

# Enjeux sociétaux et linguistiques des Humanités numériques au Maghreb

## *Digital Humanities: social and linguistic challenges for the Maghreb*

Alain VAUCELLE (1), Henri HUDRISIER (2)

(1) Département ARTEMIS, TELECOM SudParis, Institut Mines-Télécom, Evry, France / Maison des Sciences de l'Homme Paris Nord, Saint-Denis, France

Alain.Vaucelle@telecom-sudparis.eu

(2) Laboratoire Paragraphe Université Paris 8, Saint-Denis, France

henri.hudrisier@wanadoo.fr

**Résumé.** La création de grands corpus numériques multilingues au Maghreb rassemblant des corpus oraux et leur transcription, leur traductions et interprétations, ainsi que les ressources manuscrites ou imprimés est plus complexe que celle des pays d'ancienne et de large culture de l'imprimé. Cet article s'intéresse à des questions de méthode concernant la mise en œuvre de bibliothèques numériques de l'existant. Ce document s'interroge aussi sur les méthodes des Humanités numériques que les projets futurs de numérisation et de mise à disposition des ressources devraient idéalement intégrer.

**Mots-clés.** TEI, langues, bibliothèques, normes, ISO.

**Abstract.** The setting-up of large, multilingual, digitalised entities in French-speaking North Africa, bringing together oral entities and their transcription, translation, interpretation, as well as handwritten or printed resources, is more complex than that of countries with a broad and print-based culture. This article examines issues of methodology, related to the transformation of digital libraries from their extant state. This document also seeks answers to questions about methods for digital humanities, that future digital projects and installations of resources should ideally include.

**Keywords.** TEI, languages, libraries, standard norms, ISO.

Toutes les langues du monde (y compris les langues mortes) peuvent potentiellement devenir des constituants du patrimoine numérique mondial. Cependant, à l'évidence, toutes les communautés linguistiques ne sont pas égales dans leur appropriation des savoir-faire leur permettant de réussir cette mutation. Cette question, très liée à celle de la fracture numérique Nord-Sud, nécessite d'être analysé avec soin et sans a priori notamment en dépassant la doxa politiquement correcte de l'égalité « autre que de droit » de toutes les

langues du monde. Cette communication vise à répondre à des questions liées mais distinctes en ce qui concerne la mise en œuvre des Humanités numériques dans la collecte et la dissémination des corpus lors de la constitution de bibliothèques numériques.

La réalisation de grands corpus linguistiques numériquement organisés est devenue non seulement un enjeu culturel mais aussi économique, sociétal, et cognitif fondamental. En effet l'accumulation de ces patrimoines numériques linguistiques conditionne la disponibilité de référentiels indispensables pour réaliser la traductique du futur et le web sémantique. Les pays du nord sont eux-mêmes confrontés à des disparités dans l'appropriation d'une collégialité numérique instrumentalisée (bibliothèques numériques normalisées) et conforme aux objectifs qui sont ceux des Humanités numériques.

Les populations au Maghreb, au Sahel ou en diaspora sont confrontées à un multilinguisme et un multiculturalisme beaucoup plus complexe que ceux des pays d'ancienne et large culture de l'imprimé. La situation culturelle linguistique du Maghreb est traversée par une richesse de disparités, des hiérarchies langagières complexes : langue maternelle, langue dominante du milieu, cultures orales, cultures écrites, polyglossies fonctionnelles : arabe dialectal/arabe classique, franco/arabe ou arabo-berbère, etc.

Tout cela pose des questions stratégiques qui touchent à la compétitivité directement cognitive des individus mais surtout, cela rend urgent la réalisation pratique de grand corpus linguistiques numériques référentiels multilingues s'attachant à la fois à la collecte de corpus oraux et de leur transcription, traduction, interprétation, mais aussi au rassemblement numérique et interopérable de ressources manuscrites ou imprimés (voire pétrographiques) qui constituent le patrimoine commun du Maghreb. Ces ressources pourraient venir compléter les fonds déjà considérables détenus par les grandes bibliothèques nationales (pas seulement au Maghreb) et des centaines d'autres endroits, comme les bibliothèques du désert (Gaudio, 2002). La question n'est pas dans cet article de s'intéresser à cette collecte mais à des questions de méthode plus actuelles, celles de la mise en réseau numérique sous forme de bibliothèques numériques de l'existant et celles des méthodes des Humanités numériques que les projets pilotes à venir devraient idéalement intégrer.

Les méthodes des Humanités numériques ont pour objet de codifier les processus d'appropriation d'une nouvelle culture numérique en émergence, de façon à ce que les chercheurs en Humanités puissent raccourcir le temps nécessaire pour s'adapter, et reconstituer leurs habitus de collégialité tout en y associant toutes les nouvelles facilités offertes par la culture numérique. Dit autrement, les Humanités digitales cherchent à réaliser un travail de réorganisation de la culture savante en Arts, Lettres et Sciences sociales par rapport aux nouveaux paradigmes du numérique, de façon similaire à ce que leurs prédécesseurs les « Humanistes » avaient fait par rapport aux nouveaux paradigmes sociétaux et techniques qu'induisaient l'imprimerie. Ce sont donc non seulement les enjeux sociotechniques des technologies de l'information et de la communication qui sont pris en compte, mais aussi les enjeux socio-informatiques dans l'appropriation des outils et des pratiques, la dissémination, et la pérennisation des résultats. Notre enjeu n'est pas, bien sûr, de redéfinir ce rôle des Humanités digitales, mais plutôt de définir un contexte. Notre ambition se limite à chercher concrètement et à affiner les conditions spécifiques

d'appropriation des Humanités numériques dans un contexte multilingue et historiquement complexe que constitue le Maghreb.

Les pays du Maghreb (et plus largement la communauté savante s'intéressant aux enjeux linguistiques précités) sont donc confrontés à un triple défi :

Le premier est de pouvoir s'approprier les méthodes des Humanités numériques dans des communautés linguistiques et territoriales beaucoup plus complexes que la communauté académique anglophone. C'est cette communauté anglophone qui a inventé ce concept en s'appuyant sur un partage académique collégial qui s'organisait avec une culture largement unilingue et une histoire du territoire nord américain beaucoup plus consensuelle<sup>1</sup>.

Le deuxième écueil réside dans l'appropriation de ces Humanités numériques en partageant certes la notion fondatrice d'Humanités, mais sans oublier que cette notion est nécessairement perçue comme un apport académique de l'ex-colonisateur, et non pas comme cela pourrait l'être idéalement comme un bien commun universel.

Enfin, le Maghreb est de plus confronté à la réalité d'un historique culturel très ancien mais moins univoque que les Humanités européennes liées à la révolution guttembérienne de l'imprimerie, à la Renaissance européenne et à la culture latine.

### **Une situation culturelle et linguistique complexe**

Le Maghreb, en comparaison à l'Europe connaît une histoire culturelle et linguistique beaucoup plus complexe jusqu'à nos jours : carthaginoise, libyque, berbère, punique, romaine, chrétienne, arabe, musulmane, turque, française, espagnole et donc une histoire traversée par ces diversités. Un seul exemple peut faire entrevoir les enjeux auxquels sont confrontées les élites humanistes maghrébines. Il est connu que c'est largement grâce aux traductions arabes (à partir de textes latins et grecs) que la philosophie et la littérature antique a pu être transmise et reconstitué par les humanistes de la Renaissance européenne. Or, pour des raisons multiples<sup>2</sup>, le patrimoine culturel du monde arabe amorce alors ce que l'on pourrait appeler un déclin relatif par rapport à l'Europe, qui à partir de la Renaissance<sup>3</sup> accumule littérature et savoir grâce à l'imprimerie et à l'organisation savante et académique qui en découle.

L'extrême complexité du patrimoine culturel maghrébin s'origine dans ces multiples contradictions : chaque couche de colonisation (romaine, arabe, ottomane, française) ou même les dernières reconquêtes d'indépendance nationale induisant des difficultés voire des dénis de légitimités culturelles et linguistiques. Ce dernier point est très sensible, il est délicat à développer mais il n'en est pas moins réel. Le multilinguisme maghrébin, qui se complique du fait de ses facettes bi ou triglossiques se surajoute à cette complexité culturelle.

---

<sup>1</sup> Tout au moins en apparence au niveau académique car il est bien évident que le « melting pot » est bien sûr complexe sur le plan historique, linguistique et territorial.

<sup>2</sup> Refus de se plier à la segmentation en lettres isolées de l'écriture arabe retardant l'imprimerie de 2 siècles, large pratique du latin et du grec dans l'Europe chrétienne contrairement au monde musulman pratiquant d'évidence l'arabe (et qui plus est un arabe canoniquement stable).

<sup>3</sup> Le monde arabe avait connu des sommets civilisationnels dès le 8<sup>e</sup> et 9<sup>e</sup> siècle.

Quelle langue parle t'on selon son époque, son statut social, sa situation de communication, son appartenance tribale ?

Les situations de bi ou triglossies fonctionnelles sont très connues chez les universitaires arabes qui les vivent quotidiennement mais moins connues chez les européens. L'article de Ferguson (Fergusson, 1959), constitue un jalon conceptuel fondateur et indéniablement neutre par rapport aux premières publications scientifiques encore lourdement marquées par le contexte colonial (Marçais, 1930) : « La DIGLOSSIE est une situation langagière relativement stable dans laquelle, en plus des dialectes de base de la langue (qui peuvent inclure une forme standard ou des standards régionaux), il existe une variété superposée, très divergente, hautement codifiée (et souvent grammaticalement plus complexe), véhicule d'une littérature écrite vaste et respectée, soit à une époque antérieure soit dans une autre communauté linguistique. Cette variété est apprise essentiellement par l'enseignement et est utilisée pour la plupart des fonctions écrites et des fonctions orales à caractère formel, mais n'est pratiquée par aucun groupe de la communauté pour la conversation ordinaire. »

L'histoire de la plupart des nations connaît de semblables situations, complexités et dénis. Ainsi, un physicien français intervenant dans un colloque international à Paris est-il libre de s'exprimer en français ? Bruxelles ou Strasbourg sont des villes francophones mais les discussions d'experts<sup>4</sup> de la Commission ou du Parlement se font-elles en français ? L'Angleterre par exemple a construit difficilement sa maturité linguistique actuelle en connaissant un affrontement entre le français (ou plutôt sa variante anglo-normande parlée par l'aristocratie). Ainsi, les devises royales actuelles « Dieu est mon droit » et « Honni soit qui mal y pense » sont, parmi bien d'autres, les traces encore vivantes d'une longue période de biglossie dans les Iles britanniques (qui n'est évidemment pas propre à la seule réalité arabe (Suleiman, 1982) (Talmoudi, 1984), pour arriver à la situation actuelle de l'anglais authentique langue saxonne, mais disposant d'un large lexique issu du français.

À l'évidence le Maghreb a connu jusqu'au 20<sup>e</sup> siècle des bouleversement politiques, culturels et linguistiques qui conditionnent la réalité et la légitimité de prise en compte de ses patrimoines notamment linguistiques : les patrimoines berbères, juifs, arabes, francophones, hispanophones sont-ils l'objet d'une attention patrimoniale apaisée ?

Certes le français était la langue des colonisateurs, mais c'est aussi avec l'anglais une langue de large diffusion, une langue largement partagée dans certains milieux ou professions, une langue d'enseignement pour certaines disciplines, la langue aussi d'un certain nombre d'auteurs maghrébins.

Autre défi, le Maghreb est évidemment un territoire dont l'encadrement scientifique et académique est moindre, malgré l'excellence de ses niveaux scolaires et universitaires. D'autre part les Lettres et les Humanités ont été relativement mal traitées du fait des nombreux ordres et contre-ordres qui ont été donnés depuis les débuts des indépendances : francophonie, arabophonie, bilinguisme, interdiction ou promotion des langues berbères. La situation est très contrastée suivant les territoires et suivant la classe d'âge des individus.

---

<sup>4</sup> Nous précisons discussions d'experts car pour les parlementaires ou les ministres le recours à la traduction simultanée est un luxe coûteux qu'ils peuvent exiger.

Certains professeurs non (ou difficilement) arabophones<sup>5</sup>, se voyant quelquefois imposer d'enseigner leur discipline en arabe ce qui entraîne fatalement une expression argumentative de faible qualité qui se transmet fatalement aux étudiants qui suivent les cours dans ces disciplines brutalement et administrativement arabisées<sup>6</sup>.

### **Les enjeux des bibliothèques numériques et des Humanités numériques**

La fausse évidence de l'unilinguisme en communication a été largement décrite (Vannini, Le Crosnier, 2012). Le multilinguisme est sans nul doute beaucoup plus répandu que le monolinguisme qui serait l'exception. À l'évidence, depuis 2005, la suprématie monolingue anglophone sur l'Internet est sur le déclin, au fur et à mesure que la population des pays émergents en ligne augmente (notamment face au chinois)<sup>7</sup>.

On sait aussi, notamment depuis Louis-Jean Calvet (Calvet, 1974) que les guerres des langues est une réalité que la mutation numérique accélère d'évidence dans un premier temps mais qui peut fort bien au contraire redistribuer les cartes et renverser certaines situations acquises.

Ainsi, comme c'est souvent le cas lors d'une mutation technologique, des pays faiblement encadrés, faiblement équipés en bibliothèque peuvent profiter de la « redistribution des cartes » qu'induit l'opportunité de la numérisation des bibliothèques, et de la fantastique opportunité que représente la synergie mondiale de la recherche en Humanités numériques et de la mise en place des bibliothèques numériques. À condition de disposer d'un noyau même modeste de prise en compte locale de ces nouveaux paradigmes de l'information et de la communication académique, on peut potentiellement compter sur l'aide collégiale du monde entier pour aménager le patrimoine franco-arabo-berbère<sup>8</sup>. C'est sur ces hypothèses que s'appuient les projets dans lesquels sont engagés les auteurs<sup>9</sup> de cette communication. Les auteurs de ce texte sont tous les deux engagés dans des projets directement liés à ces objectifs en termes d'expertise

---

<sup>5</sup> Notamment la question importante des élèves ou même de professeur maîtrisant l'arabe dialectal mais pas l'arabe classique à l'écrit. Un des auteurs peut témoigner de certains étudiants appartenant à certaine classe d'âge ne maîtrisant ni le français, ni l'arabe comme langue dans laquelle ils peuvent rédiger des textes argumentatifs.

<sup>6</sup> Un déficit de la qualité d'expression similaire existe aussi bien sûr, dans les pays du Nord. Par exemple, certains professeurs de faible niveau anglophone se voient parfois contraints d'enseigner uniquement en anglais. Ils s'adaptent évidemment mais la qualité de leur enseignement en pâtit considérablement. Tous les spécialistes conviennent que la communication multilingue par écoute passive de la langue de l'autre est bien plus formatrice. De plus l'étudiant (plus jeune) s'adapte plus facilement et plus utilement à la maîtrise multilingue qui deviendra pour lui un atout. C'est par exemple le choix des écoles fédérales polytechniques de Lausanne et Zurich où les cours sont indifféremment donnés en allemand, français et anglais.

<sup>7</sup> Voir les statistiques : <http://www.internetworldstats.com/stats7.htm>

<sup>8</sup> Ou plutôt en ajoutant les communautés déjà largement engagé anglo-hispano-néerlandais-franco-arabo-berbère

<sup>9</sup> Les principaux projets sont la Bibliothèque Numérique Berbère (financement Ile de France), la Bibliothèque Numérique Franco-Berbère (financement OIF-FFI), HumanitéDigitMaghreb (financement CNRS-ISCC) ainsi que l'organisation de colloques CNPLET, IRCAM, ERCILIS.

technique (bibliothèques franco-berbère ; normes et standards pour des ressources linguistiques et pédagogiques ; Text Encoding Initiative<sup>10</sup>). Depuis longtemps, ils collaborent étroitement avec de nombreux intellectuels de ces communautés. Ils savent d'expérience que leur méconnaissance tant du berbère que de l'arabe est un facteur à même de favoriser l'analyse des blocages et d'accélérer des processus d'appropriation.

### **Le chercheur en sciences de l'information et de la communication face à l'émergence des Humanités numériques**

Il est complexe d'analyser en termes de sciences de l'information et de la communication les nombreuses questions qui touchent à l'aménagement linguistique des grands corpus patrimoniaux numériques ; difficile aussi de trouver une légitimité des Humanités numériques et d'illégitimer parallèlement son refus par les acteurs scientifiques traditionnels des SHS et des Humanités. La raison en est la différence flagrante d'industrialisation des langues et d'appropriation académique des potentiels qui en dérivent : bibliothèques numériques, pratiques des Humanités numériques. De ce fait, les grands ensembles territorio-linguistiques sur lesquels se répartissent les chercheurs sont directement concernés et sont de ce fait, à la fois objets et sujets. Il est donc difficile pour des chercheurs anglophones ou ceux d'une grande langue européenne dans un pays développé, de comparer sereinement leurs états de l'art respectif et d'observer les niveaux d'appropriation pour ce qui est des bibliothèques numériques, des Humanités numériques et des enjeux de performance en ingénierie des langues qui y sont directement associés : interopérabilité normative des documents et des ressources, performance éditoriale, documentaire ou traductique.

La deuxième grande question vient de ce que, comme le décrit si bien Bertrand Gille (Gille, 1978), un nouvel ensemble technique ne s'installe vraiment qu'à deux conditions consubstantielles : la mise en place effective d'un environnement technique cohérent (ici des bibliothèques numériques intercompatibles et normalisées<sup>11</sup>) et l'installation d'une technoculture correspondante (ie. l'appropriation de méthodes des Humanités numériques<sup>12</sup>). C'est cette étape, émergence universelle de la mutation vers de nouveaux habitus méthodologiques et de nouveaux environnements techniques s'articulant avec les derniers progrès de l'information structurée et balisée, qui sont loin d'être

---

<sup>10</sup> La TEI (Text Encoding Initiative) est un standard d'échange et de traitement par balisage de corpus numériques prioritairement en Sciences humaines et sociales (SHS), mais c'est aussi le nom d'une communauté mondiale, fonctionnellement organisée en collèges partageant ces conventions d'échange et de traitement (non pas uniformes, c'est très important), mais standardisés et donc intercompatibles et interopérables.

<http://www.tei-c.org>

<sup>11</sup> Conformément notamment aux instructions de l'OCLC (Online Computer Library Center).

<sup>12</sup> L'Alliance of Digital Humanities Organisations (ADHO) a adopté comme publication principale, le journal officiel de l'ALLC « Journal of Digital Scholarship in the Humanities » publié par les Oxford University Press et deux autres publications de portée mondiale « DHQ, Digital Humanities Quarterly » et « Digital Studies / Le champ numérique » Voir aussi une cartographie conceptuelle de l'univers des Humanités digitales disponibles sur <http://www.allc.org/publications/mind-map-digital-humanities>

complètement adoptés et installés dans le milieu académique des SHS et des Humanités et avec de surcroît des disparités importantes suivant les degrés de développement, les continents et les langues.

On constate ainsi que les bibliothèques virtuelles numériques ne s'imposent pas de facto en synergie avec les bibliothèques traditionnelles selon la même évidente synergie fonctionnelle qui s'est naturellement installée entre la téléphonie fixe et la téléphonie mobile. De même, les habitus académiques traditionnels dans les sciences humaines et la recherche littéraire (les Humanités), ne se transforment pas naturellement pour s'élargir et adopter en synergie les propositions des Humanités numériques. Des retards s'installent, confortés par des certitudes scientifiques ou professionnelles établies depuis longtemps, dans la plupart des disciplines de SHS (linguistique, recherche littéraire, ethnologie), souvent bloqués aussi par certains professionnels des bibliothèques ou même des informaticiens, voire même certains chercheurs appartenant plus aux mentalités des « anciens » contre celles des « modernes ». Dans le monde anglophone, les grands progrès des Humanités numériques, des bibliothèques numériques, du Web sémantique, de la traductique du futur en émergence ont été précisément rendus possibles parce que des grandes Fondations Nord Américaines, des grandes bibliothèques du monde développé se sont coordonnées, ont investi des fonds considérables. Ces mêmes sponsors subventionnent encore aujourd'hui le cadre institutionnel et les équipes de chercheurs correspondant. Cela permet la mise en place de ce nouveau contexte technique. Cela offre aussi aux communautés de chercheurs en Humanités numériques un accompagnement depuis maintenant près de 20 ans pour s'approprier ces nouveaux paradigmes et réussir ainsi à constituer des corpus structurés (surtout anglophones) relativement conséquents.

Dans le monde francophone européen par exemple (celui dans lequel est écrit cet article) les Humanités numériques sont un concept émergeant encore bien flou. Il n'en est pas de même au Canada (même francophone). Et pourtant depuis 2007, la France s'est dotée du TGE Adonis<sup>13</sup> qui devrait lui permettre de mettre fin aux fausses polémiques interdisciplinaires et d'accélérer l'appropriation dans le milieu des SHS de méthodes intercompatibles et standardisées de traitement des documents et des corpus linguistiques.

Car telle est bien la question centrale, depuis une trentaine d'années les scientifiques des sciences exactes et expérimentales ont mis en place une collégialité numérique mondiale. Les acteurs des SHS tardent à s'y mettre car ils acceptent beaucoup plus difficilement que des normes d'interopérabilité, de réusabilités et de diffusion de leurs documents (et surtout de leurs corpus de documents) puissent être mises en place. Il est cependant absolument clair que ces normes distinguent clairement 3 types de balisage ou d'aménagement de métadonnées : (a) référentiel et documentaire (b) structurel et formel (c) sémantique. Ces trois niveaux garantissent que le chercheur n'est aucunement soumis à une quelconque entrave de sa liberté de chercheur. Le champ sémantique lui est largement ouvert pour instrumentaliser numériquement ses hypothèses innovantes, par contre le champ référentiel ainsi qu'un minimum de structuration consensuelle avec les autres communautés de chercheurs de sa

---

<sup>13</sup> Le TGE Adonis est une plateforme de recherche en sciences humaines et sociales qui permet d'assurer l'accès et de préserver l'accès aux données numériques produites par les SHS. <http://www.tge-adonis.fr>

discipline lui permettront d'assurer la communicabilité, la réusabilité de ses documents et corpus donc la possibilité (comme en sciences expérimentales et exactes) de faire grandir au niveau planétaire de très grand corpus de documents (majoritairement des textes<sup>14</sup>), d'y accéder collégialement et donc de pouvoir développer des recherches en SHS au niveau mondial, notamment dans des disciplines comme la recherche littéraire. Notons que jusqu'à présent, ces chercheurs n'avaient pas d'autre moyen que de collaborer par l'échange de publications dans un contexte peu normalisé et donc très faiblement réutilisable. Certes, depuis les débuts de l'informatique des nombreux chercheurs en sciences humaines avaient eu l'idée d'informatiser des documents et des corpus. Ils avaient ainsi produit quantité de travaux passionnants : analyser des lexiques d'auteurs, structurer des corpus et par là conforter des hypothèses d'analyse structurale et/ou stylistique. Les linguistes avaient assez souvent produit des corpus très conséquents. Les archéologues, les historiens avaient produit d'importants travaux sur quantités de corpus plus originaux les uns que les autres. Cependant l'interopérabilité intrinsèque des documents associés aux couches successives d'analyse et de recherche n'étaient pas standardisée.

### **La traductique et le e-sémantique du futur**

On peut concevoir facilement que le rassemblement de gros corpus de documents, notamment des textes de patrimoine littéraire, philosophiques, sociologiques, linguistiques, etc... est souvent considéré comme non-prioritaire pour nombre de décideurs politiques et économiques. Il est patent, que même dans les arbitrages scientifiques les SHS et les Humanités sont trop souvent considérés comme une priorité seconde. On ignore en cela que dans un futur proche l'efficacité de la traductique et du e-sémantique dans telle ou telle langue sera très directement fonction du volume et de la diversité des grands corpus de ressources textuelles rassemblés et structurés de façon normative. Dès lors les enjeux deviennent beaucoup plus stratégiques car ces deux progrès prospectifs sont très importants pour assurer la prospérité économique, industrielle et économique des communautés linguistiques.

Nous en sommes actuellement, tant pour la traductique que pour l'e-sémantique à une véritable mutation des méthodes. Jusqu'à un passé très proche tant la traductique que l'e-sémantique fonctionnaient (et fonctionnent encore) grâce à des dictionnaires, des règles syntaxiques, des grammaires, des contextes relativement limités rassemblés par les équipes qui ont conçu ces outils techniques.

Or, l'extrême efficacité du Cloud computing, les progrès fantastiques des ordinateurs tant en puissance de calcul qu'en rapidité, font qu'il devient de plus en plus réaliste de désambiguïser le sens des énoncés ou des traductions en le comparant au contexte exhaustif des occurrences dans la totalité d'un patrimoine linguistique. Mais pour proposer ces nouvelles opportunités, il importe que les ressources textuelles (mais aussi les ressources orales voire multimédias) présentes sur le Web, soient normalisées, interopérables et balisées d'un point de vue linguistique.

---

<sup>14</sup> Mais pas seulement : les photographies, les documents filmiques ou audio peuvent être associés à ces corpus ainsi que des graphiques des formules mathématique ou chimiques, des plans, etc.



L'ISO TC37 (terminologie et ressources textuelles) développe ainsi des familles de normes (Organisation Internationale de Normalisation 24610, 2006), (Organisation Internationale de Normalisation 24611, en développement), (Organisation Internationale de Normalisation 24613, 2008), (Organisation Internationale de Normalisation 24614, 2010), (Organisation Internationale de Normalisation 24616, en développement). Ces normes sont développées pour qu'à terme la totalité de l'information textuelle (voire même des corpus audios) ne soit pas comme aujourd'hui des masses de documents structurées au minimum, n'ayant pour agencement que leur seule logique éditoriale propre, mais qu'elles puissent être structurées et balisées linguistiquement de façon intercompatibles et normalisés. Autre point fondamental, cette structuration linguistique est prévue pour pouvoir être réalisée avec un maximum d'automatismes, donc un minimum d'intervention humaine.

Pour comprendre pleinement cette révolution attendue, il est nécessaire de revenir sur la mutation majeure que représentait déjà le Web lors de son arrivée au début des années 90. C'est parce que le HTML, structure de façon normalisée tous les documents présents sur la toile d'un point de vue référentiel et du point de vue de sa structure éditoriale que le Web a permis une révolution documentaire impensable auparavant. Le balisage linguistique normalisé est donc la deuxième mutation du Web qui permettra un e-sémantique et une traductique jusqu'ici inenvisageable.

Mais pour que cette deuxième mutation « e-sémantique » et traductique puisse advenir, il est indispensable que les communautés linguistiques concernées (pour nous ici francophones, arabophones et berbérophone), prennent en compte le développement de ces progrès et réalisent concrètement de très grands corpus littéraires et linguistiques (tendant à l'exhaustivité des ressources), structurés (ou structurables) selon ces nouvelles normes de balisage linguistique. Notons que ces normes ne sont surtout pas une nouvelle façon de pratiquer l'analyse. Elles sont avant tout le résultat d'un consensus mondial pour décrire et structurer de façon intercompatible et donc communicante des ressources dans toutes les langues du monde. Il importe donc que des acteurs maghrébins de la linguistique s'approprient rapidement ces nouvelles normes, balisent rapidement « et manuellement » des corpus tests pour que des balisages automatiques puissent être réalisables au plus tôt ; dès que des corpus de masses seront disponibles.

### **L'impérieuse nécessité des bibliothèques numériques normalisées**

On voit bien que les questions d'Humanités digitales, d'e-sémantique et de traductique du futur sont étroitement liées au développement de très grandes bibliothèques virtuelles qui (mieux que le Web actuel), ouvriront « l'accès intelligent » à la quasi-exhaustivité des ressources textuelles et audios du monde entier.

Mais on sait bien aussi que ce progrès ne se réalisera pas aussi rapidement dans toutes les langues.

Selon le degré d'initiative proactive de telle ou telle communauté linguistique, leur aménagement des ressources sera plus ou moins avancé. Ce sont donc de nouvelles fractures numériques qui pourraient advenir.

Autant on pouvait considérer (bien à tort) que les Humanités digitales n'étaient pas une activité scientifique prioritaire, autant la performance traductique et e-semantic aura un impact beaucoup plus direct, économique et industriel, sociétal sur le devenir du Maghreb multilingue.

La question actuelle est donc de savoir comment et quand le monde maghrébin s'investira dans ces enjeux actuels. Les projets dans lesquels nous sommes engagés sont bien modestes, ne sont pas suffisamment larges pour dépasser le premier niveau de bibliothèque numérique et dessiner un cadre véritable d'Humanités numériques.

Ce démarrage maghrébin local nous apparaît indispensable car il permettrait à une collégialité Nord-Sud de se mettre en place : de nombreux documents appartenant d'évidence au Maghreb et aussi à la France existent par exemple à la BNF.

Ces nouvelles méthodes d'Humanités numériques, mais aussi pour la linguistique (et son aspect d'ingénierie pour l'e-semantic et la traductique) suggèrent que non seulement elles dépendent d'une mobilisation de la société civile, académique et politique, mais aussi elles sont étroitement liées au cadre normatif d'intégration et d'interopérabilité de la numérisation des documents. L'enjeu est déterminant pour les pays en voie de développement afin de garantir un accès et un partage des ressources du savoir pour le plus grand nombre.

## Bibliographie

- Ben Henda M., Hudrisier H. (janvier 2012). Les normes et standards des TICE, des enjeux primordiaux pour le Sud. *Revue Frantice*, n° 4.
- Ben Henda M. (2012). *Vision historique, technique et prospective des systèmes d'information et de communication: interopérabilité normative globalisée*. Mémoire de HDR sous la dir. de Roland Ducasse, Université Bordeaux III, à paraître.
- Ben Henda M., Hudrisier H. (2009). Normalisation et terminologies multilingues pour les TICE. In *Forum Terminologique International*, Université de Sousse 20 au 23 novembre 2009.
- Calvet L.-J. (1974). *Linguistique et colonialisme : petit traité de glottophagie*. Payot, Paris.
- Ferguson C. A. (1959). Diglossia. In *Word*, 15 (3), 325-340.
- Gaudio A. (sous la dir. de). (2002). *Les bibliothèques du désert : Recherches et études sur un millénaire d'écrits*. L'Harmattan, Paris.
- Gille B. (1978). *Histoire des techniques*. Collection Encyclopédie de la Pléiade, Gallimard, Paris.
- Hudrisier H., Romary L. (2003). Le balisage normalisé des concepts et documents en liaison avec les normes de l'EAD. In *Colloque Normes & standards pour l'apprentissage en ligne*, Versailles, 19 mars 2003. En ligne (consulté le 16/03/04) : <http://www.initiatives.refer.org/Initiatives-2003/notes/notes/henri.htm>. Reprint in *Études et documents berbères*, N° 19, 20, éd. La boîte à documents, MSH Paris Nord, 2004.
- Hudrisier H., Ben Henda M. (2008). Cartago : une terminologie large langue de l'enseignement électronique à distance. In *Les outils d'aide à la traduction*. Séminaire de l'Union Latine, Bucarest, février 2008.
- Hudrisier H. (2011). Normalisation et prospérité multiculturelle. Chap. 3. In *La norme numérique ; Savoir en ligne et Internet*. Sous la dir. de Perriault, J. & Vagner, C. CNRS éd. Paris, 63-87.
- Hudrisier H., Ben Henda M. (2009). Enjeux normatifs des TICE de l'enseignement des langues dans le contexte arabo-berbère. *Colloque international Les TICE et les méthodes d'enseignement/apprentissage des langues*, CNPLET, Alger et Paragraphe, Paris, 30 mai-1er juin 2009, Tipaza, Algérie.
- Hudrisier H., Vaucelle A. (2009). Technical and normative scenarios in the medium. In *International Preservation News*. n°47, May 2009, IFLA PAC, ISSN 0890 – 4960, [http://www.ifla.org/files/pac/IPN\\_47\\_web.pdf](http://www.ifla.org/files/pac/IPN_47_web.pdf)
- Hudrisier H. (2009). La nécessité d'adapter Internet à la mondialisation linguistique. In *Critique de la société de l'information* (coordonné par J.-P. LaFrance). Les Essentiels d'Hermès, CNRS éditions, Paris, 115-134.
- Hudrisier H., Ould Braham O., Saleh I. (2008). La numérisation (BNB) et le e-learning (workshop de Tipaza, Algérie, 28-29 mai 2008). In *Études et documents berbères*, N° 27, éd. La boîte à documents, Paris, 175-183.

- Organisation Internationale de Normalisation. *ISO 24610 : Structures de traits*, 2006.
- Organisation Internationale de Normalisation. *ISO/FDIS 24611 : Cadre d'annotation morphosyntaxique, en développement*, en développement.
- Organisation Internationale de Normalisation. *ISO 24613 : Cadre de balisage lexical*, 2008.
- Organisation Internationale de Normalisation. *ISO 24614 : Segmentation des mots dans les textes écrits*, 2010.
- Organisation Internationale de Normalisation. *ISO/FDIS 24616 : Plateforme d'informations multilingues*, en développement.
- Marçais W. (1930). La diglossie arabe. In *L'Enseignement Public*, Revue pédagogique, tome CIV, n° 12, 401-409, tome CV, 20-39.
- Ould Braham O., Hudrisier H. (2006). La bibliothèque numérique berbère. In *La langue française dans l'aventure informatique*. Colloques Lexipraxis 2005 & 2006, Paris, éd. AUF & AILF.
- Ould Braham O., Hudrisier H. (2004). Le berbère et les nouvelles technologies de l'information. In *Études et documents berbères*, n° 19, 20, éd. La boîte à documents, MSH Paris Nord, 293-294.
- Ould Braham O., Hudrisier H. (2008). Recueil et constitution de corpus oraux dans le domaine berbère, (Salle du CNPLET, Alger, lundi 2 juin 2008). In *Études et documents berbères*, n° 27, éd. La boîte à documents, Paris, 193-204.
- Romary L., Hudrisier H. (2004). TEI : Text Encoding Initiative. In *Études et documents berbères*, n° 19, 20, éd. La boîte à documents, MSH Paris Nord.
- Suleiman S. M. (1982). *Jordanian Arabic between Diglossia and Bilingualism Linguistic Analysis*. Amsterdam.
- Talmoudi F. (1984). *The Diglossia Situation in North Africa, a Study of Classical Arabic/Dialectal Arabic Diglossia with a Sample Text in 'Mixed Arabic'*. *Orientalia Gothburgensia*, 8.
- Vannini L., Le Crosnier H (dir.)(2012). *Net.laNg, réussir le cyberspace multilingue*. éd. C&F, Paris.
- Vaucelle A., Hudrisier H., Ben Henda M., Klett F., (2009). *ComMPEG-SCORM : Rapport final & Livre Blanc, APO ISCC 2009 13/01/2009*. <http://www.alain-vaucelle.fr/archives/945>
- Vaucelle A., Hudrisier H. (2010). Langages structurés & lien social. In *Tic & société*, volume 4, n°1, Interactivité. <http://ticetsociete.revues.org/790>