

ATELIER FONTE GAIA (4/12/2015)

INTERFACE POUR LES ÉDITIONS NUMÉRIQUES :
TECHNOLOGIES, CHERCHEURS ET USAGERS

Présents : Elena Pierazzo (Université Grenoble-Alpes), Peter Stokes (King's College London), Geoffroy Noël (King's College London), Emmanuelle Morlock (UMR 5189 HiSoMA, MOM, Lyon), Marjorie Burghart (CIHAM, Lyon), Jean-Claude Zancarini (Triangle, Lyon), Séverine Gedzelman (Triangle, Lyon), Bruno Bureau (UMR 5189, HiSoMA/MOM, Lyon), Christian del Vento (Université Sorbonne, Paris 3), Claire Huan (société ZEDD), Isabelle Cogitore (MSH-Alpes), Christine Noaille (Université Grenoble-Alpes), Francis Goyet (Université Grenoble-Alpes), Claire Mouraby (Université Grenoble-Alpes), Héloïse Faivre (Université Grenoble-Alpes), Filippo Fonio (Université Grenoble-Alpes), Hervé Blanchon (LIG, Université Grenoble-Alpes), Francis Brunet-Manquat (LIG, Université Grenoble-Alpes), Laura Antonietti (Université Grenoble-Alpes), Élise Leclerc (Université Grenoble-Alpes), Andrea Penso (Université Grenoble-Alpes), Elina Leblanc (Université Grenoble-Alpes).

INTRODUCTION, ELENA PIERAZZO (UNIVERSITÉ GRENOBLE-ALPES)

Le projet franco-italien *Fonte Gaia*, né en 2009, a pour objectif de réunir des chercheurs et des bibliothécaires, au sein d'« un réseau d'échange d'informations et de débats d'idées autour de l'italianisme contemporain et international¹ ». Il s'appuie pour cela sur deux éléments, à savoir un blog (*FGBlog*), lieu de réflexions autour de l'italianisme à l'ère du numérique, et une bibliothèque numérique scientifique et collaborative (*FGBib*), qui permettra aux utilisateurs d'être à la fois lecteurs et auteurs des contenus qu'ils consultent. Dans ce cadre, le projet souhaite mettre au point une interface collaborative et enrichie, qui intégrerait des systèmes de commentaires, des propositions de numérisation ou encore de moissonnage.

Tout cela fait intervenir des réflexions sur les rapports entre l'homme et la machine : quelles infrastructures choisir ? Comment s'engager avec les technologies ? Qu'en est-il des utilisateurs finaux : est-ce nous, quelqu'un comme nous ou des personnes externes au monde de la recherche ? À chaque début de projet, nous pensons savoir ce qu'il veut. Mais comment savoir ce qu'il veut *réellement* ? Faut-il adopter la technique du *bottom-up* et demander directement aux utilisateurs potentiels ce qu'ils souhaitent ? Ou bien faut-il adopter la technique du *top-down* et proposer et faire découvrir aux utilisateurs des usages ? Entre *bottom-up* et *top-down*, quelle attitude choisir pour répondre aux besoins des utilisateurs ?

¹ Présentation du projet *Fonte Gaia* : <http://fontegaia.hypotheses.org/projet-fonte-gaia-2> [Consulté le 27/06/2016].

LES INFRASTRUCTURES

DIGIPAL : INFRASTRUCTURE, GEOFFROY NOËL (KING'S COLLEGE LONDON)

DigiPal (<http://www.digipal.eu/>) est à la fois un projet de paléographie numérique, conduit de 2011 à 2014, qui étudie les écritures médiévales en Angleterre entre 1000 et 1100, et une plateforme logicielle *open source* pour la recherche collaborative et la publication en ligne. Cette plateforme est aujourd'hui utilisée dans le cadre d'autres projets, à l'exemple du projet *EXON*, consacré à l'étude de différentes versions du *Domesday Book*, ou encore le projet *Models of Authority*, qui s'intéresse à l'étude des chartes écossaises. Les possibilités d'utilisation de *DigiPal* sont multiples :

- Importer, cataloguer et catégoriser des images en haute résolution.
- Isoler, décrire et comparer des lettres à l'intérieur d'un même manuscrit.
- Chercher, collectionner et comparer les lettres de plusieurs manuscrits.
- Éditer et baliser les textes à l'aide d'outils inspirés de la TEI, qui permettent de montrer en parallèle l'image du manuscrit, le texte balisé (« vue d'édition ») et le texte tel qu'il sera affiché sur l'interface (« vue publique »).

Quels outils et quelle architecture adopter pour permettre de telles utilisations ? *DigiPal* est bâti sur une architecture à trois niveaux. Le premier est le *client side* (ou côté *browser*), qui permet à l'utilisateur d'interagir avec la base de données. Le *client side* communique avec le *server side*, deuxième niveau de cette infrastructure. Ce dernier se construit autour du CMS *Django*, auquel sont ajoutés plusieurs applications, à l'exemple de *Woosh*, un moteur de recherche à facettes, qui fonctionne en Python, tout comme *Django*, ce qui évite d'avoir un serveur séparé et donc facilite la programmation. Enfin, le *server side* communique avec les bases de données relationnelles, dernier niveau de l'infrastructure. À cette architecture est associée une API, qui autorise n'importe quelle application à communiquer avec la plateforme, dont le système est ouvert et transparent.

LE PROJET IG LOUVRE, EMMANUELLE MORLOCK (UMR HISOMA/MOM, LYON)

Comment construire des systèmes maintenables ? C'est avec cette problématique à l'esprit qu'Emmanuelle Morlock a commencé sa présentation, en prenant pour exemple le projet *IG Louvre* (<https://cahier.hypotheses.org/ig-louvre>). Ce dernier est un projet d'édition épigraphique, dont les problématiques sont liées à la nature même des textes. En effet, les épigraphies sont des textes ostentatoires, à la fois destinés à être lus et à être vus et dont la typologie est indissociable du contexte d'exposition. Par conséquent, la manière dont sont présentés ces documents change l'allure de l'interface : si le texte est mis en avant, c'est une édition ; si c'est l'objet, c'est un catalogue.

Un travail de modélisation des objets a fait apparaître deux types de texte : un premier texte destiné à être transcrit et balisé en TEI EpiDoc et un deuxième, abstrait qui correspond à la structure de la source. Des correspondances doivent être établies entre ces deux types de

texte, de façon pérenne. Au texte transcrit et au texte abstrait correspondent deux modèles : l'un éditorial et l'autre abstrait. L'enjeu est donc de trouver une interface générique qui puisse prendre en compte ces différents modèles.

Le choix des outils est actuellement en cours de réflexion. Plusieurs solutions ont déjà été envisagées. Ainsi, le CMS Omeka aurait permis de gérer l'architecture du site, mais pas XML. La plateforme SynopsX, qui permet de publier et d'exposer facilement des corpus en TEI, a l'avantage de gérer les feuilles de style XSLT 2, propres à Epidoc. Cependant, SynopsX est en cours de développement. Il n'est donc pas encore tout à fait stable. Il existe bien d'autres solutions techniques envisageables (TEI Boiler Plate, Kiln), chacune devant être analysées afin de vérifier leur adéquation avec les besoins spécifiques du projet *IG Louvre*.

DISCUSSIONS

À la suite de ces deux interventions, deux points ont été mis en évidence. Il est tout d'abord apparu que les infrastructures doivent permettre aux chercheurs un contrôle qualité sur leur travail, en passant d'une vue éditeur à une vue du résultat final rapidement. Ce point a été développé avec plus de précision au cours de la deuxième partie de la matinée.

La discussion s'est ensuite tournée vers Omeka, la solution *open source* choisie par *Fonte Gaia*. Omeka s'avère être une bonne solution pour les bibliothèques et les expositions. Il fait cependant face à d'importants problèmes de mise à jour. En effet, lors du passage à une nouvelle version, certains plugins soit ne sont pas portés à la nouvelle version et deviennent donc inutilisables ; soit sont mis à jour après la structure générale et donc ne fonctionnent plus pendant un certain temps. Cela est en grande partie dû au fait que les plugins sont développés par des prestataires et qu'ils ne gèrent qu'une couche du site. Cela ne veut pas dire qu'Omeka n'est pas une solution envisageable pour un projet numérique : tout dépend de la complexité du projet et du temps et des compétences que l'on est prêt à engager pour développer l'interface.

LE CHERCHEUR ET SES BESOINS

HYPERDONAT, BRUNO BUREAU (UMR HiSoMA/MOM, LYON)

Le projet *Hyperdonat* (<http://hyperdonat.huma-num.fr/>) propose une édition bilingue latin-français du commentaire de Donat à Térence. L'objectif n'est pas l'établissement d'un texte, comme c'est le cas pour les publications analogiques, mais *des* textes avec différentes strates correspondant aux états successifs du texte. Le numérique apporte une attention renouvelée aux sources et à toutes leurs dimensions pour mettre en avant l'histoire du texte. L'idéal est alors d'aboutir à un texte multi-support qui joue de la complémentarité entre le livre et le numérique. *Hyperdonat* propose ainsi un outil d'aide à la réalisation d'une édition critique :

- Rassembler et structurer les données en TEI à travers un travail de collations, au cours duquel le collationneur est guidé afin d'obtenir un résultat homogène.

- Permettre aux chercheurs de marquer la présence de signes et de représenter le matériau. Pour faciliter ce travail, le texte est éclaté sous la forme d'une liste. Le chercheur se concentre alors uniquement sur ce qu'il voit, et non sur ce que l'auteur a voulu dire. Tout le travail d'apparat est, quant à lui, réservé à la machine.
- Créer une édition critique : la machine rassemble toutes les variantes et produit un apparat critique. Dans un premier temps, les variantes sont mises aux mêmes niveaux. C'est ensuite aux chercheurs d'opérer des choix éditoriaux.
- Comparer une édition critique avec d'autres manuscrits pour vérifier sa validité.

C'est avant tout un outil de travail, dont l'objectif est de transmettre des éditions facsimilés (montrer la transcription et sa version normalisée) ; de permettre aux utilisateurs d'isoler des traditions et de les comparer ; de représenter la tradition manuscrite *et* l'histoire du texte. Le chercheur doit comprendre les enjeux de ce qu'il produit. Pour cela, il doit avoir la possibilité de visualiser immédiatement son travail, dans une optique de contrôle qualité de l'encodage et des collations. Au cours de la réalisation de l'outil, chercheurs et techniciens s'apportent un éclairage mutuel : les contraintes numériques amènent une réflexion épistémologique sur son propre travail, qui entraîne à son tour une réflexion technique.

HYPERMACHIAVEL ET HYPERPRINCE, JEAN-CLAUDE ZANCARINI ET SÉVERINE GEDZELMAN (TRIANGLE, LYON)

HyperMachiavel (<http://hyperprince.ens-lyon.fr/hpermachiavel>), né au cours du projet ANR « Guerres 16/17 » (2007-2011), est un logiciel d'alignement, de comparaison et d'annotation lexicale de traductions. À partir d'outils structurés en XML-TEI, il permet de comparer l'édition *princeps* du *Prince* de Machiavel avec cinq de ses traductions françaises afin d'étudier la complexité du vocabulaire utilisé par Machiavel, mais aussi afin d'étudier l'acte de traduction et ses effets sur la connaissance même du texte. La mise en ligne de ce corpus porte le nom d'*HyperPrince* (<http://hyperprince.ens-lyon.fr/>).

L'enjeu a été d'imaginer un outil qui permette une analyse linguistique précise des textes, tout en étant plaisant et facilement manipulable. Pour cela, différentes outils ont été créés :

- Un outil de visualisation synoptique qui permet de comparer plusieurs textes, dont le nombre est fixé par le chercheur lui-même, et de vérifier l'alignement des textes et des mots.
- Des outils linguistiques (dictionnaires) et lexicographiques (index, concordanciers), afin d'étudier la forme, la fréquence et le contexte des mots à travers les différentes éditions.
- Des outils de visualisation, qui proposent deux types de graphes. Le premier, que l'on peut qualifier de graphe proportionnel, permet d'identifier les traductions majeures d'un mot. Le second présente l'ensemble des traductions françaises d'un mot, ainsi que les autres mots italiens traduits par les mots français qui traduisent le mot d'origine.

Cette organisation permet au chercheur de faire des expériences, dans la mesure où c'est lui seul qui contraint son outil, mais également de s'extraire des textes, de les transcender afin de visualiser les données autrement.

TEI CRITICAL APPARATUS TOOLBOX, MARJORIE BURGHART (CIHAM, LYON)

Pour présenter les rapports entre le chercheur, ses besoins et leurs expressions techniques, Marjorie Burghart s'est appuyée sur sa propre expérience. La réalisation d'une édition critique peut s'avérer être une opération périlleuse, surtout lorsque le chercheur dispose de nombreux textes témoins. En effet, la gestion de l'apparat critique devient rapidement ingérable avec un simple logiciel de traitement de texte : le volume des notes de bas de page a tendance à prendre le pas sur le texte lui-même.

Travailler en XML TEI est apparu à Marjorie Burghart comme une solution possible à la gestion de l'apparat critique. Elle a donc choisi de développer son propre outil, pour répondre aux besoins des chercheurs lors de l'établissement d'une édition critique et faciliter leur travail. C'est ainsi qu'est née la *TEI Critical Apparatus Toolbox* (<http://ciham-digital.humanum.fr/teitoolbox/>), une application en ligne qui d'une part permet de faire le contrôle qualité de son encodage et d'autre part de voir les textes en parallèle, de les comparer et de prendre des décisions éditoriales.

DISCUSSION

Ces trois projets accordent une place centrale à la visualisation et au contrôle qualité : le chercheur doit avoir la possibilité de vérifier à tout moment que ce qu'il voit est en accord avec ce qu'il a encodé. La visualisation est un aspect important pour les chercheurs, puisqu'elle apporte une distance critique par rapport à leur travail. Par conséquent, en tant qu'outils de recherche et de compréhension, les interfaces ne doivent pas uniquement se concentrer sur les données, mais également sur la mise en page qui permet de les comprendre.

La corrélation entre expression des besoins et interface apparaît comme le point le plus épineux d'un projet. L'enjeu est en effet de trouver des outils qui s'adaptent aux besoins précis des chercheurs, tout en sachant que la complexité des outils choisis peut conduire à l'abandon de choix épistémologiques, voire à une inutilisation de l'outil. Le dialogue entre chercheurs et techniciens prend alors tout son sens, afin de mettre en contact toute une communauté avec des outils qu'elle ne connaît pas toujours.

LES UTILISATEURS FINAUX

SCHOLA RHETORICA, CHRISTINE NOILLE ET FRANCIS GOYET (UNIVERSITÉ STENDHAL GRENOBLE 3)

Schola Rhetorica (<http://rare.u-grenoble3.fr/spip/>) est un projet collaboratif dédié à la rhétorique francophone. Il se compose de quatre bases de données, chacune correspondant aux quatre buts de la rhétorique : nommer, décrire, définir et écrire. L'interface se présente comme une « Fabrique de Rhétorique », divisée en trois tâches distinctes :

- Un établi pour la composition, où le chercheur-*relecteur* peut repérer, comparer et confronter les textes.
- Un atelier encyclopédique, où le chercheur-*collaborateur* peut apporter des références sur la rhétorique et les partager.
- Une bibliothèque, où le chercheur-*éditeur* peut éditer, établir et annoter les textes.

Chaque besoin est exprimé par des outils spécifiques : une interface de consultation dynamique pour étudier les textes, un wiki collaboratif pour l'encyclopédie et un établissement des textes en TEI pour les éditer.

La TEI éclate les textes en données. Ces dernières sont désormais « lues » par des machines extracteurs de données à des fins statistiques et quantitatives. On parle alors de *distant reading*, voire de *non reading* : les textes ne sont plus lus, mais quantifiés. Ce phénomène de *non reading* est ce que le projet veut éviter : l'objectif est d'être lu réellement. Pour cela, il faut que la mise en scène des données dans l'interface reproduise le parcours de lecture critique d'un chercheur dans le milieu analogique. Ainsi, l'interface de *Schola Rhetorica* se compose de trois colonnes inter-agissantes qui permettent de naviguer aisément d'un texte à un autre, d'une séquence de texte à une autre afin de les comparer ou de les analyser. Il est, par exemple, possible d'afficher un texte et son commentaire ou encore d'afficher un traité face à un terme du glossaire. Cette interface propose alors un parcours sur-mesure pour chaque utilisateur, afin de favoriser la lecture et la relecture humaine des données.

BiPRAM, CHRISTIAN DEL VENTO (UNIVERSITÉ SORBONNE, PARIS 3)

L'objectif de *BiPram* (Bibliothèques Privées à l'Âge Moderne - <http://cahier.hypotheses.org/bipram>) est de transcrire et analyser les catalogues et les inventaires des bibliothèques privées à l'époque moderne, puis de faire des liens avec les éditions existantes afin de créer une base documentaire des volumes identifiées. À terme, cette base de données permettra l'identification de réseaux d'influence entre auteurs, ainsi que l'étude de la circulation des idées à partir de celle des livres.

BiPram a pour vocation d'accompagner les chercheurs dans leur travail : l'enjeu est donc de définir un outillage en fonction des pratiques et indépendamment des technicités. Le chercheur doit pouvoir utiliser l'outil sans être confronté à la technique, ce qui ne peut se faire

qu'à l'aide d'un système *User Friendly*. Christian del Vento a présenté quelques exemples de services proposés par BiPraM en accord avec les pratiques des chercheurs :

- Créer des champs interrogeables par l'inventaire.
- Utiliser différentes couleurs pour distinguer les différents états du texte.
- Annoter les textes.
- Enrichir les textes : chaque entrée dans un catalogue ou dans un inventaire est associée à une fiche bibliographique pour retrouver l'édition correspondante.

Ainsi, l'utilisateur-chercheur nourrit la base de données que l'utilisateur vient ensuite consulter et interroger.

DIGIPAL : LES UTILISATEURS FINAUX, GEOFFROY NOËL (KING'S COLLEGE LONDON)

Parmi les utilisateurs de *DigiPal*, il est possible de distinguer deux grandes catégories : l'équipe de recherche et les utilisateurs externes, parmi lesquels se trouvent des enseignants, des étudiants, ainsi que des « visiteurs publics » (*public visitors*) dont le profil et les besoins sont inconnus. Les utilisateurs externes peuvent participer à la description d'un manuscrit en profondeur et partager leurs recherches sur Twitter ou par mail *via* un URL. Ils ont également la possibilité de créer des collections d'images, ce qui permet de décroiser et d'exploiter d'une autre manière les résultats d'une recherche. La particularité de toutes ces activités, et donc de *Digipal*, est que tout se passe du côté client. Les utilisateurs n'ont pas besoin de créer de compte pour utiliser les outils : cela permet d'encourager la participation et de créer un effet d'immédiateté. Cependant, il est impossible de savoir qui est l'utilisateur et ce qu'il fait.

Les compétences et les dispositions sont différentes entre les utilisateurs, mais il est possible d'identifier ces problèmes. L'équipe de *DigiPal* a pour cela réalisé un important *feedback*. Parmi les différentes méthodes employées, Geoffroy Noël a choisi de nous en présenter deux en détail : l'*Usability Lab* et l'*Usability Evaluation*. La première est une enquête réalisée de manière ponctuelle, à la moitié du projet, auprès d'un ensemble d'utilisateurs, afin de les connaître et d'identifier plus précisément leurs besoins. La deuxième méthode consiste en un travail individuel, réalisé par une personne extérieure au projet, sur toutes les parties publiques de l'interface. Ce rapport, à la fois subjectif et objectif, a permis de corriger l'interface pour la rendre plus intuitive.

De ce retour d'expérience, il est apparu que pour comprendre les besoins des utilisateurs, il est parfois nécessaire de changer la structure du site. La page d'accueil peut ainsi offrir des parcours différents aux utilisateurs. C'est un moyen d'identifier leur profil, mais avec le risque important de les ségréger. Ces *feedbacks* ont également eu l'effet d'un « rappel à l'ordre » : en effet, au fil du projet, on oublie certains éléments ou bien on ne les voit plus, ce que seul un œil extérieur permet de corriger. L'interface est alors améliorée, en accord avec les besoins des utilisateurs.

DISCUSSION

De ces trois interventions, quatre points importants ont été mis en évidence :

- L'un des plus grands écueils des projets en humanités est de faire de son principal public soi-même, c'est-à-dire de transposer sur les utilisateurs potentiels ses propres exigences et besoins, qui ne seront probablement pas ceux des utilisateurs finaux. Ne faudrait-il alors pas faire des transferts de compétences de communautés en communautés pour répondre au mieux aux besoins de chacun ?
- Les utilisateurs ont beau connaître les contenus proposés par un portail, ils n'en connaissent pas forcément leur modélisation. Il semble donc essentiel de guider l'utilisateur, de ne pas le laisser seul face à une page au risque de le décourager. Ce sentiment de découragement est le plus susceptible d'apparaître lors d'une recherche par le biais d'un formulaire. Une quantité trop importante de champs à remplir peut créer un véritable blocage chez l'utilisateur, blocage qu'il convient de ne pas sous-estimer. C'est dans cette optique que de plus en plus de projets proposent des recherches à facettes, qui permettent à l'utilisateur de filtrer progressivement les contenus du site. Il est important que l'utilisateur n'emprunte pas forcément les chemins qu'on lui a assignés. Il faut qu'il se perde, mais au bon sens du terme, afin de s'approprier l'interface et d'apprendre à naviguer.
- Les projets en humanités doivent proposer des outils qui engagent d'autres publics, moins spécialisés. Le *crowdsourcing* apparaît comme l'un des moyens pour parvenir à cet objectif, soit engager le public extérieur et ensuite réintégrer les productions dans la plateforme. Certains participants ont cependant émis des réserves à l'égard du *crowdsourcing* : ce dernier est-il réellement suffisant pour engager le public extérieur ? Il ne faut pas que cette solution diminue l'importance d'une interface *User Friendly* ou bien *responsive* dans l'engagement des utilisateurs. De plus, il semblerait qu'une interface collaborative demande des investissements massifs et une gestion de collaborateurs extérieurs aux cercles académiques. Il ne faut également pas oublier les enjeux éthiques que soulève une telle interface. Dans tous les cas, proposer des outils ne suffit pas : il faut les accompagner, créer un mouvement pour inciter l'utilisateur à interagir avec eux.
- Une autre difficulté est liée à la pérennité des interfaces : après un échange dynamique, les participants sont tombés d'accord sur le fait que face à la fragilité des technologies, il est important de sauvegarder la logique, la modélisation qui a permis la construction des interfaces, à travers la documentation, les articles, mais aussi et peut-être surtout à travers les métadonnées.

Finalement, qui sont les utilisateurs finaux ? Quelle est leur relation avec la recherche ? Traditionnellement, les utilisateurs sont les mêmes personnes qui produisent le projet, c'est-à-

dire les chercheurs. Et même en s'engageant dans la voie du large public, on répond souvent à nos besoins. Il semble donc important de renforcer la communauté, de monter progressivement en échelle, pour faire vivre les interfaces.