



**HAL**  
open science

## Pour un atlas numérique sonore des langues vernaculaires du monde arabophone (ANSOLVAR)

Laura Abou Haidar

► **To cite this version:**

Laura Abou Haidar. Pour un atlas numérique sonore des langues vernaculaires du monde arabophone (ANSOLVAR). Jean-François FAU. De la pierre au papier, du papier au numérique., Geuthner, pp.31-48, 2020, 978-2-7053-4062-9. hal-03103934

**HAL Id: hal-03103934**

**<https://hal.univ-grenoble-alpes.fr/hal-03103934>**

Submitted on 8 Jan 2021

**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.

# Pour un atlas numérique sonore des langues vernaculaires du monde arabophone (ANSOLVAR)

Laura ABOU HAIDAR

Maitre de conférences HDR – Université Grenoble Alpes, Laboratoire LIDILEM

[Laura.Abou-Haidar@univ-grenoble-alpes.fr](mailto:Laura.Abou-Haidar@univ-grenoble-alpes.fr)

## 1 – Introduction

Les langues constituent une composante fondamentale du patrimoine immatériel de l'humanité, comme le mentionne la *Convention pour la sauvegarde du patrimoine culturel immatériel*<sup>1</sup> de l'UNESCO (2003) qui évoque les « *mesures visant à assurer la viabilité du patrimoine culturel immatériel, y compris l'identification, la documentation, la recherche, la préservation, la protection, la promotion, la mise en valeur, la transmission, essentiellement par l'éducation formelle et non formelle, ainsi que la relativisation des différents aspects de ce patrimoine* » (UNESCO, 2003 : 3). Dans un certain nombre de pays, les langues font l'objet de plans de sauvegarde et de préservation, et le concept de « patrimoine immatériel » peut être intégré à travers des dispositifs incitatifs ou contraignants : les Etats n'hésitent pas à avoir recours à des textes de loi ou à inscrire la protection de telle ou telle langue dans leur constitution. La valorisation de la diversité linguistique dans une perspective patrimoniale constitue une réelle avancée dans de nombreux pays, même si beaucoup reste à faire et si des résistances peuvent perdurer : c'est le cas par exemple de la France, qui d'un côté admet la diversité linguistique sur son territoire, comme mentionné par Baude et Alessio : « *à côté du français langue nationale, langue commune, présente sur les cinq continents, les langues de France constituent un patrimoine culturel unique : il y a sur le territoire de la République (...) plus de 75 langues (...) reconnues comme « langues de France », c'est-à-dire parlées par des citoyens français en France depuis assez longtemps pour faire partie du patrimoine culturel national* »<sup>2</sup> ; mais qui parallèlement, n'a toujours pas ratifié la *charte européenne des langues régionales et minoritaires*<sup>3</sup> qui date de 1992.

Quel que soit le niveau de « vitalité » d'une langue, elle est susceptible d'être concernée par des actions de préservation. Celles-ci concernent aussi bien les langues dites « sûres », portées par des politiques officielles volontaristes, que des langues considérées en péril. Sur le plan mondial, les « langues en danger » occupent chercheurs et organismes depuis quelques décennies, et dans le cadre des travaux de l'UNESCO, ces langues en danger sont répertoriées dans un atlas<sup>4</sup> qui met en lumière les différents « niveaux de vitalité » qui permettent d'apprécier le degré de sécurité ou les menaces qui pèsent sur elles (Tableau 1).

Qu'en est-il du monde arabophone ? Celui-ci présente un paysage sociolinguistique très riche et particulièrement complexe, marqué par une instabilité géopolitique extrême, qui affecte profondément l'évolution des langues qui font partie intégrante d'un patrimoine immatériel qui peine à bénéficier d'une réelle reconnaissance par les Etats, sans même parler d'une quelconque

---

<sup>1</sup> Consultable sur le site internet suivant : <https://ich.unesco.org/doc/src/18440-FR.pdf>

<sup>2</sup> Baude & Alessio, 2008, p. 44.

<sup>3</sup> Accessible sur le site internet suivant : <https://www.coe.int/fr/web/european-charter-regional-or-minority-languages/text-of-the-charter>

<sup>4</sup> Edition interactive de l'*Atlas UNESCO des langues en danger dans le monde* : <http://www.unesco.org/languages-atlas/fr/atlasmap.html>

politique de préservation ou de protection des langues minoritaires ou des variétés régionales, à l'instar de la charte européenne des langues régionales ou minoritaires par exemple. Dans cette étude, en partant d'un bref panorama des particularités sociolinguistiques de cette immense aire géographique, nous expliquerons pourquoi une solution d'atlas numérique sonore des langues vernaculaires, qui constitue certes un projet de très grande ampleur et de longue haleine, pourrait se révéler particulièrement pertinente dans une optique de sauvegarde du patrimoine linguistique du monde arabophone.






Niveau de vitalité		Transmission de la langue d'une génération à l'autre
	sûre	la langue est parlée par toutes les générations ; la transmission intergénérationnelle est ininterrompue >> not included in the Atlas
	vulnérable	la plupart des enfants parlent la langue, mais elle peut être restreinte à certains domaines (par exemple: la maison)
	en danger	les enfants n'apprennent plus la langue comme langue maternelle à la maison
	sérieusement en danger	la langue est parlée par les grands-parents ; alors que la génération des parents peut la comprendre, ils ne la parlent pas entre eux ou avec les enfants
	en situation critique	les locuteurs les plus jeunes sont les grands-parents et leurs ascendants, et ils ne parlent la langue que partiellement et peu fréquemment
	éteinte	il ne reste plus de locuteurs >> l'Atlas contient les références des langues éteintes depuis les années 1950

Tableau 1 – Les niveaux de vitalité auxquels fait référence l'atlas UNESCO des langues en danger dans le monde

## 2 – Un paysage sociolinguistique riche et complexe

Précisons d'emblée que dans une perspective sociolinguistique, l'appellation « monde arabophone » nous paraît bien plus appropriée que celle de « monde arabe » : en effet, si « la » langue arabe constitue un socle très robuste qui fédère les pays rassemblés dans le cadre de la Ligue Arabe, et en particulier en tant que langue officielle inscrite dans les constitutions de ces différents pays, il n'en reste pas moins qu'une multitude de locuteurs de cette vaste zone géographique possède des origines ethniques autres, et ne s'identifient pas fondamentalement et principalement comme « arabes » alors que ce sont des habitants autochtones de cette région. Cette région est caractérisée par une très grande diversité de langues en usage et en contact, et la multitude de langues qui coexistent possèdent des statuts variés.

On y trouve tout d'abord les différentes variétés d'arabe, les principales étant les suivantes :

- Les dialectes maternels, dont la dénomination varie selon les pays (darija, aammiya,...), sont présents dans leur écrasante majorité sous leur forme exclusivement orale, mais dans certains pays (le Liban ou le Maroc par exemple) on observe un militantisme plus ou moins ancien en faveur de leur préservation et de leur diffusion sous une forme graphique : de nombreux intellectuels, écrivains, poètes, journalistes... sont convaincus de la nécessité de valoriser ces variétés de langue dans une démarche de

« démarginalisation »<sup>5</sup> et d'affirmation identitaire et statutaire. Certains de ces dialectes font l'objet de descriptions (socio)linguistiques importantes : on peut citer les travaux d'auteurs tels que Dominique Caubet, Madiha Doss, Djamel Kouloughli, Catherine Miller, Catherine Taine-Cheikh, etc. ; citons également les travaux de l'association internationale de dialectologie arabe (AIDA) et en particulier les actes de sa 11<sup>ème</sup> conférence internationale, qui sont consacrés aux variétés arabes<sup>6</sup> et qui rendent compte de l'extrême vitalité du champ.

- Vient ensuite la variété « arabe standard moderne » ou ASM, qui constitue la référence scolaire formalisée et écrite, et qui est la langue d'enseignement dans les systèmes éducatifs du monde arabophone. Cette langue est issue d'un vaste projet de réforme mené au XIX<sup>e</sup> siècle par des intellectuels dont *Ahmad Faris Al-Shidiahq* et *Nasif Al-Yaziji* pour ne citer que les plus renommés, basé sur la langue littéraire classique qui a subi des opérations de « modernisation »<sup>7</sup> y compris sur le plan morphosyntaxique et lexical, et marquées notamment par « des influences étrangères occidentales sur le plan de l'organisation des énoncés ou du choix des lexèmes, ainsi que des influences régionales dialectales »<sup>8</sup>. Pour ces raisons, nous considérons l'ASM comme une variété distincte des autres. Du point de vue des usages, pour reprendre les propos de Ziamari et De Ruiter<sup>9</sup>, il s'agit d'un « arabe que l'on ne parle pas ».
- Viennent également les variétés liturgiques : les trois monothéismes dans lesquels l'usage de l'arabe est attesté depuis des siècles sont à prendre en compte, aussi bien l'Ancien que le Nouveau Testament qui possèdent des versions en langue arabe, que le Coran dont la variété (l'arabe coranique) est la plus traditionnellement citée.
- Citons pour finir d'autres variétés qui couvrent un large spectre aussi bien diachronique que synchronique : pour ne retenir que deux exemples extrêmement éloignés sur le plan temporel, citons l'arabe antéislamique, qui est très familier aux usagers arabophones à travers une production poétique<sup>10</sup> foisonnante ; ou encore, sur un plan plus contemporain, le concept de *Educated Spoken Arabic (ESA)* que l'on doit à Mitchell (1979) et qui « recouvre tout un continuum de pratiques avec un dosage plus ou moins important du rapport dialectal-AMS. Il ne constitue pas une variété stabilisée autonome mais plutôt « un style », propre aux échanges formels »<sup>11</sup>.

La variabilité régionale des dialectes arabes est prise en compte dans la littérature par la délimitation de cinq zones géodialectales<sup>12</sup> [Figure 1].

---

<sup>5</sup> Ziamari & de Ruiter, 2015.

<sup>6</sup> Grigore & Bituna, 2016.

<sup>7</sup> Owen, 2015.

<sup>8</sup> Abou Haidar, 2018, p. 40.

<sup>9</sup> Voir note 5.

<sup>10</sup> Citons les fameuses Muallaqat qui datent du VI<sup>e</sup> siècle, ou le poète Imru'ul Qays natif d'Arabie au début du VI<sup>e</sup> siècle.

<sup>11</sup> Miller, 2010, p. 7.

<sup>12</sup> Cohen, 1970 & Versteegh, 1997.



Figure 1 – Représentation des zones géodialectales du monde arabophone d’après les travaux de Cohen et Versteegh, figure empruntée à Salam (2012).

Dans le monde arabophone, l’arabe et ses variétés sont dans une situation de contact avec une multiplicité d’autres langues, de nature très différente. Nous avons tout d’abord des langues historiques parmi les plus anciennes au monde. On y trouve aussi des langues issues de la colonisation héritée des empires coloniaux français, britannique ou espagnol. Citons enfin des langues issues de phénomènes migratoires plus récents, datant pour la plupart du XX<sup>ème</sup> siècle. Si aucun relevé exhaustif n’est disponible, une source numérique précieuse consacrée à « *L’aménagement linguistique dans le monde* »<sup>13</sup> établie par Leclerc permet d’estimer à plus de 240 le nombre de langues vernaculaires en usage dans le monde arabophone. En nous basant sur les travaux de Leclerc, et à l’exclusion des dialectes arabes et du paysage sociolinguistique complexe du Soudan<sup>14</sup>, le tableau 2 présente un premier relevé des langues en usage dans le monde arabophone, qui demande à être affiné, enrichi, nuancé. Ces données constituent une base de travail qui pourrait servir de point de départ au projet ANSOLVAR comme cela sera évoqué plus loin.

### 3. Des risques sur les langues

Ce chiffre estimatif et provisoire de 240 langues ne doit pas masquer une réalité inquiétante. Dans son « *Atlas des langues en danger dans le monde* »<sup>15</sup>, l’UNESCO répertorie 121 langues dans le monde arabophone comme étant « vulnérables, en danger, sérieusement en danger, en situation critique », voire pour une partie d’entre elles, « éteintes ».

Les facteurs de risque sur les langues dans le monde arabophone sont de nature diverse. Comme partout ailleurs, on peut identifier quelques facteurs habituels qui font qu’une langue est vulnérable : pour Leclerc (2018)<sup>16</sup> « De façon générale, on peut dire qu’une langue est menacée dans sa survie dès qu’elle n’est plus en état d’expansion, dès qu’elle perd de ses fonctions de communication dans la vie sociale ou n’est plus pratiquée quotidiennement pour les besoins usuels de la vie, dès qu’elle n’est plus rentable au plan économique, ou dès qu’il n’y a plus suffisamment de locuteurs pour en assurer la diffusion ».

<i>Les dialectes arabes</i>	Bambara	Goudjarati	Maltais	Somali
<i>ASM et ses variétés</i>	Bengali	Grec	Mehri	Songhay

<sup>13</sup> Site internet conçu par Jacques Leclerc <http://www.axl.cefan.ulaval.ca/>

<sup>14</sup> Par manque de données fiables, les dialectes arabes et les langues couchitiques, tchadiques et nigéro-congolaises ne sont pas répertoriées dans cette étude.

<sup>15</sup> Préalablement cité, voir note 3.

<sup>16</sup> Source [http://www.axl.cefan.ulaval.ca/Langues/2vital\\_mortdeslangues.htm](http://www.axl.cefan.ulaval.ca/Langues/2vital_mortdeslangues.htm)

<i>Langues couchitiques</i>	Berbère : 27 <sup>17</sup> langues dont Chaoui, Ghadamès, Mozabite, Nafusi, Sawknah, Siwi, Tachelih, T, Touarègues, etc.	Hébreu	Mushungulu	Soninké
<i>Langues tchadiques</i>	Biélorusse	Italien	Langues Nubiennes	Swahili
<i>Langues nigéro-congolaises</i>	Chinois mandarin	Jibbali	Ourdou	Tamoul
Adyguéen	Cingalais	Jiiddu	Pachtou	Tchéchéne
Afar	Circassien	Hausa	Penjabi	Télougou
Albanais	Comorien	Judéo-arabe	Persan	Tigrigna
Allemand	Coréen	Judéo-espagnol	Portugais	Turc
Amharique	Dabarré	Kenuzi-dongola	Poular	Turkmène
Araméen : Néo-araméen occidental, Néo-araméen oriental (dont syriaque, chaldéen..) Néo-araméen central,...	Domari	Kurde	Russe	Langues tsiganes
Arménien	Espagnol	Kurmandji	Serbe	Wolof
Azer	Filipino	Maay Maay	Sindhi	
Azerbaïdjanais	Français	Malayalam	Singhali	
Baloutchi	Garré	Malgache	Sokotori	

Tableau 2 – Répertoire provisoire des langues du monde arabophone

Leclerc recense les causes suivantes :

- Les conquêtes militaires
- La faiblesse numérique
- La domination socio-économique
- L'impuissance politique (dont les langues sans Etats)
- L'impérialisme culturel.

L'impact de ces facteurs est sans doute amplifié par l'instabilité géopolitique chronique du monde arabophone ; le couplage de ces phénomènes peut avoir un effet dévastateur sur le paysage sociolinguistique traditionnel. Les guerres qui se déchaînent d'une manière récurrente

<sup>17</sup> Répertoire par Glottolog <https://glottolog.org/resource/languoid/id/berb1260>

dans la région, et qui ont pour conséquence des populations décimées, ou bien des mouvements migratoires massifs internes ou externes, sont susceptibles de provoquer des bouleversements majeurs sur le plan linguistique. Ceux-ci peuvent se manifester sous plusieurs formes en fonction de diverses configurations. Marshak, qui a travaillé sur *Les représentations sociolinguistiques des variétés de l'arabe parlé à Beyrouth*, évoque « la restructuration de l'espace socioculturel de la capitale libanaise »<sup>18</sup> comme conséquence entre autres de la guerre qui a ravagé le Liban à partir de 1975. D'autres phénomènes sont à prendre en compte. Lorsque des mouvements massifs de migrations vers différentes destinations sont observés, cela peut se traduire par des évolutions très distinctes d'un substrat unique commun dans les diasporas, et l'éloignement des variétés linguistiques peut être particulièrement marqué que ce soit sur les plans phonétique, phonologique, lexical, morphosyntaxique ou sémantique. Ce phénomène n'est pas inquiétant en soi, puisqu'il rend compte de la vitalité d'une langue et de son adaptabilité y compris dans des contextes hostiles ou des situations *a priori* critiques. Dans des cas extrêmes de massacres, cela peut entraîner la disparition de variétés autochtones historiques minoritaires de telle ou telle zone géographique, ou bien une mise en péril extrême d'une variété minoritaire jusque-là relativement préservée. Dans d'autres cas encore, la préservation d'une variété de langue est possible grâce à un engagement fort dans la diaspora, ce qui éloigne la langue de son « berceau » d'origine : c'est le cas par exemple de l'araméen en Suède, avec une diaspora importante arrivée dans la deuxième moitié du XX<sup>e</sup> siècle et implantée principalement dans la région de Södetäljje, surnommée « Mésopotäljje, Suède moyen-orientale »<sup>19</sup>. Mais dans tous les cas, il ne faut perdre de vue le fait que « Depuis que l'espèce humaine parle, les langues n'ont cessé de se remplacer les unes les autres, de « mourir » et de « naître » pour rester dans cette métaphore. Ce qui inquiète en fait aujourd'hui, ce n'est pas que des langues disparaissent (cela a toujours été le cas) mais que leur nombre diminue, que ces disparitions ne soient pas compensées par des apparitions »<sup>20</sup>.

#### **4. Collecter et stocker des données langagières orales : les corpus sonores**

Les menaces qui pèsent sur l'existence des langues font que leur conservation peut devenir un vrai enjeu patrimonial. La question se pose de savoir sous quelle modalité, écrite et/ou orale, cette conservation peut-elle se faire. Jusqu'à la fin du XIX<sup>e</sup> siècle, seule la modalité graphique était envisageable, dans les projets ou dispositifs de préservation ou collecte ou conservation des données langagières. Avec les innovations technologiques qui ont accompagné la révolution industrielle, que ce soit l'invention du téléphone en 1876, du phonographe en 1877, ou de la photographie en 1880, une des conséquences les plus spectaculaires de cette révolution technologique dans le domaine des sciences du langage se traduira par l'apparition des 'sciences de la parole' : il devient enfin possible de « fixer » le son, de le stocker, de le manipuler ou de le décrire, que ce soit dans une visée d'archivage ou de recherche.

Dans son discours d'inauguration des Archives de la parole donné en 1911, archives accessibles au grand public grâce à la plateforme Gallica de la BNF<sup>21</sup>, Fernand Brunot déclare : « La création d'appareils qui enregistrent et reproduisent la voix humaine complète une série d'inventions qui a commencé le jour où l'homme a imaginé de dessiner un premier symbole

---

<sup>18</sup> Marshak, 2008, p. 85.

<sup>19</sup> Conte, 2015.

<sup>20</sup> Calvet, 2004, p. 288.

<sup>21</sup> <https://gallica.bnf.fr>



pour représenter sa pensée. Après l'écriture, après l'imprimerie, il restait encore un progrès essentiel à faire, car ni l'une ni l'autre ne fixent ni ne transmettent la parole dans son intégrité absolue »<sup>22</sup>. Les archives sonores ainsi constituées au début du XX<sup>e</sup>, et qui seront numérisées grâce à la BNF plusieurs décennies plus tard, ont bénéficié dans les premières années de leur constitution, de dons variés, tels que ceux de la firme Pathé, qui a mis à disposition des disques de son catalogue. Ces archives comportent des voix d'anonymes ou de célébrités, des poètes, des écrivains, des politiciens, etc. (Apollinaire ou Poincaré pour ne donner que ces deux exemples).

Progressivement, dans le courant du XX<sup>e</sup> siècle, on va voir se constituer des *corpus sonores* : « A corpus is a collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language »<sup>23</sup>. Des perspectives nouvelles seront offertes aux chercheurs grâce à l'émergence et au développement du numérique. En effet, compte tenu de la multiplicité de contraintes liées aux données sonores, qui sont bien plus longues à collecter et à traiter, et bien lourdes à stocker ou à archiver que les données écrites, le numérique apporte une réelle plus-value dans le champ des sciences du langage. On verra ainsi se développer pour le français des corpus linguistiques sonores de grande ampleur : sans prétendre à l'exhaustivité, citons le corpus du Laboratoire Ligérien de Linguistique consacré aux *enquêtes sociolinguistiques de la ville d'Orléans* ou ESLO<sup>24</sup>, collecté dans une première période entre 1968 et 1974 (ESLO 1) puis dans une deuxième étape à partir de 2008 (ESLO 2), et qui comprend des centaines d'heures d'enregistrement sonores qui permettent de décrire un « portrait sonore » d'une ville française à travers la production orale de plusieurs centaines d'orléanais s'exprimant dans une multitude de contextes. Citons également les bases de données sonores telle que PFC ou « Phonologie du français contemporain »<sup>25</sup> et son extension IPFC ou « Interphonologie du français contemporain »<sup>26</sup> : ces deux corpus ont pour objectif de décrire les variations observées à l'oral dans la langue française aussi bien telle qu'elle est réalisée par des locuteurs natifs dans plusieurs pays francophones (pour PFC), que par des locuteurs de français langue étrangère ou seconde (pour IPFC). Citons enfin les différents corpus multimodaux collectés et constitués par l'équipe ICAR, tel que le « Corpus de langues parlées en interaction » (CLAPI)<sup>27</sup> qui a l'avantage d'intégrer la composante vidéo, et qui permet de mener des analyses multidimensionnelles du langage parlé en interaction, dans des contextes personnels et professionnels variés.

Pour l'arabe, il existe quelques rares corpus ou bases de données sonores (voir Groupe ARAPI, 2018a, pour une revue du web répertoriant des corpus ou bases de données sonores librement accessibles). Un corpus d'« arabe parlé en interaction » est en cours de constitution dans le cadre de l'IFPO (institut français du Proche-Orient) par le même Groupe ARAPI, constitué de chercheuses confirmées dans le domaine de l'analyse des interactions<sup>28</sup>.

---

<sup>22</sup> Source : <https://gallica.bnf.fr/html/und/enregistrements-sonores/enregistrements-la-sorbonne-paris-1911-1914>  
<https://gallica.bnf.fr/ark:/12148/bpt6k1310983j/f2.media>

<sup>23</sup> Sinclair, 1996.

<sup>24</sup> <http://eslo.huma-num.fr/>

<sup>25</sup> Durand & al., 2002.

<sup>26</sup> Detey & al., 2010, Racine & al., 2012.

<sup>27</sup> CLAPI, <http://clapi.icar.cnrs.fr>

<sup>28</sup> Traverso, Pinon, Dimachki et Choueiri ; voir Groupe ARAPI 2018a, 2018b.



Les corpus sont en général conçus selon des protocoles expérimentaux spécifiques et ciblés en fonction des objectifs des projets de recherche dans le cadre desquels ils sont constitués. Ils sont numérisés, transcrits, codés, et accompagnés de métadonnées plus ou moins fournies : ce sont donc des ressources particulièrement précieuses aussi bien pour des études pluridimensionnelles et pluridisciplinaires des échantillons des langues considérées, qu'en tant que composante d'un patrimoine culturel immatériel représentatif d'une époque, d'un lieu, d'une communauté linguistique, d'usages ou d'un état de langue, entre autres facteurs.

### **5. Intégrer la dimension spatiale : les atlas linguistiques**

Parallèlement aux corpus, les atlas linguistiques sont des outils de stockage et de sauvegarde des données langagières particulièrement intéressants car ils ont l'avantage d'intégrer l'espace comme donnée fondamentale. Comme l'écrit Séguy (1973), dans le cas des atlas linguistiques, la représentation des faits langagiers se fait « sous forme de cartes géographiques » à travers des enquêtes de terrain dans lesquelles des informateurs sont interrogés dans de multiples lieux préalablement identifiés. Pour Séguy, « les atlas linguistiques constituent un département de l'histoire, limitrophe de l'archivistique et de la muséographie ; une autre version du musée ethnographique et surtout du musée de la parole, mais conçue sur un plan tout autre, collectionnant des faits bien plus nombreux et reflétant une prospection géographique plus serrée »<sup>29</sup>. La dimension patrimoniale des atlas linguistiques est d'autant plus importante qu'elle permet de rattacher des variétés à des territoires.

Aboutissement de recherches engagées dès le XIX<sup>e</sup> siècle, le premier atlas linguistique a vu le jour au début du XX<sup>e</sup> siècle : l'*Atlas linguistique de la France (Paris, 1902-1910)* de Jules Gilliéron<sup>30</sup> fixe un cadre méthodologique et de collecte de données à partir d'enquêtes sur le terrain, et ouvre la voie à l'élaboration de multiples atlas linguistiques des régions de France.

Dans le monde arabophone, après plusieurs tentatives allant dans le sens de la constitution d'atlas linguistiques (Bergsträßer en 1915, puis Cantineau en 1940<sup>31</sup>), un premier atlas des dialectes arabes d'Égypte est édité en 1983 par Behnstedt et Woidich, et ouvre la voie à des travaux similaires de Behnstedt sur d'autres zones géographiques, telles que le Yémen en 1985 ou la Syrie en 1997<sup>32</sup>.

Les atlas existent traditionnellement sous une forme écrite et scripturale, mais des atlas linguistiques sonores voient le jour depuis quelques années. Citons « l'atlas parlant » du *ladin des Dolomites et des dialectes limitrophes*<sup>33</sup> de l'université de Salzbourg, qui a été en gestation dès le milieu des années 1980, et qui a fortement inspiré l'atlas parlant VIVALDI (VIVAio Acustico delle Lingue e dei Dialetti d'Italia)<sup>34</sup>. Citons également « l'atlas sonore des langues régionales de France »<sup>35</sup> basé sur des échantillons sonores d'une fable en français et dans plusieurs langues régionales, qui s'est élargi « aux langues et dialectes italiens »<sup>36</sup> en partenariat avec l'université de Turin. Les protocoles expérimentaux adoptés par les chercheurs sont extrêmement variés, et les données collectées de nature diverse : poèmes lus ou enquêtes par

---

<sup>29</sup> Séguy, 1973, p. 85.

<sup>30</sup> Voir Séguy note précédente.

<sup>31</sup> Voir Behnstedt & Woidich, 2013, p. 306.

<sup>32</sup> Voir Behnstedt & Woidich (2013 : 300-325) pour une revue de la littérature.

<sup>33</sup> <http://ald.sbg.ac.at/ald/ald-i/index.php?id=0002&lang=fr>

<sup>34</sup> Bauer, 2010.

<sup>35</sup> Boula de Mareüil & al., 2017, <https://atlas.limsi.fr/>

<sup>36</sup> <https://atlas.limsi.fr/?tab=it>

questionnaires, les échantillons audio, qu'ils s'accompagnent ou non de métadonnées incluant la transcription phonétique et le codage, rendent compte de la variabilité phonétique et phonologique (entre autres) en fonction d'un ancrage dans un espace géographique donné, clairement identifié, et qui est considéré comme facteur de variation.

## **6. Conclusion : pour un atlas numérique sonore des langues vernaculaires du monde arabophone (ANSOLVAR)**

Les organisateurs du colloque PPN « de la pierre au papier, du papier au numérique » dont ces actes sont issus ne s'y sont pas trompés à travers le titre de la session dans laquelle cette communication a été proposée : la langue est bien le « premier des patrimoines ». Et c'est bien dans une perspective de préservation des langues en tant que patrimoine oral culturel immatériel, dans une zone géographique particulièrement instable sur le plan géopolitique, et dans laquelle les langues sont soumises à des tensions particulièrement fortes, qu'il nous semble opportun de poser les bases d'un atlas numérique sonore des langues vernaculaires du monde arabophone.

Quelques principes fondamentaux guident notre action à ce stade embryonnaire dans la conception du projet.

Tout d'abord, il serait souhaitable qu'un tel projet fédère de multiples compétences scientifiques, afin qu'il soit un projet porté par un collectif, au service du patrimoine oral immatériel. La valorisation et la préservation de la diversité linguistique des usages nous semble un fondement absolu à poser. Ensuite, s'inspirant de projets anglo-saxons tels que CHILDES<sup>37</sup>, le « Child Language Data Exchange System », il serait judicieux que ce projet soit de nature participative et ouverte, de manière à ce qu'il bénéficie d'une réelle dynamique internationale : un protocole méthodologique de collecte de données pourrait être mis à disposition de la communauté scientifique, et la plateforme pourrait idéalement être conçue de sorte que le téléversement des données langagières collectées sur la base d'un protocole méthodologique unique soit accessible à tous. En outre, il serait utile que l'outil soit conçu de manière dynamique, évolutive et non figée, que ce soit au niveau des données sonores ou des métadonnées elles-mêmes. Il sera important d'anticiper largement et bien en amont l'extrême diversité sociolinguistique du monde arabophone. Enfin, une dimension diachronique pourrait être utilement intégrée, de manière à laisser la voie ouverte à des études longitudinales dont on connaît l'importance pour la prise en compte de la vitalité des langues.

### **Bibliographie**

Abou Haidar, Laura, *De la linguistique à la didactique, regards croisés en phonétique. Oralité – Variabilité – Corpus*. Mémoire d'Habilitation à diriger des recherches, Université Grenoble Alpes, Grenoble, 2018.

Baude, Olivier & Alessio, Michel, « Les corpus de la parole : patrimoine immatériel et langues de France », *Culture et recherche*, Ministère de la Culture et de la Communication, Paris, 2008, p. 42-43.

Bauer, Roland, « Le projet VIVALDI : présentation d'un atlas linguistique virtuel », in Aurrekoetxea, Gotson & Ormaetxea, Jose Luis (dir.), *Tools for Linguistic Variation*, Universidad del País Vasco/Euskal Herriko Unibertsitatea, Bilbao, 2010, p. 71-88.

---

<sup>37</sup> <https://talkbank.org/share/principles.html>

- Behnstedt, Peter & Woidich, Manfred, « Dialectology », in Owens, Jonathan (ed.), *The Oxford Handbook of Arabic Linguistics*, Oxford University Press, Oxford, 2013, p. 300–325.
- Boula de Mareuil, Philippe, Vernier, Frédéric, Rilliard, Albert, « Enregistrements et transcriptions pour un atlas sonore des langues régionales de France », *Géolinguistique* 17, 2017, p. 23-48
- Calvet, Luis-Jean, « La diversité linguistique : quel enjeu pour la francophonie ? », *Hermès, La Revue* 4/40, CNRS Editions, Paris, 2004, p. 287-293.
- Cohen, David, *Etudes de Linguistique Sémitique et Arabe*, Mouton, The Hague-Paris, 1970.
- Conte, Maria Laura, « Bienvenue à Mesopotamie, Suède moyen-orientale », *Oasis. Chrétiens et musulmans dans le monde global* n°22, 2015.
- Detey, Sylvain & Racine, Isabelle, « Interphonologie, corpus et français langue étrangère : le projet IPFC », *1<sup>ère</sup> Journée IPFC*, Paris-Maison des Sciences de l'Homme, en ligne : 2010, <https://www.projet-pfc.net/2010/03/09/ipfc/>
- Grigore, George & Bituna, Gabriel (éds.), « Arabic Varieties: Far and Wide », *Proceedings of the 11th International Conference of AIDA 2015*, Editura Universității din București, Bucarest, 2016, en ligne, <https://aidabucharest2015.ils.unibuc.ro/conference-archives/conference-information/previous-aida-conferences/>
- Groupe ARAPI (Véronique Traverso, Catherine Pinon, Loubna Dimachki, Lina Choueiri), « Corpus d'arabe parlé (1) : Quels corpus d'arabe parlé en libre accès ? », *Les Carnets de l'Ifpo*, en ligne : 2018, <http://ifpo.hypotheses.org/8865>
- Groupe ARAPI (Véronique Traverso, Catherine Pinon, Loubna Dimachki, Lina Choueiri), « Corpus d'arabe parlé (2) : Contraintes et problèmes liés à la réalisation des corpus d'arabe parlé en interaction », *Les Carnets de l'Ifpo*, en ligne : 2018, <http://ifpo.hypotheses.org/9039>
- Leclerc, Jacques, *L'aménagement linguistique dans le monde*, en ligne : 2019, <http://www.axl.cefano.ulaval.ca/>
- Marshak, Mikhail, « Représentations sociolinguistiques des variétés de l'arabe parlé à Beyrouth : observations préliminaires », in Mermier, Franck (dir.), *Liban, espaces partagés et pratiques de rencontre*, Publications de l'Institut français du Proche-Orient, Beyrouth, 2008, p. 85-100.
- Miller, Catherine, « Langues et Médias dans le monde arabe/arabophone. Entre idéologie et marché, convergences dans la glocalisation », en ligne : 2010, [halshs-00578851](https://halshs.archives-ouvertes.fr/halshs-00578851)
- Mitchell, T. F., « Educated Spoken Arabic in Egypt and the Levant, with special reference to participle and tense », *Journal of Linguistics* 14, 1979, p. 227-259.
- Owens, Jonathan (éd.), *The Oxford Handbook of Arabic Linguistics*, Oxford University Press, Oxford, 2013.
- Racine, Isabelle, Detey, Sylvain, Zay, Françoise & Kawaguchi, Yuji, « Des atouts d'un corpus multitâches pour l'étude de la phonologie en L2 : l'exemple du projet « Interphonologie du français contemporain » (IPFC) », in Kamber, Alain & Skupien Dekens, Carine (dir.), *Recherches récentes en FLE*, Peter Lang, Berne, 2012, p. 1–19.
- Salam, Fathi, *Espace Acoustique et Patrons Coarticulatoires : les Voyelles de l'Arabe Libyen de Tripoli en Contexte Pharyngalisé*, Thèse de Doctorat, université de Franche-Comté, Besançon, 2012.
- Séguy, Jean, « Les Atlas linguistiques de France par régions », *Langue française* 18, 1973, p. 65-90.

Sinclair, John, « Preliminary recommendations Corpus Typology », en ligne : 1996  
<http://www.ilc.cnr.it/EAGLES/corpus/corpus.html>.

UNESCO, Convention de 2003 pour la sauvegarde du patrimoine culturel immatériel, en ligne :  
2003, <https://ich.unesco.org/doc/src/18440-FR.pdf>

Versteegh, Kenneth, *The Arabic Language*, Edinburgh University Press, Edinburgh, 1997.

Ziamari, Karima & De Ruyter, Jan Jaap, « Les langues au Maroc : Réalités, changements et évolutions linguistiques », in Dupret, Baudoin, Rhani, Zakaria, Boutaleb, Assia et Ferrié, Jean-Noël (dir.), *Le Maroc au présent. D'une époque à l'autre, une société en mutation*, Centre Jacques Berque, Rabat, 2015.