



# Séminaire du LaDHUL

3 décembre 2013

## **Charlotte Touati, « Mémoires Falashas »**

Cet exposé de Charlotte Touati est issu de son projet Post-Doc FNS visant à créer une base de données multimédia accessible sur Internet et permettant d'explorer les écrits des Falashas, une communauté juive d'Ethiopie. Le projet n'ayant commencé que depuis quelques mois, la présentation traite principalement des questions liées à l'encodage des données en vue d'obtenir une base de données multimédia conforme à un certain nombre d'attentes.

### **Quelques éléments d'histoire**

Un bref retour sur l'histoire des Falashas s'impose : Qui sont-ils ? D'où viennent-ils ? Pourquoi est-il intéressant et important de construire une base de données multimédia afin d'explorer leurs écrits ?

#### ***Christianisme éthiopien, Zara Yaqob et mise en exode des Falashas***

L'Ethiopie adopte le christianisme au début du 4<sup>e</sup> siècle et devient ainsi le deuxième royaume-Etat officiellement chrétien après l'Arménie. De par sa proximité avec plusieurs royaumes judéo-chrétiens du Yémen voisin et, plus tard, son relatif isolement au sein d'une corne de l'Afrique à confession majoritairement musulmane, l'Ethiopie développe un christianisme tout à fait particulier et fortement imprégné d'éléments judaïques. Au sein de ce royaume chrétien vit également une communauté juive (dont l'arrivée est difficile à dater et la provenance difficile à situer) du nom de « Beta-Israël ». La judaïté de cette communauté est assez particulière car non rabbinique.

Au XVI<sup>e</sup> siècle, le roi éthiopien Zara Yaqob, qui voyait les Beta-Israël comme de possibles concurrents religieux à son christianisme d'Etat à consonance judaïque, prononce un décret stipulant que toute personne non-baptisée serait dorénavant *Falasian*, littéralement « paysan sans terre ». Le sens du terme *Falasian* s'est progressivement rétréci pour ne concerner que la communauté Beta-Israël, bientôt appelée « *Falasha* ». Le terme *falasha* est ainsi récent et exogène à la communauté Beta-Israël. Sous la contrainte de Yaqob, les « désormais-Falashas » sont forcés de s'exiler en des lieux retranchés au nord de l'Ethiopie, notamment autour du lac Tana et sur les hauteurs du Simien. De par cet exode forcé en des lieux difficilement accessibles,



1

UNIL | Université de Lausanne

LADHUL - Laboratoire  
de cultures et humanités  
digitales de l'UNIL

les Falashas sont longtemps restés ignorés du monde occidental ainsi que des autres communautés juives.

### ***Antoine d'Abbadie et Joseph Halévy***

L'existence des Falashas est révélée au monde occidental par les écrits de voyage du cartographe, linguiste et géologue français Antoine d'Abbadie qui sillonne l'Ethiopie entre 1838 et 1849. Passablement intrigué par cette communauté juive non rabbinique, il organise la copie systématique des manuscrits falashas qu'il découvre au cours de ses déambulations. Ces copies sont prises en charge par des scribes éthiopiens sur des supports préalablement formatés.

Durant les années 1860 en France, les échos du voyage d'Antoine d'Abbadie auprès d'une communauté juive d'Ethiopie pendant longtemps isolée interpellent Joseph Halévy, membre de la toute nouvelle « Alliance israélite universelle ». En 1867, Joseph Halévy parvient à convaincre l'Alliance de l'envoyer au nord de l'Ethiopie afin d'établir la judaïté présumée des Falashas. En ressort de multiples notes et textes sur leur mode de vie, leurs habitudes et leur judaïté, qui ne sera finalement reconnue par le gouvernement israélien qu'en 1975.

### ***Mengistu, la grande famine et le « rapatriement » vers Israël***

En 1984, le dictateur éthiopien Mengistu coupe l'arrivée de céréales jusqu'aux zones montagneuses du nord de l'Ethiopie afin d'affamer les rebelles du Front de libération du peuple du Tigray. S'ensuit un exode massif des habitants du nord de l'Ethiopie – dont les Falashas – vers le Soudan et ses camps de réfugiés. Pour la première fois, les Falashas sont comptés et regroupés. Emu par la mort de plus de 4000 Falashas tout en étant désireux de se racheter une crédibilité internationale après son soutien affiché à l'Apartheid, le gouvernement israélien met en place – plus ou moins secrètement – un pont aérien entre le Soudan et l'Israël afin de « rapatrier » la communauté falasha, récemment reconnue officiellement juive et bénéficiant donc de la loi du retour. L'opération tourne court car le Soudan – membre de la ligue arabe – est rapidement mis au courant et refuse de cautionner une initiative participant au peuplement d'Israël. En 1991, sous l'impulsion d'Ariel Sharon, les opérations de transfert des Falashas reprennent pour se terminer durant l'été 2013. Aujourd'hui, il n'y a officiellement plus de Falashas en Ethiopie.

### ***Assimilation au judaïsme rabbinique et importance d'une mémoire accessible et consultable***

La politique israélienne d'intégration est basée sur l'assimilation : les Falashas se doivent d'apprendre l'hébreu et d'intégrer des éléments rabbiniques à leur judaïsme. Les Falashas sont ainsi de plus en plus hébraïsés et la seconde génération n'apprend généralement plus l'amharique ou le tigrigna (les deux langues traditionnellement parlées par les Falashas en Ethiopie). Cette politique suscite des réactions de contestation qui voient certains Falashas d'Israël revendiquer un pluralisme judaïque pouvant laisser une place à leur déclinaison particulière et chargée d'histoire.

Mais l'ont-ils à leur disposition, cette histoire ? C'est là – entre autres – qu'intervient le projet de Charlotte Touati. En effet, en plus de constituer un sujet de recherche inédit (la culture falasha a été très peu étudiée et leurs textes religieux jamais traduits dans leur intégralité), le projet vise à rendre disponibles et consultables *à moindre frais* les écrits falashas. Rendre disponible de la mémoire bientôt perdue tout en y joignant des instruments pour la consulter et la faire vivre : telle est en somme la performance souhaitée pour ce projet de recherche.

## Rencontre avec les informaticiens

Pour une jeune chercheuse en histoire des religions, il est très difficile de créer une base de données sans aide. Aussi, Charlotte Touati nous explique qu'elle a mandaté une entreprise de conseil en informatique ayant déjà travaillé sur des bases de données multimédia. La relation est de type prestation de service et pourrait être schématisée comme suit :

CT : « Je souhaite créer une base de données multimédia permettant d'explorer les écrits sacrés des Falashas. J'ai une certaine somme d'argent et un certain nombre d'attentes. Qu'est-il possible de faire et comment ? ».

INF : « Soit. Présentez-nous vos attentes et discutons-en afin de trouver les solutions les plus viables ».

### ***Attente 1 : La base de données devra tendre à l'exhaustivité et donc être conçue de façon modulable***

La colonne vertébrale de cette base de données est les textes religieux falashas recopiés par différents scribes éthiopiens à la demande d'Antoine d'Abbadie lors de ses déambulations en Ethiopie. Il s'agit donc de traiter des textes manuscrits que la Bibliothèque nationale de France a regroupés sous l'appellation « Manuscrit Ethiopien d'Abbadie 107 ». Tous les 11 textes de ce manuscrit 107 sont écrits en guèze, la langue liturgique classique de l'Ethiopie qui possède le même alphasyllabaire que l'amharique et le tigrigna. Le manuscrit 107 ne pouvant être déplacé hors de la BNF, il s'agit ici de se baser sur des scans noir et blanc en format Jpeg.

Ce manuscrit 107 contient tous les écrits falashas connus à ce jour. Pour autant, il est difficile de croire à l'exhaustivité du recensement d'Antoine d'Abbadie et rien n'indique que de nouveaux manuscrits n'émergeront pas. Il est donc souhaitable de pouvoir inclure ces futurs textes « potentiels » dans la base de données sans trop d'efforts d'actualisation.

### ***Attente 2 : La base de données devra mettre à disposition des traductions tout en permettant de remonter jusqu'aux originaux***

Une autre *raison d'être* de la base de données est de rendre lisible et compréhensible le contenu du manuscrit 107. Peu de gens – Falashas y compris – sont capables de lire le guèze et encore moins de le comprendre ; de la traduction est nécessaire, avec son lot de possibles trahisons. A ce stade du projet, seule une traduction en français est en marche. Mais l'ambition est de décliner ce manuscrit en plusieurs langues modernes (notamment en amharique et en hébreu, les langues principales des Falashas aujourd'hui), ce qui rejoint l'impératif de modularité stipulée plus haut.

*Traduttore, traditore !* Ce problème bien connu implique ici une nouvelle exigence : donner la possibilité à l'utilisateur de remonter facilement jusqu'aux scans du manuscrit 107. Un exemple (parmi d'autres possibles) est relatif à la ponctuation. En guèze, il y a deux points entre chaque mot et quatre points entre chaque phrase. Mais l'ennui, la fatigue, l'empressement ou le zèle maniaque des scribes mandatés par Antoine d'Abbadie (ou leurs prédécesseurs) ont fait de ces



3

UNIL | Université de Lausanne

LADHUL - Laboratoire  
de cultures et humanités  
digitales de l'UNIL

quatre points de délimitation des éléments plutôt aléatoires... Très souvent, pour pouvoir donner du sens à une phrase en guèze et la traduire de manière intelligible, il faut s'éloigner quelque peu du manuscrit et changer ces quatre points de place. Mais où les remettre ? Et est-ce toujours nécessaire de le faire ? N'y a-t-il pas là une erreur de la traductrice ? Peut-être, des fois, certainement. Aussi est-il primordial de laisser la possibilité aux utilisateurs de la base de données de « remonter » jusqu'au manuscrit pour se faire eux-mêmes une opinion.

### ***Attente 3 : La base de données devra proposer une exploration souple et agréable du manuscrit***

La qualité souvent médiocre des manuscrits scannés rend leur lecture peu agréable et il est généralement préférable de passer par du format texte, pixélisé de manière plus adéquate. De plus, pour permettre une exploration souple du manuscrit ; pour permettre à l'utilisateur de sauter d'un passage à l'autre, de rechercher ceux qui l'intéressent pour en laisser tomber d'autres, bref d'être *agile* dans sa manipulation du manuscrit 107, il faut un moteur de recherche performant capable de reconnaître avec précision les caractères du guèze, ce qu'une version uniquement manuscrite-scannée rendrait terriblement compliqué. Le mieux serait un format Unicode et son armée de *caractères nommés*, ses *points de code* et autres *codets* conçus collectivement et à grands frais. Cela rendrait la lecture plus agréable et le moteur de recherche à coup sûr plus performant.

Heureusement pour le projet, le guèze possède le même alphasyllabaire que l'amharique, la langue administrative de l'Ethiopie. De par son statut de langue administrative d'un pays comptant plus de 90 millions d'habitants, l'amharique possède un format Unicode standardisé depuis plus de 15 ans et permettant une représentation claire et opérable de ses nombreux caractères.

Une des particularités du guèze et de l'amharique est que ce sont des langues alphasyllabaires pour lesquels chaque signe renvoie à une syllabe. Il est donc difficile pour un moteur de recherche de prendre appui sur des mots (délimités par deux points) puisque ceux-ci comprennent plusieurs morphèmes. Il est également difficile pour le moteur de recherche de prendre appui sur les signes-syllabes largement trop nombreux. Après discussion avec les conseillers en informatique, il semble plus sage de choisir une unité de balisage plus grande, comme par exemple les phrases, qui sont toutes différentes, délimitées (même si la délimitation est parfois problématique) et que l'organisation du manuscrit rend aisément repérables (comme on le verra par la suite). Il semble ainsi que pour pouvoir être agile dans la manipulation de ce manuscrit, le balisage doive s'opérer en fonction des phrases.

A partir de là, ces quelques attentes (modularité, traduction, accès aux originaux, lisibilité et souplesse) confèrent déjà une certaine *figuration* à la base de données. L'informaticien et sa cliente voient en effet se dessiner trois niveaux qui seront autant d'*interfaces* : l'interface 1 sera constituée des originaux scannés par la BNF ; l'interface 2 sera constituée du même contenu mais retranscrits en format guèze-Unicode ; l'interface 3 sera constituée de la traduction française en format lui-aussi Unicode. Et pour lier entre elles ces trois interfaces ; pour basculer de l'une à l'autre, il y aura besoin d'un moteur de recherche capable de retrouver des phrases et les faire corrélérer aux mêmes phrases situées sur une autre interface.

#### ***Attente 4 : La base de données devra pouvoir servir d'outil pédagogique pour l'étude du guèze***

Une autre ambition du projet est de créer un outil performant permettant de progresser dans l'apprentissage du guèze. Comment satisfaire au mieux ce désir ? Une idée serait de donner la possibilité à l'utilisateur de mieux comprendre la composition des mots en guèze de telle sorte qu'en sélectionnant n'importe quel mot de l'interface guèze-Unicode, le moteur de recherche soit capable de lui proposer la définition de chaque morphème présent au sein du mot sélectionné. On pourrait même imaginer que pour certains mots problématiques ayant plusieurs orthographes différentes, leur sélection activerait un moteur de recherche suffisamment puissant pour proposer toutes les écritures alternatives présentes au sein du manuscrit. En somme, l'idée est d'inclure la possibilité de faire apparaître des *info-bulles* qui permettraient de mieux saisir la composition des mots en guèze ainsi que leurs éventuelles variations. Quelle belle fonctionnalité ! Un outil capable de déployer les morphèmes du guèze tout en respectant la variance propre au manuscrit 107 (rappelons que les textes ont été copiés par des scribes provenant d'endroits différents et aux habitudes d'écriture différentes). Mais pour qu'un tel outil se mette en place, il faudrait implémenter un moteur de recherche permettant une analyse morphologique de chaque terme, « ce qui peut s'avérer lourd à mettre en place. Mais c'est faisable. Ce qui est sûr, c'est qu'il faudra de toute façon normaliser quelque part : soit au niveau de l'encodage des données, soit au niveau du moteur de recherche. Ça n'est qu'au prix d'une normalisation qu'on pourra aboutir à un respect de la variance » explique l'informaticien, pour le coup largement romancé.

#### ***Attente 5 : La base de données devra permettre de déambuler dans le manuscrit 107 et dans le monde (perdu) auquel il renvoie***

C'est un élément important qui justifie l'expression « base de données *multimédia* ». On l'a vu, le manuscrit 107 est l'œuvre collective de multiples scribes éthiopiens qui, à la demande d'Antoine d'Abbadie, ont recopiés divers textes falashas à des endroits et des moments différents. Le manuscrit est donc *pluriel* et porte les traces silencieuses de différents lieux, paysages, visages et cultures. Cette déambulation peut parfois se sentir dans le texte, notamment par l'utilisation de tel morphème plutôt que tel autre, donnant ainsi des indications sur son scribe et sa région d'origine. Mais c'est surtout dans les lettres et les notes personnelles d'Antoine d'Abbadie et de Joseph Halévy (qui ont visité les mêmes villages falashas à 15 ans d'intervalle) que ces *métadonnées* peuvent être recueillies. L'enjeu ici est véritablement de nature historique car aujourd'hui, les villages falashas visités par les deux compères sont soit vides, soit repeuplés par d'autres habitants aux coutumes différentes. Ainsi donc, en recueillant les éléments de contextes qui entourent les textes du manuscrit 107 et en les incluant à la base de données sous forme de métadonnées, on se donne la possibilité de recréer quelque peu le monde falasha avant son exode en Israël. De multiples possibilités sont envisageables : cela peut passer par des coordonnées géographiques renvoyant à des cartes interactives ou encore des annotations de carnets de voyage en lien avec certains passages. L'important ici est d'exploiter au mieux les

The logo of the University of Lausanne (UNIL) is a stylized, handwritten-style wordmark in blue, consisting of the letters 'U', 'n', 'i', 'l' connected together.

5

UNIL | Université de Lausanne

LADHUL - Laboratoire  
de cultures et humanités  
digitales de l'UNIL

éléments de contexte disponibles afin de faire revivre *digitalement* le monde des juifs d’Ethiopie.

***Attente 6 : Une première version de la base de données devra rapidement être mise en ligne***

CT : « Une dernière chose encore. Un de mes sponsors souhaite qu’une première version de la base de données soit rapidement mise en ligne ».

INF : « Hum, rapidement vous dites ? Je vois. Le temps presse apparemment, c’est embêtant. Le plus *simple* serait de fonctionner sur la base de feuilles Excel. Ça implique un protocole d’encodage non-standardisé mais on n’a pas le temps de faire autrement. Mais comme la base de données n’est pas très grande, ça ne devrait pas poser trop de problèmes. Pour la première version, le site ne concernera certainement que le manuscrit, sans les métadonnées permettant la déambulation. Mais conservez-les tout de même et on formalisera leur encodage plus tard ».

## **Encodage des données**

Parallèlement à ces séances préliminaires, il y a l’encodage manuel des données. Car les bases de données ont souvent des origines modestes ; une construction protocolaire, codifiée, manuelle et laborieuse des données vouées à être ensuite implémentées, triturées, croisées, traitées par un logiciel afin d’obtenir un certain rendu à l’écran.

### ***11 feuilles Excel de 10 colonnes chacune***

Ici, l’encodage ne part pas de zéro : Antoine d’Abbadie avait déjà mis en place un protocole suffisamment élaboré pour être utilisé comme base. Le manuscrit 107 est divisé en 11 textes et numéroté par folio. Chaque folio correspond à deux pages ; une page recto et une page verso. Chaque page est divisée en 2 colonnes, elles-mêmes divisées en 22 lignes. Si on y ajoute la numérotation des phrases en fonction de chaque texte (par exemple « phrase 22 du texte Y »), le moteur de recherche pourra disposer d’informations « géographiques » suffisamment solides pour lui permettre un balisage par phrase et basculer ainsi d’une interface à l’autre (cf. attente 2 et 3).

Mais cela ne suffit évidemment pas : il faut également transformer le guèze Jpeg en guèze Unicode afin de le rendre opérable par le moteur de recherche (cf. attente 3). Il faut également traduire chaque phrase en « bon » français (cf. attente 2) et en traduction mot à mot afin de rendre possible l’implémentation d’une analyse morphologique de chaque terme (cf. attente 4). Il faut finalement inscrire un certain nombre de commentaires qui pourront plus tard servir de métadonnées (cf. attente 5).

Si on assemble tous ces éléments, on obtient un format d’encodage permettant une mise en ligne rapide d’une première base de données « minimale » (cf. attente 6). Chacun des 11 textes du manuscrit correspond à une feuille Excel constituée de 10 colonnes, de A à J :

**A** : titre du texte / **B** : Folio où se situe la phrase / **C** : Colonne où se situe la phrase (1 ou 2) / **D** : Numéro de la phrase / **E** : Ligne du début de la phrase / **F** : Ligne de la fin de la phrase / **G** : transcription de la phrase en guèze Unicode / **H** : traduction française / **I** : traduction mot à mot / **J** : commentaires.



Rien n'est moins sûr puisqu'au contraire, on voit ici qu'elle provoque de l'irréversibilité. Cette obligation de rendre des comptes *au plus vite* ordonne de prendre une certaine direction qui – après coup – ne pourra être modifiée qu'à grands frais. Paradoxalement, au lieu de donner des garanties, cette injonction en élimine : elle contribue, en fait, à ce que le respect du cahier des charges initial du projet *ne soit pas garanti*. Mais peut-être est-ce là un élément constitutif des bases de données : une forte propension à se rendre irréversibles.

Florian Jaton