

## Partage d'expériences

### *Ethiopian manuscript Archives :*

Rencontre entre la diplomatie, les archives du royaume chrétien d'Éthiopie et les humanités numériques



Les églises orthodoxes d'Éthiopie sont jusqu'à aujourd'hui les institutions qui conservent la majorité des documents d'archive. Ici, l'église de Dabra Warq, dans le Gojjam © Anaïs Wion, 2007

### Des archives éparpillées dans les bibliothèques monastiques

Qu'appelle-t-on les archives manuscrites éthiopiennes ? Il s'agit d'un qualificatif général englobant les textes de nature administrative, juridique et historique qui furent produits par les pouvoirs politiques et religieux éthiopiens pour témoigner de leurs lois, de leurs règles, de leurs traditions : écrits légaux et pseudo-légaux, documents historiographiques ou contractuels, chartes de donation de terres, règles concernant les nominations de dignitaires religieux ou fonctionnaires royaux, etc. Les producteurs de ces documents, entre le XIII<sup>e</sup> et le XX<sup>e</sup> siècle, sont l'administration royale et dans une moindre mesure les administrations religieuses. Ils sont rédigés en *ge'ez*, la langue ritualisée de l'Éthiopie chrétienne, une langue sémitique qui possède son propre système d'écriture, un alpha-syllabaire. Les actes privés ne sont issus que tardivement, à partir de la mi-XVIII<sup>e</sup> siècle et ils sont eux rédigés en *amharique*, la langue véhiculaire des hauts-plateaux éthiopiens, écrite avec le même système d'écriture que le *ge'ez*, son ancêtre indirect. Ce terme « archives » est donc à considérer de manière très large.



Fin d'un manuscrit de patristique comportant, après le texte théologique, un folio portant des documents d'archives d'époques différentes. Église de Maḥḍara Maryam, région du Bagemḍer © Anaïs Wion, 2007

L'une des particularités de la conservation de ces documents en Éthiopie est qu'ils sont copiés, dans la majeure partie des cas, dans les espaces blancs laissés dans les manuscrits bibliques ou liturgiques. Ainsi les manuscrits des Évangiles, des Miracles de Marie contiennent dans leurs marges et entre leurs chapitres, les documents d'archive du royaume. Cette dispersion des archives au sein des livres des bibliothèques monastiques est l'une des raisons du peu de considération qui leur fut accordé par les chercheurs jusqu'à aujourd'hui. Quelques milliers, peut-être même centaines de milliers, de documents de nature diverse constituent un corpus conséquent de sources largement sous-exploitées à ce jour.

### Les enjeux scientifiques : construire une diplomatie et une plateforme d'édition

Ces documents sont pourtant des sources fondamentales pour faire l'histoire des structures de pouvoirs, de la construction des territoires, des rapports sociétaux, etc. Il s'agit à la fois de faire connaître ces documents et de se doter d'outils d'analyse pour comprendre les modalités de leur production, de leur copie, de leur préservation. En un mot, de créer une diplomatie éthiopienne, c'est-à-dire une méthode pour analyser et éditer les documents d'archive manuscrits, à l'instar de ce qui se fait depuis plusieurs siècles dans les études médiévales.

La mise en place de nouveaux outils de publication et d'analyse à travers le projet *EMA (Ethiopian manuscript Archives)* s'inscrit dans une double démarche, à la fois novatrice et classique. Novatrice car elle accompagne les réflexions menées autour des nouveaux paradigmes des humanités numériques et s'inscrit ainsi dans une véritable dynamique de création scientifique. Classique par ailleurs, car l'analyse proposée par EMA s'inscrit dans la tradition de la diploma-

tique et de la critique textuelle. Or cette science érudite qu'est la diplomatique est à la pointe des avancées des outils de recherche et de publications électroniques depuis près de vingt ans désormais.

### Trouver les partenaires institutionnels

Ce projet s'est développé dans le cadre du projet « CORNAFRIQUE – Écrire l'histoire de la Corne de l'Afrique : textes, réseaux & sociétés », financé de 2006 à 2010 par l'ANR. EMA est l'axe documentaire de ce programme de vaste ampleur visant à reconsidérer les fondements de notre connaissance de l'histoire de la Corne du Moyen Âge à la période contemporaine.

Lors du montage du projet EMA, plusieurs partenaires institutionnels ont été évoqués. Le choix s'est finalement porté sur TELMA (Traitement électronique des manuscrits et des archives), l'un des Centres de ressources soutenu par le TGE Adonis. Ce projet fut à l'origine commun à l'Institut de recherche sur l'histoire des textes (IRHT, UPR841) et à l'École nationale des Chartes, deux institutions aux traditions anciennes en matière de traitement et d'analyse des sources archivistiques et tout autant impliquées dès les années 1990 dans la réflexion sur le traitement électronique des documents manuscrits. Pour notre projet, le label de TELMA et les ressources qui s'y sont mobilisées sont un gage de dynamisme et de pérennité. Surtout, la structuration des données proposée par TELMA est basée sur le langage XML et les recommandations de balisage de la TEI (Text Encoding Initiative).

### Le choix d'un outil technologique et d'un modèle de structuration de l'information

Étant donné que les sources que nous traitons sont des sources textuelles dont nous voulons conserver toute la richesse, l'encodage en XML/TEI paraît la meilleure solution. La TEI est en effet un outil polyvalent pouvant rendre compte des trois états d'un texte :

- ▶ son contenu textuel, bien entendu ;
- ▶ son état physique et le contexte matériel dans lequel il existe ;
- ▶ les intentions de ses auteurs comme les appréciations des générations de lecteurs qui se sont appropriés le texte, ce qui inclut les notes et commentaires des auteurs des corpus électroniques eux-mêmes.

Au-delà de la possibilité de représenter dans ses diverses dimensions un document textuel, la TEI est aussi un langage de structuration informatique qui permet de gérer des images et des bases de données (un « mashup »). La TEI a donc semblé être le choix le plus pertinent, même s'il faut reconnaître un réel effort d'investissement pour en maîtriser et en comprendre les possibilités.

Concrètement, le travail s'est déroulé selon les étapes suivantes. Il a fallu d'abord définir les typologies documentaires, ce qui a permis de distinguer trois niveaux de description des documents et de définir la structuration globale du projet.

- ▶ Le premier niveau, le plus englobant, est celui du **corpus**, terme générique par lequel on désigne un ensemble nécessairement artificiel de documents. Leur réunion peut être légitimée soit par un regroupement déjà opéré par les institutions éthiopiennes en charge de ces documents, soit par un choix scientifique de la part du chercheur qui est porteur de ce corpus au sein de l'édition EMA. Cette notion de « corpus » préserve la responsabilité scientifique de chaque auteur et les corpus fonctionnent comme des chapitres interdépendants d'une œuvre collective.
- ▶ Le deuxième niveau est celui du **manuscrit**, puisque les documents d'archives éthiopiens sont copiés dans des *codex*. C'est à ce

niveau qu'est fixée la granularité du fichier XML qui correspond à un manuscrit. L'unité numérique respecte ainsi l'unité codicologique et physique de la documentation initiale. Les manuscrits sont décrits suivant les principes classiques d'un catalogue descriptif.

▶ Enfin, le troisième niveau est celui du **document textuel**, qui est transcrit, traduit, annoté et encodé sémantiquement. Contrairement à de multiples projets d'édition de sources, ce n'est pas ici l'image numérisée du document qui prime, même si celle-ci est accessible à chaque fois qu'elle est disponible. C'est la transcription du texte qui est au cœur du processus de numérisation. L'action de transcrire en caractères éthiopiens le texte source permet de mener des recherches directement dans les textes. Chaque texte est présenté accompagné d'une (ou plusieurs) traduction(s) permettant aux lecteurs non spécialistes d'avoir accès à leur contenu. Enfin, il est précédé d'un résumé et d'une analyse, ainsi que d'un appareil critique et d'un appareil de notes (références et notes scientifiques).

À partir du moment où cette structure a été validée, les ingénieurs chargés de TELMA au sein de l'IRHT ont pu créer un formulaire de saisie permettant de renseigner les principaux champs nécessaires à la description de ces trois niveaux. Bien sûr, de nombreux aller-retour entre les propositions des ingénieurs et celles des chercheurs ont eu lieu, et ce formulaire s'est construit progressivement. Au bout du compte, les chercheurs peuvent générer un fichier XML/TEI bien structuré et déjà très complet sans avoir besoin de passer, dans un premier temps, par un éditeur de texte en XML. Ils ont à disposition un outil qui permet de saisir les données de façon complète ou progressive. Ils peuvent donc mener le travail d'analyse des documents sans être trop accaparés par l'apprentissage d'une technologie autant qu'utiliser des outils complexes sans s'autolimiter ni ressentir d'infériorité technique.

L'expérience a montré que le formulaire permet d'éviter deux écueils majeurs. Tout d'abord, sans ce type d'accompagnement, l'utilisateur risque de générer de nombreuses erreurs dans la structuration du fichier TEI/XML. Mais aussi, et de façon plus surprenante, il permet d'éviter que les chercheurs découvrant les possibilités de TEI veuillent encoder trop finement leurs matériaux et se perdent dans l'infini des possibles de ce langage.

### ETHIOPIAN MANUSCRIPT ARCHIVES, un projet collectif

▶ **Coordinatrice** : Anaïs WION (CEMAf)

▶ **Chercheurs contribuant en tant qu'auteurs** : Claire BOSCH-TIESSE (CFEE), Donald CRUMMEY (Université de Urbana-Champaign, USA); Manfred KROPP (Université de Mainz, Allemagne); SHIFERAW BEKELE (Université d'Addis Abeba); Marie-Laure DERAT (CFEE); DERESSE AYNATCHEW (Université de Dabra Berhan, Ethiopie)

▶ **Conception scientifique et technique** : Fabrice MELKA (CEMAf); Richard WALTER (IRHT et ADONIS); Cyril MASSET (IRHT); Elsa DEVIENNE (webmaster CDD IRHT)

▶ **Chercheurs participant à la maturation scientifique et institutionnelle du projet** : Denise OGILVIE (Archives de France); Paul BERTRAND (IRHT); Lou Burnard (Adonis)

▶ **Doctorants ayant collecté des données** : HABTAMU MENGISTIE (université de Urbana-Champaign); Margaux HERMAN (Paris 1); François LE CADRE (Paris 1 et Hamburg Universität); YOHANNES G. SELLASE (Paris 1)

## Nos balises nous appartiennent !

Néanmoins, tout ne peut être encodé au travers du formulaire de saisie, et en particulier le texte lui-même doit être balisé manuellement. L'un des enjeux de ce projet était donc de confier le balisage en TEI du contenu aux chercheurs. Pour ce faire, un workshop réunissant les participants a eu lieu à Addis Abeba, au [Centre français des études éthiopiennes](#) (UPR3137) en octobre 2009. Une fois réglées les multiples questions liées à la gestion des applications logicielles et à l'utilisation des polices unicode de transcription et de translittération sur les ordinateurs des uns et des autres, la réflexion commune sur les balises TEI s'est avérée extrêmement créative et fructueuse pour tous. Le guide d'encodage a ainsi pu être mis en chantier, pour rassembler et fixer les décisions prises collectivement à cette occasion.

Qu'est-ce qui est ainsi balisé, et surtout, sur quelle(s) représentation(s) du texte fait-on porter l'effort de balisage ? Le balisage permettant la génération d'index (noms propres ou termes spécifiques) est réalisé sur le texte traduit afin que les lecteurs ne maîtrisant pas les langues éthio-sémitiques puissent bénéficier de la consultation des outils de recherche mis en place. En revanche, le balisage des parties structurantes du discours et des termes techniques (liés au vocabulaire juridique et foncier par exemple) s'effectue sur les deux représentations du texte : la **transcription** et la **traduction**.

Il faut garder à l'esprit que si les choix scientifiques priment, ils sont indissociables de contraintes techniques pour lesquelles il faut mesurer le coût en temps de travail, aussi bien pour les développeurs que pour les utilisateurs. Concrètement, chaque nom propre, terme spécifique, segment de discours dans la transcription et dans la traduction, ainsi que dans les éléments para-textuels (résumé, note scientifique...) serait l'idéal. Mais cela serait par trop chronophage, aussi faut-il choisir sur quoi va porter l'effort, particulièrement dans un premier temps du travail lorsque s'élabore le choix des balises à employer et les modalités de leur utilisation.

## La publication électronique : de quelques avantages

La phase suivante est celle de la construction de l'interface de publication. Les choix opérés pour la structuration des ensembles documentaires privilégient naturellement trois modes d'entrée principaux dans les textes : par le niveau corpus (ill. 01), par le niveau manuscrit et enfin au niveau de chaque document (ill. 02). Les index permettent de naviguer dans l'ensemble des documents. Les données géographiques généreront aussi des cartes qui seront autant de points de départ pour arriver aux informations textuelles.

L'interface est encore en phase de test, aussi les captures d'écran présentées ici sont-elles des aperçus de ce que sera le projet final.

La première vague de publication réalisée par une équipe de huit chercheurs, historiens et philologues, concerne une trentaine de



Illustration 01



Illustration 02

manuscrits et plusieurs centaines de documents. Si certains textes ont déjà été publiés ailleurs mais nécessitent une nouvelle traduction et un nouvel appareil critique, la plupart sont à ce jour inédits.

L'intérêt du travail d'édition électronique est qu'il permet une publication progressive de la documentation, qui enrichit constamment l'édition des sources. Cela montre au passage le besoin de disposer d'outils souples pouvant répondre aux évolutions d'un projet. Enfin, l'hypertextualité permet de faire des liens directs entre les sources publiées et les articles issues de l'analyse de celles-ci.

Par ailleurs, l'encodage des métadonnées et des données en TEI doit permettre d'exploiter ce matériau enrichi. Actuellement, les outils d'exploitation restent le maillon faible de l'édition électronique, par manque de visibilité et d'adaptabilité des outils existants autant que parce que l'effort s'est d'abord concentré sur les enjeux liés au travail éditorial. Il est donc à souhaiter que dans un avenir proche, les travaux des acteurs des humanités numériques se dirigent vers une plus grande accessibilité des outils susceptibles de manipuler, visualiser, faire parler les textes encodés.

## Documenter les procédures de travail et favoriser l'interopérabilité

La plasticité des outils nécessite que soient bien documentées les procédures de travail. Les équipes changent, les besoins évoluent, aussi faut-il toujours pouvoir revenir sur ses traces et comprendre les choix scientifiques qui furent opérés en amont. Il est donc important de fournir aux participants du projet mais aussi aux lecteurs, l'ensemble de la documentation produite (guide de saisie dans le formulaire, guide d'encodage et guide des balises) ainsi que les fichiers XML/TEI natifs. La TEI est en effet une initiative fonctionnant sur le mode de l'*open source*. Ainsi chaque projet, qui opère nécessairement des choix au sein du langage TEI, peut bénéficier à l'ensemble de la communauté des utilisateurs de ce langage. Il est toujours très utile de consulter la documentation de projets similaires afin d'opérer ses propres choix en ayant conscience de ceux effectués par d'autres. C'est aussi une garantie d'interopérabilité que de permettre à un projet d'opérer ses choix à partir d'autres projets existants. Cette capitalisation de l'expérience assure une cohérence de l'évolution de l'ensemble des projets menés en TEI, en passe de devenir le langage pivot pour l'édition électronique textuelle.

**Anaïs Wion**

Historienne, responsable du projet EMA,  
Centre d'études des mondes africains (UMR8171)  
Avec la collaboration de Richard Walter (TGE Adonis)

contact&info

► Anaïs Wion

Centre d'études des mondes africains (CEMAF)

anaïs.wion@univ-paris1.fr

► [Ethiopian Manuscripts Archive](#)