

Corpus : annotation et codage

Les humanités numériques et le projet de Bibliothèque Numérique Franco-Berbère

Alain Vaucelle¹, Tassadit Yacine²

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

Alain.Vaucelle@telecom-sudparis.eu

²Laboratoire d'Anthropologie Sociale, Ecole des Hautes Etudes en Sciences Sociales, Paris France yacine@msh-paris.fr

Résumé

Le projet Bibliothèque Numérique Franco-Berbère (BNFB) est un projet financé par l'Organisation Internationale de la Francophonie. La BNFB a pour objectif de créer des dynamiques d'élaboration de ressources numériques franco-berbères normalisées.

Ce projet a pour ambition de fonctionner comme un portail documentaire de ressources berbères et de faire communiquer entre elles des bases de données diverses et hétérogènes, et donc de réaliser des partenariats entre plusieurs établissements que rapprochent leurs collections (complémentarité des fonds) ou leurs publics (services culturels d'une même collectivité).

Ce projet de corpus linguistique nous amène à nous interroger sur les méthodes des Humanités numériques. Cet article propose des éléments de réflexion sur les enjeux socio-économiques des processus numériques d'appropriation et d'organisation de la culture savante en Arts, Lettres et Sciences sociales par rapport aux nouveaux paradigmes du numérique.

Le projet BNBF

La BNFB a pour objectif de syndiquer et de mettre en réseau toutes les dynamiques d'élaboration de ressources numériques berbères. Ces ressources s'adresseront à la fois à des communautés académiques (revues savantes ou spécialisées, colloques, fonds libyques, fonds ethnolinguistiques et ethnomusicologiques), mais aussi à des communautés d'apprenants (formation initiale ou d'adultes, baccalauréat berbère), et au grand public.

En territoire touareg nigérien le projet se déploie sur un cycle d'ateliers pilotes d'enseignement traditionnel par les femmes de contes en français et en langue touareg. Les outils pédagogiques de ces ateliers seront à la fois classiques mais aussi multimédia (captation sonore et tablette graphique numérique).

La réalisation de contenus Franco-berbères (voire franco-arabo-berbères¹) est stratégique pour la Francophonie car ces langues sont une composante fondamentale du monde francophone du Maghreb et du Sahel ainsi que de la diaspora vivant en Europe. Les langues berbères relèvent majoritairement de la culture orale (bien qu'il s'agisse d'une culture écrite plusieurs fois millénaire), ce qui rend difficile leur transmission intergénérationnelle avec des outils traditionnels alors que des solutions numériques multimédias innovantes seraient parfaitement pertinentes (capacité de numériser des corpus oraux, de les structurer, de les traiter, de les «feuilleter» à l'instar d'un texte). Nous proposons dans ce projet d'éditer numériquement des ressources en direction d'un lectorat et d'auditeurs (car nombre de ces ressources peuvent être sonores), disparates, et souvent pauvres. Ces ressources doivent pouvoir être déclinées sous forme orale ou écrite, sous plusieurs alphabets au choix : la coexistence de trois écritures (latine, arabe et tiffinagh) correspondant à trois traditions justifie aussi ce choix, le passage de l'une à l'autre s'opérant potentiellement d'un clic de souris. Le choix du tout numérique est dicté par la dispersion des locuteurs (tant sur leurs territoires d'origine qu'en diaspora), par la pluralité des langues berbères se référant à un patrimoine commun : kabyle, rifain, chleuh,

chaouia, touareg... mais aussi par l'émergence d'un marché en forte expansion (ordinateurs portables, tablettes, mais surtout ordiphones en anglais *smartphone* et livres électroniques).

La BNFB vise donc à rassembler les corpus numériques (textuels, audiovisuels, pédagogiques ou patrimoniaux) permettant de préfigurer les usages du futur mais aussi ceux des populations numériquement «branchées» qui grandissent d'années en années notamment chez les jeunes quelque soit leur origine sociale ou territoriale.

Les partenaires sur ce projet sont : Télécom SudParis (département ARTEMIS-France), Alliance Cartago (France), AUF (Agence Universitaire de la Francophonie), Université Evry Val Essonne (laboratoire Analyse & probabilité EA 2172, France), la Faculté des Sciences Humaines d'Agadir et l'Université d'Oujda (Maroc), l'Université de Tizi Ouzou (laboratoire LAELA : Aménagement et Enseignement de la Langue Amazighe -Algérie), Université Abdou Moumouni de Niamey-Faculté des Lettres et Sciences Humaines-Département de Linguistique (Niger).

Ce projet s'adresse à la Communauté linguistique et la diaspora berbère pour des usages pédagogiques ou patrimoniaux, ainsi qu'à la communauté savante des études berbères et libyques et un public cultivé généraliste. Dans ce projet les jeunes sont ciblés en priorité car il est fondamental que l'enfant ayant le berbère pour langue maternelle puisse aussi utiliser des outils numériques adaptés pour déployer son intelligence. Bien que l'handicap d'appartenir à une langue dominée comme le berbère puisse déclencher parfois l'excellence intellectuelle, il s'avère que c'est souvent une cause d'échec. Des investissements relativement modestes dans la réalisation de ces fonds numériques, outils patrimoniaux et ressources d'enseignement en ligne sont à même de répondre aux besoins de ces jeunes. L'épreuve du baccalauréat berbère en France destinée à 2 à 3000 candidats (sous cinq langues berbères distinctes) est emblématique de ce besoin non satisfait pour des jeunes. Là aussi la syndication des ressources s'avère rentable pour répondre aux besoins des jeunes, mais aussi à ceux des adultes.

Notons que la cible adulte est particulière pour ce qui est des femmes souvent moins confrontées au monde extérieur, statistiquement moins alphabétisées² et pour lesquelles il est pertinent de pouvoir s'alphabétiser dans la langue qu'elles pratiquent, et au-delà, grâce au multilinguisme, dans les langues partenaires du berbère : le français mais aussi l'arabe et d'autres langues sahéliennes. Le cas du berbère en territoire sahélien est particulièrement ciblé dans le projet. Du fait de son statut, la femme Touareg enseigne depuis des millénaires la langue et l'écriture (tiffinagh). Alliant tradition et modernité ces ateliers de contes français et berbères sont expérimentés dans le Sahel nigérien.

Des enjeux sociétaux pour les langues berbères

La réalisation de grands corpus linguistiques est un enjeu sociétal, et cognitif fondamental. En effet l'accumulation de ces patrimoines numériques linguistiques conditionne la disponibilité de référentiels indispensables pour réaliser et optimiser la traductique du futur et le web sémantique et donc la présence et le référencement de ces corpus sur le web. Les populations au Maghreb, au Sahel ou en diaspora sont confrontées à un multilinguisme et un multiculturalisme beaucoup plus complexe que ceux des pays de 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.

Cela pose donc des questions d'ordre stratégique qui touchent à la compétitivité cognitive des individus et à la reconnaissance culturelle des populations. La réalisation pratique de grands corpus linguistiques numériques multilingues s'attachant à la fois à la collecte de corpus oraux et de leur transcription, leur traduction, leur interprétation est complexe. Le rassemblement numérique et interopérable de ressources manuscrites ou imprimées (voire pétrographiques) qui constituent le patrimoine commun du Maghreb sont des enjeux socio-économiques et culturels d'autant plus

urgent que l'on a pu assister en Afrique à la destruction de patrimoines millénaires pour cause de conflits. Une des principales difficultés est de pouvoir numériser des ressources qui ne sont pas référencées au sein des fonds détenus par les grandes bibliothèques nationales (pas seulement au Maghreb) mais dans des centaines d'autres endroits, comme les bibliothèques du désert (A. Gaudio, 2002). La question dans ce projet est donc aussi de s'intéresser à la méthode de collecte associée à 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 numériques d'appropriation et d'organisation 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 constituent le Maghreb et l'Afrique subsaharienne.

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 consensuelle³. 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'excolonisateur, 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.

Sciences de l'information et de la communication et 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 «interopérables» et normalisées⁴) et

l'installation d'une technoculture correspondante (*i.e.* l'appropriation de méthodes des Humanités numériques⁵). 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 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 Adonis⁶ 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 «réussabilité» et de diffusion de leurs documents (et surtout de leurs corpus) puissent être mises en place. Il est cependant absolument clair que ces normes distinguent clairement 3 types de balisage ou d'aménagement en 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 discipline lui permettront d'assurer la communicabilité, la «réussabilité» de ses documents et corpus donc la possibilité (comme en sciences expérimentales et exactes) de faire grandir au niveau planétaire de très grands corpus de documents (majoritairement des textes⁷), 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 de 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.

Langue et sémantique en devenir

La mise en œuvre de gros corpus de documents, notamment des textes de patrimoine littéraire, philosophiques, sociologiques, linguistiques, est souvent considérée comme très complexe et très couteuse pour les décideurs politiques et économiques. Pourtant il est certain que dans un futur proche l'efficacité de la traductique et de l'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, dans 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.

L'ISO TC37 (terminologie et ressources textuelles) développe ainsi des familles de normes pour qu'à terme la totalité de l'information textuelle (voire même des corpus audio) 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 normalisées, et ce de façon automatique donc avec un minimum d'intervention humaine.

Mais pour que cette 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.

La plateforme retenue pour le BNF

C'est dans ce contexte des Humanités numériques que nous avons retenu la plateforme open source Omeka⁸ comme solution de balisage (Dublin Core) et de diffusion de nos ressources. Cette plateforme est simple d'utilisation tant en gestion administrateur qu'en partage des contenus. Omeka est au croisement de logiciels spécialisés dans la gestion de collections comme Greenstone, et de la gestion de contenus comme Wordpress. Omeka est un projet développé par le Centre pour l'Histoire et les nouveaux Médias Roy Rosenzweig, et l'université George Mason⁹.

Cette plateforme est dotée d'outils d'importation et de gestion de multiples formats (images, audio, vidéo, textes). Son interopérabilité est garantie par la l'importation et la gestion des métadonnées Dublin Core¹⁰, et de l'OAI-PMH qui permet le

moissonnage des ressources entrantes et sortantes, ainsi que la gestion de corpus TEI (Text Encoding Initiative)¹¹.

La compatibilité avec Zotero¹², le logiciel de gestion des données bibliographiques et des documents de recherche est assurée à l'aide d'un module d'extension. Plusieurs modules d'extensions sont à la disposition des utilisateurs afin d'optimiser la gestion et la diffusion des contenus.

Dans Omeka, le balisage Dublin Core peut se faire directement en ligne, lors de l'importation des données au sein de la plateforme. Le Dublin Core est un schéma de métadonnées générique. Ce schéma permet de décrire les ressources numériques à l'aide de 15 éléments de description formels (titre, créateur, éditeur), intellectuels (sujet, description, langue,) et relatifs à la propriété intellectuelle. Le Dublin Core est une norme internationale ISO 15836, reconnue par le W3C.

La plateforme est compatible avec l'Open Archives Initiative (OAI, initiative pour des archives ouvertes). L'OAI est un protocole facilitant l'échange et la valorisation d'archives numériques. A l'aide de ce protocole des fournisseurs de services peuvent moissonner les métadonnées sur les sites de fournisseurs de données. Il est ainsi possible d'utiliser un protocole OAI pour créer un outil de recherche simultanée moissonnant des données dans plusieurs bases de données bibliographiques et affichant le résultat dans une seule fenêtre de recherche (indépendamment du lieu physique des métadonnées). L'OAI a un moteur de recherche spécifique pour le contenu en libre accès, BASE. Archives Ouvertes *Protocol for Metadata Harvesting* - L'OAI-PMH (Open Archives Initiative's Protocol for Metadata Harvesting) ou protocole OAI facilite donc l'échange de données entre des fournisseurs de données (par exemple des bibliothèques ou des musées...) et un fournisseur de service (qui peut être aussi une bibliothèque, un centre de documentation, un portail thématique ou local désirant rassembler des données). Ce protocole d'échange permet de créer, d'alimenter et de tenir à jour, par des procédures automatisées, des réservoirs d'enregistrements qui signalent, décrivent et rendent accessibles des documents, sans les dupliquer ni modifier leur localisation d'origine.

Grâce au protocole OAI, une bibliothèque agissant en tant que fournisseur de données a la possibilité d'offrir une visibilité accrue à ses documents, notamment à ses publications électroniques ou à ses fonds spécialisés. Réciproquement, en tant que fournisseur de service, une bibliothèque peut réaliser une base de données ou un portail documentaire dans son domaine de spécialité ou sur un thème quelconque, en collectant les données descriptives de ressources et documents de tous types, accessibles sur l'Internet dans des entrepôts OAI. Enfin, le protocole OAI permet de faire communiquer entre elles des bases de données diverses et hétérogènes, et donc de réaliser des partenariats entre plusieurs établissements que rapprochent leurs collections (complémentarité des fonds) ou leurs publics (services culturels d'une même collectivité).

Des protocoles et des normes pour des bibliothèques numériques

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 audio du monde entier. Mais on sait bien aussi que ce progrès ne se réalisera pas aussi rapidement dans toutes les langues. Le risque est de voir certaines communautés linguistiques ne pas prendre en compte les enjeux de ces Humanités numériques. Le danger est grand de voir se créer de nouvelles fractures numériques.

Autant on pouvait considérer (bien à tort) que les Humanités numériques n'étaient pas une activité scientifique prioritaire, autant la performance traductique et e-sémantique aura un impact beaucoup plus direct, économique, industriel, et 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-sémantique 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.

Références

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.

Fergusson 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. (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., 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., 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.

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., 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

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). *ConvMPEG-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>

1-Voire aussi de contenus multilingues prenant en compte des langues africaines sahéniennes

2-A l'exception du contre exemple touareg, société dans laquelle les femmes enseignent traditionnellement la langue et l'écriture tfinagh aux enfants.

3-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.

4-Conformes notamment aux instructions de l'OCLC (Online Computer Library Center).

5-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>

6 -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>

7-Mais pas seulement : les photographies, les documents filmiques ou audio peuvent être associés à ces corpus ainsi que des graphiques des formules mathématiques ou chimiques, des plans, etc.

8- www.omeka.org

9- <http://www.gmu.edu/>

10- <http://dublincore.org/>

11- www.tei-c.org

12- <http://www.zotero.org/>