



HAL
open science

De HumanitéDigitMaghreb à HD Muren

Mokhtar Ben Henda

► **To cite this version:**

Mokhtar Ben Henda. De HumanitéDigitMaghreb à HD Muren : Dissémination de la TEI en milieu linguistique franco-maghrébin puis expérimentation de la synergie enseignement-recherche-patrimoine. Les interfaces numériques dans la recherche aujourd'hui, IXe Congrès de l'Association des Francoromanistes allemands, Sep 2014, Münster, Allemagne. hal-04546699

HAL Id: hal-04546699

<https://hal.science/hal-04546699>

Submitted on 15 Apr 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

**IXe Congrès de l'Association des Francoromanistes allemands
« Les interfaces numériques dans la recherche aujourd'hui »
Münster, du 25 au 27 septembre 2014**

De HumanitéDigitMaghreb à HD Muren :

Dissémination de la TEI en milieu linguistique franco-maghrébin puis expérimentation de la synergie enseignement-recherche-patrimoine.

Henri Hudrisier¹, Mokhtar Ben Henda²

Résumé

Nous rendons ici compte de nos recherches-action pour la dissémination de la TEI dans un contexte multiculturel et multilingue du Maghreb à travers le projet HumanitéDigitMaghreb. Nous présentons ensuite HD Muren, une expérimentation de synergie participative entre enseignement, recherche et patrimoine toujours en TEI, dans le domaine des Lettres et des Humanités en général. L'évolution de notre approche entre les deux projets, nous permet d'ouvrir quelques pistes d'une dialectique entre le document et ses possibles interfaces ; une promenade historique des réflexes d'appropriation des technologies de l'information partant des pionniers du SGML jusqu'aux *XMLnatives*.

Mots clés

Humanités digitales, synergies participatives pour l'élaboration de corpus linguistiques, XML, ergonomie, interface.

Abstract

In this paper we make report on our research concerning the dissemination of TEI in the multicultural and multilingual context of the Maghreb throughout the HumanitéDigitMaghreb project. We then present a crowdsourcing experiment named "HD-Muren" as a synergy between education, research and heritage (always using TEI), in the field of Arts and Humanities in general. The evolution of our approach between the two projects allows us to open some dialectical tracks between the document and its possible interfaces; between a historical overview of information technology ownership starting from SGML pioneers to *XMLnatives*.

Keywords

Digital humanities, crowdsourcing for linguistic corpora, XML, ergonomic interface.

¹ Maître de conférences HDR, Chaire Unesco-ITEN, Université Paris 8

² Maître de conférences HDR, Chaire Unesco-ITEN, MICA, Université de Bordeaux Montaigne

« Quand on considère un objet en lui-même & dans son être propre, sans porter la vûe de l'esprit à ce qu'il peut représenter, l'idée qu'on en a est une idée de chose, comme l'idée de la terre ou du soleil. Mais quand on ne regarde un certain objet que comme représentant un autre, l'idée qu'on en a est une idée de signe, et ce premier objet s'appelle signe. C'est ainsi que l'on regarde d'ordinaire les cartes et les tableaux. Ainsi, le signe enferme deux idées, l'une de la chose qu'il représente, l'autre de la chose représentée ; & sa nature consiste à exciter la seconde par la première. »

Antoine Arnauld & Pierre Nicole³

« Il faut élever les enfants non pas en vue de l'état actuel de l'espèce humaine, mais en vue de son état futur amendé, c'est à dire adapté à l'idée de d'humanité et à sa destination future »

Emanuel Kant⁴

1 Présentation

Notre article s'attachera à décrire les différents projets de recherche dont nous avons été acteurs ces dernières années à des niveaux de responsabilités divers et qui visaient tous au rassemblement en TEI de textes littéraires et d'autres documents (corpus oraux, musique). Selon les projets, nous soulignerons leurs disparités techniques mais aussi d'objectifs et de format financier. Nous justifierons les raisons qui nous ont poussées, comme bien d'autres collègues, à faire jouer la synergie entre projets et rechercher bien sûr à nous investir dans des projets successifs pour affiner nos problématiques. Nous pouvons constater d'ailleurs que nous avons sans doute sous-estimé la question des interfaces dans nos premiers projets (la considérant comme technique et contingente), et souligner que le colloque de Munster nous a sans nul doute aidé à reconsidérer cette question beaucoup plus fondamentale et théorique que nous aurions pu le penser il a quelques années.

De ce point de vue, notre projet actuel, HD Muren⁵, consacre une part importante de ses tâches et de son budget à analyser, puis maquetter ces questions d'interfaces. Soulignons aussi que la Chaire ITEN-Unesco à laquelle nous appartenons, nous a permis de nous associer à un autre projet de recherche sur « le livre et bibliothèque augmentés » supporté par la BnF⁶. Soulignons enfin, toujours grâce au colloque de Munster que nous avons élargi notre partenariat puisque Sacha Diwersy est devenu partenaire de notre projet HD Muren.

2 Les objectifs d'HumanitéDigitMaghreb et les projets avec lesquels il était en synergie

Le projet HumanitéDigitMaghreb avait des vocations multiples :

³ Arnauld, Antoine & Nicole, Pierre (1683)

⁴ Kant, I. (1724-1804). (1886).

⁵ HD Muren Humanités digitales multiculturelles et multidisciplinaires en mode crowdsourcing pour la recherche, l'enseignement et le patrimoine.

⁶ Voir notamment [LAB 2014-1] ; [LAB 2014-2] ; [LAB 2015]

- En tant que recherche action, il se focalisait sur des objectifs de dissémination des Humanités digitales et plus particulièrement la TEI (*Text Encoding Initiative*) pour que se développent dans différentes universités maghrébines des équipes expertes en la matière.
- Il se voulait une contribution au renforcement du potentiel d'appropriation culturelle du patrimoine littéraire, artistique et des SHS dans les pays du Sud, notamment au Maghreb.
- En synergie avec le Projet BNBf de l'OIF⁷, mais aussi avec MEI&TEI⁸ Nous avons aussi le projet plus ambitieux d'amorcer la constitution de corpus d'études codés en TEI, ou pour le moins numérisés à même de s'associer avec les grandes bibliothèques nationales ou universitaires du Maghreb, pour créer un contexte d'études littéraires et de recherches en SHS ou en art bénéficiant des avantages technologiques des documents structurés et normalisés.

Il est indéniable que nous avons réussi à « convertir » un certain nombre d'universitaires au bien-fondé théorique et pratique de la TEI. Des séminaires et même un mastère spécialisé se sont mis en place en Tunisie, en Algérie et au Maroc, en suite directe de HumanitéDigitMaghreb. Diverses disciplines ont été directement impactées par notre démarche et nombre d'universités ont bénéficié des formations pratiques que nous avons organisées : TEI pour l'Art à Gabes (TU), TEI et études littéraires à Sfax (TU), Musique et musicologie à Tunis et Sousse (TU), Méthodologie de la TEI, incidence de la création de corpus sur la documentation et la recherche littéraire tant française qu'arabe à l'université de la Manouba (Tunis), TEI et corpus oraux pour des études linguistiques berbères (Tizi-Ouzou, Algérie), TEI pour l'analyse des contes et de la littérature orale berbère à Oujda (Maroc), TEI, travaux sur des corpus littéraires et sur la terminologie berbère à Agadir (MA). Ces pôles de compétences que nous avons grandement contribué à amorcer, pour ce qui est de l'aspect TEI, restent encore aujourd'hui liés à nos activités à Paris (EHESS, Chaire Unesco ITEN-Paris 8) et à Bordeaux. De ce point de vue, l'objectif de dissémination théorique et pratique a bien fonctionné pour le petit nombre d'universitaires que nous avons cooptés et certains de leurs étudiants. Il est cependant indéniable que, comme en Europe (et pire qu'en Europe), il y a loin de l'adhésion inconditionnelle de quelques *aficionados* de la TEI à une appropriation académique de masse qui devrait logiquement se produire. Nous sommes aussi loin d'une prise de conscience des décideurs de patrimoines et il faudrait rapidement mettre en place d'importants corpus numériques structurés en langue arabe, berbère et française⁹ sur lesquels pourraient travailler un nombre croissant de chercheurs et divers usagers (scolaires, médias, etc.). En effet, nous pensions dès le début du projet qu'il est très utile et très opportun, pour le Maghreb, comme pour l'Europe et pour le Moyen Orient, que de tels corpus puissent exister pour contribuer à rééquilibrer le volume relatif des corpus numériques (structurés si possible) en langue arabe ou berbère face aux corpus en langues européennes pour l'étude littéraire, les SHS et l'art.

⁷ BNBf (Bibliothèque Numérique Franco Berbère), un projet soutenu par l'OIF-FFI (Organisation de la Francophonie, Fonds Francophone des Inforoutes) 2012-2014.

⁸ Auquel succède actuellement le projet coordonné par Sylvaine Leblond-Martin : MEI Æ NORMA, MEI, Auralités Et Nouvelles Oralités des Musiques Actuelles, un projet lui aussi soutenu par Idéfi-CréaTIC. Voir notamment [MAR 2013-1] ; [MAR 2013-2] ; [ROL 2014] ; [HAN&all 2014]

⁹ Et d'autres langues partenaires modernes ou anciennes du Maghreb : anglais, espagnol, italien, maltais, diverses langues du monde juif, grec et latin.

Induire plus d'égalité culturelle et idéologique: on fait l'hypothèse que la mise en place de tels corpus aurait des effets bénéfiques pour contrebalancer par des textes littéraires, philosophiques¹⁰, scientifiques modernes ou anciens, les contenus dominants actuels : dans le meilleur des cas, la profusion numérique de textes religieux en provenance presque exclusivement de l'Arabie Saoudite ou du Golfe, contribue grandement à radicaliser les citoyens du Maghreb. Dans le pire des cas, il s'agit de propagande terroriste mise en ligne par *Daech* et autres *Front Al Nosra*. Cet objectif de rééquilibrage culturel et linguistique faisait partie des objectifs sociétaux du projet HumanitéDigitMaghreb et dans une moindre mesure de ceux du projet HD Muren. Par contre, pour ce qui est de l'importante question de l'appropriation de la TEI, il nous paraît clair que dès que nous voulons élargir le public d'utilisateurs de la TEI, il est indispensable de penser à l'ergonomie des interfaces.

Pour ce qui est du projet BNBf, financé par l'OIF, il est évident qu'il était plus facile de se l'approprier, dans la mesure où les documents de cette bibliothèque numérique n'étaient pas supposés être nécessairement structurés, ni en TEI, ni même en XML. Il suffisait qu'ils soient numérisés et catalogués, ce qui fut réalisé en Dublin Core sur une plateforme OMEKA¹¹. Malgré sa focalisation par construction, sur les langues berbères et le français¹², le projet a bien rempli ses objectifs : rassembler des documents pour réaliser une Bibliothèque Numérique Franco-Berbère. L'interface OMEKA qui avait été choisie est conviviale, mais les chercheurs berbérissants impliqués dans le projet partageaient tous globalement le même sentiment : « des corpus relativement importants, directement accessibles en numérique, cela augmente bien sûr, notre visibilité interne et même internationale ! Mais pourquoi faire, si cela n'augmente pas notablement notre capacité à pouvoir mieux travailler ? »

Ce qui nous importe comme enseignant-chercheur berbérissant, c'est pouvoir mieux traiter et plus efficacement nos corpus d'étude qu'avec nos documents analogiques traditionnels : les transcriptions papier (ou sur bandes magnétiques audio) de nos recueils linguistiques kabyles, les recueils de contes et de littérature berbères, le foisonnement médiatique de presse, de radio, des disques et des très nombreuses émissions de télévision en langue berbère ? Comment aussi mieux s'organiser en France pour organiser une épreuve optionnelle de langue berbère¹³ (orale et écrite) au Baccalauréat ? Il faut d'ailleurs poser des problèmes similaires, notamment en ce qui concerne l'épreuve optionnelle de langue arabe ? ».

Dans le projet BNBf, la spécificité franco-berbère constituait le fondement même du projet et de

¹⁰ Notamment les textes de la Renaissance arabe, avant-coureurs de la Renaissance européenne ou encore les traductions de textes antiques en arabe qui ont servi de conservatoire des grands textes grecs ou latins.

¹¹ Omeka est un logiciel de gestion de bibliothèque numérique mis à disposition sous licence libre.

¹² Dans la limite de ses moyens et de ses déboires politiques. En effet, le Niger était partenaire du projet à travers l'étude et le rassemblement de corpus linguistiques en langues touareg (les langues touarègues appartiennent à la famille des langues berbères). Malheureusement le coup d'Etat advenu en plein projet, a remis en cause une part très importante du financement du projet (à l'origine un volume de financement plus important que celui additionné du Maroc et de l'Algérie – ce qui était logique vu le coût des missions). Il en résulte un appauvrissement conséquent de l'ensemble du projet, tant en termes de corpus à rassembler, que de synergies des partenaires et bien sûr de non attribution d'une part très conséquente des financements.

¹³ On trouve profusion de publications sur la diversité des parlers et langues berbères. Voir par exemple : [ELM 2005]

son financement. Soulignons que tous les chercheurs liés aux contenus étaient naturellement trilingues au minimum : arabophones, berbérophones et francophones. On pourrait d'ailleurs souligner que ce trilinguisme « quasi-simultanément maternel » se doublait chez tous d'une compétence tri-scripturale : écrire et lire en caractères latin, arabes et tifinagh¹⁴. De plus, tous ces chercheurs étaient des passionnés de la « cause berbère ». C'est la raison pour laquelle, les chercheurs du projet BNF étaient tous intéressés et demandeurs pour s'associer au projet HumanitéDigitMaghreb¹⁵, apprendre avec leurs collègues arabisants à dépasser le premier niveau du catalogue Dublin Core et organiser des corpus structurés en TEI et notamment maîtriser les modalités de balisage spécifiques pour des corpus oraux (*TEI-speech*), des terminologies (*TEI-dictionary*), des poésies (*TEI-verse*), baliser des corpus chantés et musicaux, (*TEI&MEI*), savoir s'organiser pour coder des contes selon des approches savantes et sémiotiques, coder des variantes de textes, leurs interprétations savantes plurielles (*TEI apparats critiques*), gérer des manuscrits anciens¹⁶ (*TEI-manuscrit*).

On comprend dès lors, combien la soif d'un rééquilibrage patrimonial culturel et linguistique dans des pays longtemps colonisés, constitue en fait un moteur puissant de mobilisation dans le secteur qui nous intéresse. Malheureusement le Maghreb est relativement sous équipé (et sous encadré) pour ce qui est des institutions patrimoniales. De plus, dans nombre de pays ex-colonisés, les grandes collections patrimoniales qui les concernent (notamment les fonds anciens) sont pour une large part conservée dans les grandes bibliothèques et les grands musées des puissances ex-coloniales. Face à ce déséquilibre, la demande n'est pas tant celle du rapatriement des documents primaires, cela paralyserait leurs capacités aujourd'hui assez faibles de conservation. Il s'exprime surtout une demande de numérisation massive de ces fonds (au Nord comme au Sud), mais cela n'a de sens que si se développe simultanément une forte compétence d'appropriation conceptuelle de ces fonds numériques qui peut induire ultérieurement une dissémination numérique de ces patrimoines au Maghreb, dans le monde arabe et au-delà.

Dans cet article, nous nous garderons bien de rentrer dans la polémique souvent stérile des rapatriements patrimoniaux symboliques organisés en guise de repentance néocoloniale¹⁷. Il est beaucoup plus intéressant à l'âge de la mondialisation numérique de s'interroger sur la capacité d'appropriation technique, conceptuelle, culturelle et linguistique de ces fonds. Si les fonds orientaux des grandes bibliothèques européennes sont fantastiquement riches, il faut bien sûr que la totalité des chercheurs, locuteurs natifs de ces aires linguistiques, puissent y accéder et puissent

¹⁴ Il est remarquable que l'écriture traditionnelle berbère, les caractères tifinagh n'ont été que tardivement normalisés en UNICODE. [ZEN 2004]. Sur Unicode voir aussi [AND&HUD 2002] et notamment un article de Rachid Zghibi [ZGH 2002]

¹⁵ Prendre en compte toutes les composantes culturelles et linguistiques modernes ou anciennes de la société maghrébine et de ses diasporas.

¹⁶ Textuels mais aussi musicaux. Voir sur ce point : [OUR&HAS 2012] ; [HAN&all 2014] ; [GUE 2008]

¹⁷ On doit d'ailleurs noter qu'un pays comme la France n'a pas la capacité juridique de vendre ou de rétrocéder des objets constituant ses collections nationales, contrairement au Royaume Uni qui peut vendre certaines pièces de ses collections nationales pour faciliter le fonctionnement de ses établissements culturels. Lorsque des rétrocessions ont été faites, par exemple pour la restitution de têtes maories à la Nouvelle Zélande (2009) ou au retour sur le sol coréen des manuscrits royaux de la dynastie Joseon (2010), il faut savoir que cela n'était qu'une mascarade diplomatique au cours de laquelle les autorités culturelles françaises sont convenues avec leurs homologues étrangers d'un simple « prêt à long terme » des dites pièces patrimoniales qui restent formellement propriété de la France.

localement « aménager conceptuellement » ces fonds numériques, ce qui justifie la réalisation de fonds numériques structurés¹⁸.

Les deux projets BNBf et HumanitéDigitMaghreb étant maintenant terminés, nous devons prendre acte que si l'interface OMEKA du projet BNBf était ergonomiquement facile à s'approprier par des chercheurs en Sciences humaines ayant un niveau moyen de l'usage informatique, par contre, il en était tout autrement pour ce qui concerne les difficultés ergonomiques d'appropriation de la TEI dans le projet HumanitéDigitMaghreb.

3 Quelles ergonomies d'interfaces TEI pour répondre aux objectifs de nos projets ?

Avant de répondre à cette question, il est important de faire un retour historique pour comprendre le rapport complexe que la communauté TEI entretient avec la notion d'interface.

3.1 Le TEI consortium non concerné par les interfaces ?

On connaît tous les causes ontologiques de la faiblesse de la TEI et des Humanités Digitales en général en matière d'interface. Rappelons en effet que les objectifs généraux de la TEI ont été définis par la résolution finale de la conférence préparatoire tenue au Vassar College à Poughkeepsie (NY, USA) en novembre 1987. Cette résolution connue sous le nom de « Principes de Poughkeepsie » fut peu à peu précisée à travers une série de documents de travail. D'après ces documents les « Recommandations [en anglais *Guidelines*] devaient :

- Être suffisamment précises pour représenter les propriétés des textes intéressants pour les chercheurs ;
- Être simples, claires et concrètes ;
- Être utilisables facilement par les chercheurs et ne pas nécessiter l'utilisation de logiciels spécifiques¹⁹ ;
- Permettre une définition rigoureuse des textes en vue de traitements efficaces ;
- Être modifiables par l'utilisateur ;
- Respecter les normes en vigueur ou sur le point d'être adoptées.

Le monde de la recherche étant large et divers, pour que ces recommandations aient une large audience, il était important de s'assurer que :

- Les descriptions des caractéristiques fondamentales d'un texte puissent être facilement échangées ;
- Les descriptions spécialisées puissent être facilement ajoutées (ou supprimées) d'un texte ;
- La même caractéristique puisse être encodée en parallèle de plusieurs façons ;

¹⁸ Il est très difficile de ré-intervenir plusieurs fois et depuis des lieux divers sur des fonds numériques non structurés. En TEI, par exemple l'en-Tête TEI, permet de gérer des pluralités de versions successives, les campagnes de balisages peuvent être multiples, tout cela sans risquer de détruire la cohérence des données.

¹⁹ Ce qui signifie si on prend la recommandation à la lettre que toute la facilité ergonomique doit résider dans l'intelligence structurelle des documents et des corpus.

- La richesse du balisage puisse être déterminée par l'utilisateur de la façon la plus simple possible ;
- Une documentation relative au texte et à la façon dont il a été codé puisse être fournie.²⁰ »

C'est précisément grâce à l'indépendance de principe de son métalangage de codage vis à vis de toutes sortes d'interfaces techniques, qu'un si grand nombre de chercheurs peuvent encore, presque 30 ans après sa fondation, se reconnaître encore dans la TEI. Les Guidelines (Recommandations) TEI²¹ se sont sophistiquées et enrichies, mais la globalité du système de codage est restée le même assurant rigoureusement son interopérabilité interne. L'arrivée du HTML, puis du XML qui se sont tous deux rigoureusement construits comme des DTD de SGML n'ont eu pour conséquence que de conforter les options d'origine de la communauté TEI qui s'était dès l'origine basée sur SGML. Contrairement à quantité d'autres créateurs ou gestionnaires de données qui adoptèrent (mais aussi subirent) la révolution d'HTML puis de XML, la communauté TEI n'eût seulement qu'à s'organiser à la marge pour intégrer cette progression évolutive allant de SGML à HTML puis XML.

3.2 Les langages à balises de la famille « ML », une révolution ergonomique historique des langages informatiques : des informations « Men & Machine Readable »

Pour bien comprendre, du point de vue techno-historique, la pertinence du choix de la communauté TEI qui avait choisi SGML, puis bien sûr intégré très vite HTML et XML, il faut prendre le temps de considérer combien ces langages de balisages constituèrent une « révolution ergonomique de l'informatique ». Rappelons que cette transformation radicale de la philosophie techno-ergonomique du traitement de l'information textuelle s'origine presque simultanément (milieu des années 60) chez les bibliothécaires et chez les éditeurs de grosses documentations techniques (aéronautiques notamment). Pour ces deux communautés professionnelles, il était nécessaire de définir une ergonomie de leurs masses de données qui puissent être lues à la fois par des humains et des machines, mais aussi en interactions itératives²² : c'est la philosophie des *Men & Machine Readable Languages*. Remarquons d'ailleurs au passage que cette exigence ergonomique du balisage s'est développée selon deux directions inverses de facilitation ergonomique :

1) *Les grandes bibliothèques voulaient continuer d'utiliser la compétence des catalogueurs, pas forcément enclins à se lancer dans l'informatique mais qui pour la plupart savaient déjà dactylographier leurs fiches de catalogage qui comportaient des séparateurs de zones normalisés*²³.

²⁰ C'est la traduction française de TEI Lite : An Introduction to Text Encoding for Interchange [CAH 1996]

²¹ [TEI 2015]

²² Lorsque le code est présenté sous forme de balises explicitement distinctes du contenu du document grâce à des séparateurs faciles à repérer : par exemple [...] en SGML, HTML, XML, TEI ou aussi \...{...} en Latex, il devient possible pour un humain de lire un document en oubliant mentalement le code ou au contraire de contrôler à l'œil le positionnement correct de telle ou telle balise facilement isolable du contenu du document. Si par ailleurs les balises sont rédigées dans un code proche d'énoncés abrégés des commandes ou de l'étiquetage du contenu, cela facilite d'autant plus le potentiel de lecture du document balisé et codé directement à l'œil. Complémentairement, on lance de façon itérative, le contrôle par la machine de la qualité des documents « bien formés », mais aussi du bien-fondé d'une méthode de codage, puis de sa formalisation définitive dans un schéma XML ou TEI.

²³ En fait, une syntaxe des zones de catalogage (et de leurs séparateurs) normalisée de façon internationale au sein de l'IFLA (International Federation of Library Associations and Institutions). L'IFLA a été fondée en 1927.

Dès 1965, à la Bibliothèque du Congrès, l'informaticienne Henriette Avram propose de développer des formats MARC (acronyme de *MACHine-Readable Cataloging*) qui deviendront des formats d'échange de données bibliographiques permettant d'informatiser les catalogues de bibliothèques²⁴, que ceux-ci soient manuscrits, dactylographiés ou informatisés. Chaque zone de la fiche catalographique devait apparaître dans un ordre conventionnel (par exemple : nom d'auteur, prénom ou initiale, titre, ville d'édition, édition, date), avec un style de typographie prédéfini (capitale pour nom d'auteur, italique pour titre) et avec une ponctuation de séparation elle aussi normalisée et signifiante à base de tirets, de points, de virgules, de deux points et autres barres obliques. En fait les bibliothécaires avaient inventé depuis longtemps ce langage de balisage parfaitement ésotérique mais efficace. Le format MARC n'a été en fait que sa traduction informatique et a permis de définir dans des systèmes informatiques des masques de saisie de zones bibliographiques correspondant au savoir-faire des bibliothécaires : pouvoir créer des fiches catalographiques traditionnelles lisibles par des machines, mais surtout, échanger des catalogues à l'échelle internationale ou plus modestement interbibliothèques :

2) *Sur un autre versant, la communauté des imprimeurs* avait disposé assez tôt de « logiciels maison » leur permettant de fabriquer des bandes de photocompositions qui malheureusement étaient pratiquement incompatibles dès que l'on changeait de fournisseur informatique. Il fallait donc, non pas inventer UN langage de balisage d'informations textuelles à vocation éditoriale (il en existait déjà quelques-uns), mais en imposer UN en lui donnant une grande valeur ajoutée concurrentielle, notamment en permettant aux éditeurs de documentations techniques complexes (aéronautique, annuaires, notices techniques, indicateurs de chemin de fer) de pouvoir gérer au quotidien leurs modifications éditoriales (sans recourir à chaque fois aux informaticiens-typographes) et de pouvoir utiliser dans la même application informatique des bases calculables et/ou imprimables. Il s'agissait dès lors pour les informaticiens et les typographes-informaticiens de mettre à la disposition de leurs utilisateurs, un langage leur permettant de structurer de façon pérenne et universelle leurs grandes quantités d'informations textuelles pour les modifier, gérer des versions, structurer typographiquement, logiquement ou sémantiquement leurs données. Ce nouveau langage fût IBM-GML qui se standardisa progressivement à partir du moment où il devint SGML²⁵. Ces communautés d'utilisateurs (éditeurs, responsables de documentations techniques, responsables de catalogues commerciaux ou industriels avec articles, prix et références, etc.) devaient pouvoir lire leurs textes ou leurs données, mais en pouvant optionnellement et à tout moment intervenir sur des codes de balisages suffisamment explicites pour être directement compréhensibles, hiérarchisés par fonctions et familles d'ordres ou d'étiquetages (logiques, structurels, référentiels, sémantiques). Le progrès ergonomique suivait donc une direction inverse de celui des bibliothécaires. Il fallait que la logique informatique partagée de façon ergonomiquement disparate chez des imprimeurs, des éditeurs, des ingénieurs ou des gestionnaires

²⁴ Grâce à ce langage balisé catalographique les premiers services d'échange sont lancés en 1969 à partir de la Bibliothèque du Congrès. Ces échanges catalographiques établiront durablement leur suprématie en la matière.

²⁵ « En 1969, Charles Goldfarb, qui est alors chef de projet chez IBM, fait lancer par cette compagnie un langage descriptif GML (Generalized Markup Language), destiné à encapsuler l'ancien langage Script trop lié physiquement aux possibilités techniques des imprimantes. Goldfarb, que GML rend vite célèbre, quitte alors cette compagnie pour développer un successeur de GML appelé SGML (Standard Generalized Markup Language), publié en 1986 comme norme ISO (ISO 8879:1986). » [WIK 2015]

de données complexes, puisse s'unifier grâce à un langage indifféremment lisible par des hommes ou par des machines (*man&machine readable*) : ce fut SGML.

Rappelons au passage une plaisanterie en français sur la signification de ce sigle qui avait cours dans la communauté des utilisateurs de SGML : le sigle se déclinait non pas en « *Standard Generalized Markup Language* », mais en « Sûrement Génial Mais Laborieux ». Ce jeu de mots traduit bien la difficulté d'appropriation ergonomique d'un tel langage.

Pour la génération des *XMLnatives*, ce débat historique doit sembler bien inutile. Comme le dit si bien John Searle²⁶, la « réalité sociale » qui caractérise notamment l'environnement techno-culturel, est implicitement compréhensible pour chaque nouvelle génération. Pour les professionnels ou usagers de l'information et de la communication qui ont commencé à utiliser l'informatique après 1990 (apparition du Web), l'universalité de ces langages de balisage, leur réalité ergonomique *men&machine readable* va de soi. Ils n'imaginent pas à quel point, pour ceux qui ont vu apparaître ces langages dans les années 70 et 80 pour les mieux informés, puis dans les années 90 pour le grand public découvrant HTML, cette nouvelle génération technologique du document balisé SGML constituait déjà en soi une véritable interface. Bien sûr une interface qui nous apparaît aujourd'hui peu conviviale, mais qui est malgré tout une interface.

De plus cette génération des pionniers des langages à balises avait encore en mémoire les décennies qu'il leur avait fallu attendre pour se libérer ergonomiquement de l'omni-emprise des informaticiens. Il faut pour cela rappeler que jusque dans les années 85²⁷ (les utilisateurs de SGML constituaient un tout petit club), les usagers d'un service informatique ne pouvaient presque rien gérer par eux-mêmes. Ce qui était déjà prévu pour être « saisi » dans des cases pouvait l'être, mais la moindre adaptation, le moindre paramétrage exigeait absolument l'intervention des informaticiens professionnels²⁸. La nouvelle philosophie technique des langages à balises qui se popularisèrent largement avec HTML (début des années 90), sonna comme le gong de la fin d'un round : celui de la suprématie des exigences (ou plutôt du dictat) de l'informaticien sur l'utilisateur. Désormais, l'utilisateur pouvait avoir la main sur le traitement de ses documents. L'exigence de séparation drastique et intangible du document d'avec l'environnement informatique et matériel était le prix à payer pour bénéficier de cette liberté de l'utilisateur, malheureusement les générations des *XMLnatives* pourraient avoir tendance à l'oublier.

De ce fait, pour la communauté TEI, il était et il reste à juste titre inenvisageable de revenir sur cette résolution de séparation drastique sous peine de risquer de se voir confisquer les savoir-faire savants que de milliers de spécialistes en littérature, en SHS, en linguistique, en archivistique, en

²⁶ [SEA 1998]

²⁷ Période à laquelle les PC commencent à se généraliser.

²⁸ Dans ses Essais, Montaigne situe bien l'état de compétence cognitive par rapport au calcul chez l'honnête homme de la Renaissance : « Je suis né et nourri aux champs parmy le labourage ; J'ay des affaires et du ménage en main depuis que ceux qui me devançaient en la possession des biens que je jouis m'ont quitté leur place. Or je ne scay compter ny à jet ny à plume. » MONTAIGNE (Michel de), Essais, livre II. En fait, cette dépendance du « calculateur de profession », jusqu'à tard dans l'âge classique s'est rejouée, mutatis mutandis dans la première période de l'informatique et jusqu'à l'apparition des micro-ordinateurs ET des langages à balises

paléographie ont consensuellement accumulés et formalisés depuis presque 30 dans les TEI guidelines²⁹ (avec l'aide précieuse et indispensable de Sperberg-McQueen et Lou Burnard).

Malgré la détermination passionnée des leaders de la TEI à disséminer leurs méthodes, force est de constater que transparaît parfois encore aujourd'hui un certain élitisme et un certain ésotérisme académique de certains membres de la communauté TEI. Il est naturel en effet que des chercheurs très spécialisés (étude des manuscrits, apparats pour édition critique, codage des corpus oraux, balisage des dictionnaires) considèrent, à bon droit, qu'ils peuvent :

- 1) Investir quelques semaines de leur vie de chercheur pour maîtriser sans l'aide d'interfaces WYSIWYG³⁰, le sous-ensemble des balises qui leur permettra de réaliser leurs recherches spécialisées ;
- 2) Éventuellement faire développer des interfaces ad hoc pour leur application TEI, ou plutôt comme cela se généralise de plus en plus, recourir à des instances spécialisées comme DARIAH ou en France HumaNum ;
- 3) Considérer que leurs outils et leurs méthodes sont trop spécialisés pour valoir la peine d'être présentés, même face à un « grand public cultivé ».

Néanmoins les faits sont têtus. Encore aujourd'hui, on doit constater que malgré son extraordinaire pertinence, non seulement pour la recherche académique mais pour l'éducation, les bibliothèques numériques, les archives, la linguistique, la gestion industrielle des textes (notamment la gestion de multiples versions), la TEI a encore la réputation d'être difficile à s'approprier. Cela reste vrai.

Et pourtant, venant en contradiction avec les remarques ci-dessus, rien que pour la France, il ne se passe pas de semaine sans qu'un appel à formation TEI soit publié dans des réseaux académiques généralistes. Les Archives de France, l'École des Chartes, l'ENS (École Normale Supérieure) et maintenant un certain nombre de filières universitaires organisent des sessions de formations, voire des cycles de master spécialisés sur les Humanités digitales et la TEI en particulier.

On peut en tirer la conclusion que la TEI a réussi son difficile pari de départ : s'imposer mondialement comme un métalangage universel et standardisé de description de contenus et de corpus indépendants de toute technologie d'interface.

Néanmoins, il devient d'année en année de plus en plus évident que la communauté TEI a réussi à convertir à sa cause nombre d'instances décisionnaires européennes et américaines de la recherche en littérature, en Art et en SHS. Cela se concrétise en Europe avec DARIAH (*Digital Research Infrastructure for Arts and Humanities*) et localement en France, où s'est mis en place d'abord ADONIS et maintenant Huma-Num qui est une très grande infrastructure de recherche (TGIR) visant

²⁹ En fait les chapitres de spécialité des TEI guidelines notamment : Chapitre 6 Verse, Ch. 7 Performance Texts (ie. théâtre, scénario), Ch. 8 Transcriptions of Speech, Ch. 9 Dictionaries, Ch. 10 Manuscript Description, Ch. 12 Critical Apparatus.

³⁰ « what you see is what you get »

à faciliter le tournant numérique de la recherche en Sciences humaines et sociales³¹.

Vu de l'extérieur, il nous apparaît que les leaders du TEI Consortium peuvent considérer à bon droit qu'ils ont réussi à développer avec la TEI P5 un univers de balisage très complet pour répondre aux besoins de la communauté savante des Humanités digitales. Ils doivent maintenant pouvoir se consacrer à répondre à des enjeux de dissémination au-delà de leurs *aficionados*. Sans être dans le secret des stratégies du TEI Consortium, il semble qu'aujourd'hui un mouvement s'amorce pour que le partenariat du TEI Consortium avec *oXygen XML Editor*³² s'amplifie, permettant sans doute d'offrir des interfaces de plus en plus ergonomiques et conviviales (*oXygen* ++ ?) tout en restant fidèle aux résolutions de séparation absolue du document avec les plateformes³³.

4 Évolution de nos objectifs de HumanitéDigitMaghreb à HD Muren

HumanitéDigitMaghreb était un projet à très petit budget (10 000 euros sur 2 ans) à travers lequel nous avons la « volonté militante » de disséminer la TEI au Maghreb et dans la Diaspora maghrébine en France. Nous voulions par ailleurs répondre à des questions de « recherche-numérique en Humanités », donc initier au balisage TEI, nos partenaires du projet BNFB et nombre de nos amis universitaires au sud de la Méditerranée. Au-delà de la création de ce réseau de compétence académique, il nous semblait important de pérenniser ces premiers résultats et d'amorcer la réalisation de corpus sur le Maghreb et en langues maghrébines. En l'absence de nouveaux projets « du Nord » venant conforter ces premiers résultats, le suivi de ces premiers acquis était loin d'être assuré.

Soulignons cependant, que pour des raisons qui sont à la fois anecdotiques³⁴ mais qui tiennent aussi à l'excellence techno-informatique des musiciens, nous avons pu constater que poésie et musique (qui ne font qu'un dans la culture arabo-berbère), constituaient un excellent terrain d'application et d'expérimentation de la TEI, puis de la TEI associée à la MEI (Music Encoding Initiative). D'abord parce que la poésie mais aussi le théâtre, qui sont des textes « naturellement structurés » dans lesquels fond et forme sont intimement liés pour l'analyse savante (ou même scolaire) des textes. De plus, les poèmes sont des textes souvent courts (tout en étant très riches pour l'analyse). Il est donc assez facile de maîtriser le petit univers des balises de bases permettant de démontrer l'excellence du procédé. Il est relativement aisé d'adapter TEIverse au codage de poèmes arabes et

³¹ Pour remplir cette mission, la TGIR Huma-Num est bâtie sur une organisation originale consistant à mettre en œuvre un dispositif humain (concertation collective) et technologique (services numériques pérennes) à l'échelle nationale et européenne en s'appuyant sur un important réseau de partenaires et d'opérateurs. Voir : <http://www.huma-num.fr/la-tgir-en-bref>

³² *oXygen XML Editor* est un excellent éditeur XML. Il fournit une suite complète d'outils de création et de développement XML et il est notamment compatible avec la totalité des balises TEI. Il est conçu pour accueillir un grand nombre d'utilisateurs, allant des débutants (ergonomie moyenne mais efficace) jusqu'aux experts XML.

³³ Lors du congrès de la TEI à Lyon en octobre 2015, on peut noter une communication importante qui devrait aller dans ce sens : « Extending the TEI support and new features in *oXygen* 17.0 - by Alexandru Jitianu and George Bin ».

³⁴ Rencontre entre Anas Ghrab (Enseignant-chercheur à l'École de musique de Sousse et Directeur du Centre des Musiques Arabes et Méditerranéennes à Sidi Bou Saïd -près de Tunis- spécialiste des appareils critiques dans les manuscrits musicaux arabes) et Sylvaine Leblond-Martin (Compositrice, ingénieure de recherches à IDEFI CréaTIC de l'Université Paris 8, désormais initiatrice d'un projet compagnon de HD Muren, un autre projet Idéfi-Créatic : MEI Æ NORMA)

d'en découvrir les exigences de paramétrage du balisage TEI. Cela fût fait grâce notamment à un enseignant chercheur de l'Université Tunis La Manouba³⁵.

Le projet M&TEleuroMed était un très petit projet (3 000 euros sur 2 ans, 2014-2015) initié par les co-auteurs de cet article à la MSH Paris Nord. Il fût essentiellement animé par Sylvaine Leblond-Martin (post-doctorante à la Chaire ITEN-Unesco) et Josée April (Conservatoire de Rimouski au Québec). Sylvaine Leblond-Martin a maintenant conçu et anime le projet MEI Æ NORMA beaucoup plus ambitieux.

Remarquons, que pour les poésies courtes mais aussi en musique, le rythme poétique, la cadence et les mesures musicales tiennent en partie lieu de facilitateur d'interface. Pour la musique, il est évident que l'informatique est devenue omniprésente dans le travail des musiciens et notamment des compositeurs. Habités à jongler avec divers logiciels, virtuoses des claviers, les usagers de la MEI (et de la synergie TEI & MEI) n'exigent pas vraiment des interfaces ergonomiques³⁶.

De son côté, le projet HD Muren est en cours de mise en place dans le cadre d'Idéfi-Créatic-Paris 8³⁷, sous la responsabilité scientifique de la Chaire ITEN-Unesco³⁸, en partenariat avec les universités de Bordeaux 3, de L'ESPE d'Aquitaine (Ecole supérieure du professorat et de l'éducation), des universités de Cologne, de Tunis la Manouba, du Rectorat de Créteil et d'autres partenaires en cours d'intégration. L'équipe de la Chaire ITEN-Unesco est particulièrement active pour animer et développer des projets dans le cadre d'Idéfi-Créatic. Nos collègues avec lesquels nous travaillons en synergie étroite développent notamment des projets avec la BnF (projet de livre augmenté sur l'Odyssée³⁹) et les Archives de France.

Ce projet HD Muren s'origine certes comme la suite logique de nos travaux antérieurs. Il s'appuie aussi sur le constat de nos difficultés à ce que se construisent des corpus, notamment des corpus dans des langues maghrébines dont nous avons déjà signalé l'importance, particulièrement la réalisation de ressources numériques renforçant des Humanités ouvertes sur toutes les cultures.

Pour ce qui concerne la France, par rapport à il y a 4 ou 5 ans, la réputation d'excellence et de

³⁵ Zghibi R. (2013)

³⁶ A priori rien n'est moins ergonomique qu'un violon ou même qu'un clavier de piano. Et pourtant, les musiciens continuent de s'en servir avec bonheur.

³⁷ Fondé en 2012, le projet CréaTIC (innover, créer, former avec le numérique) de l'Université de Paris 8 (à laquelle s'associent des partenaires tels que l'Université Paris Ouest Nanterre-La Défense, la Maison des Sciences Humaines Paris Nord, le Conservatoire National Supérieur d'Art Dramatique, les Archives Nationales et 37 partenaires étrangers) est doté d'un budget de 5,2 M€ sur huit ans. Ce grand projet cadre repose sur cinq principes fondamentaux : la création comme moteur épistémologique de pédagogies innovantes (dans le cadre d'Ateliers-Laboratoires), le recours systématique aux technologies numériques de pointe pour l'enseignement et le travail collaboratif, l'accompagnement personnalisé des étudiants (du recrutement au diplôme) dans le cadre de pédagogies par projets, l'évaluation et la transparence des résultats, l'insertion professionnelle. Aujourd'hui, CréaTIC compte 17 formations et 24 ateliers-laboratoires ouverts à plus de 500 étudiants.

³⁸ Les deux auteurs de cet article sont enseignants-chercheurs à la Chaire ITEN-Unesco de la Fondation MSH animée par Ghislaine Azémard. La Fondation Maison des Sciences de l'Homme est par ailleurs l'institution qui héberge HumaNum (Unité Mixte de Services du CNRS) voisin au même étage de la Chaire Unesco.

³⁹ Voir notamment [LAB-1 2014] ; [LAB-2 2014] ; [LAB 2015]

pertinence de la TEI s'est affirmée d'année en année dans les instances académiques. Le cadre plus vaste des Humanités Numériques devient quant à lui un véritable effet de mode, ce qui est excellent pour la dissémination de la TEI. Si nous pouvions penser, il y a 3 ou 4 ans, que la dissémination de la TEI en milieu académique au Maghreb était un objectif de recherche-action en Sciences de l'information digne d'intérêt, aujourd'hui l'excellente réputation dont la TEI jouit désormais rendent sans doute le seul objectif de dissémination de la TEI moins pertinent.

Par contre, le relatif isolement académique de la communauté TEI⁴⁰ par rapport à son fantastique potentiel d'aide à l'intelligence (la compréhension, l'analyse, l'interprétation, la critique) des textes au-delà du milieu étroit des chercheurs déjà initiés, nous semble constituer un vrai sujet pour le chercheur en sciences de l'information, mais aussi pour le chercheur en Sciences de l'éducation qui étudie la pertinence de la TEI notamment pour la pédagogie de la littérature⁴¹. En collaboration avec Laurent Romary (un des pionniers français de la TEI), un des deux auteurs avait déjà évoqué dès le début des années 2000 le sujet de l'usage de la TEI comme outil de pédagogie pouvant fonctionner en synergie avec la recherche⁴².

Nous avons aussi émis l'hypothèse qu'une synergie enseignement/recherche/patrimoine pouvait être mise en mouvement dans un cercle de synergies vertueuses dans lequel chacun de ces 3 pôles est gagnant-gagnant.

La synergie participative enseignement/recherche/patrimoine est l'objectif princeps du projet HD-Muren⁴³ que nous mettons en œuvre dans le cadre d'IDEFI-CréaTIC, un programme en formations innovantes qui lui sert de cadre institutionnel et conforte certains des objectifs d'HD Muren. En effet, l'objectif de ce grand programme-cadre est de mettre en œuvre de nouvelles méthodes pédagogiques associant enseignement et recherche notamment dans des formations liées à l'usage des dispositifs et technologies numériques pour l'enseignement, la création et le travail collaboratif. Dans ce cadre institutionnel structurant, HD-Muren vise à proposer la TEI comme cadre d'une pédagogie active liant enseignement et recherche en Lettres, Sciences Humaines et Arts. Nous faisons aussi l'hypothèse qu'à l'ère du numérique, les patrimoines culturels, les ressources pédagogiques, les exercices des apprenants, les travaux divers des chercheurs ou encore les œuvres des créateurs, constituent potentiellement un même continuum à condition que l'ensemble de ces corpus soient balisés et structurés de façon standardisée, interopérable et normalisée. Ces corpus pourront ainsi s'associer de façon cumulative en utilisant les matériaux réalisés comme exercices pédagogiques en leur donnant une valeur scientifique ajoutée et en faisant circuler des ensembles structurés de corpus numériques. Cette collégialité respectueuse de ces diversités d'objectifs (enseignement, recherche ou gestion de patrimoines, inter-compatibilité normative) est précisément ce que permet la TEI.

⁴⁰ Que l'on a mainte fois justifié plus haut.

⁴¹ Mais aussi pour d'autres disciplines qui relèvent de la textualité (philosophie, linguistique, apprentissage des langues, paléographie, étude des manuscrits, des dictionnaires, de la musique avec la MEI, mais aussi une synergie TEI&MEI). Cette pédagogie interdisciplinaire peut concerner l'enseignement secondaire comme le supérieur.

⁴² [ROM&HUD 2003]

⁴³ HD-Muren : « Humanités Digitales multiculturelles en mode crowdsourcing via la recherche et l'enseignement »

Notre projet HD Muren s'appuie sur trois premiers constats :

- 1) La transformation du fait numérique, de son économie, de sa prise en compte professionnelle et institutionnelle est grandement conditionnée par son appropriation de masse (digital natives) et par l'évolution d'une informatique surtout locale à la maîtrise mondiale des réseaux ;
- 2) La synergie enseignement/recherche/patrimoine, qui de notre point de vue avait été la règle depuis l'invention de la science (Antiquité puis Renaissance) et jusqu'au milieu du XX^{ème} siècle, avait perdu beaucoup de son importance après la Deuxième guerre mondiale mais redevient d'actualité grâce aux réseaux numériques ;
- 3) Le constat directement lié au point précédent, le « patrimonial numérique » mais aussi des « ressources pédagogiques en ligne » sont deux domaines dont on ne peut continuer à garantir la croissance exponentielle actuelle sans recourir à des participations externes à leurs pôles métiers respectifs : cela conforte notre hypothèse de faire jouer les synergies entre enseignement, recherche et gestion ou conservation des patrimoines.

Si nous néglignons trop longtemps d'activer ces potentiels de symbiose entre ces trois pôles, nous augmenterions indubitablement les risques de marchandisation Google de ces biens *communs* patrimoniaux. Face à ces risques, le recours à une production participative des utilisateurs (*crowdsourcing*) est indispensable, les deux candidats-partenaires potentiels étant d'une part les seniors et d'autre part les élèves, les étudiants et leurs professeurs. Dans le cas de notre projet, il nous paraît parfaitement légitime que les élèves et les étudiants (mais aussi les enseignants) qui sont les premiers bénéficiaires mais aussi utilisateurs de la recherche en Lettres et Sciences humaines soient incités à devenir contributeurs. Sur l'autre versant, les chercheurs en Lettres, Arts et SHS⁴⁴, mais aussi les gestionnaires de patrimoines ont, eux aussi, grand intérêt à partager en synergie la construction de leurs corpus. Par contre, cette synergie doit s'accompagner de règles éthiques : les élèves, les étudiants de première année et les professeurs doivent pouvoir bénéficier de nombreuses rencontres avec les chercheurs et les professionnels du patrimoine pour se sentir ainsi acteurs à part entière des progrès scientifiques et de l'augmentation des patrimoines.

Ces contacts doivent susciter des vocations et permettre de contribuer à la plus-value scientifique ou patrimoniale. Les élèves et les professeurs utilisateurs doivent ressentir à quel point ces corpus sont des biens communs précieux, difficiles et dispendieux à mettre en œuvre, et constater par l'expérience participative que les musées et les bibliothèques numériques peuvent aussi croître grâce à leurs potentielles contributions en nature. Il nous apparaît aussi que cette conscience contributive peut devenir le gage de leur future appétence pour utiliser ces patrimoines numériques. A l'inverse, les chercheurs en littérature par exemple, qui pourront apprécier d'être déchargés de nombreuses tâches, pour eux fastidieuses, de numérisation et du balisage référentiel et structurel de leurs corpus d'étude, pourront constater combien il peut leur être utile de bénéficier du travail de plusieurs classes pilotes, encadrées par leurs professeurs, qui auront pu baliser

⁴⁴ Qui sont très faiblement subventionnés par rapport aux sciences exactes et expérimentales,

systématiquement et exhaustivement tout l'ensemble de l'œuvre d'un auteur : par exemple, des figures de style, des types de versification, des hypothèses d'interprétation sémantique des textes. On peut même faire travailler les élèves sur la commande explicite d'un chercheur.⁴⁵

5 Nos premières hypothèses dans HD Muren

Pour conclure cet article, nous aimerions revenir sur les questions d'expérimentation pédagogique et d'autres besoins d'interfaces qui peuvent en découler.

Dans la phase de démarrage de HD Muren, une professeure de Français⁴⁶ du secondaire, mais impliquée aussi à « Baccalauréat +1 » développe actuellement des pilotes de plans de cours, des méthodologies d'ateliers participatifs pour l'usage de la TEI en enseignement du Français ou de la littérature depuis la 1^{ère} classe du secondaire jusqu'après le Baccalauréat, notamment lettres supérieures et la formation des professeurs. Son travail est coordonné par l'équipe de l'ESPE d'Aquitaine⁴⁷ avec pour l'ESPE des objectifs multiples :

- Participer à la création de schémas TEI qui, liés à oXygen, serviront de *Framework* pour supporter les activités
- Déterminer et modéliser également sur des *Frameworks* la création des corpus ainsi produits et accumulés (bien sûr en collaboration avec des professeurs documentalistes dirigeant les CDI (Centres de documentation et d'information) dans les lycées ou collèges français mais en collaboration aussi avec Canopé (le réseau français de documentation pédagogique)
- Discuter ces premiers plans d'activités pédagogiques, de créations et d'usages de corpus et proposer des améliorations, des intégrations, etc.
- Former de maîtres pour ces activités en vue de classes pilotes
- Réaliser des simulations de classes pilotes et les évaluer.
- Étudier ces activités en tant que recherche pédagogique
- En lien avec d'autres partenaires étudier les interactions et les apports des synergies croisées enseignement/recherche/patrimoine.
- Faire évaluer ces travaux
- En collaboration avec d'autres partenaires, fabriquer des maquettes de livres (ou bibliothèques) pédagogiques augmentés (à partir des schémas TEI).

⁴⁵ Les 2 dernières pages sont tirées de la présentation de HD Muren au COSSI 2015. La thématique du COSSI 2015, « Communication, information et savoir : quel management pour une organisation durable ? », nous a semblé particulièrement bien ciblée pour traiter de la composante collaborative du projet grâce à la TEI. [HUD&all 2015]

⁴⁶ Garance Hudrisier. Notons qu'à la fin des années 90 elle avait balisé en TEI le théâtre de Claudel qui était l'objet de sa thèse [HUD-G. 2002].

⁴⁷ Vincent Liquète, [LIQ&KOV 2013], Anne Lehmanns [LEH&LIQ 2015] et Camille Capelle.

6 Retour sur la notion d'interface pour l'apprentissage des langues étrangères et la littérature comparée

Nous avons déjà souligné combien *l'Atelier Interfaces numériques* de Münster nous a énormément servi à prendre du recul sur la question des interfaces qui s'avère centrale dans notre projet HD Muren.

Nous n'oublions pas aussi, combien nous avons été touchés de rencontrer dans une petite ville allemande qui accueillait le Congrès des Franco-romanistes, autant de Francophones plus puristes et plus amoureux du français que nous. Hors de notre Atelier Interfaces, nous avons constaté qu'une grande majorité de ces participants étaient en fait des professeurs de français tant du secondaire que du supérieur.

Cela rebondit évidemment sur l'importante question de l'apprentissage des langues, sur la mise en parallèle et l'alignement de corpus multilingues⁴⁸, sur les dictionnaires bilingues, toutes activités pour lesquelles la TEI est extrêmement utile et les interfaces indispensables.

Nous espérons qu'au-delà de la collaboration avec votre collègue de Cologne (Sasha Diwersi), ce sujet du « français langue étrangère » nous permettra très vite de renforcer des liens de collaboration et d'amitié avec la communauté des Franco-romanistes numériques.

Bibliographie

1. André, Jacques & Hudrisier, Henri, (sous la dir. de), (2002), *Unicode, écriture du monde ?* Paris, ed. Hermès, Lavoisier (Document numérique, vol 6, n°3-4/2002, 365p
2. Abbès, R. (2000). *Encodage des corpus de textes arabes en conformité à la TEI, outils et démarche technique*. Rapport final de projet DIINAR-MBC.
3. Arnauld, Antoine & Nicole, Pierre (1683), *La logique ou L'art de penser : contenant outre les règles communes, plusieurs observations nouvelles, propres à former le jugement*. Paris : ed. G. Desprez, P 55. <http://visualiseur.bnf.fr/Visualiseur?Destination=Gallica&O=NUMM-57444>
4. Ben Henda, Mokhtar, (2015). *Interdisciplinarity dynamics using TEI for building multilingual digital corpora: a Maghreb-based case study*. In Digital Humanities Institute-Beirut, American University of Beirut, 2-6 March 2015.
5. Burnard, Lou (2012). *Encoder l'oral en TEI : démarches, avantages, défis*. In Conférence à la Bibliothèque Nationale de France, Paris: Abigaël Pesses.
6. Cahier GUTenberg (1996) *TEI Lite : An Introduction to Text Encoding for Interchange*, numéro 24, juin 1996 (pages 23-151). <http://www-tei.uic.edu/orgs/tei/intros/teiu5.tei>. Voir aussi <http://www.univ-rennes1.fr/pub/GUTenberg/publicationsPS/24-teilite.ps.gz>
7. Conseil national du numérique (2014). Jules Ferry 3.0, *Bâtir une école créative et juste dans un monde numérique | Conseil national du numérique* (No. 3.0). Paris. Consulté à l'adresse

⁴⁸ Mais aussi par exemple un alignement « français vieilli/français moderne ».

<http://www.cnumerique.fr/education-2/>

8. Dacos, M. (2010). *Manifeste des Digital humanities* [Billet]. consulté à l'adresse: <http://tcp.hypotheses.org/318>
9. Elmedlaoui, Mohamed, (2005), *Le berbère et l'histoire du plurilinguisme au Maghreb (le cas du Maroc)*, in revue Études et Documents Berbères vol. 23, 2005, pp. 153-178.
10. Guesdon, Marie-Geneviève (2008). *Bibliothèque nationale de France : Manuscrits catalogue 'Archives et manuscrits'*. In, the Fourth Islamic Manuscript Conference, Cambridge
11. Hankinson, A., Pugin, L. and Fujinaga. I. (2014). *Optical music recognition for navigating and retrieving music manuscript images*. In The Medieval and Renaissance Conference, 3–6 July, Birmingham, UK.
12. Hudrisier, Henri, (1999). *La « lecture assistée par ordinateur » et ses applications savantes ou pédagogiques dans un contexte interactif normalisé : la TEI (Text Encoding Initiative)*. In Passerelles (Revue de l'Université de Paris 8), (24), pp 57-64.
13. Hudrisier, Henri, Azémard, Ghislaine, Ben Henda, Mokhtar, Diwersy, Sasha, Lehmans, Anne, Liquète, Vincent, Romary, Laurent, (2015), *Synergie enseignement-recherche pour l'aménagement numérique structuré (TEI) de patrimoines littéraires multilingues et multiculturels*, in COSSI 2015, « Communication, information et savoir : quel management pour une organisation durable ? », Université de Montréal (à paraître)
14. Hudrisier H., Bezina S., Ismail H., Zghibi R., Zghidi S., Romary L., Ben Henda M., Abd-Elrazak, L., Laborderie A., Azemard G. (2014), *La stylistique des notes chez Rétif de la Bretonne : un hypertexte avant le numérique*. In Colloque du CIDE.17 Le Livre post-numérique, historique, mutations et perspectives, Fès, Maroc, 19-20 novembre 2014 (sous la direction de Khaldoun Zreik, Ghislaine Azémard, Stéphane Chaudiron et Gaétan Darque), ed. Europia, Paris, pp 127 à 152.
15. Hudrisier, Garance et Henri, (2000), *Les enjeux culturels et dialectiques de la lecture assistée par ordinateur*, in Le français aujourd'hui n°129 mars 2000 : « ordinateur et textes : une nouvelle culture ? » Sous la dir. de J. Huynh et S Plane. Pp 40 à 49
16. Hudrisier, Garance, (2002), *Les didascalies dans le théâtre de Paul Claudel*, sous la direction de Michel Autrand, Paris, Sorbonne.
17. HumaNum. (s. d.). *TGIR Huma-Num | Très Grande Infrastructure de Recherche dédiée aux humanités numériques*. Consulté 30 mai 2015, à l'adresse <http://www.huma-num.fr/>
18. Ide, N., & Véronis, J. (1996). *Présentation de la TEI*. In Cahiers GUTenberg, TEI: Text Encoding Initiative, N°24, 4-10.
19. Ide, N. (1996). *Representation schemes for language data: the Text Encoding Initiative and its potential impact for encoding African languages*. In CARI'96
20. Ide, N. M., & Véronis, J. (1995). *Text Encoding Initiative: Background and Contexts*. Springer.
21. Jungen, C. (2012). *Quand le texte se fait matière*. In Terrain, n° 59(2), 104-119.
22. Kant, I. (1724-1804). (1886). *Traité de pédagogie / Emmanuel Kant ; (traduction Jules Barni) ; avec une préface, des sommaires analytiques et un lexique, par Raymond Thamin*, Consulté à l'adresse <http://gallica.bnf.fr/ark:/12148/bpt6k943071>
23. Laborderie Arnaud, (2014), *Lectures plurielles : discontinuités et ruptures sémantiques*, in Livre post-numérique : historique, mutations et perspectives, actes du 17e colloque international sur le document électronique, CIDE.17, Paris, Europia, 2014.
24. Laborderie Arnaud, (2014), *Expositions virtuelles et valorisation patrimoniale : le cas des collections islamiques de la BnF*. Sou-Dufrène B. N. (dir.), Patrimoine du Maghreb à l'ère numérique, actes des Rencontres du numérique d'Alger, Paris, Hermann, 2014.

25. Laborderie Arnaud, (2015), *Éditorialisation des bibliothèques numériques : le cas des Essentiels de Gallica*, in Documents et dispositifs d'information à l'ère post-numérique, actes du 18e colloque international sur le document électronique, CiDE.18, Paris, Europa, (à paraître)
26. Lehmans, Anne & Liquète, Vincent. (2015) *Conditions for a sustainable information transculture*. Actes du colloque Media and information Futures, 8-9 mai 2014, Tampere (Finlande), à paraître.
27. Liquète, V. & Kovacs, S. (coord.), (2013), *La lutte des classements. Introduction générale*. Hermès, n°66, 2013, p. 9-15.
28. Martin de Guise, Sylvaine, (2013), *La MEI (Music Encoding Initiative): un standard au service de la musique kabyle*, in Iles d'Imesli, N°5, "Les Actes du Colloque Les Langues et la Culture Amazighe du Département des Lettres et des Langues de l'Université Mouloud Mammeri" (Tizi-Ouzou, Algérie), p.245-277.
29. Martin de Guise, Sylvaine, (2013), *MEI-Music Encoding Initiative - et TEI - Text Encoding Initiative, deux standards au service de la musique tunisienne*, in Les Actes du Colloque Technologies numériques dans l'art et les sciences humaines à l'Institut Supérieur des Arts et Métiers de Gabès (Gabès, Tunisie), mars 2013 (en ligne et sous-presse).
30. Montaigne, Michel de, (1580), *Essais*, livres I et II, Bordeaux, Simon Millanges. http://bibliotheque.bordeaux.fr/in/faces/details.xhtml?id=BordeauxS_B330636101_S4754_REScoffre&jscheck=1
31. Ourabah Mohammed, S., & Hassoun, M. (2012). *A TEI P5 Manuscript Description Adaptation for Cataloguing Digitized Arabic Manuscripts*. Journal of the Text Encoding Initiative,
32. Roland, P., A. Hankinson, and L. Pugin. (2014). *Early music and the Music Encoding Initiative*. Early Music (forthcoming).
33. Romary, Laurent, & Hudrisier, Henri, (2003). *Le balisage normalisé des concepts et documents en liaison avec les normes de l'EAD*. Colloque Normes & standards pour l'apprentissage en ligne, Versailles, 19/03/03. Consulté 30 mai 2015, à l'adresse : <http://www.initiatives.refer.org/Initiatives-2003/notes/notes/henri.htm>
34. Searle, John (1998), *La Construction de la réalité sociale*, Paris, éd. Gallimard
35. TEI Consortium. (2015). *Guidelines for Electronic Text Encoding and Interchange*. (Originally edited by Sperberg-McQueen, C.M. and Burnard, L., Version 2.8.0. avril 2015). Consulté 30 mai 2015, à l'adresse <http://www.tei-c.org/Guidelines/P5/>
36. Université Paris 8. (s. d.). *IDEFI-CréaTIC* - Université Paris 8. Consulté 30 mai 2015, à l'adresse <http://www.univ-paris8.fr/IDEFI-CreaTIC>
37. Wikipédia, (2015), SGML, <https://fr.m.wikipedia.org/wiki/SGML> consultée le 15/10/2015
38. Zenkouar, Lahbib (2004), *L'Écriture amazighe Tifinagh et Unicode*, in revue Études et Documents Berbères vol. 22, pp. 175-173.
39. Zghibi Rachid (2013). *Création et indexation des documents numériques en texte intégral par la TEI*, Symposium Les nouveaux outils d'indexation de nouvelles données dans un environnement numérique du Web 2.0, Tunis, 3 avril 2013
40. Zghibi, Rachid, (2002), *Le codage informatique de l'écriture arabe : d'ASMO 449 à Unicode et ISO/CEI 10646*, in André, Jacques & Hudrisier, Henri, (sous la dir. de), (2002), *Unicode, écriture du monde ?* Paris, ed. Hermès, Lavoisier (Document numérique, vol 6, n°3-4/2002, p 155 à 182