51, 59, 77, 168
 Sijilmasa, 9
 Siwa, 22
 Sokna, 22
 Sonorité, 194, 212, 214
 Spectrogramme, 197, 236, 237, 243, 249, 255, 261, 267, 269, 274, 275, 283, 284, 288, 290, 295
 Bande large, 236
 Bande étroite, 236, 267, 269
 Statistique, 49, 50, 52, 53, 56, 58, 61, 175, 233, 237-239, 241, 292, 295, 296
 Sud-Oranais, 4, 7, 9-11, 21, 22, 24, 64, 135, 186, 187
 Asla, 32, 64, 87, 91-95, 109, 121, 125, 127, 128, 139, 142, 158, 166, 185
 Beni Abbes, 21
 Beni Ounif, 32, 139, 140, 167
 Berrebi, 21, 32, 98, 111
 Boukais, 32, 103, 104, 112, 115-117,

- 166
- Boussemghoun, 9, 11, 16, 32, 66, 93,
94, 111, 166, 185
- Béchar, 15, 18, 21, 32
- Chellala, 6, 108, 118, 121, 125, 127,
128, 139, 143, 158, 166, 167
- Chellala dahrana, 32
- Chellala gueblia, 32
- Fendi, 32, 139, 140, 167
- Gharbi, 6
- Igli, 21, 23, 32, 109–111
- Kalaat Bouamama -
Moghrar tahtani, 32
- Kalaat Bouamama - Moghrar n
wadday, 9
- Lahmar, 22, 32, 103, 116, 117, 166
- Mazzer, 32, 110, 112
- Mogheul, 32
- Moghrar fougani, 32
- Namous, 6
- Ouakda, 21, 32
- Sfissifa, 32, 139, 140
- Souara, 5
Oued Souara, 6
- Taghit, 6, 15, 32
- Tiout, 32, 185
- Wakda, 10
- Zaouia Tahtania, 32
- Zoubia, 6
- Zousfana, 6
- Suisse, 60
- Syllabe, 112, 189, 212, 237, 248,
249, 253, 255, 256, 264, 269,
274, 277, 278, 283–286, 290,
291, 293–296
- Ambisyllabique, 215, 256
- Hétérosyllabique, 143, 215
- Tautosyllabique, 215
- T**
- Tchèque, 216
- Terrain d'étude, 3, 7, 9, 21, 29, 30,
33, 34, 46, 52, 60, 116, 148, 153,
170, 176, 184, 221
- Point d'enquête, 5, 10–12, 15, 24,
29, 33, 42, 44, 45, 47, 52, 54, 55,
59, 64, 65, 69, 74, 76, 79, 80, 86,
99, 108, 109, 112, 118, 119, 129,
138, 148, 153, 160, 169, 184,
185
- Timimoun, 211
- Tizi Ouzou, 45
- Touat, 176
- Tumulus, 9
- U**
- Unité d'analyse, 3
Objet d'étude, 3, 20, 23–25, 29, 34,
49, 116, 148, 170, 184, 185, 191,
193, 209, 220
- Population d'étude, 3, 11, 29, 170,
184, 221
- Unité morphémique, 25, 110, 114,
124, 130, 140, 146
- Unité phonique, 25, 90, 92, 93, 108,
112, 119, 129, 144, 145, 186,
257, 259, 264, 265, 267, 269,
272, 276, 290, 295
- Unité significative, 72, 93, 110

V

- Variable linguistique, 44, 92, 118
Variation linguistique, 34, 39, 44,
46, 48, 49, 60, 64, 86, 95, 116,
119, 133, 138, 149, 153, 168
Vesunes, 8
Vocoide, 188, 189, 216, 217, 241,
272, 283, 288–291, 293–296

W

- Weinreich, Uriel, 47
Wilaya, 5
Adrar, 176
Béchar, 5, 6, 11, 17–20
El-Bayadh, 5, 11
Naâma, 5, 11

X

- X-ray, 196

Y

- Yeou, Mohamed, 23

Z

- Zianides, 10

Bibliographie

- [Abondolo, 2015] Abondolo, D. (2015). *The Uralic Languages*. Routledge.
- [Ahmed and El-Aiachi, 1846] Ahmed, M. and El-Aiachi (1846). *Voyages dans le sud de l'Algérie et des états barbaresques de l'ouest et de l'est par El-Aiachi et Moula-Ah'med*. Number 9 in *Exploration scientifique de l'Algérie*. Imprimerie royale, Paris.
- [Aikhenvald, 1988] Aikhenvald, A. (1988). A Structural and Typological Classification of berber Languages. In Brauner, S. and Wolff, E., editors, *Progressive Tradition in African and Oriental Studies*, volume 21 of *Asien, Afrika, Latein Amerika : Sonderheft*, pages 36–43. Akademie Verlag, Berlin.
- [Alderete, 1997] Alderete, J. (1997). Dissimilation as local conjunction. In Kusumoto, K., editor, *Proceedings of the North East Linguistics Society*, volume 27, pages 17–32. GLSA.
- [Alpaydin, 2014] Alpaydin, E. (2014). *Introduction to Machine Learning*. Adaptive computation and machine learning. MIT Press.
- [Alvarez López, 2010] Alvarez López, L. (2010). A alternância /d/, /l/ e /r/ em variedades linguísticas afrolatinas. *Journal of Portuguese and Spanish lexically-based creoles*, 2(1) :1–27.
- [Ameur, 1990] Ameur, M. (1990). A propos de la classification des dialectes berbères. *Etudes et Documents Berbères*, (7) :15–27.
- [Asnaghi and Geeraerts, 2016] Asnaghi, C. and Geeraerts, D. (2016). Geographical patterns of formality variation in written standard California English. *Digital Scholarship in the Humanities*, 31(2) :244–263.

- [Auer and Schmidt, 2010] Auer, P. and Schmidt, J. E. (2010). *Language and Space : An International Handbook of Linguistic Variation*. Number 30 in Handbooks of linguistics and communication science. De Gruyter Mouton.
- [Aurrekoandxea, 2013] Aurrekoandxea, G. (2013). Is a scientific measurement of linguistic boundaries possible? In *Current Approaches to Limits and Areas in Dialectology*, pages 123–142. Ernestina Carrilho, Catarina Magro and Xosé Álvarez.
- [Aurrekoandxea, 2016] Aurrekoandxea, G. (2016). Analysis of the morphological variation of Basque. *Dialectologia and Geolinguistica*, 24(1) :21–41.
- [Aurrekoetxea, 1995] Aurrekoetxea, G. (1995). *Bizkaieraren egituraketa geolinguistikoa*. PhD thesis, Universidad del País Vasco.
- [Avram, 1993] Avram, A. (1993). Cercetări experimentale asupra consoanelor lichide din limba română. *Fonetică și dialectologie*, XII :7–20.
- [Backley, 2011] Backley, P. (2011). *Introduction to Element Theory*. Edinburgh University Press.
- [Baltazani, 2009] Baltazani, M. (2009). Acoustic characterization of the Greek rhotic in clusters. In *Proceedings of the 18th International Symposium on Theoretical and Applied Linguistics*, volume 1, pages 87–95.
- [Baltazani and Nicolaidis, 2013] Baltazani, M. and Nicolaidis, K. (2013). The many faces of /r/. *Rhotics : New data and perspectives*, pages 125–144.
- [Baltazani et al., 2011] Baltazani, M., Nicolaidis, K., et al. (2011). Production of the Greek rhotic in initial and intervocalic position : An acoustic and electropalatographic study. In Gavriilidou, Z. and Efthymiou, A., editors, *Selected paper of 10th International Conference on Greek Linguistics*, pages 141–152, Democritus University of Komotini.
- [Bandyopadhyay and Saha, 2012] Bandyopadhyay, S. and Saha, S. (2012). *Unsupervised Classification : Similarity Measures, Classical and Mandaheuristic Approaches, and Applications*. Springer Science & Business Media.
- [Bart et al., 2013] Bart, G., Glaser, E., et al. (2013). Analysis of Swiss German syntactic variants using spatial statistics. In Álvarez, X. and Carrilho, E., editors, *Current Approaches to Limits and Areas in Dialectology*, pages 143–169. Cambridge Scholars Publishing, Newcastle.
- [Basset, 1929] Basset, A. (1929). *Etudes de géographie linguistique en Kabylie (sur quelques termes berbères concernant le corps humain)*. Leroux, Paris.

- [Basset, 1935] Basset, A. (1935). Parlers touaregs du Soudan and du Niger notes and cartes linguistiques. *Bulletin du Comité d'études historiques and scientifiques de l'Afrique occidentale française*, XVIII(2-3) :336–352.
- [Basset, 1936] Basset, A. (1936). *Atlas linguistiques des parlers berbères, Algérie, Territoires du Nord, Noms d'animaux domestiques : Fascicule I*. Faculté des Lettres de l'Université d'Alger, Alger.
- [Basset, 1938] Basset, A. (1938). Présentation des premières cartes d'un atlas linguistique - en cours de réalisation - des parlers du Sahara and du Soudan. In Barr, K., Brøndal, V., et al., editors, *Actes du 4ème Congrès international des linguistes*, pages 177–182, Copenhage.
- [Basset, 1939] Basset, A. (1939). *Atlas linguistiques des parlers berbères, Algérie, Territoires du Nord, Noms d'animaux domestiques : Fascicule II*. Faculté des Lettres de l'Université d'Alger, Alger.
- [Basset, 1941] Basset, A. (1941). La langue berbère dans les territoires du Sud : la répartition, les études, remarques. *Revue africaine*, 85(LXXXV 386-387) :62–71.
- [Basset, 1942] Basset, A. (1942). Etudes de géographie linguistique dans le Sud marocain. *Hespéris*, 29:3–22.
- [Basset, 1948] Basset, A. (1948). La Langue berbère au Sahara. *Cahiers Charles de Foucauld*, 10:115–127.
- [Basset, 1951] Basset, A. (1951). L'enquête linguistique. In *Conférences de l'Institut de linguistique de l'Université de Paris*, Paris. Institut de linguistique de l'Université de Paris.
- [Basset, 1959] Basset, A. (1959). *Articles de dialectologie berbère*, volume 58 of *Collection linguistique / Société de linguistique de Paris*. Klincksieck, Paris.
- [Basset, 1886] Basset, R. (1886). Notes de lexicographie berbère. Spécimen du dialecte des k'çours oranais. *Journal Asiatique*, 1886:67–85.
- [Ben-David and Bat-El, 2016] Ben-David, A. and Bat-El, O. (2016). Paths and stages in acquisition of the phonological word in Hebrew. In Berman, R., editor, *Acquisition and Development of Hebrew : From infancy to adolescence*, volume 19 of *Trends in Language Acquisition Research*, pages 39–68. Benjamins.
- [Ben Si Said, 2014] Ben Si Said, S. (2014). *De la nature de la variation diatopique en kabyle : étude de la formation des singulier et pluriel nominaux*. PhD thesis, Université Nice Sophia Antipolis.

-
- [Bennett, 2013] Bennett, W. G. (2013). *Dissimilation, consonant harmony, and surface correspondence*. PhD thesis, Rutgers University of New Jersey-New Brunswick.
- [Berque, 1978] Berque, J. (1978). *L'intérieur du Maghreb (15ème-19ème siècle)*. Bibliothèque des Histoires. Gallimard.
- [Blecua, 2001] Blecua, B. (2001). *Las vibrantes del español : manifestaciones acústicas y procesos fonéticos*. PhD thesis, Universitat Autònoma de Barcelona.
- [Bolla, 1981] Bolla, K. (1981). *A conspectus of Russian speech sounds*, volume 32 of *Slavistische Forschungen*. Böhlau Verlag, Köln.
- [Boothroyd, 2004] Boothroyd, A. (2004). Room acoustics and speech perception. In *Seminars in Hearing*, volume 25, pages 155–166. Thieme Medical.
- [Boudjellal, 2015] Boudjellal, M. (2015). *Contribution à la géographie linguistique du berbère chaouià*. PhD thesis, INALCO, Paris.
- [Boudot-Lamotte, 1964] Boudot-Lamotte, A. (1964). Notes ethnographiques et linguistiques sur le parler berbère de Timimoun. *Journal asiatique*, 252(4) :488–558.
- [Bradley, 2001] Bradley, T. G. (2001). *The phonetics and phonology of rhotic duration contrast and neutralization*. PhD thesis, Pennsylvania State University, University Park.
- [Bradley and Schmeiser, 2003] Bradley, T. G. and Schmeiser, B. S. (2003). On the Phonandic Reality of Spanish /r/ in Complex Onsets. In *Theory, practice, and acquisition : Papers from the 6th Hispanic Linguistics Symposium*, pages 1–20.
- [Bren, 1998] Bren, M. (1998). Dandermining the Distance Index. In Chikio, H., Keiji, Y., et al., editors, *Data Science, Classification, and Related methods*, Studies in Classification, Data Analysis, and Knowledge Organization, pages 460–467. Springer.
- [Camps, 1983] Camps, G. (1983). Comment la Berbérie est devenue le Maghreb arabe. *Revue de l'Occident musulman et de la Méditerranée*, 35(1) :7–24.
- [Camps, 1991] Camps, G. (1991). Bazinas. In Camps, G., editor, *Encyclopédie berbère*, pages 1400–1407. Peeters.
- [Catford, 1977] Catford, J. (1977). *Fundamental problems in phonetics*. Indiana University Press.
- [Cathcart, 2015] Cathcart, C. A. (2015). *Iranian Dialectology and Dialectomandry*. PhD thesis, UC Berkeley.
- [Celata et al., 2014] Celata, C., Bertini, C., et al. (2014). Proprietà acustiche e articolatorie di /r/ nella Toscana occidentale.

- [Celata and Ricci, 2017] Celata, C. and Ricci, I. (2017). The physiology of voiced apical trills : implications for sound change. In *XIII Convegno Nazionale AISV*, Pisa, Italy.
- [Chaker, 1973] Chaker, S. (1973). *Le système dérivationnel verbal berbère : (dialecte kabyle)*. PhD thesis, Université de Paris-V (René Descartes).
- [Chaker, 1984] Chaker, S. (1984). *Textes en linguistique berbère : introduction au domaine berbère*. Éd. du C.N.R.S., Paris.
- [Chaker, 1995] Chaker, S. (1995). *Linguistique berbère : études de syntaxe et de diachronie*. Peeters.
- [Chaker, 1998] Chaker, S. (1998). Géographie linguistique. In Camps, G., editor, *Encyclopédie berbère*, pages 3059–3061. Peeters.
- [Chambers and Trudgill, 1998] Chambers, J. K. and Trudgill, P. (1998). *Dialectology*. Cambridge University Press.
- [Cheurfi, 2007] Cheurfi, A. (2007). *L'encyclopédie maghrébine*. Casbah éditions.
- [Chomsky, 1965] Chomsky, N. (1965). *Aspects of the theory of syntax*, volume Special technical report of the Research Laboratory of Electronics of *MIT Press Paperback Series*. MIT Press.
- [Chomsky and Halle, 1968] Chomsky, N. and Halle, M. (1968). *The Sound Pattern of English*. MIT Press.
- [Clark and Hillenbrand, 2007] Clark, M. J. and Hillenbrand, J. M. (2007). Quality of back vowels before /r/ in the American Inland North. *Journal of the International Phonetic Association*, 37(03).
- [Clements, 1990] Clements, G. N. (1990). The role of the sonority cycle in core syllabification. *Papers in laboratory phonology*, 1:283–333.
- [Cohen and Ben-David, 2016] Cohen, E.-G. and Ben-David, A. (2016). The role of allophony and frequency in the acquisition of the Hebrew rhotic. *Clinical linguistics & phonetics*, 30(2) :101–118.
- [Colantoni and Steele, 2005] Colantoni, L. and Steele, J. (2005). Liquid asymmetries in French and Spanish. *Toronto Working Papers in Linguistics*, 24.
- [Coleman, 1999] Coleman, J. (1999). The nature of vocoids associated with syllabic consonants in Tashlhiyt Berber. In *Proceedings of the XIVth international congress of phonetic sciences*, volume 1, pages 735–738.

- [Coleman, 2001] Coleman, J. (2001). The phonetics and phonology of Tashlhiyt Berber syllabic consonants. *Transactions of the Philological Society*, 99(1) :29–64.
- [Cominardi, 1991] Cominardi, F. (1991). Au coeur des monts des Ksour : le ksar de Chellala Dahrania. In *Études et documents berbères*, volume 8, pages 135–158. Édisud.
- [Cour, 2004] Cour, A. (2004). *L'établissement des dynasties des Chérifs au Maroc et leur rivalité avec les Turcs de la Régence d'Alger : 1509-1830*. publishers Bouchène.
- [Cutting and Day, 1975] Cutting, J. E. and Day, R. S. (1975). The perception of stop-liquid clusters in phonological fusion. *Journal of Phonetics*, 3(2) :99–113.
- [Delattre and Freeman, 1968] Delattre, P. and Freeman, D. C. (1968). A dialect study of American r's by x-ray motion picture. *Linguistics*, 6(44) :29–68.
- [Dell and Elmedlaoui, 1997] Dell, F. and Elmedlaoui, M. (1997). La syllabation et les géminées dans la poésie berbère du Maroc (dialecte chleuh). *Cahiers de grammaire*, 22:1–95.
- [Dell and Tangi, 1992] Dell, F. and Tangi, O. (1992). Syllabification and empty nuclei in Ath-Sidhar Rifian Berber. *Journal of African Languages and Linguistics*, 13(2) :125–162.
- [Dell and Tangi, 1993] Dell, F. and Tangi, O. (1993). On the vocalization of /r/. *Ath-Sidhar Rifian Berber, Linguistica Communicatio*, 5:53.
- [Demolin, 2001] Demolin, D. (2001). Some phonetic and phonological observations concerning /R/ in Belgian French. *Etudes & travaux - Institut des langues vivantes et de phonétique*, (4) :63–73.
- [Desanges, 1990] Desanges, J. (1990). Autolatae / Autololes / Autoteles. In Camps, G., editor, *Encyclopédie berbère*, pages 1175–1176. Peeters.
- [Desanges, 1998] Desanges, J. (1998). Gétules. In Camps, G., editor, *Encyclopédie berbère*, pages 3063–3065. Peeters.
- [Dhananjaya et al., 2012] Dhananjaya, N., Yegnanarayana, B., et al. (2012). Acoustic analysis of trill sounds. *The Journal of the Acoustical Society of America*, 131(4) :3141–3152.
- [Dickey, 1997] Dickey, L. W. (1997). *The phonology of liquids*. PhD thesis, University of Massachusetts Amherst.
- [Doutté and Gautier, 1913] Doutté, E. and Gautier, E. F. (1913). *Enquête sur la dispersion de la Berbère en Algérie*. Jourdan.

- [El Mountassir, 2003] El Mountassir, A. (2003). *Dictionnaire des verbes Tachelhit - Français : (Parler berbère du sud du Maroc)*. L'Harmattan.
- [Fant, 1971] Fant, G. (1971). *Acoustic Theory of Speech Production : With Calculations based on X-Ray Studies of Russian Articulations*, volume 2 of *Description and Analysis of Contemporary Standard Russian*. De Gruyter.
- [Galand, 1960] Galand, L. (1960). Berbère (Phonétique et phonologie). *Encyclopédie de l'Islam*, 1:1216–1217.
- [Galand, 1964] Galand, L. (1964). Géographie linguistique dans la région d'Imin-Tanout (Grand Atlas marocain). In *Communications and rapports du Premier Congrès international de dialectologie générale*, pages 49–63, Louvain - Bruxelles. Pop, Sever and Van Windekens, Albert J.
- [Gan and Ma, 2007] Gan, G. and Ma, C. (2007). *Data Clustering : Theory, Algorithms, and Applications*. SIAM.
- [Gibson, 2015] Gibson, M. (2015). A stochastic approach to rhotic variation in Spanish codas. *Loquens*, 2(1).
- [Gili Gaya, 1921] Gili Gaya, S. (1921). La "r" simple en la pronunciación española. *Revista de Filología Española*, (8) :271–280.
- [Gillian, E. and Jaworski, S., 2014] Gillian, E. and Jaworski, S. (2014). A Discussion of Walpiri and Polish rhotics. In *Unity in Diversity*, volume 2 of *Cultural and Linguistic Markers of the Concept*, pages 133–148. Newcastle upon Tyne : Cambridge Scholars Publishing, sabine asmus and barbara braid. edition.
- [Gilliéron and Edmont, 1912] Gilliéron, J. and Edmont, E. (1912). *Atlas linguistique de la France*. Atlas linguistique de la France. Forni.
- [Giovanni et al., 2014] Giovanni, A., Lagier, A., et al. (2014). Physiologie de la phonation. *Emc Oto Rhino Laryngologie*, 9(2) :1–15.
- [Glynn and Fischer, 2010] Glynn, D. and Fischer, K. (2010). *Quantitative Methods in Cognitive Semantics : Corpus-driven Approaches*. Walter de Gruyter.
- [Goebel, 2010] Goebel, H. (2010). Dialectometry and quantitative mapping. In Lameli, A. and Kehrein, R., editors, *Language and Space*, volume *Language Mapping of An International Handbook of Linguistic Variation*, pages 433–457. De Gruyter Mouton.
- [Goebel, 2012] Goebel, H. (2012). Introduction aux problèmes et méthodes de l'École dialectométrique de Salzbourg". In Álvarez Pérez, A., Carrilho, E., et al., editors,

- Proceedings of the International Symposium on Limits and Areas in Dialectology*, pages 117–166, Lisboa.
- [Grieve, 2014] Grieve, J. (2014). A comparison of statistical methods for the aggregation of regional linguistic variation. *Aggregating dialectology, typology, and register analysis : Linguistic variation in text and speech*, 28:53.
- [Grieve, 2016] Grieve, J. (2016). *Regional Variation in Written American English*. Cambridge University Press, Cambridge.
- [Guarisma, 1986] Guarisma, G., editor (1986). *La Méthode dialectométrique appliquée aux langues africaines*. Peeters.
- [Gudschinsky, 1956] Gudschinsky, S. C. (1956). The ABC's of lexicostatistics (glottochronology). *Word*, 12(2) :175–210.
- [Guemide and Benachiba, 2012] Guemide, B. and Benachiba, C. (2012). Using ICT to Improve Local Dialect of Béchar. *Timsal n Tamazight*, 3.
- [Guerrab, 2014] Guerrab, S. (2014). *Analyse dialectométrique des parlers berbères de Kabylie*. PhD thesis, Institut National des Langues and Civilisations Orientales, Paris.
- [Guerssel, 1977] Guerssel, M. (1977). Constraints on phonological rules. *Linguistic Analysis*, 3(3) :267–305.
- [Gussenhoven, 2015] Gussenhoven, C. (2015). Zwara Tamazight (Berber, Afroasiatic).
- [Hachid, 1992] Hachid, M. (1992). *Les Pierres écrites de l'Atlas saharien : El-Hadjra el-mektouba*. Entreprise nationale des arts graphiques, Alger.
- [Hall, 2006] Hall, N. (2006). Cross-linguistic patterns of vowel intrusion. *Phonology*, 23(03) :387–429.
- [Hall, 2003] Hall, N. E. (2003). *Gestures and segments : Vowel intrusion as overlap*. PhD thesis, University of Massachusetts.
- [Hall, 1997] Hall, T. A. (1997). *The phonology of coronals*. Number 149 in Current Issues in Linguistic Theory Series. John Benjamins Publishing.
- [Halle et al., 2000] Halle, M., Vaux, B., et al. (2000). On feature spreading and the representation of place of articulation. *Linguistic Inquiry*, 31(3) :387–444.
- [Hankamer and Aditi, 1988] Hankamer, J. and Aditi, L. (1988). The timing of geminate consonants. *Journal of Phonetics*, 16.

- [Hanson and Stevens, 2003] Hanson, M. H. and Stevens, K. N. (2003). Models of aspirated stops in English. In *Proceedings of the 15th international congress of phonetic sciences*, pages 783–786.
- [Harris, 1984] Harris, J. W. (1984). La espirantización en castellano y la representación fonológica autosegmental. *Working Papers in Linguistics*, 1:149–167.
- [Hayes, 1986] Hayes, B. (1986). Inalterability in CV Phonology. *Language*, 62(2) :321–351.
- [Heeringa and Johnson, 2009] Heeringa, W. and Johnson, K. (2009). Measuring Norwegian dialect distances using acoustic features. *Speech Communication*, 51(2) :167–183.
- [Heeringa and Kleiweg, 2006] Heeringa, W. and Kleiweg, P. (2006). Evaluation of String Distance Algorithms for Dialectology. In *Proceedings of the Workshop on Linguistic Distances*, LD '06, pages 51–62, Stroudsburg, PA, USA. Association for Computational Linguistics.
- [Henriksen and Willis, 2010] Henriksen, N. C. and Willis, E. W. (2010). Acoustic characterization of phonemic trill production in Jerezano Andalusian Spanish. In *Selected proceedings of the 4th Conference on Laboratory Approaches to Spanish Phonology*, pages 115–127. Cascadia Somerville, MA.
- [Heselwood et al., 2011] Heselwood, B., Howard, S., et al. (2011). Assimilation of /l/ to /r/ in Syrian Arabic : An electropalatographic and acoustic study. In *Instrumental Studies in Arabic Phonetics*, pages 63–98. John Benjamins Publishing, Philadelphia.
- [Howson et al., 2014] Howson, P., Komova, E., et al. (2014). Czech trills revisited : An ultrasound EGG and acoustic study. *Journal of the International Phonetic Association*, 44(02) :115–132.
- [Howson et al., 2013] Howson, P., Neufeld, C., et al. (2013). An electromagnetic articulography investigation of the Czech trill-fricative. *The Journal of the Acoustical Society of America*, 133(5) :3609.
- [Hu and Kokkinakis, 2014] Hu, Y. and Kokkinakis, K. (2014). Effects of early and late reflections on intelligibility of reverberated speech by cochlear implant listeners. *The Journal of the Acoustical Society of America*, 135(1) :EL22–EL28.
- [Hult, 1986] Hult, G. (1986). Speaker-independent recognition of connected Swedish digits. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '86.*, volume 11, pages 1093–1096.

- [Ibn Khaldun, 1968a] Ibn Khaldun, A.-R. (1968a). *Histoire des Berbères et des dynasties musulmanes de l'Afrique septentrionale*, volume 3. P. Geuthner, Paris.
- [Ibn Khaldun, 1968b] Ibn Khaldun, A.-R. (1968b). *Histoire des Berbères et des dynasties musulmanes de l'Afrique septentrionale*, volume 1. P. Geuthner, Paris.
- [Šimáčková, 2001] Šimáčková, r. (2001). Variability in the realization of Czech trilled r. In Kosta, P. and Blaszczyk, J., editors, *Investigations into formal Slavic linguistic : contributions of the fourth European Conference on Formal Description of Slavic Languages–FDSL IV*, Potsdam Linguistik International, pages 28–30. Peter Lang.
- [Inouye, 1995] Inouye, S. B. (1995). *Trills, Taps and Stops in Contrast and Variation*. PhD thesis, University of California, Los Angeles.
- [IPA, 1999] IPA, A. (1999). *Handbook of the International Phonetic Association : A Guide to the Use of the International Phonetic Alphabet*. Cambridge University Press.
- [Jaworski and Gillian, 2011] Jaworski, S. and Gillian, E. (2011). On the phonetic instability of the Polish rhotic /r/. *Poznań Studies in Contemporary Linguistics PSiCL*, 47:380.
- [Katsika and Kavitskaya, 2015] Katsika, A. and Kavitskaya, D. (2015). The Phonetics of r-Deletion in Samothraki Greek. *Journal of Greek Linguistics*, 15(1) :34–65.
- [Katz et al., 2015] Katz, W. F., Mehta, S., et al. (2015). A kinematic analysis of Japanese /r/ : Effects of syllable position and vowel context. *The Journal of the Acoustical Society of America*, 137(4) :2382.
- [Kavitskaya and Iskarous, 2009] Kavitskaya, D. and Iskarous, K. (2009). Trills and palatalization : Consequences for sound change. *Proceedings of the formal approaches to Slavic Linguistics*, 17:97–110.
- [Kenstowicz, 1994] Kenstowicz, M. (1994). *Phonology in Generative Grammar*. Wiley.
- [Kluge, 2005] Kluge, A. (2005). A Synchronic Lexical Study of Gbe Language Varieties : The Effects of Different Similarity Judgment Criteria. *Linguistic Discovery*, 3(1).
- [Kochetov and Colantoni, 2009] Kochetov, A. and Colantoni, L. (2009). Articulatory characteristics of coronal consonants in Argentine Spanish : An electropalatographic study. *Journal of the Acoustical Society of America*, 126(4) :2222.
- [Kokkinakis and Loizou, 2011] Kokkinakis, K. and Loizou, P. C. (2011). The impact of reverberant self-masking and overlap-masking effects on speech intelligibility by

- cochlear implant listeners (L). *The Journal of the Acoustical Society of America*, 130(3) :1099–1102.
- [Kossmann, 1997] Kossmann, M. (1997). *Grammaire du parler berbère de Figuig (Maroc oriental)*. Peeters, Louvain & Paris.
- [Kossmann, 1999] Kossmann, M. (1999). *Essai sur la phonologie du proto-berbère*, volume 12 of *Grammatische Analysen afrikanischer Sprachen*. Koppe Verlag.
- [Kossmann, 2010] Kossmann, M. (2010). Grammatical notes on the Berber dialect of Igli (Sud oranais, Algeria). pages 69–120. Rüdiger Köppe Verlag, Köln.
- [Kruskal, 1983] Kruskal, J. B. (1983). An overview of sequence comparison : Time warps, string edits, and macromolecules. *SIAM review*, 25(2) :201–237.
- [Kumagai, 2016] Kumagai, Y. (2016). Developing linguistic Atlas of Japan database and advancing analysis of geographical distributions of dialects. *The future of dialects : Selected papers from methods in Dialectology XV*, 1:333–362.
- [Labov, 1976] Labov, W. (1976). *Sociolinguistique*. Le sens commun. Éditions de Minuit.
- [Labov, 2007] Labov, W. (2007). Transmission and Diffusion. *Language*, 83(2) :344–387.
- [Labov William et al., 2008] Labov William, Ash Sharon, and Boberg Charles (2008). *The Atlas of North American English, Phonetics, Phonology and Sound Change*. De Gruyter Mouton, Berlin, Boston.
- [Lacoste, 1960] Lacoste, Y. (1960). *L'Algérie passé et présent*. Sociales.
- [Ladefoged and Maddieson, 1996] Ladefoged, P. and Maddieson, I. (1996). *The sounds of the world's languages*. Phonological Theory. Wiley-Blackwell, Oxford.
- [Lafkioui, 2007] Lafkioui, M. (2007). *Atlas linguistique des variétés berbères du Rif*. Rüdiger Köppe.
- [Lafkioui, 2009] Lafkioui, M. (2009). Analyses dialectométriques du lexique berbère du Rif. In Vossen, R., Ibrizimow, D., et al., editors, *Études berbères IV – Essais lexicologiques and lexicographiques et autres articles*, volume 25 of *Berber Studies*, pages 133–150, Köln. Rüdiger Köppe.
- [Lafkioui, 2015] Lafkioui, M. (2015). Méthodologie de recherche en géolinguistique. *Corpus*, (14) :139–164.
- [Lahrouchi, 2005] Lahrouchi, M. (2005). Regular and Irregular Imperfective conjugations in Berber languages.

- [Lahrouchi, 2009] Lahrouchi, M. (2009). La structure interne des racines triconsonantiques en berbère tachelhit. In Chaker, S. and Mettouchi, A., editors, *Etudes de phonétique et de linguistique berbères. Hommage à Naïma Louali 1961-2005*, volume Maghreb-Sahara of SELAF - Société d'Études Linguistiques et Anthropologiques de France, pages 177–193. Peeters, Paris-Louvain.
- [Lawson et al., 2013] Lawson, E., Scobbie, J. M., et al. (2013). Bunched /r/ promotes vowel merger to schwa : An ultrasound tongue imaging study of Scottish sociophonetic variation. *Journal of Phonetics*, 41(3) :198–210.
- [Lee and Krantzschmar, 1993] Lee, J. and Krantzschmar, W. A. (1993). Spatial analysis of linguistic data with GIS functions. *International Journal of Geographical Information Systems*, 7(6) :541–560.
- [Lhote, 1970] Lhote, H. (1970). *Les gravures rupestres du Sud-Oranais*, volume 16 of *Mémoires du centre de recherches anthropologiques préhistoriques et ethnographiques*. CNRS, Paris.
- [Lindau, 1980] Lindau, M. (1980). The story of /r/. *The Journal of the Acoustical Society of America*, 67(S27) :S27–S27.
- [Léon L'Africain, 1898] Léon L'Africain, L. (1898). *Description de l'Afrique, tierce partie du monde*. Number XV in Recueil de voyages et de documents pour servir à l'histoire de la géographie, depuis le XIIIe jusqu'à la fin du XVIe siècle. Ernest Leroux, Paris.
- [Léonard, 2012] Léonard, J.-L. (2012). *Éléments de dialectologie générale*. Michel Houdiard, Paris, France.
- [Louali, 2002] Louali, N. (2002). L'ambissyllabité des consonnes géminées : le cas du berbère (tachelhit). In *Actes du colloque 25ème journée d'étude sur la parole*.
- [Lounissi, 2010] Lounissi, S. (2010). *Étude de géographie linguistique chaoui sur les plans phonético-phonologique et lexical*. PhD thesis, Université Mouloud Mameri, Tizi Ouzou.
- [Maddieson and Disner, 1984] Maddieson, I. and Disner, S. F. (1984). *Patterns of sounds*. Speech Science and Communication. Cambridge university press.
- [McCarthy, 1979] McCarthy, J. J. (1979). *Formal problems in Semitic phonology and morphology*. PhD thesis, MIT, Massachusetts.
- [McGowan, 1992] McGowan, R. S. (1992). Tongue-tip trills and vocal-tract wall compliance. *The Journal of the Acoustical Society of America*, 91(5) :2903–2910.

- [Mitchell, 2012] Mitchell, H. B. (2012). *Data Fusion : Concepts and Ideas*. Springer Science & Business Media.
- [Mitterer and Ernestus, 2008] Mitterer, H. and Ernestus, M. (2008). The link between speech perception and production is phonological and abstract : Evidence from the shadowing task. *Cognition*, 109(1) :168–173.
- [Mitterer and Scharenborg, 2013] Mitterer, H. and Scharenborg, O. (2013). Phonological abstraction without phonemes in speech perception. *Cognition*, 129(2) :356–361.
- [Müller, 2011] Müller, D. (2011). *Developments of the lateral in Occitan dialects and their Romance and cross-linguistic context*. PhD thesis, Universität Heidelberg.
- [Mouili, 2012] Mouili, F. (2012). *Aspects of Sociolinguistic Variation in the Berber Speech of Igli*. Lambert Academic Publishing.
- [Métoz, 2006] Métoz, L. (2006). L'émergence de la lexicostatistique et de la glottochronologie en France au 19ème s. : Constantine Samuel Rafinesque, Jules Dumont d'Urville and Paul Broca. *Histoire Épistémologie Langage*, 28(2) :125–155.
- [Narayanan et al., 1999] Narayanan, S., Byrd, D., et al. (1999). Geometry, kinematics, and acoustics of Tamil liquid consonants. *The Journal of the Acoustical Society of America*, 106(4) :1993–2007.
- [Naït-Zerrad, 1998] Naït-Zerrad, K. (1998). *Dictionnaire des racines berbères (formes attestées). I. A-BEZL*, volume Maghreb-Sahara of SELAF - Société d'Études Linguistiques et Anthropologiques de France. Peeters, Paris-Louvain.
- [Naït-Zerrad, 1999] Naït-Zerrad, K. (1999). *Dictionnaire des racines berbères (formes attestées). II. C-DEN*, volume Maghreb-Sahara of SELAF - Société d'Études Linguistiques et Anthropologiques de France. Peeters, Paris-Louvain.
- [Naït-Zerrad, 2001] Naït-Zerrad, K. (2001). *Dictionnaire des racines berbères (formes attestées). III. D-GEY*, volume 3. Peeters.
- [Naït-Zerrad, 2004] Naït-Zerrad, K. (2004). *Linguistique berbère et applications*. Paris, France.
- [Naït-Zerrad, 2005] Naït-Zerrad, K. (2005). Essai d'analyse dialectométrique appliquée au berbère. *Studi Magrebini*, 3:229–238.
- [Naït-Zerrad, 2006] Naït-Zerrad, K. (2006). Classification des parlers berbères et analyse dialectométrique. In Allati, A., editor, *Linguistique amazighe : Les nouveaux horizons*, pages 64–73. Faculté des Lettres et des sciences Humaines de Tétouan, Tétouan.

- [Naït-Zerrad, 2009] Naït-Zerrad, K. (2009). Le calcul de la distance linguistique appliqué au berbère : Exploration des concepts et méthodes. In Chaker, S. and Mettouchi, A., editors, *Etudes de phonétique et de linguistique berbères. Hommage à Naïma Louali 1961-2005*, volume Maghreb-Sahara of SELAF - Société d'Études Linguistiques et Anthropologiques de France, pages 341–354. Peeters, Paris-Louvain.
- [Naït-Zerrad, 2012] Naït-Zerrad, K. (2012). Le berbère, l'aménagement linguistique et les linguistes. *Cahiers de l'Observatoire des pratiques linguistiques*, 3:153–160.
- [Naït-Zerrad, 2013] Naït-Zerrad, K. (2013). Oranie / Oran. *Encyclopédie berbère*, 35:5802.
- [Naumann, 2012] Naumann, C. (2012). *Acoustically based phonemics of Siwi (Berber)*, volume Berber studies. Köppe.
- [Nerbonne, 1999] Nerbonne, J. (1999). Edit distance and dialect proximity. In Sankoff, D. and Kruskal, J., editors, *Time Warps, String Edits and Macromolecules : The theory and practice of sequence comparison*, volume 15 of *The David Hume Series*, pages V–XV. Cambridge University Press.
- [Nerbonne and Heeringa, 1997] Nerbonne, J. and Heeringa, W. (1997). Measuring dialect distance phonetically. In Coleman, J., editor, *Proceedings of the Third Meeting of the ACL Special Interest Group in Computational Phonology (SIGPHON-97)*, pages 11–18, Madrid. Association for Computational Linguistics.
- [Nerbonne and Krandzschmar, 2006] Nerbonne, J. and Krandzschmar, W. (2006). Progress in dialectometry : toward explanation. *Literary and Linguistic Computing*, 21(4) :387–397.
- [Nicolaidis and Baltazani, 2011] Nicolaidis, K. and Baltazani, M. (2011). An electropalatographic and acoustic study of the Greek rhotic in /Cr/ clusters. In *Proceedings of the 17th International Congress of Phonetic Sciences*, pages 1474–1478, City University of Hong Kong. ICPHS.
- [Ouhalla, 1988] Ouhalla, J. (1988). *The Syntax of Head Movement : A Study of Berber*. PhD thesis, University of London.
- [Parker, 2012] Parker, S. (2012). Sonority distance vs. sonority dispersion—a typological survey. *The sonority controversy*, 18:101–165.
- [Pavlík, 2014] Pavlík, R. (2014). On Some Quantitative Properties of Czech ř in Newsreading. *Phonetica*, 71(1) :22–49.

- [Penchoen, 1973] Penchoen, T. G. (1973). *Étude syntaxique d'un parler berbère : Ait Frah de l'Aurès*, volume Studi Magrebini. Instituto Universitario Orientale, Centro di Studi Magrebini.
- [Pi, 2006] Pi, C.-Y. T. (2006). Beyond the Isogloss : Isographs in Dialect Topography. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, 51(2-3) :177–184.
- [Pons Moll, 2008] Pons Moll, C. (2008). The sonority scale : categorical or gradient. In *Poster presented at the CUNY Conference on the Syllable*.
- [Pop, 1950] Pop, S. (1950). *La Dialectologie : Aperçu historique et Méthodes d'enquêtes linguistiques*, volume fascicule 38 & fascicule 39 of *Recueil de travaux d'histoire and de philologie*. Publication Universitaires de Louvain.
- [Prasse, 2003] Prasse, K.-G. (2003). *Dictionnaire touareg-français : A-L*, volume 1. Museum Tusulanum Press.
- [Proctor, 2009] Proctor, M. I. (2009). *Gestural characterization of a phonological class : The liquids*. PhD thesis, Yale University.
- [Péronnet, 1998] Péronnet, L. (1998). *Atlas linguistique du vocabulaire maritime acadien*. Presses Université Laval.
- [Punnoose, 2011] Punnoose, R. (2011). *An auditory and acoustic study of liquids in Malayalam*. PhD thesis, Newcastle University.
- [Quilis, 1981] Quilis, A. (1981). *Fonética acústica de la lengua española*. Number 49 in *Biblioteca románica hispánica III : Manuales*. Gredos, Madrid.
- [Rabanus, 2011] Rabanus, S. (2011). The State of the Art in Linguistic Cartography. In Reinhammar, M. and Elmevik, L., editors, *Studier i dialektologi och språksociologi. Föredrag vid Nionde nordiska dialektologkonferensen i Uppsala 18-20 augusti 2010*, volume Acta Academiae Regiae Gustavi Adolphi 116, pages 31–52.
- [Rasoloniaina, 2014] Rasoloniaina, B. (2014). Réflexions sur les problèmes de recueil de données dans un espace urbain complexe. *Cahiers internationaux de sociolinguistique*, 6(2) :103–119.
- [Raymond and Parker, 2005] Raymond, M. and Parker, S. (2005). Initial and medial geminate trills in Arop-Lokep. *Journal of the International Phonetic Association*, 35(01) :99–111.
- [Recasens, 1991] Recasens, D. (1991). On the production characteristics of apicoalveolar taps and trills. *Journal of phonetics*, 19(3-4) :267–280.

- [Recasens and Espinosa, 2006] Recasens, D. and Espinosa, A. (2006). Articulatory, positional and contextual characteristics of palatal consonants : Evidence from Majorcan Catalan. *Journal of Phonetics*, 34:295–318.
- [Recasens and Espinosa, 2007] Recasens, D. and Espinosa, A. (2007). Phonetic typology and positional allophones for alveolar rhotics in Catalan. *Phonetica*, 64(1) :1–28.
- [Recasens and Pallarès, 1999] Recasens, D. and Pallarès, M. D. (1999). A study of /j/ and /r/ in the light of the “DAC” coarticulation model. *Journal of Phonetics*, 27(2) :143–169.
- [Ridouane, 2003] Ridouane, R. (2003). *Suites de Consonnes en Berbère : Phonétique and Phonologie*. phdthesis, Université de la Sorbonne nouvelle - Paris III.
- [Ridouane, 2006] Ridouane, R. (2006). Investigating speech production : A review of some techniques.
- [Ridouane, 2008] Ridouane, R. (2008). L'inaltérabilité des géminées et la spirantisation. *Etudes et documents berbères*, 27:127–149.
- [Ridouane, 2010] Ridouane, R. (2010). Geminate at the junction of phonetics and phonology. *Papers in laboratory phonology*, 10:61–90.
- [Ridouane, 2014] Ridouane, R. (2014). Tashlhiyt Berber. *Journal of the International Phonetic Association*, 44(02) :207–221.
- [Ridouane, 2015] Ridouane, R. (2015). *Questions amazighes (Tachelhit) à l'interface de la phonétique et de la phonologie*. PhD thesis, Univeristé Denis Diderot Paris 7.
- [Rumpf and Pickl, 2009] Rumpf, J. and Pickl, S. (2009). Structural analysis of dialect maps using methods from spatial statistics. In *Zeitschrift für Dialektologie und Linguistik*, number 76, pages 280–308. Steiner, Stuttgart.
- [Saa, 1995] Saa, F. (1995). *Aspects de la morphologie et de la phonologie du berbère dans le Ksar Zenaga à Figuig (Maroc)*. PhD thesis, Université Panthéon-Sorbonne . Paris, Paris.
- [Sadiqi, 1986] Sadiqi, F. (1986). *Studies in Berber Syntax : The Complex Sentence*. Königshausen and Neumann, Würzburg.
- [Savu, 2014] Savu, C. F. (2014). More on the rhotic tap and the implications of its structure. *Academic Journal of Modern Philology*, 3:75–87.
- [Scheer, 2011] Scheer, T. (2011). *A Guide to Morphosyntax-phonology Interface Theories : How Extra-phonological Information is Treated in Phonology Since Trubandzkoy's Grenzsignale*. Walter de Gruyter.

-
- [Scherrer and Stoeckle, 2016] Scherrer, Y. and Stoeckle, P. (2016). A quantitative approach to Swiss German – Dialectometric analyses and comparisons of linguistic levels. *Dialectologia and Geolinguistica*, 24(1) :92–125.
- [Schiller, 1999] Schiller, N. (1999). The phonetic variation of German /r/. In Butt, M. and Fuhrhop, N., editors, *Variation und Stabilität in der Wortstruktur. Untersuchungen zu Entwicklung, Erwerb und Varietäten des Deutschen und anderen Sprachen*, number 141-142 in Germanistische Linguistik, pages 261–287. Olms, Zurich.
- [Schiltz, 1996] Schiltz, G. (1996). German dialectometry. In *Data Analysis and Information Systems*, pages 526–539. Springer.
- [Schmeiser, 2009] Schmeiser, B. (2009). Prosodic and segmental effects on vowel intrusion duration in Spanish /rC/ clusters. *Phonetics and Phonology : Interactions and Interrelations*, 306:181–202.
- [Scobbie et al., 2013] Scobbie, J. M., Punnoose, R., et al. (2013). Articulating five liquids : a single speaker ultrasound study of Malayalam. In Spreafico, L. and Vietti, A., editors, *Rhotics : New Data and Perspectives*, pages 99–124. BU Press, Bozen-Bolzano.
- [Séguy, 1971] Séguy, J. (1971). La relation entre la distance spatiale and la distance lexicale. *Revue de linguistique romane*, 35(139-140) :336–357.
- [Sibler et al., 2012] Sibler, P., Weibel, R., et al. (2012). Cartographic Visualization in Support of Dialectology. Columbus, Ohio, USA.
- [Solé, 1998] Solé, M.-J. (1998). Phonological universals : Trilling, voicing, and frication. In *Annual Meeting of the Berkeley Linguistics Society*, volume 24, pages 403–416.
- [Solé, 2002] Solé, M.-J. (2002). Aerodynamic characteristics of trills and phonological patterning. *Journal of Phonetics*, 30(4) :655–688.
- [Spajić et al., 1994] Spajić, S., Ladefoged, P., et al. (1994). The rhotics of Toda. In *Fieldwork Studies of Targeted Languages II*, pages 35–66, Los Angeles. Department of Linguistics, UCLA.
- [Spajić et al., 1996] Spajić, S., Ladefoged, P., et al. (1996). The trills of Toda. *Journal of the International Phonetic Association*, 26(01) :1–21.
- [Steele and Chandler, 2016] Steele, B. and Chandler, J. (2016). *Algorithms for Data Science*. Springer.
- [Stoeckle, 2016] Stoeckle, P. (2016). Horizontal and vertical variation in Swiss German morphosyntax. In Côté, M.-H. and Knooihuizen, R., editors, *The future of dialects*.

- Selected papers from methods*, number XV in *Dialectology*, pages 195–215. Language Science Press.
- [Stone, 2010] Stone, M. (2010). Laboratory Techniques for Investigating Speech Articulation. In Hardcastle, W. J., Laver, J., et al., editors, *The Handbook of Phonetic Sciences*, pages 7–38. Blackwell Publishing Ltd., Oxford, UK.
- [Swadesh, 1950] Swadesh, M. (1950). Salish internal relationships. *International Journal of American Linguistics*, 16(4) :157–167.
- [Szmrecsanyi, 2012] Szmrecsanyi, B. (2012). *Grammatical Variation in British English Dialects : A Study in Corpus-Based Dialectometry*. Cambridge University Press.
- [Tillinger, 2013] Tillinger, G. (2013). Langues, dialectes and patois—Problèmes de terminologie dialectologique. *Argumentum*, 9:1–18.
- [Uffmann, 2007] Uffmann, C. (2007). Intrusive [r] and optimal epenthetic consonants. *Language Sciences*, 29(2–3) :451–476.
- [Vago and Gósy, 2007] Vago, R. M. and Gósy, M. (2007). Schwa vocalization in the realization of /r/. In *Proceedings of the 16th International Congress of Phonetic Sciences*, pages 505–509.
- [Viaut, 2007] Viaut, A. (2007). *Variable territoriale et promotion des langues minoritaires*. Maison des Sciences de l’Homme d’Aquitaine.
- [Vihman, 1996] Vihman, M. M. (1996). *Phonological development : The origins of language in the child*. Applied language studies. Blackwell Publishing, Malden.
- [Waller and Fawcandt, 2013] Waller, M. A. and Fawcandt, S. E. (2013). Data Science, Predictive Analytics, and Big Data : A Revolution That Will Transform Supply Chain Design and Management. *Journal of Business Logistics*, 34(2) :77–84.
- [Warner and Tucker, 2011] Warner, N. and Tucker, B. V. (2011). Phonetic variability of stops and flaps in spontaneous and careful speech. *The Journal of the Acoustical Society of America*, 130(3) :1606–1617.
- [Wattel and Reenen, 2010] Wattel, E. and Reenen, P. (2010). Probabilistic maps. In Lameli, A. and Kehrein, R., editors, *Language and Space*, volume *Language Mapping of An International Handbook of Linguistic Variation*, pages 495–505. De Gruyter Mouton.
- [Webb and Bradley, 2009] Webb, E. R. and Bradley, T. G. (2009). Rhotic Metathesis Asymmetries in Romance Formalizing the Effects of Articulation and Perception on

- Sound Change. In Masullo, P. J., editor, *Romance Languages : Structure, Interfaces, and Microparametric Variation. Selected Papers from the 37th Linguistic Symposium on Romance Languages*, pages 321–337, Amsterdam. John Benjamins.
- [Weinreich, 1968] Weinreich, U. (1968). Is a structural dialectology possible. *Readings in the Sociology of Language*, pages 305–319.
- [Wheeler, 1979] Wheeler, M. (1979). *Phonology of Catalan*. Number 28 in Philological Society. Blackwell.
- [Wieling and Nerbonne, 2011] Wieling, M. and Nerbonne, J. (2011). Bipartite spectral graph partitioning for clustering dialect variandies and dandecting their linguistic features. *Computer Speech & Language*, 25(3) :700–715.
- [Wiese, 2003] Wiese, R. (2003). The unity and variation of (German) /r/. *Zeitschrift für Dialektologie und Linguistik*, 70(1) :25–43.
- [Williams and Merwe, 1996] Williams, C. H. and Merwe, I. V. D. (1996). Mapping the Multilingual City : A Research Agenda for Urban Geolinguistics. *Journal of Multilingual and Multicultural Development*, 17(1) :49–66.
- [Yang, 2010] Yang, C. (2010). *Lalo regional variandies : Phylogeny, dialectometry, and sociolinguistics*. PhD thesis, La Trobe University.
- [Yeou, 1997] Yeou, M. (1997). Locus equations and the degree of coarticulation of Arabic consonants. *Phonetica*, 54(3-4) :187–202.
- [Yeou, 2011] Yeou, M. (2011). Compiling of a Berber-French Dictionary (Figuig dialect). *Les ressources langagières : construction et exploitation*, pages 137–151.
- [Yeou and Honda, 2011] Yeou, M. and Honda, K. (2011). A palatographic, photoglottographic and airflow study of some coronal geminates in Figuig Berber. In *Proceedings of the 17th International Congress of Phonandic Sciences*, pages 2216–2219, City Univ. Hong Kong. ICPHS.
- [Yeou and Honda, 2015] Yeou, M. and Honda, K. (2015). An instrumental study of some coronal geminates in Figuig Berber.
- [Żygis, 2005] Żygis, M. (2005). (Un) markedness of trills : the case of Slavic r-palatalisation. *Zeitschrift für Slawistik*, 50(4) :383–407.
- [Ylinen et al., 2005] Ylinen, S., Huotilainen, M., and Näätänen, R. (2005). Phoneme quality and quantity are processed independently in the human brain. *NeuroReport*, 16(16) :1857–1860.

- [Zeroual and Clements, 2015] Zeroual, C. and Clements, G. N. (2015). The feature [pharyngeal]. In Rialland, A. and Ridouane, R., editors, *Features in Phonology and Phonetics*, Phonology and phonetics, pages 247–276. De Gruyter Mouton, Berlin.
- [Zeroual et al., 2014] Zeroual, C., H. Esling, J., et al. (2014). Contribution à l'étude physiologique des possibilités articulatoires dans la cavité pharyngale. In *Journées d'Etude sur la Parole (JEP 2014)*, Nantes, France.
- [Zeroual et al., 2008] Zeroual, C., Hoole, P., et al. (2008). Réalisations phonétiques des suites /tr/ en arabe marocain : données par EMA et par endoscopique. *JEP 2008*.

Mohamed EL IDRISSE

Description des variétés berbères en danger du Sud-Oranais (Algérie) - Étude dialectologique, phonologique et phonétique du système consonantique

Résumé

Il existe dans le sud ouest algérien plusieurs variétés de berbère. Certaines d'entre elles sont situées dans la région dite du Sud-Oranais et peuvent être cataloguées comme des langues en danger. Nous avons donc entrepris de décrire ces variétés avant qu'elles ne disparaissent. Cela a été mené à bien en réalisant plusieurs enquêtes de terrain. Par ailleurs, ce travail de documentation linguistique et de conservation du patrimoine culturel n'est qu'un des aspects de cette thèse. Ce travail académique est à la croisée de différents domaines disciplinaires. Nous avons eu recours aux méthodes en usage en Sciences de l'Information Géographique (SIG) et en Sciences Des Données (SDD) pour mener une étude dialectologique. Grâce aux SIG, nous avons réalisé une étude géolinguistique qui nous a permis de visualiser sur des cartes linguistiques la distribution de la variation linguistique de certaines consonnes. À partir de ces données, nous avons discuté de la réalité phonologique de ces consonnes simples et géminées. Dans le prolongement, une étude dialectométrique a été effectuée en nous basant sur des méthodes de partitions des données. Nous avons utilisé les méthodes d'Apprentissage Non Supervisé (PHA, k-moyenne, MDS, ...) et les méthodes d'Apprentissage Supervisé (CART) connues en SDD. Les résultats ont été affichés sous la forme de figures (cartes linguistiques, dendrogramme, heatmap, arbre, ...) à des fins d'exploration visuelle des données. L'ensemble de ces études a été accompli par le biais d'un traitement informatique (langage R). Puis, nous avons entrepris une analyse phonétique fondée sur une étude acoustique des rhotiques alvéolaires : [r], [r̥], [rʰ] et [rʰ̥]. Ces unités phoniques se distinguent par leur temporalité et leur réalisation articulo-voiciale. Ainsi, les spectrogrammes nous ont permis d'examiner la distribution de ces sons. Puis cela a nous permis de distinguer ce qui relevait de la phonétique et de la phonologie. Ensuite, nous avons achevé cette thèse par une étude phonétique et statistique. Ces dernières ont porté sur l'obstruction réalisée par la pointe de la langue et sur la nature des vocoïdes qui accompagnent les rhotiques alvéolaires dans l'environnement d'une consonne.

Apprentissage Non Supervisé, Apprentissage Supervisé, R, Phonétique, Phonologie, Carte linguistique, SIG, Informatique, Algérie, Berbère, Rhotique, Trill, Tap, Langue en danger

Résumé en anglais

There are several Berber languages in the south west of Algeria. Some of them are situated in the so-called Sud-Oranais and they can be categorized as endangered languages. So I have decided to describe them before they disappear. That's why, I have carried out several fieldworks. But, this linguistic documentation work and cultural heritage conservation are just one of aspects of our thesis. This PhD are transdisciplinary. I have used the methods which are applied in Geographic Information Science (GIS) and in Data Science (DS) to carry out a dialectological study. A geolinguistic study has been undertaken and has enabled to visualize the expansion of the linguistic variation of certain consonants through GIS. Based on these data, I have debated the phonological reality of the simple and geminate consonants. From this research, a dialectometric study was carried out on the basis of data partitioning methods. I have used the Unsupervised Learning Methods (HAC, k-mean, MDS, ...) and the Supervised Learning Methods (CART) known in DS. A visual exploration (linguistic maps, dendogram, heatmap, tree, ...) approach is proposed in order to analyse the results which have been realized through computer processing (R language). Then, I have undertaken a phonetic analysis, which is based on an acoustic study of alveolar rhotics : [r], [r̥], [rʰ] and [rʰ̥]. These phonic unities are distinguished by their temporality and their articulatory realization. Thus, the spectrograms enabled to examine the distribution of these sounds and to distinguish what was related to phonetic and phonology. Then, this thesis with a phonetic and statistical study has reinforced this research focused on the obstruction made by the tip of the tongue and on the nature of the vocoids which goes along with the alveolar rhotic in the area of the consonants.

Unsupervised Learning, Supervised Learning, R, Phonetic, Phonology, Linguistic map, GIS, Computer processing, Algeria, Berber, Rhotic, Trill, Tap, Endangered language