Transcrire un corpus audio dans la perspective de la préservation du patrimoine culturel immatériel

Les obstacles rencontrés, les méthodologies retenues, les questions restées en suspens

Marie-Madeleine BERTUCCI(1),
Mounia ILLOURMANE(2)

(1) (2) Cy Cergy Paris université

Résumé de Marie-Madeleine Bertucci : L'article a pour objet « l'étude de la mise en mots de la mémoire des « Chibanis » du quartier de la Goutte d'or à Paris » à partir d'un corpus d'entretiens biographiques audio. Ces entretiens ont été réalisés par Jean-Marc Bombeau de 2010 à 2012, dans les locaux de la galerie associative l'Échomusée, située rue Cavé dans le XVIIIe arrondissement de Paris, avec un groupe de Chibanis ou anciens du quartier. Cette étude discursive mémorielle est le sujet de la thèse de Mounia Illourmane. Ces entretiens font émerger la mémoire de ces acteurs sociaux en les conduisant à jeter un regard rétrospectif et réflexif sur leur passé, de leur arrivée à Paris dans les années soixante, dans le cadre d'une immigration de travail jusqu'à leur retraite qui prend effet dans la période contemporaine. Ils autorisent l'émergence d'un processus de patrimonialisation de données mémorielles qui constituent le socle d'un patrimoine immatériel. Les entretiens ont visé à la fois le passé et le présent et ont cherché à mettre en lumière, à travers le travail sur la mémoire, le retour rétrospectif sur le passé, l'historicité du présent et les types de liens sociaux tissés à travers l'expérience de l'exil. Au plan théorique, la réflexion s'inscrit dans la perspective des *cultural* et des *subaltern studies* pour mettre en évidence le caractère fragmentaire et discontinu de ces récits mémoriels par le biais d'une étude du lexique privilégié par les énonciateurs. Le patrimoine culturel immatériel, porté par la convention de l'UNESCO du 17 octobre 2003, et sa conservation supposent deux types de questionnements. D'une part, il s'agit de concevoir différemment les politiques patrimoniales, en les envisageant du point de vue des détenteurs du patrimoine et non plus seulement du point de vue des experts, d'autre part d'élaborer des modalités de préservation des données autres que celles du patrimoine matériel, ce sur quoi va porter l'article.

Résumé de Mounia Illourmane : Ce sujet m'a été proposé par ma directrice de thèse Marie-Madeleine Bertucci. Jean-Marc Bombeau a mis à notre disposition un ensemble de données brutes (séquences vidéo, bandes son et photos), sauvegardées sur un disque dur. Avec ma directrice de recherche, nous avons conçu un outil de classement, d'indexation et d'arborescence des données pour déterminer la nature et la quantité des matériaux à exploiter. À partir d'une démarche muséale et ethnographique, il s'agit de transcrire un corpus audio dans une perspective de conservation d'archives et de valorisation de données peu accessibles

matériellement et s'inscrivant dans le cadre du patrimoine culturel immatériel. Au cours de cette tâche néanmoins, sont apparues certaines difficultés d'ordre technique. Étant donné que le corpus n'est pas destiné à cet effet, la prise de son était artisanale ce qui a rendu la transcription plus complexe. Face à des locuteurs non natifs avec un accent prononcé, les logiciels de reconnaissance vocale et de transcription automatique sont inadaptés. Par conséquent, le choix des outils s'est porté sur deux logiciels informatiques, « Express Scribe » et « E-Speaking ». Le premier est un logiciel professionnel de lecture audio conçu pour faciliter la transcription des enregistrements audio. Grâce à cet outil on peut contrôler la vitesse de lecture audio et avoir ainsi une lecture à vitesse variable. « E-Speaking » est un logiciel de contrôle et de reconnaissance vocale qui a pour fonction la dictée et l'exécution des tâches sur ordinateur avec la voix de l'utilisateur. On l'a utilisé notamment pour accélérer le travail de transcription des bandes son. L'emploi de ces deux outils a permis la transcription en restant le plus fidèlement possible à l'original tout en tenant compte des nombreuses contraintes d'ordre syntaxique, orthographique et phonologique résultant du caractère oral des données.

Mots-clés : corpus audio, transcription, reconnaissance vocale, indexation, archivage, conservation du patrimoine immatériel.

Abstract of Marie-Madeleine Bertucci*: The purpose of the paper is "the study of the mise en mots of the memory of the "Chibanis" of the Goutte d'or district in Paris" based on a corpus of audio biographical interviews conducted by Jean-Marc Bombeau, in the premises of the Echomusée gallery, located rue Cavé in the 18th arrondissement of Paris, with a group of Chibanis or former members of the district from 2010 to 2012. This discursive memory study is the subject of Mounia Illourmane's thesis. These interviews bring out the memory of these social actors by leading them to take a retrospective and reflexive look at their past, from their arrival in Paris in the 1960s, in the context of labour immigration, to their retirement which takes effect in the contemporary period. They allow the emergence of a process of patrimonialization of memory data, which constitutes the basis of an intangible heritage. The interviews focused on both the past and the present and sought to highlight, through the work on memory, the retrospective look back at the past, the historicity of the present and the types of social ties forged through the experience of exile. On the theoretical level, the reflection is part of the perspective of cultural and subordinate studies to highlight the fragmentary and discontinuous nature of these memory narratives through a study of the lexicon favoured by enunciators. The intangible cultural heritage, supported by the UNESCO Convention of 17 October 2003, and its conservation involve two types of questioning. On the one hand, it is a question of conceiving heritage policies differently, by considering them from the point of view of heritage holders and no longer only from the point of view of experts, and on the other hand of developing methods for preserving data other than those of the tangible heritage, which is what the paper will cover.*

Abstract of Mounia Illourmane*: This subject was proposed to me by my thesis supervisor Marie-Madeleine Bertucci. Jean-Marc Bombeau provided us with a set of raw data (video sequences, soundtracks and photos), saved on a hard disk. With my research director, we developed a data classification, indexing and tree structure tool to determine the nature and quantity of materials to be used. Based on a museum and ethnographic approach, it is a question of transcribing an audio corpus from the perspective of archive conservation and the enhancement of data that are not easily accessible in material terms and that are part of the intangible cultural heritage. Nevertheless, during this task, some technical difficulties appeared, since the corpus is not intended for this purpose, the sound recording is artisanal, which made the transcription more complex. In the case of non-native speakers with a pronounced accent, speech recognition and automatic transcription software are inadequate. Consequently, the choice of tools was made between two computer programs, "Express Scribe" and "E-Speaking". The first is a professional audio playback software designed to facilitate the transcription of audio recordings. Thanks to this tool we can control the speed of audio playback as well as have a variable speed playback. "E-Speaking" is a speech control and recognition software that is used to dictation and perform tasks on a computer with the user's voice. It has been used, in particular, to speed up the transcription of soundtracks. The use of these two tools allowed the transcription to be made as close as possible to the original while taking into account the many syntactic, orthographic and phonological constraints resulting from the oral nature of the data.*

Keywords*: audio corpus, transcription, voice recognition, indexing, archiving, conservation of intangible heritage.*

L'article a pour objet l'étude discursive mémorielle[1] « de la mise en mots de la mémoire des « Chibanis » du quartier de la Goutte d'or à Paris » effectuée à partir d'un corpus d'entretiens biographiques audio réalisés par Jean-Marc Bombeau, dans les locaux

[1] Cette étude est le sujet de la thèse de Mounia Illourmane. Elle s'inscrit dans le travail conduit dans le sous-axe *Contextes plurilingues de l'espace francophone* du laboratoire Lexique Textes Discours Dictionnaires de Cy Cergy Paris Université.

Published in : Louise Ouvrard et Odile Farge (dir.), *Corpus audiovisuels*, Editions des archives contemporaines, ISBN : 9782813003799, doi : 10.17184/eac.9782813003799

de la galerie associative l'Échomusée[2] située dans le XVIII[e] arrondissement de Paris, avec un groupe de Chibanis ou *anciens* du quartier, de 2010 à 2012. Ces entretiens font émerger la mémoire de ces acteurs sociaux en les conduisant à jeter un regard rétrospectif et réflexif sur leur passé, de leur arrivée à Paris dans les années soixante, dans le cadre d'une immigration de travail jusqu'à leur retraite qui prend effet dans la période contemporaine. Ils autorisent l'émergence d'un processus de patrimonialisation de données mémorielles, qui constituent le socle d'un patrimoine immatériel. Les entretiens ont visé à la fois le passé et le présent et ont cherché à mettre en lumière, à travers le travail sur la mémoire, le retour rétrospectif sur le passé, l'historicité du présent et les types de liens sociaux tissés à travers l'expérience de l'exil. Au plan théorique, la réflexion s'inscrit dans la perspective des *cultural* et des *subaltern studies* pour mettre en évidence le caractère fragmentaire et discontinu de ces récits mémoriels par le biais d'une étude du lexique employé par les énonciateurs. Le patrimoine culturel immatériel, porté par la convention de l'UNESCO du 17 octobre 2003[3], et sa conservation supposent deux types de questionnements. D'une part, il s'agit de concevoir différemment les politiques patrimoniales, en les envisageant du point de vue des détenteurs du patrimoine et non plus seulement du point de vue des experts, d'autre part d'élaborer des modalités de préservation des données autres que celles du patrimoine matériel, ce sur quoi va porter ce chapitre. Cette recherche a donc pour objet de sauvegarder une expression culturelle encore vivante mais néanmoins menacée, la mémoire des Chibanis, et de mettre à la disposition de la communauté scientifique[4] un ensemble de données peu accessibles matériellement et s'inscrivant dans le cadre du patrimoine culturel immatériel. Par ailleurs, ce projet s'inscrit dans le contexte de la réflexion menée actuellement en SHS sur la constitution de corpus d'archives numériques et de la pérennisation de leur accessibilité par la numérisation et l'indexation des sources.

L'article se développera en deux temps. Dans un premier temps, Marie-Madeleine Bertucci présentera la problématique générale des modalités de conservation du patrimoine mémoriel d'une communauté minorée, les Chibanis du quartier de la Goutte d'Or dans une perspective sociolinguistique. Dans un second temps, Mounia Illourmane abordera les questions liées à la transcription du corpus qui se sont posées au fil de l'avancée de son travail de transcription[5].

Marie-Madeleine Bertucci[6]

[2]https://gouttedor-et-vous.org *Echomusee* L'adresse postale de l'Échomusée se situe 21, rue Cavé, 75018 Paris.

[3]On se reportera au site de l'UNESCO sur ce point : www.unesco.org

[4]Sans exclure de les communiquer à un plus large public si la demande se manifeste.

[5]Par convention, il a été décidé que lorsque Mounia Illourmane décrirait dans cet article ce qui relève de sa démarche de recherche personnelle et des initiatives qu'elle a prises de son propre chef, elle utiliserait la première personne du singulier, à savoir *je*. Quand il s'agira d'une réflexion menée conjointement avec Marie-Madeleine Bertucci, elle utilisera *nous*.

[6]Par convention, de manière à distinguer ce qui relève de la réflexion générale dans laquelle s'inscrit le travail de thèse de Mounia Illourmane, laquelle est prise en charge par Marie-Madeleine Bertucci et le travail conduit par Mounia Illourmane dans le cadre de sa thèse, il a été décidé d'indiquer les noms en tête des parties respectivement rédigées par l'une et l'autre des chercheures et les deux noms pour la conclusion élaborée en commun.

Published in : Louise Ouvrard et Odile Farge (dir.), *Corpus audiovisuels*, Editions des archives contemporaines, ISBN : 9782813003799, doi : 10.17184/eac.9782813003799

1 Problématique générale des modalités de conservation du patrimoine mémoriel d'une communauté minorée, les Chibanis du quartier de la Goutte d'Or

1.1 Position du problème. L'émergence du sujet migrant à travers le prisme de la mémoire

Le patrimoine culturel immatériel désormais (PCI), porté par la convention de l'UNESCO du 17 octobre 2003, et sa conservation supposent deux types de questionnements. D'une part, il s'agit d'élaborer des modalités de préservation des données autres que celles du patrimoine matériel, d'autre part, le PCI suppose de concevoir différemment les politiques patrimoniales, en les envisageant du point de vue des détenteurs du patrimoine et non plus seulement du point de vue des experts. Pour ce faire, on soulignera d'emblée la nécessité de faire émerger une mémoire de l'immigration qui permette à la fois de changer le regard sur les migrants et d'éviter l'obstacle de l'enfermement dans des récits mémoriels qui constitueraient une assignation identitaire clivante, au sujet de laquelle on fera l'hypothèse qu'elle est un frein à l'intégration.

Les entretiens, qui constituent le corpus permettent de mettre en mots la mémoire de ces existences modestes et de les rendre visibles et audibles. Ils constituent un ensemble de connaissances, un savoir sur une forme d'immigration discrète, voire invisible, alors que cette génération de migrants a été le témoin de transformations importantes et notamment de l'émergence d'une néo-traditionnalisation des mœurs, voire d'un retour du religieux chez les jeunes générations. L'un des intérêts des entretiens, et non des moindres, est de mettre en évidence ces effets de contraste entre ces deux générations. Par ailleurs, les entretiens constituent un témoignage sur le quartier de la Goutte d'Or et montrent en quoi celui-ci est un lieu à la fois privé et public auquel s'attachent les mémoires de ces migrants, et qui fait l'objet de discours. Ils constituent une méthodologie pertinente pour appréhender certaines trajectoires de migrants et pour faire émerger leur degré d'insertion dans la société française contemporaine à travers les types de liens sociaux qu'ils ont noués. La recherche tentera de mettre en évidence la présence d'éléments tangibles récurrents dans les entretiens, qui prennent la valeur de traces ou de supports de mémoire, comme la valeur des lieux, les formes de vécu, le potentiel émotionnel et thématique qui s'y rapporte (Bertucci, 2009, 2017). On fera l'hypothèse que l'identité des migrants se dit à travers des lieux et thèmes (topoï) privilégiés : par exemple, le rapport à la langue et aux lieux d'origine et le rapport à la langue française (Bertucci, 2012 : 97), la mobilité, l'arrivée en France, l'insertion professionnelle et l'intégration, les liens avec les autres migrants, avec les Français, l'hypothèse du retour au pays, la mémoire et l'oubli du passé… (Bertucci, 2006 : 128). On postulera donc des similitudes dans l'expérience des sujets concernés (Bertucci, 2014 : 43).

1.2 Les entretiens biographiques

On supposera qu'une impulsion narrative est sous-jacente aux entretiens et qu'elle fait apparaître chez les informateurs un désir de se représenter en se racontant (Bertucci, 2014 : 42). Les informateurs se situent parfois dans des situations de marginalité, voire de conflit ou de crise et sont confrontés à l'altérité. On peut penser que cette

Published in : Louise Ouvrard et Odile Farge (dir.), *Corpus audiovisuels*, Editions des archives contemporaines, ISBN : 9782813003799, doi : 10.17184/eac.9782813003799

situation de distance peut les amener à mettre en mots et à relier des éléments isolés de leur histoire. Les entretiens biographiques rendent possible l'émergence d'un sujet qui ne peut être visible que par un détour par sa vie personnelle, laquelle en s'articulant au champ du social, fait émerger une forme spécifique d'identité. Or, l'entretien biographique, donne à lire la biographie des individus. Il est un moyen d'accéder à la connaissance de l'Autre parce qu'il met en place un ordre du sens (Adam, 1995) et lui donne une forme qui fait sens. En cela, il est particulièrement apte à exprimer l'altérité. Se saisissant comme objet et *quasi personnage*, le narrateur ou *quasi narrateur* (Ricœur, 1983 : 320) se pose à la fois en sujet et en objet, donnant de ce fait au lecteur ou à l'auditeur une possibilité d'appréhender des éléments identitaires au nombre desquels figurent les lieux selon notre hypothèse. La finalité de la recherche est de mettre en lumière l'expérience vécue par le sujet et le type d'identité qui en résulte dans une démarche herméneutique (Bertucci, 2014). Le processus réflexif est au centre de la méthodologie dans la mesure où il permet la construction et la représentation de l'Autre, la confrontation avec lui dans un espace social, historicisé et contextualisé. Cette approche paraît d'autant plus pertinente lorsque la recherche porte, comme ici, sur des identités potentiellement clivées entre les langues, les cultures, les appartenances, car la dynamique réflexive de l'entretien biographique amène les informateurs à construire le sens par eux-mêmes (Bertucci, 2012 : 95). On admettra également que l'organisation des discours mémoriels, malgré la diversité des informateurs, n'est pas forcément aléatoire et qu'il existe un ordre narratif signifiant, résultant d'une expérience de l'espace et du temps singulière, qui est celle de la mobilité, qu'elle soit exil ou migration (Bertucci, 2012 : 95). Enfin, on supposera que, de par leur caractère rétrospectif, les discours des informateurs impliquent un processus mémoriel particulier signifiant par ses souvenirs mais aussi par ses silences et qu'il y a une part de l'histoire du narrateur qui n'est pas remémorée (Bertucci, 2014 : 43).

1.3 Principes méthodologiques

La démarche prendra appui de manière très souple sur les méthodes mises en œuvre par la socio-histoire (Noiriel, 2006 a), les *cultural studies* (Thompson, 1963), les *subaltern studies* (Merle, 2004) avec l'objectif d'aller vers une conception de l'histoire perçue comme « une histoire anthropologique qui s'intéresse surtout aux liens que tissent les individus entre eux dans le cadre de leur vie quotidienne » (Noiriel, 2006 b : 40). Elle s'appuiera également sur une forme d'observation participante et une typologie des médiations sociales (Ferrarotti, 1990 : 61).

Cette démarche suppose un engagement particulier de l'observateur, une réflexivité en retour, par la pratique d'une observation participante, supposant un décentrement socioculturel et ce que Lévi-Strauss appelle une « distanciation sociale » (1958 : 416). Comprendre l'Autre et le connaître implique de se distancier de soi pour appréhender son expérience de l'exil, interprétée comme une expérience de rupture avec la culture d'origine. Les entretiens biographiques peuvent être interprétés comme trace de cette expérience. On postulera donc qu'un individu reflète un ou plusieurs pôles sociaux et inversement, que le système social se projette vers l'individu et ce à travers des médiations sociales comme on l'a vu plus haut. Chaque comportement individuel est à la fois la synthèse d'une structure sociale et une individualisation de l'histoire sociale

Published in : Louise Ouvrard et Odile Farge (dir.), *Corpus audiovisuels*, Editions des archives contemporaines, ISBN : 9782813003799, doi : 10.17184/eac.9782813003799

collective (Ferrarotti, 1990 : 21). Il est la représentation, l'attestation de l'altérité, et donc il affirme ce que Bakhtine appelle la relation *exotopique* (1978), laquelle rend possible de poser en tant que telle l'existence de l'Autre et son aptitude à parler en son nom.

Le cadre théorique et la démarche méthodologique ayant été posés, il est désormais possible d'aborder dans un second temps les différents problèmes rencontrés par Mounia Illourmane au fil de l'avancée de la transcription du corpus.

Mounia Illourmane

2 La transcription du corpus. Les obstacles rencontrés, les méthodologies retenues, les questions restées en suspens

« La mise en mots de la mémoire des Chibanis du quartier de la Goutte d'Or à Paris » est un sujet qui m'a été proposé par Marie-Madeleine Bertucci, ma directrice de thèse. L'idée est de sauvegarder et de mettre en mots la mémoire de ces « vieux immigrés », en partant d'un corpus d'archives audio, de documents authentiques et d'enregistrements vocaux qui n'ont fait l'objet d'aucune manipulation au préalable. En effet, l'identification de certaines pistes de recherches est une étape qui s'avère nécessaire, suite à l'exploitation des données transcrites avec l'objectif de mettre au point des méthodologies de recherches et des approches pour mieux appréhender le traitement et l'analyse du corpus.

2.1 Les données. Mise au point d'un outil de classement, d'indexation et d'arborescence

La récupération des données auprès de Jean-Marc Bombeau a eu lieu en décembre 2014 dans les locaux de la galerie de « l'Échomusée » dans le XVIII^{e} arrondissement de Paris. Il s'agit de documents authentiques, séquences vidéos, bandes son et photos, peu accessibles matériellement. Ces données sont brutes et représentent une manne de ressources et de matériaux pour la constitution de notre corpus.

Les données sont sauvegardées sur un disque dur. Afin de trouver une méthodologie d'organisation des données reçues, ma directrice de recherche et moi avons réalisé un pré-test sur un échantillon. Par la suite, j'ai conçu un outil de mise en ordre des données en utilisant une méthodologie d'inventaire qui consiste à établir une liste exhaustive des éléments figurant dans le disque dur, en élaborant un inventaire physique qui correspond à un comptage manuel de chaque type de document dont les procédés sont le regroupement, le classement, l'indexation et l'arborescence. Le disque dur est présenté ici comme le support de stockage des données à exploiter, sauvegardant ces dernières d'une manière hiérarchique, en subdivision. C'est le principe de l'arborescence : des dossiers peuvent contenir à leur tour d'autres sous-dossiers subdivisés. Ce schéma de stockage et d'arborescence peut être recréé autant de fois que nécessaire. Afin de déterminer avec exactitude la nature et la quantité des matériaux détenus, l'arborescence permet donc de voir la hiérarchisation du contenu. Il s'agit de schémas qui permettent de mieux appréhender l'organisation des données. L'inventaire

Published in : Louise Ouvrard et Odile Farge (dir.), *Corpus audiovisuels*, Editions des archives contemporaines, ISBN : 9782813003799, doi : 10.17184/eac.9782813003799

est donc composé de 132 pages format A4 avec 191 tableaux, présentés sous la forme suivante : colonnes de gauche à droite :

Document, type du document, remarque

Bande son, durée, remarque

Séquence vidéo, durée, remarque

Photo, remarque

À l'issue de cette démarche, nous avons sélectionné les données à exploiter dans notre étude, ce qui a orienté notre choix sur les bandes son.

2.2 Génération d'un corpus

La recherche à partir des corpus oraux est très répandue en sociolinguistique. C'est une démarche empirique traitant des données authentiques. L'étape suivante consiste à transcrire et à baliser le corpus. Les bandes son sont transcrites en orthographe standard. La partie transcrite comporte 213 pages et repose sur des enregistrements d'archives audio composés de 18 bandes son. La durée totale des enregistrements est de 6 heures et 58.32 minutes. La durée varie d'une bande son à une autre, la plus courte est de 03.58 minutes et la plus longue est de 1 heure et 2.04 minutes. Dans le corpus on distingue au total 33 « quasi personnages » (Ricœur, 1983 : 320)[7], dont 4 intervieweurs et 29 Chibanis. Toutefois, une anonymisation totale du corpus s'avère nécessaire pour la protection des données personnelles des Chibanis. De ce fait, nous avons choisi de ne pas constituer d'archive sonore. L'exploitation des bandes son se limite seulement à la version transcrite de celles-ci, ce qui sous-entend une non-utilisation de la version sonore. La voix constitue en effet une donnée personnelle identifiable.

2.3 Prise en main des logiciels

Il s'agit d'une transcription mot à mot : j'ai mis par écrit exactement tout ce qui est dit à l'oral incluant des phrases incomplètes, des reformulations et des intonations, on y lira aussi les remarques d'ordre général « tu voulais que je te filme aussi », les mots répétés à l'oral « avait des problèmes, des problèmes avec leur enfant ». Des arrêts fréquents sont nécessaires pour ajouter les éléments extralinguistiques comme les rires et les hésitations. La transcription manuelle reste donc incontournable pour notre corpus. Ce type de transcription a principalement pour but la réalisation d'une retranscription authentique, tout en respectant les enchaînements et les chevauchements de l'enregistrement oral : « Les transcriptions qu'utilisent les linguistes conservent soigneusement toutes les particularités des productions orales : répétitions, hésitations, amorces de mots, retouches » (Baude, 2006 : 30).

C'est une technique très utile mais coûteuse en temps : « Selon les estimations courantes, un minimum de trente minutes de travail est nécessaire pour transcrire une minute d'enregistrement » (Baude, 2006 : 30). Cette opération nécessite l'utilisation d'un lecteur audio et d'un traitement de texte, entre lesquels il est essentiel de faire

[7] Voir plus haut sur ce point.

Published in : Louise Ouvrard et Odile Farge (dir.), *Corpus audiovisuels*, Editions des archives contemporaines, ISBN : 9782813003799, doi : 10.17184/eac.9782813003799

des va-et-vient successifs de *play/stop*, pour les transcrire par saisie sur l'éditeur de texte « Word Office ». Néanmoins, j'ai également utilisé les logiciels « Express Scribe » et « E-Speaking ».

2.3.1 Express Scribe[8]

Express Scribe est un logiciel professionnel de lecture audio qui permet de lire des fichiers audio au format Wav ou MP3. Ce programme ne permet pas la conversion automatique des fichiers audio en texte, mais il contient d'autres fonctionnalités pour rendre le processus plus facile, telle que la transcription, outil conçu pour faciliter la transcription des enregistrements audio au clavier dans un éditeur de texte. La fonction de contrôle permet de maîtriser la vitesse de lecture, qui se gère entièrement à partir des touches d'accès rapides du clavier. On peut sauvegarder, lire ou revenir en arrière en utilisant les touches retour, avance rapide. Le logiciel dispose d'une option de chargement automatique des fichiers audio. Elle est accessible à partir d'un périphérique (lecteur cd ou dvd, clé USB…) ou d'un réseau local.

2.3.2 E-Speaking[9]

E-Speaking est un logiciel *shareware*[10] de contrôle et de reconnaissance vocale, disponible en téléchargement. Il a pour fonction la dictée et l'exécution de tâches sur ordinateur par la voix de l'utilisateur. Il assure de meilleurs résultats en s'intégrant à Word Office et contient également des commandes intégrées comme *Voice Dictation* pour la dictée. Celle-ci consiste à reformuler sous forme intelligible et grammaticalement correcte les propos de l'utilisateur.

2.3.3 Pourquoi l'emploi de ces deux outils ?

Olivier Baude note que : « Les systèmes de transcription doivent être adaptés en fonction des corpus à transcrire » (2006 : 31). Partant des différentes difficultés techniques rencontrées, liées à la transcription du corpus, nous abordons ici certaines solutions pour y remédier. L'emploi de ces deux logiciels a permis la transcription des enregistrements des Chibanis en restant le plus fidèle possible à l'original, tout en tenant compte des nombreuses contraintes d'ordre syntaxique, orthographique et phonologique résultant du caractère oral des données.

2.4 Difficultés rencontrées

La transcription des entretiens présente des difficultés d'ordre pratique. La première difficulté est liée au caractère oral des données d'un point de vue syntaxique, orthographique et phonologique. Les logiciels de transcription automatique ne correspondent pas à notre corpus, il a donc fallu tirer profit des logiciels à notre disposition. Cependant nous avons été obligées de recourir à un traitement majoritairement manuel, même si ces logiciels facilitent grandement la tâche de transcription des données. Afin

[8] L'URL du site est : https ://www.nch.com.au/scribe/fr/index.html

[9] On se reportera à l'URL suivante pour le site : https ://e-speaking.com/, ainsi qu'à l'adresse : https ://www.nchsoftware.com/fr/index.html

[10] Payant, disponible avec souscription à un abonnement.

Published in : Louise Ouvrard et Odile Farge (dir.), *Corpus audiovisuels*, Editions des archives contemporaines, ISBN : 9782813003799, doi : 10.17184/eac.9782813003799

de contrôler au mieux la transcription intégrale, une telle démarche est nécessaire, le travail manuel et l'assistance des outils informatique est incontournable.

De plus, la reconnaissance vocale n'a pas pour finalité d'interpréter et de retranscrire d'autres voix que celle de l'utilisateur. Dès lors, la seule utilisation possible est de répéter à voix haute devant un microphone les propos tenus par les Chibanis pour que le logiciel les retranscrive en texte. Étant donné que les logiciels de reconnaissance vocale ne prennent pas en compte les différents accents des locuteurs, il est essentiel de revoir et de corriger les textes retranscrits.

2.5 Analyse et traitement des données recueillies

Il s'agit de réaliser une analyse de contenu des entretiens. Cette étude est menée selon une méthodologie d'analyse visant à extraire les éléments pertinents qui constitueront la base du sujet. Selon Bardin : « l'analyse de contenu cherche à mettre au jour les implicites dissimulés derrière les paroles. La linguistique est une étude de la langue, l'analyse de contenu est une quête, à travers des messages, de réalité autre » (1977 : 48).

Une fois les entretiens transcrits, j'ai procédé au décorticage des données brutes, le discours des personnes interviewées est porteur d'informations à exploiter : « tout document parlé, écrit ou sensoriel contient potentiellement une quantité d'informations sur la personne qui en est l'auteur, sur le groupe auquel elle appartient, sur les faits et événements qui y sont relatés, sur les effets recherchés par la présentation de l'information, sur le monde ou sur le secteur du réel dont il est question » (Mucchielli, 1984 : 17). En effet, l'analyse de contenu s'appuie sur la catégorisation quantitative et qualitative. : « c'est une technique de recherche pour la description objective, systématique et quantitative du contenu manifeste de la communication » (Berelson 1952 cité dans Ghiglione & Matalon, 1978 : 155). Elle permet de concilier les deux méthodes et de réaliser un tri statistique de l'ensemble des données recueillies.

En effet, ce travail de synthèse, a permis d'établir un inventaire des mots les plus occurrents et des thèmes les plus saillants, pour réaliser une analyse lexicale d'une part et une analyse thématique d'autre part. À partir des résultats de cette première démarche, j'ai réalisé l'étude thématique. Ainsi, l'analyse qualitative se fonde sur les résultats quantitatifs : « l'analyse de contenu peut être une analyse de « signifiés » (ex : l'analyse thématique) mais elle peut être aussi une analyse des « signifiants » (analyse lexicale) » (Bardin 1977 : 39).

Marie-Madeleine Bertucci et Mounia Illourmane

Conclusion. Les questions restées en suspens

L'intérêt de ce travail est qu'il permet de conserver les traces de voix minorées, celles des Chibanis, du point de vue du patrimoine culturel immatériel et d'analyser l'intérêt du support audio dans cette perspective. Ce type de support ne va pas sans poser de problèmes comme a essayé de le montrer Mounia Illourmane. La question restée en suspens est celle de l'utilisation de l'arborescence réalisée pour indexer et classer les

Published in : Louise Ouvrard et Odile Farge (dir.), *Corpus audiovisuels*, Editions des archives contemporaines, ISBN : 9782813003799, doi : 10.17184/eac.9782813003799

données de départ dans la perspective de l'analyse du corpus et l'impact éventuel de ce type de description sur les modalités de conceptualisation de la thématique et le déroulement de la recherche.

Bibliographie

Adam Jean-Michel, 1995, « Aspects du récit en anthropologie », in Adam Jean-Michel, Borel Marie-Jeanne, Calame Claude, Kilani Mondher, *Le discours anthropologique, description, narration, savoir*, Lausanne : Payot, p. 251-283.

Bakhtine Mikhaïl, 1978, *Esthétique et théorie du roman*, Paris : Gallimard.

Bardin Laurence, 2003, *L'analyse de contenu,* Paris : Presse universitaire de France.

Baude Olivier (dir.), 2006, *Corpus oraux. Guide des bonnes pratiques,* Paris : CNRS Éditions.

Berelson Bernard, 1952, *Content Analysis in Communication Research*, New York : The Free Press.

Bertucci Marie-Madeleine, 2006, « La fonction du récit de vie dans la constitution de l'identité plurilingue chez les migrants : quelques hypothèses », in Molinié Muriel, Bishop Marie-France (coord.), *Autobiographie et réflexivité,* Amiens : UCP / CRTF, Encrage éditions, p. 127-140.

Bertucci Marie-Madeleine, 2009- 2017 (avec la collaboration de Assier Julie et Chemblette Emma pour la transcription du corpus), *Mémoires de l'immigration, vers un processus de patrimonialisation ?* Rapport de recherche universitaire pour le Ministère de la Culture et la Cité nationale de l'Histoire de l'immigration, université de Cergy-Pontoise, CRTF-pôle LaSCoD, 1 volume, 170 p. , disponible à la BNF, URL http://catalogue.bnf.fr/ark:/12148/cb42101812c. Rapport mis en ligne en 2017 sur le site du Ministère de la Culture et de la Communication, page du Patrimoine Ethnologique. http://www.culturecommunication.gouv.fr/Thematiques/Patrimoine-ethnologique/Soutien-a-la-recherche/Travaux-de-recherche/Liste-des-travaux-de-recherche-par-mots-cles/Migration

Bertucci Marie-Madeleine, 2012, « Le récit de vie, un processus réflexif à l'œuvre dans la production des savoirs », in Blanchet Philippe, Kebbas Malika, Kara Attika-Yasmine (dir.), *Pluralité linguistique et démarche de recherche. Vers une sociolinguistique complexifiée*, *Cahiers internationaux de sociolinguistique* 2, Paris : L'Harmattan, p. 83-102.

Bertucci Marie-Madeleine, 2014, « Le récit de vie : lieu d'expression de la pluralité linguistique et culturelle. Proposition pour une méthodologie de la recherche », in Feussi Valentin (dir.), *Pluralité linguistique et culturelle. Actualité de la recherche en sociolinguistique*, *Cahiers internationaux de sociolinguistique* 5, Paris : L'Harmattan, p. 41-47.

Ferrarotti Franco, 1990, *Histoire et histoires de vie. La méthode biographique dans les sciences sociales*, Paris : Méridiens Klincksieck.

Ghiglione Rodolphe, Matalon Bennjamin, 1978, *Les Enquêtes Sociologiques,* Paris : Armand Colin.

Lévi-Strauss Claude, 1958, *Anthropologie Structurale*, Paris : Plon.

Merle Isabelle, 2004, « *Les subaltern studies*. Retour sur les principes fondateurs d'un projet historiographique de l'Inde coloniale », in *Genèses* 56, p. 131-147.

Mucchielli Roger, 1984, *L'analyse de Contenu des Documents et Communications »*, Paris : ESF.

Noiriel Gérard, 2006 a, *Le creuset français. Histoire de l'immigration XIX^{e}-XX^{e} siècles*, Paris : Seuil.

Noiriel Gérard, 2006 b, *Introduction à la sociohistoire*, Paris : La Découverte.

Ricœur Paul, 1983, *Temps et récit*, T.1, Paris : Seuil.

Thompson Edward P., 1975, *The making of the english working class,* Londres : Penguin book.

Published in : Louise Ouvrard et Odile Farge (dir.), *Corpus audiovisuels*, Editions des archives contemporaines, ISBN : 9782813003799, doi : 10.17184/eac.9782813003799