

Normalisation et TIC : enjeux stratégiques du multilinguisme et du multiculturalisme numérique dans la société de l'information

Mokhtar Ben Henda

Mcf, ISD, Université La Manouba, TUNISIE
Chercheur Associé, CEM-GRESIC, MSHA, Université Bordeaux 3, FRANCE
Mokhtar.Benhenda@msha.fr

Avril 2005

Qui ne connaît pas de langues étrangères ne sait rien de la sienne.
Johann Wolfgang von Goethe (1749-1832)

Résumé :

Le particularisme linguistique et culturel des entités dites "minoritaires" dans l'accès et l'échange de l'information scientifique et technique (IST) sur les réseaux mondiaux prend progressivement une place importante dans les axes de recherche en développement et dans les efforts de normalisation des contenus, services et systèmes d'information scientifique et technique. Ceci s'observe désormais à travers le penchant progressif des acteurs du domaine de la normalisation des technologies de l'information, des instances internationales d'aide au développement et des utilisateurs eux-mêmes qui réclament des solutions plus adaptées à leurs profils culturels et linguistiques. Un accent particulier est donné dans ce document au concept des réseaux sémantiques multilingues construits sur la base de métadonnées, terminologies et ontologies multilingues et multiculturelles. Le questionnement principal se situe autour des alternatives linguistiques et culturelles offertes et des rôles des entités dites « minoritaires » pour en tirer le meilleur profit en tant qu'acteurs potentiels dans le rééquilibrage de la société de l'information en terme d'usage des TIC et d'accès à l'IST.

Mots clés :

Normalisation, TIC, Multilinguisme, Réseaux sémantiques, Terminologie, Ontologie

1. Introduction

La généralisation des TIC et de leurs modes d'application à échelle planétaire bute parfois contre des usages non-conformes au particularisme culturel et linguistique des milieux qui les adoptent. Enfouis dans les rouages des dispositifs de programmes d'aide au développement, des programmes de coopération, ou tout simplement véhiculés par de simples inductions mercantiles et commerciales, des standards de fait ou de consensus parviennent à s'ériger comme alternatives universelles unanimement acceptées sans trop de résistance ni de révocation particulières. Il a fallu pour certains pays en voie de développement de voir s'accroître cette diversité culturelle et linguistique au sein même des pays développés comme en Europe ou en Amérique du Nord, pour entamer une remise en question de leur présence dans la société de l'information à travers une technologie et un contenu certes, mais aussi à travers une identité et une équité de production, d'accès et de diffusion de l'information scientifique et technique, numérique, multilingue et multiculturelle. Cette remise en cause générale, et cette ébullition identitaire soudaine qui reprend le dessus d'un penchant originel d'unipolarité, en l'occurrence anglo-saxonne et plus précisément américaine, véhiculée par les

concepts de la mondialisation et de la globalisation, a su engendrer autour d'elle un besoin transversal de standardisation et de normalisation avec des objectifs « déclarés » comme la complémentarité, l'équité de partage et la coopération multilatérale mais qui couve, selon d'autres, un acte et une réaction indéniablement marchande à base d'intérêts économiques et de profits. Même si la tendance actuelle vers une normalisation des pratiques et des usages des TIC et de l'IST dans des domaines d'application stratégiques bute encore contre cette diversité culturelle et linguistique, parfois voulue pour des raisons identitaires, elle reste par contre porteuse d'un jeu de rôles inégalé entre les acteurs potentiels d'un modèle économique dominant, que les Etats-Unis consacre à travers la liberté des échanges et la diminution des barrières tarifaires douanières de la société de l'information (producteurs et consommateurs) : inégalité d'accès aux ressources, manque de sensibilisation des entités minoritaires, etc.

2. La réplique stratégique des PVD

L'histoire repose entre autres sur des paradigmes évidents justifiant, à tout bon ou mauvais escient, l'antagonisme entre les entités à valeurs différentes. Les coalitions stratégiques ont toujours été la solution à la résistance et à l'influence de l'autre. C'est même un procédé devenue politiquement correct dans des mécanisme de démocratisation qui respectent les droits des minorités : l'essentiel est de se défendre et si l'on dispose de conjonctures favorables, de jouer même l'offensive pour dominer. Il suffit de regarder la scène internationale contemporaine ou la société humaine dans son parcours historique sous ses différentes facettes sociétales pour observer la récurrence du mécanisme de coalitions ou de regroupements stratégiques à buts défensifs ou belliqueux. A ne considérer que l'époque moderne, afin de minimiser l'errance dans les arcanes interminables et bien parlants de l'histoire de l'humanité en général, pour voir se consolider cette tendance qui confirme encore une fois la théorie d'Ibn Khaldoun préconisant la sociabilité de l'être humain. Il suffirait de voir les soubassements stratégiques des regroupements locaux, nationaux et régionaux qui ont toujours marqué les sociétés humaines pour comprendre la variabilité des mobiles et des raisons stratégiques qui ont toujours poussé les individus, les groupes et les nations à se regrouper. Sans aucune doute que le facteur économique ressortira à chaque fois comme le facteur le plus déterminant dans ces stratégies unificatrices. Des organisations comme le Commonwealth (1920), l'OCDE (1948), L'Union Européenne (1993), les pays non alignés (1955), le COMECON (1949), l'Union Africaine (2002) et tant d'autres qui ont marqué le siècle dernier et qui marqueront pour toujours la communauté internationale, sont des marques d'évidence de cette course interminable vers l'acquisition des pouvoirs ; des pouvoirs essentiellement économiques, politiques et militaires.

Or, il ne s'agit pas toujours de ces uniques mobiles stratégiques déclarés comme catalyseurs de regroupements humains. Des considérations culturelles et linguistiques, bien que parfois cloîtrées dans les replis des considérations économiques, ont également eu leur part de contribution dans la définition de l'échiquier mondial. Des institutions phares reflètent cette tendance qui s'est constituée timidement mais qui a su revenir en force avec l'avènement du concept de la mondialisation étiqueté au label anglo-saxon. L'Union Européenne, est certes une réaction économique défensive à l'unipolarité américaine voire même une anticipation à une nouvelle bipolarité, cette fois-ci américano-chinoise qui se profile. Mais elle restera toujours une unification à valeurs culturelles évidentes. L'Européanité est un concept qui prend ses marques dans le foisonnement culturel mondial. La résistance modérée à l'intégration des pays de l'Est dans l'Union puis la résistance farouche à l'intégration de la Turquie sont des marqueurs de résistance culturelle évidente à un melting-pot potentiel qui risquerait de perturber l'harmonie d'une culture européenne historique. Dans le même ordre d'idée, l'exemple de l'ALECSO (Alter Ego de l'Unesco pour le monde arabe) est quant à lui révélateur d'un ensemble de soucis à soubassements culturels et linguistiques à la fois qui, bien qu'ils s'inscrivent dans une dynamique de coalition économique régionale (Ligue arabe), dénotent de l'importance que ces valeurs peuvent prendre en marge des considérations économiques. Les cas de la francophonie aujourd'hui (pays parlant totalement ou partiellement le français) ou de l'Union latine (pays hispanophones ou lusophones), sont également des exemples qui traduisent l'importance des facteurs culturels et linguistiques dans les stratégies de développement.

La mondialisation a été au fait une arme à double tranchant : si elle a contribué à la genèse d'un modèle à dominance culturelle et linguistique anglo-saxonne, elle a su très rapidement éveiller les complexes d'usurpation culturelle et linguistique enfouis depuis très longtemps dans les basses couches du sous développement économique et technologique, de l'analphabétisme, du colonialisme et de la pauvreté. Alors que le monde était partagé entre quelques langues dominantes qui veillaient à protéger des zones d'influences historiques et de dépendances linguistiques (i.e. franco-anglo-italo-portugaise en Afrique), les langues vernaculaires et les identités culturelles nationales sont progressivement revenues au devant de la scène en tant que moyen d'expression de beaucoup de mouvements nationalistes comme le panarabisme, les mouvements basque Euskadi, kurde, arménien, afrikaans etc. Ce revirement s'est même transformé ensuite en initiatives souvent promulguées par des langues dominantes dans une optique de coalition contre une hégémonie anglo-saxonne rampante et toujours menaçante.

C'était d'ailleurs à l'Organisation Internationale de la Francophonie (OIF), durant son cinquième sommet en octobre 1993 à l'Île Maurice, que l'idée revient de lancer l'initiative de la défendre la diversité culturelle dans le monde. Le sommet était d'ailleurs identifié sous le label *L'unité dans la diversité* ^[1]. En 1994, l'OIF propose à l'Unesco la mise en place d'une convention qui protégerait les productions et les échanges culturels des seules lois du marché. Le concept est vite adopté par l'Unesco pour aboutir en première phase, en 2001, à la « Déclaration universelle de l'Unesco sur la diversité culturelle ». Selon le directeur général de l'UNESCO, M. Koïshiro Matsuura « *la préservation de la diversité culturelle s'inscrit dans le mandat constitutionnel de l'UNESCO* ». La question était depuis lors d'élaborer un « *instrument normatif de nature contraignante* » qu'une simple déclaration de principes. Le 22 mai 2003, et à l'occasion de la tenue de la journée mondiale sur la diversité culturelle, le groupe des ambassadeurs francophones de l'Unesco a organisé un débat en vue de « *faire adopter par la Conférence générale d'octobre 2003 le principe de l'élaboration d'un projet de convention internationale sur la diversité culturelle pouvant être finalisé d'ici 2005* » ^[2] capable de garantir la protection de la diversité des contenus culturels et des expressions artistiques.

Cette initiative fut largement appuyée à échelle internationale. Le monde hispanophone et lusophone, des grands pays du Commonwealth comme l'Inde, des pays africains anglophones, l'Union Européenne, la Chine et tant d'autres pays ont appuyé cette initiative qui acquiert une approbation internationale quasi unanime. L'Assemblée générale des Nations Unies s'est même alignée à la Déclaration universelle de l'Unesco et a proclamé le 21 mai Journée Mondiale de la diversité culturelle pour le dialogue et le développement ^[3]. L'un des rares pays à s'y opposer sont les Etats-Unis d'Amérique pour des raisons tout du moins compréhensibles pour une entité qui s'est longuement installée dans le confort de l'hégémonie et la dominance linguistique et culturelle dans le monde. Jean Musitelli, ancien ambassadeur de France auprès de l'Unesco rapporte que « *l'attitude résolument hostile qu'ont affichée d'emblée les représentants américains (déclarant, avant tout débat, que la convention était une " mauvaise idée ") a beaucoup fait pour cristalliser la volonté des partisans de la convention et les inciter à s'exprimer avec vigueur* » ^[4].

Or, le revirement du contexte ne fut pas uniquement à l'ancienne : sentiments nationalistes, propagandes médiatiques, alphabétisation et scolarisation de masse. La réforme fut aussi progressiste et à l'ère du temps. Modernisme oblige, les technologies furent l'un des acteurs clés de la réforme linguistique et culturelle. Sans pouvoir échapper à la règle, les défenseurs de la diversité culturelle et

^[1] V^e Sommet de la Francophonie : Grand Baie, Ile Maurice. 16/18 octobre 1993.

[http://www.francophonie.org/francophonie/sommets/maurice_1993.html] (Visité le 5 mars 2005)

^[2] Allocution de Monsieur Abdou Diouf, secrétaire général de l'OIF, lors du débat organisé par le Groupe des Ambassadeurs francophones de l'UNESCO à l'occasion de la "Journée mondiale sur la diversité culturelle"

[<http://www.francophonie.org/presse/discours/infos.cfm?num=170>] (Visité le 5 mars 2005)

^[3] Voir Journée Mondiale de la Diversité Culturelle. [http://portal.unesco.org/culture/fr/ev.php-URL_ID=12508&URL_DO=DO_TOPIC&URL_SECTION=201.html]

^[4] Voir Jean MUSITELLI. La convention sur la diversité culturelle. Forum permanent Planetagora.org sur le pluralisme culturel, 2004. [http://www.planetagora.org/theme2_suj3_note.html] (Visité le 15 juin 2005)

linguistique font désormais usage de l'outil technologique pour prendre position et défendre leurs objectifs. La réplique tiers mondialiste, à l'instar de beaucoup de pays développés, autant elle partait sur des considérations économiques et politiques, autant elle reposait sur une appropriation des technologies et des savoir faire associés. Mais peut être que l'avance est importante et que le déséquilibre s'intensifie d'avantage ? De quoi justifier le non moins célèbre fossé numérique. Une prééminence, voire une exclusivité historique anglo-saxonne et particulièrement américaine en matière de technologie numérique est-elle encore justifiée ?

3. La contre offensive technologique : un repositionnement des PVD

Nous voyons tous, ou presque, les empreintes immuables de l'américanisme technologique dans nos usages quotidiens des TIC. Depuis le simple ordinateur de bureau jusqu'aux plus infimes protocoles du réseaux Internet, tout porte en clair ou en filigrane les traces d'une source originale américaine. Le codage binaire électronique et l'information numérique dans sa structure moderne sont issus des laboratoires américains des années cinquante et soixante du siècle dernier. Les protocoles de transmission des données comme ISO-OSI, TCP-IP, MIME etc. sont des produits des laboratoires de recherche américains. Au début, tout portait cette étiquette au point de voir toute la scène technologique mondiale dominée par un seul modèle venu du nouveau monde. Les bases conceptuelles des systèmes d'information étaient elles mêmes d'une inspiration américaine. Il suffit de revoir les noms de personnages qui ont marqué ces disciplines à leur genèse pour s'en rendre compte : Mac Luhan, Von Neumann, Shannon, Wiener, Moles, Dewey (la liste est longue !). Le détail qui nous intéresse le plus à ce stade est sans doute les origines du codage binaire des données numériques qui allait ensuite profondément marquer, et pour toujours, la société de l'information dans ses bases les plus élémentaires.

3.1 Le codage numérique de l'information

Le codage binaire remonte, en fait, à très loin dans l'histoire et revient sans doute à l'œuvre de l'Anglais (encore un !) Sir Francis Bacon (1561-1626) dont le souci de crypter des messages diplomatiques l'avait poussé à inventer un système de transcription qui assigne à chaque caractère une représentation composée de deux symboles alphabétiques. Il a fallu attendre les débuts du 20^è siècle pour voir reprendre ce concept par des inventions nouvelles comme les calculateurs analogiques de Vannevar Bush (1890-1974), les calculateurs numériques électromécaniques de George Stibitz, de Howard Aiken (1900-1973) et de Konrad Zuse (qui eu le mérite en 1944 de mettre en place le premier ordinateur numérique opérationnel basée sur la technique binaire numérique) en passant par les calculateurs numériques électroniques de John V. Atanasoff, de J.Presper Eckert et John W. Mauchly [BRETON,1990].

C'est à partir des années soixante, juste après la mise en place du modèle informatique de la machine Von Neumann en 1946 et sa future concrétisation en 1954, que le modèle du codage binaire numérique a été conçu et utilisé par les américains. Ce fut d'abord le codage propriétaire EBCDIC d'IBM conçu sur les bases du modèle des cartes perforées de Hollerith, puis fut proposé ce qui est aujourd'hui très connu sous le nom de code ASCII (American Standard for Codification of Information Interchange). ASCII fut ensuite passé en norme internationale ISO sous la référence ISO 646. L'ASCII est ainsi né d'un effort de recherche américain afin de constituer un code de représentation de caractères avec lequel toutes les autres normes assureront toute la compatibilité. « D'ou la très grande importance de cette norme même si maintenant elle a vieilli » [ANDRE, 1995].

Or, les utilisateurs de l'informatique des années 70 et 80 se rappellent encore les contraintes de l'usage de cette norme dans les solutions informatiques de l'époque. Particularité singulière, sans doute, fut celle du traitement des diacritiques. Les protocoles de messagerie, même les applications bureautiques, référence anglaise oblige, ne géraient pas les caractères latins accentués. Que dire alors des caractères des autres langues non latines. Ayant été normalisées ISO et adoptées dans les solutions informatiques du matériel et du logiciel (i.e. IBM, Intel, Sun, Microsoft etc.) les tables de

codes des caractères étaient exclusivement de l'Ascii pur. Les utilisateurs non anglophones de la messagerie ou des protocoles de transfert de données étaient toujours surpris de voir leurs textes infestés de codes bizarres que les protocoles de transfert de données sur les réseaux télématiques (jusqu'aujourd'hui exclusivement Ascii) y inséraient pour remplacer les diacritiques non interprétés. Ceci a duré jusqu'à la deuxième moitié des années 80 quand les entités linguistiques non anglophones comme les langues latines, scandinaves, cyrilliques etc. avaient commencé à travailler sur une alternative pour contourner ce handicap majeur. C'est ainsi que le projet de l'extension des codes sur 8 bits est apparu pour permettre la représentation des caractères diacritiques. La norme ISO 8859-x est ainsi née pour permettre à 15 groupes de langues (dont l'arabe, l'hébreu, le finnois etc.) d'être représentés. Des protocoles comme Mime ont pu ainsi migrer vers un codage sur 8 bits qui permettait alors un transfert des données accentuées sans risque de distorsion des diacritiques.

De même, dans le monde Internet, le langage HTML, utilisé par excellence sur le Web est une implantation à origines francophones car, conçue au CERN en Suisse depuis 1989 par Tim Berners Lee, le premier serveur et le premier navigateur Web public (Nexus), mis en place en 1990 sur la base de la spécification HTML, permettait, à partir de 1993, l'usage de la norme ISO 8859-1:1987 pour un codage de caractères sur 8 bits et un codage des langues du monde, bien que non exhaustif, à travers l'application de la norme ISO 639.

Pourtant, la question qui nous intéresse à ce stade, n'est pas l'évolution du principe de codage des caractères qui d'ailleurs a su ensuite évoluer rapidement vers un codage sur 16 bits sous Unicode puis sur 32 bits sous ISO 10646. Ce qui nous intéresse dans cette histoire est la raison qui a déclenché ce processus d'évolution. Si certains pays développés comme la France et l'Allemagne n'étaient pas concernés directement par cette restriction Ascii, le codage sur plusieurs octets auraient-ils pu évoluer au stade où il est aujourd'hui ? En d'autres termes, les pays en voie de développement auraient-ils l'envergure et la latitude de prêcher pour une refonte des tables de codes pour défendre leurs particularismes linguistiques et culturels numériques ? Les chinois, les japonais et les Coréens ont fait cavaliers seuls pour être les premiers à proposer un codage sur 16 bits qui fut ensuite proposé dans le cadre de la norme Unicode comme standard régional CJK ^[5]. Des langues comme l'arabe et beaucoup de langues africaines auraient-elles pu, à elles seules, imposer une alternative technologique viable ? On se souvient toujours de l'initiative arabe de la norme ASMO 662 ayant pour but de créer l'ordinateur exclusivement arabe. Pour des considérations géopolitiques et économiques, il n'a pas été possible de mettre en place une politique d'industrialisation d'un ordinateur arabe unilingue. Le contexte d'usage multilingue de l'informatique et l'effet de la mondialisation allaient désormais à l'encontre d'une pareille initiative.

Bref, la diversité linguistique et culturelle numérique mondiale doit beaucoup à l'engrenage que « *l'exclusivement anglais* » a engendré dans les communautés industrialisées. Les alternatives de contournement à cette hégémonie anglaise ont créé des chances d'infiltration pour les langues minoritaires afin de se positionner sur l'échiquier technologique mondial. Très rarement c'était le résultat d'initiatives concrètes de la part des entités linguistiques et culturelles en question. Généralement, c'était sous l'impulsion de ces pays industrialisés, conduits par des objectifs commerciaux (i.e. ECMA ^[6]) ou d'alliances stratégiques (i.e. CJK, Latin-1), que des solutions linguistiques, au profit des langues minoritaires, allaient être élaborées.

Aujourd'hui, le phénomène de l'ouverture numérique vers les langues et les cultures minoritaires s'est largement développé dans le monde entier, bien que très souvent conditionné, de façons déclarées ou subliminales, par les raisons économiques et stratégiques. Deux concepts clés marquent désormais l'univers technologique dans ce sens : l'internationalisation (i18n) et la localisation (l10n). Internationalisation et localisation, deux facettes d'un même procédé multilingue et multiculturel au

[5] Alliance linguistique entre le Chinois, le Japonais et le Coréen. Les idéogrammes unifiés CJK (Chinese Japanese Korean) constituent un ensemble de caractères idéographiques utilisables en Chine, à Taiwan, en Corée, au Japon, et au Vietnam.

[6] European Computer Manufacture Association [<http://www.ecma.ch/>]

sein de la société de l'information, sont la source des grandes polémiques et de conflits interculturels latents.

3.2 L'internationalisation (I18n)

Par internationalisation, il est souvent fait référence, sur un plan technique, aux solutions logicielles et applicatives dont les caractéristiques et les codes ne sont pas fondés sur un seul paramètre de lieu. L'objectif de l'i18n est de développer, par des procédés normalisés et à vocation internationale, des solutions logicielles génériques caractérisées par une neutralité culturelle et linguistique aidant, dans un processus de localisation, à répondre aisément aux spécificités culturelles et linguistiques des diverses communautés mondiales.

Sauf que les degrés de perfection de l'I18n varient en fonction de plusieurs critères. Les exigences de qualité des publics cibles et les alternatives qui leurs sont données influent sur la qualité des produits internationalisés. Je cite à titre d'exemple des produits comme ceux de Microsoft. Par un souci d'universalité, les produits Microsoft ont été tous traduits vers toutes les langues à valeur commerciale évidente. La langue arabe en a eu sa part de couverture sous formes de produits logiciels bilingues entre l'arabe comme langue de base associée à l'anglais ou au français en fonction de la région concernée (français pour l'Afrique du nord, anglais pour le moyen orient). Cette dualité linguistique était par contre arrêtée en tant que solution originale durant une période de temps (avec les versions 95 et 98 de Windows) pour se limiter aux seuls produits en version arabe-anglais (des versions francisées ont été parallèlement bricolées). Parmi les raisons préconisées pour cette rupture de dualité linguistique, le déséquilibre du potentiel économique entre le Moyen Orient et le Maghreb était avancé comme argument clé. Pour des raisons d'économie d'échelle, la partie de l'Afrique du Nord, accusée, à raison ou à tort de piratage intensif, a été abandonnée par Microsoft aux produits à soubassements anglophones. Plus tard, l'arrivée massive d'Unicode allait remettre la pendule à l'heure.

3.3 La localisation (L10n)

Localiser, c'est rendre une application cohérente avec les paramètres culturels et linguistiques d'une communauté locale. Les utilisateurs de logiciels ressentent toujours le besoin de faire usage d'éléments conformes à un référentiel propre à leurs langues et à leurs pays. A titre d'exemple, un citoyen moyen orientaliste tend toujours à faire usage d'un système de dates basé sur le calendrier lunaire et d'un système de chiffres à base de chiffres hindous. Les symboles monétaires et les unités de mesures qu'il utilise dans sa vie quotidienne lui sont également nécessaires au sein des applications informatiques qu'il utilise.

Si l'internationalisation est généralement appréhendée comme une procédure d'ingénierie informatique, la localisation est par contre conçue comme un acte de traduction ou d'adaptation d'une application aux particularités locales d'un utilisateur ou d'une communautés d'utilisateurs tout en préservant la fonctionnalité de l'original. La localisation travaille généralement sur l'adaptation de l'interface (menus, aide en ligne, messages d'erreur, etc.) ou de la documentation technique aux besoins linguistiques et culturels des utilisateurs.

Ces solutions spécifient en réalité les caractéristiques culturelles des groupes diversifiés d'utilisateurs comme leurs systèmes métriques, leurs calendriers, leurs systèmes monétaires etc. Il s'agit techniquement de ressources informatiques (fichiers systèmes connus sous le nom de locales ^[7]) prévues par les concepteurs pour donner davantage de chances d'écoulement sur les marchés mondiaux. Prévoir une panoplie de paramétrages linguistiques et culturels pour un large éventail de communautés linguistiques et culturelles dans le monde est devenu un critère de qualité, une garantie marchande.

^[7] Voir Christian CASTEYDE. Les Locales : chapitre 16. [http://casteyde.christian.free.fr/online/cours_cpp/c7625.html] (Visité le 10 mars 2005)

Or, devant la multiplication des solutions de localisation proposées, la complexité des spécificités culturelles et linguistiques locales et la diversité des options qui répondraient à tous les choix et les besoins des utilisateurs, le retour à un contexte tiraillé entre l'offre et la demande, risquerait de replonger l'industrie informatique et téléinformatique dans la disparité technologique et le cloisonnement industriel d'antan. Un effort de normalisation est alors ressenti comme une parade urgente et même vitale pour pouvoir mutualiser et capitaliser les efforts et les acquis d'autrui même dans la diversité et l'hétérogénéité des acteurs industriels et des utilisateurs potentiels des systèmes d'information et de communication mis en place.

3.4 Le besoin de normaliser

L'usage des normes, dans leurs variations de formes, de domaines et d'envergures, n'est pas chose récente. Depuis toujours, et à des échelles variées, l'Homme a eu besoin de se conformer à des principes qui renvoient implicitement à des représentations particulières de ce qui est modélisé. C'est depuis l'émergence de la société industrielle que les normes avaient pris leur ampleur pour se développer progressivement dans la société civile et devenir aujourd'hui le couronnement d'un processus complexe généralement stratifié à travers des phases de recommandations, de spécifications, de standards et de normes. Pour ne reprendre qu'une définition sommaire d'un processus que les anglo-saxons n'ont pas pris la peine de nuancer, « *une norme est un ensemble de règles sanctionnées par des accords juridiques ; un standard est un ensemble de règles sanctionnées par les positions dominantes sur les marchés des produits et des services. L'une est du domaine du droit, l'autre est du domaine des faits* » [SAILLANT, 2002].

Dans le domaine des technologies de l'information, la généralisation du modèle réseautique universel, la connectique électronique qui constitue l'un des atouts majeurs de la société de l'information, la volonté économique de développer des économies d'échelle dans un système économique de libres échanges, oeuvrent désormais pour des modèles technologiques globaux. L'objectif n'en est pas uniquement de réduire la lourdeur de multiplication ou d'adaptation des solutions technologiques existantes, mais également pour faciliter les modalités de repérage des ressources sur les systèmes d'information hétérogènes et renforcer la fluidité des échanges entre les acteurs concernés. Or, la mise en place de modèles globaux ne peut exister sans développer au préalable des savoir-faire communs, des modes de fonctionnement homologués et des besoins rapprochés. Ce qui sous tend un besoin de *mobilité technologique* qui « *n'est pleinement réalisable que si cette dernière est utilisable dans différents contextes. Pour cela, certaines caractéristiques techniques doivent être officiellement reconnues afin de permettre ce partage* » [JOBERT, 2003]. C'est là l'un des objectifs de la normalisation. Yolaine Bourda, professeur à Supélec et membre de la CN 36 au sein de l'AFNOR, définit l'acte de normalisation comme un processus qui permettrait la mutualisation des ressources numériques dans des environnements technologiques différents. Elle considère que « *l'enjeu des normes se situe d'abord au niveau de l'accessibilité des ressources, en permettant de les retrouver facilement, de les réutiliser, de les transposer dans des environnements divers, de les adapter à des profils différents d'utilisateurs ou encore d'en assurer la pérennité* »^[8].

Il faut bien admettre que c'est aussi à ce stade que les normes et les standards doivent jouer leur rôle fédérateur non pas uniquement comme facteur d'harmonie entre les acteurs influents des TIC, mais aussi en tant que régulateurs d'une diversité dans laquelle certains protagonistes, et pour des raisons multiples, n'ont pas l'occasion de contribuer. La question que l'on pourrait alors se poser concerne la part de représentation dont les « have nots », ou en d'autres termes les « démunis » technologiques, généralement recensés parmi les entités culturelles et linguistiques minoritaires, disposent sur l'échiquier mondial des normes et des standards en TIC. Si la question semble poser le problème de la quantité, la qualité des solutions normatives joue quant à elle toute son importance sur le facteur

[8] Voir le compte rendu des interventions du Séminaire de Lyon et Saint-Clément-les-Places, 9 et 10 octobre 2003 organisé par le Ministère de l'éducation nationale, de l'Enseignement supérieur et de la Recherche. Direction de la technologie – SDTICE.. [<http://www.educnet.education.fr/tech/normes/lyon10-03.htm>]

d'impact et les conséquences induites dans une mondialisation désormais caractérisée par une concurrence qualitative. C'est moins donc un problème de formes que de contenus que les langues et les cultures minoritaires doivent affronter pour se repositionner sur la scène internationale. Dans leur absence très notoire des structures internationales de standardisation et de normalisation des TIC, comment les entités dites minoritaires, pourraient-elles préserver et garantir leurs spécificités propres ?

Il faudrait peut être spécifier que le parcours normatif des TIC est largement parsemé de textes conçus et édités au sein des grandes structures internationales comme l'ISO. Avec son palmarès normatif gigantesque, l'ISO prétend couvrir tous les aspects touchant les TICs. La question est par contre de savoir le taux de couverture auquel les spécificités propres aux langues et aux cultures minoritaires sont prises en compte. Car, bien que les solutions normatives universelles prêchent toujours pour une neutralité par rapport aux cultures et aux langues du monde entier, et donc œuvrent pour une équité représentative de la communauté internationale, il est parfois démontré que ces solutions reflètent implicitement des représentations à connotations identitaires propres aux experts qui les conçoivent. Selon Bernard Blandin, directeur de CESI, « *ces experts ne sont pas, pour la majorité d'entre eux, experts en sciences cognitives, ni en sciences de l'éducation, ni en ingénierie de formation, mais experts en technologie de l'information et de la communication* » [BLANDIN, 2003]. Ceci les empêcherait dans certains cas de figures d'intercepter en profondeur les imperfections et les anomalies qui pourraient résulter d'un manque d'adaptabilité entre un produit normalisé sur la base d'une universalité neutre et un usage spécifique basé sur le cognitif et l'interactif culturel de l'utilisateur. Si l'on sait de surcroît qu'une grande partie des propositions de normes en TIC est conçue au sein des laboratoires de l'IEEE (*Institute of Electrical and Electronics Engineers*), l'hypothèse des risques d'inadaptabilité culturelle des normes en TIC reste très plausible. C'est à partir de ces suspicions que Blandin recommande que la validation des propositions de modèles d'objets complexes prévus pour des usages universels, « *soit systématiquement réalisée par des experts nationaux reconnus dans ces domaines, afin de s'assurer que ces modèles reflètent bien les pratiques en cours* »^[9].

Des mesures d'appui à la diversité culturelle et linguistique des TIC se font de plus en plus entendre. Il conviendrait selon certains de « *promouvoir les normes et les standards existants lorsqu'ils correspondent à des choix techniques satisfaisants pour le plurilinguisme* »[SALAM]. Le codage des caractères multi-octets, les algorithmes bidirectionnels, les pages de codes multilingues sont autant d'acquis pour le multilinguisme que les normes et les standards ont pu progressivement mettre en place. Il est également recommandé de ne pas attendre l'apparition des normes ou des standards quand elles n'existent pas pour gérer un aspect linguistique ou culturel particulier. L'application de toute autre forme d'uniformisation capable de répondre momentanément à un besoin culturel ou linguistique préparerait le terrain pour une optimisation de ces besoins dans une phase ultérieure de normalisation. L'imbrication de toutes les formes de consensus (recommandations, spécifications, standards, normes) est désormais possible pour optimiser l'usage de tout produit technologique. L'univers Internet en est un exemple concret. A part les normes qui ont pu uniformiser, à titre d'exemple, l'usage des codes des pays (i.e. ISO 3166-1) et des codes des langues (i.e. ISO 639-2/T) ainsi que des jeux de caractères codés multilingues (i.e. ISO 8859, Unicode, ISO 10646), plusieurs *Demande de Commentaires*, plus connues sous le label de *Request for Comments*, ont été proposées pour renforcer la dimension multilingue d'Internet. A n'en citer que les deux RFC 2070 pour l'internationalisation du langage HTML et le RFC 1766 pour l'identification de langages sur la base d'un codage de deux lettres issues de l'ISO 639, suivies en option du codage du pays par deux lettres issues de l'ISO 3166. Pourtant, des standards de fait continuent toujours à imposer l'empreinte restrictive de l'ASCII, dictée par l'exclusivité historique de leurs constructeurs anglo-saxons. A n'en citer comme exemple que les standards qui régissent les noms de domaine (i.e. IETF RFC 1591), les protocoles d'administration des réseaux (i.e. TCP/IP, DHCP, SNMP, IPV6), le système de gestion des fichiers (i.e. les noms des classeurs de messages). Ce handicap n'a pas empêché pourtant le développement de plusieurs autres aspects de la technologie numérique multilingue.

^[9] Bernard BLANDIN. Op.Cit.

Seulement, et comme il est soulevé plus tôt dans ce document, le foisonnement multilingue et multiculturel international a atteint un stade où une progression qualitative devient urgente. Étant donnée que la question de la représentation informatique des données multilingues est largement résolue et même soumise aux soins des concernés directs, l'urgent est désormais de se pencher sur les outils de structuration, de référencement et de repérage des contenus multilingues. Bref, le concept des réseaux sémantiques multilingue devient une nécessité pour rattraper l'écart technologique en termes de contenus, de services et d'usages entre les langues et les cultures du monde sur les réseaux d'information.

4. Les réseaux sémantiques au profit du multilinguisme numérique

Les travaux de recherche sur les réseaux sémantiques et les produits de leur développement dans les pays développés sont en grande gestation depuis quelques années. L'application Syntax ^[10] du laboratoire français Loria est un bon exemple d'outils de définition et de développement de catégories de données, de concepts, de terminologies spécialisées et des relations sémantiques multilingues qui les associent. Beaucoup d'autres expériences pilotes ont pu engendrer des produits forts intéressants et compétitifs qui ont pu donner, à travers la technologie numérique, une nouvelle impulsion à des secteurs sensibles comme la veille stratégique et la gestion des connaissances.

Au fait, le principe des réseaux sémantiques est largement inspiré des modèles étudiés en sciences cognitives. Quand nous intégrons des connaissances dans notre système de mémorisation, elles sont systématiquement associées à des connaissances préalables créant ainsi un vaste maillage sémantique non linéaire structuré sur la base de notre interprétation et de notre compréhension des concepts et des faits. L'acte de mémorisation consiste alors à incorporer, à chaque fois, un nouveau concept (noeud) selon les liens qui lui sont attribués par chacun d'entre nous dans son propre maillage sémantique.

Les réseaux sémantiques se sont appropriés ce schéma cognitif pour le traduire en un vaste réseau électronique hypermédia où la navigation d'un réseau à un autre s'effectue en fonction des associations sémantiques entre les concepts et les idées. Les réseaux sémantiques s'apparentent aujourd'hui à des cartes conceptuelles se rapprochant d'une reproduction du réseau de connaissance en mémoire humaine. Les modèles cartographiques (i.e. Topic Maps, Umap) font désormais légion sur Internet. Or l'idée des réseaux sémantiques n'est pas de faire en sorte que les ordinateurs puissent comprendre le langage humain ou fonctionner en langage naturel. Il ne s'agit non plus de développer des systèmes intelligents capables de permettre aux machines de réfléchir. Il s'agit plutôt de « *regrouper l'information de manière utile, comme pour une gigantesque base de données, où tout est écrit en langage structuré* » [BORDERIE, 2002]. A l'ère du Web, le concept du Web sémantique reprend les mêmes préceptes pour constituer une extension du Web classique favorisant la publication, la consultation et l'automatisation des tâches de traitement des connaissances précisément formalisées.

Toujours est-il vrai que si le Web sémantique offre un cadre d'intelligence additive pour l'exploitation des ressources d'information, « *cette capacité n'est possible qu'à condition que les domaines de connaissances aient été décrits par des acteurs humains, qu'ils soient des spécialistes du domaine lors de la constitution d'ontologies de référence ou des auteurs, voire lecteurs actifs, des documents, lors de leurs activités d'annotation à l'aide de métadonnées* » [CAUSSANEL, 2003]. L'approche de description par métadonnées la plus connue consiste à faire usage d'un référentiel sémantique structuré sous forme de ressources lexicales, terminologiques, taxonomiques et ontologiques qui servent à décrire, via des schémas de métadonnées, les ressources échangées sur les réseaux sémantiques.

Tout le chapitre de la terminologie est donc remis sur scène pour reconsidérer la perspective de mettre en place des systèmes de gestion de corpus lexicaux homogènes, partageables et interopérables ; de concevoir et de mettre en place des bases de données terminologiques capables d'alimenter les

[10] Voir l'application Syntax Suite du Loria, <http://syntax.inist.fr/>

schémas de métadonnées en ressources lexicales ; plus encore d'élaborer des corpus multilingues pour les opérations de traductique ou de processus de traitement automatisé des langues.

4.1 Les bases terminologiques multilingues

Selon la définition recensée dans la base terminologique canadienne Termium, la terminologie fait référence à un ensemble de mots techniques appartenant à une science, à une activité professionnelle, ou à un groupe social. En tant qu'activité professionnelle, elle aborde scientifiquement les concepts et les termes en usage dans les langues spécialisées. Appliquée à la terminologie, la normalisation aboutit à un accord en vertu duquel des termes techniques seront utilisés dans une norme, et qui spécifie les caractéristiques selon lesquelles les termes choisis doivent être compris ^[11].

Les produits terminologiques peuvent varier des simples listes de termes plus ou moins structurées (i.e. index alphabétiques) aux thesaurus hiérarchiques, réseaux lexicaux etc. Tous sont utilisés à des affinités d'indexation (automatique ou manuelle) ou de recherche d'information. Dans des contextes plus complexes, ils peuvent servir de référentiels terminologiques plus argumentés, « où les termes sont définis et leurs conditions d'emplois décrites (terminologies de références pour les systèmes d'aide à la rédaction, bases de connaissances terminologiques multilingues pour la traduction), ou, à un niveau plus abstrait, (qui n'est plus celui des expressions linguistiques, mais celui des notions), des ontologies décrivant les objets de connaissance d'un domaine » [BOURIGAULT, 1999].

Les bases de la science de la terminologie (car c'en est une), ont été bâties à travers la *Théorie Générale de la Terminologie* qu'Eugen Wüster, de l'Ecole de Vienne, a mise en place en 1979 [CABRE, 2000]. Pour Wüster, la terminologie est normalement construite sur le sens que les termes expriment et sur l'analyse des relations qui les associent. Il focalise l'idée du concept comme entité de départ dans un champ de connaissance particulier pour en déterminer les dénominations standardisées correspondantes. C'est pour cette raison qu'il est considéré plus souple de définir les concepts dans un champ sémantique quelconque sans besoin d'entrer dans les détails plus complexes, quoique plus représentatifs, des spécificités linguistiques et culturelles qui les démarquent. La standardisation saura ensuite, selon Wüster, uniformiser les formes de dénomination des concepts selon des consensus aboutis. Cette méthodologie de travail terminologique est reconnue sous le qualifiant d'onomasiologique que Hudrisier décrit comme un procédé qui « part des concepts pour les organiser en système de concepts (1er niveau) pointant ensuite sur des registres de langues, voire des langues de spécialités ou disciplines (2e niveau) contenant des termes à un 3e niveau. Des outils comme Topic Maps (orientés comme leur nom l'indique sur la cartographie de ces réseaux de concepts ou la construction d'une ontologie électronique du métier du e-learning) viendront compléter la démarche » [HUDRISIER, 2004]. Cette approche s'oppose diamétralement à la méthodologie plutôt lexicographique dite sémasiologique.

La portée multilingue des systèmes terminologique est également sujette à controverses. La plus importante est sans doute le contexte référent d'un corpus terminologique quelconque. L'une des conjectures de la théorie de Wüster est que l'universalité du savoir le rend autonome par rapport aux langues et aux cultures. Une fois qu'un système conceptuel d'un champ d'activité est défini, une terminologie multilingue pourrait alors être construite sur la base du principe onomasiologique décrit. Cette théorie suppose également que pour chaque domaine de connaissance, il existe un seul et unique corpus terminologique consenti, utilisable dans toutes les formes d'applications possibles. Or les opposants à cette théorie argumentent que la construction d'une terminologie est tributaire du domaine de son application. Il ne suffit pas pour eux de s'entendre sur la façon de nommer les choses mais aussi sur le sens qu'on attribue aux mots qu'on utilise. A titre d'exemple, on procéderait de manière différente dans le choix des termes et de leurs descriptions pour développer « une terminologie de référence pour un système d'aide à la rédaction (dont les entrées vedettes sont les termes considérés comme « corrects » par les experts, et qui peuvent être assortis de leur définition et de contraintes sur

[11] Termium est la base de données terminologiques et linguistiques du gouvernement du Québec.
[http://www.termium.gc.ca/didacticiel_tutorial/francais/glossaire/]

leur emploi) ou un réseau lexical pour améliorer la recherche d'information sur Internet (où on cherche au contraire à indiquer le plus exhaustivement possible toutes les variantes morphosyntaxiques et sémantiques des termes, et où leur définition est accessoire) »[BERLAND, 2000].

En plus de cette contrainte de contexte référent, plusieurs constructions terminologiques multilingues se sont trouvées soumises à une difficulté supplémentaire. La correspondance directe entre des concepts exprimés dans plusieurs langues est souvent porteuse d'incohérences sémantiques profondes issues des traductions lexicales non conformes même pour des champs sémantiques qualifiés moins sensibles aux différences culturelles, comme les techniques et les sciences exactes, ou pour des langues généalogiquement et culturellement proches. On cite souvent la nuance sémantique et culturelle entre la souplesse du terme français « Demande » et sa correspondance anglaise rigoureuse (lexicale et non sémantique) « Demand ». Ce cas de figure est très fréquent entre toutes les langues du monde. Ce qui fait que la simple traduction lexicale d'une terminologie d'une langue vers une autre n'aurait pas les conséquences escomptées pour un traitement sémantique approprié. L'exemple des outils d'aide à la traduction sur Internet est révélateur de cet état de constructions terminologiques démunies de toutes formes d'analyse contextuelle pour gérer la sémantique des termes.

L'émergence des TIC devait normalement faciliter et accélérer les processus de traitement des systèmes des concepts. Les produits terminologiques sont aujourd'hui consignés dans des banques de données et exploités à des affinités multiples. Avec le développement des réseaux ouverts et distribués, de l'ingénierie linguistique, de l'intelligence artificielle et des systèmes experts, des bases de données terminologiques multilingues prolifèrent dans des aires où le multilinguisme fait pressions d'usage. Seulement, selon Jian YANG, chargé du projet CHIQ-BTML pour l'Office de la langue française au Québec, la grande majorité de ces banques de terminologies multilingues affichent des inadéquations notoires avec les prérequis des réseaux sémantiques. D'une part, elles ont été conçues sur la base d'un choix prédéterminé et fixe de langues, donc inappropriées pour l'extension des réseaux sémantiques multilingues. D'autre part, et pour une raison d'antériorité par rapport à une évolution technologique actuelle, elles affichent souvent des incohérences et des inadaptations linguistiques pour certaines langues quand il s'agit de fonctionnalités à valeurs ajoutées comme le tri, l'indexation et l'affichage. Ces banques affichent aussi un problème de rigidité quant à l'uniformité du cadre d'organisation des données qui peut parfois ne pas correspondre aux préférences d'usages pour certains partenaires linguistiques. Or, *« la globalisation de l'information a changé considérablement le contexte de production de terminologies multilingues et, par conséquent, suscite une nouvelle réflexion quant à la bonne manière de coopérer avec des partenaires de langues différentes en matière de terminologie »* [YANG, 2001].

Des recommandations d'ordre pratique sont pourtant envisageables pour optimiser l'usage des banques de données terminologiques au profit des langues minoritaires. On en cite quelques unes :

- La banque doit pouvoir intégrer, à tout moment, toute langue supplémentaire possible disposant d'un système de codage informatique sans que cela n'engendre des opérations lourdes de réécriture des codes sources de l'application,
- Elle doit respecter les caractéristiques propres de chaque langue traitée au niveau de ses particularités phonologiques, syntaxiques, morphologiques et morphosyntaxiques,
- Elle se doit d'adapter l'organisation interne des données d'une langue conformément aux préférences culturelles et linguistiques des partenaires concernés,
- Elle doit tenir compte de la nature et de la profondeur des liens d'équivalence entre les termes des langues concernées,
- Elle doit permettre une interrogation des données à partir de toutes les versions linguistiques des termes pour aboutir aux mêmes résultats obtenus à partir des liens d'équivalence linguistique prédéfinis.

La normalisation allait venir baliser cet univers complexe de la terminologie pour essayer d'aboutir à un consensus d'usage des termes basé sur le sens et le contexte ainsi que les caractéristiques selon

lesquelles les termes doivent être interprétés. La normalisation allait ainsi redéfinir ses rapports historiques avec les termes répertoriés depuis la révolution industrielle. Car il faut reconnaître que la terminologie, en tant que pratique professionnelle, a connu sa gloire dans le contexte de la normalisation industrielle. Tous les secteurs de l'industrie sont aujourd'hui quadrillés par des textes normatifs à langages techniques relativement unifiés. Sauf que tout l'arsenal normatif industriel en place se caractérise par une volonté de vouloir normaliser d'abord les objets et les procédures de leur fabrication, maintenance et organisation à travers des termes conventionnés sélectionnés autour de l'objet lui-même. Les termes et les concepts ne faisaient pas partie de leurs objectifs essentiels. C'est pour cette raison que des instances comme l'ISO, à travers leurs comités techniques, tentent de normaliser les activités conduites dans les domaines de la terminologie et du vocabulaire dans le but d'harmoniser les efforts investis dans ce sens et de permettre l'échange des données terminographiques selon des méthodes souples et perfectionnées. Des textes de normes touchant plusieurs aspects de ce domaine linguistique ont été approuvés. A n'en citer que les textes normatifs suivants :

- Travaux terminologiques-Harmonisation des notions et des termes (ISO 860:1996),
- Travaux terminologiques-Vocabulaire-Partie 1: Théorie et application (ISO 1087-1:2000),
- Mise en ordre alphabétique des données lexicographiques et terminologiques multilingues représentées dans l'alphabet latin (ISO 12199:2000),
- Aides informatiques en terminologie-Catégories de données (ISO 12620:1999),
- Applications informatiques en terminologie-Plate-forme pour le balisage de terminologies informatisées (ISO 16642:2003).

Dans le palmarès des travaux terminologiques effectués au sein des comités techniques de l'ISO, le comité technique 37 (TC 37), chargé de la terminologie et des ressources linguistiques et de contenu, a vu validée, en août 2001, la création d'un nouveau sous-comité (TC37/SC4)^[12] entièrement dédié à la normalisation dans le domaine des ressources linguistiques. Mais il reste important de préciser que l'impact des normes ISO en terminologie n'a encore de résonance et de valeur de référence qu'au sein des structures de normalisation, auprès des membres des comités techniques qui les élaborent. A ce titre, on leur reproche encore de ne pas rendre compte des divergences parfois profondes entre les différentes tendances et courants linguistiques en cours. Pour bien comprendre l'argumentaire de ces reproches, il suffit d'observer trois détails révélateurs. Primo, les deux langues dominantes de travail au sein de l'ISO sont généralement l'anglais et le français. Les versions linguistiques des documents officiels validés sont léguées aux instances nationales de normalisation, qui ne s'appliquent généralement pas à les traduire au profit de leurs communautés linguistiques locales. Secundo, seul un nombre réduit de pays développés sont membres des comités techniques de l'ISO. Ceci pose le problème de la représentativité universelle et de l'équité des chances pour figurer dans le palmarès normatif international. Tercio, le manque de motivation de la part des pays en voie de développement, pour des raisons de moyens économiques surtout, affaiblit la part de contribution de partenaires linguistiques dans des comités techniques siégeant sur des thématiques déclarées d'une priorité seconde pour leurs soucis de développement économique et social.

C'est dans cette optique d'optimiser la contribution de tous les partenaires linguistiques du monde que des appels se succèdent et des initiatives se multiplient pour que des axes de recherche en linguistique, terminologie, vocabulaire, ontologies, soient avant tout entrepris dans des contextes ouverts à la réflexion et à la recherche scientifique et universitaire. L'Agence Universitaire de la francophonie (AUF), par le biais de ses programmes d'action et de son rôle de point de liaison, catégorie A auprès du sous comité 36 de l'ISO, chargé de la normalisation des technologies éducatives, est un exemple concret de cette initiative d'ouvrir le monde de la normalisation aux entités linguistiques et culturelles partenaires encore déconnectées (i.e. arabophone, lusophone, africaines etc.) afin d'exprimer et de développer leurs particularismes culturels et linguistiques propres. L'AUF explique cette orientation par le fait que « *la langue française doit cultiver sa relation aux langues du monde, l'anglais,*

^[12] Le Comité Technique 37, Sous comité 4 de l'ISO. [<http://www.tc37sc4.org/>]

l'espagnol, le portugais, mais aussi l'arabe et les langues nationales en contact avec elle »^[13]. Mais c'est aussi à travers les interactions avec des disciplines connexes comme la lexicologie, la sémantique, la traduction automatique, l'intelligence artificielle, etc. que les langues minoritaires pourraient enrichir davantage leur part de contribution aux enjeux linguistiques et terminologiques internationaux.

En somme, l'état de l'art démontre que les champs linguistiques, terminologiques et sémantiques sont en phase de progresser vers un niveau supérieur de perfection, une fusion susceptible de générer un contexte d'une meilleure maîtrise des contenus numériques non plus dans leurs conceptions, codages et traitements informatiques uniquement, mais aussi sur le plan de leurs sens et de leurs correspondances sémantiques et culturelles. La science de la terminologie est en train d'aboutir au développement des ontologies spécialisées et multilingues qui supposent un rapprochement important entre le domaine classique de la terminologie et l'univers du Web sémantique. Laurent Romary, chairman du TC37/SC4, chercheur à l'INRIA et responsable de l'équipe Langue et Dialogue, considère que les processus de l'internationalisation de l'industrie technologique et les besoins en automatisme qui s'intensifient, particulièrement pour les langues des pays d'Europe de l'Est et du Moyen et Extrême Orient, accentuent les transactions et les volumes des informations manipulées et requièrent par conséquent des efforts particuliers pour développer des ontologies spécialisées et multilingues.

4.2 Les ontologies multilingues

En effet, le domaine des ontologies, une reprise évoluée des thésaurus et des schémas fléchés très connus dans le monde des bibliothèques et des centres de documentation, est en phase de devenir une discipline à part entière largement associée au Web sémantique. Jeff Heflin, chercheur à l'Université Lehigh aux EU, identifie les ontologies comme un mécanisme qui définit les termes utilisés pour décrire et représenter un domaine de connaissance. Contrairement aux bases terminologiques qui en dépendent dans des schémas de relations onomasiologiques et sémasiologiques, « *les ontologies incluent les définitions, informatiquement exploitables, des concepts élémentaires dans le domaine et de leurs relations* » [HEFLIN, 2004].

Zweigenbaum, chercheur au laboratoire de statistiques et d'informatique médicale-France, considère l'ontologie comme un ensemble de « *concepts, relations et contraintes effectivement utilisés pour modéliser un domaine donné. On peut considérer qu'une ontologie, dans ce sens, est l'aboutissement formel de la définition d'une terminologie* » [ZWEIGENBAUM, 1999]. Une ontologie se situe, par conséquent, à un niveau d'abstraction plus élevé que la terminologie. Elle concerne les concepts, la taxinomie, et surtout les relations entre concepts et les inférences d'un domaine du discours moyennant une spécification explicite et formelle d'une conceptualisation commune inscrite dans un esprit de partage. Le consortium du Web (W3C), une autorité en la matière, le confirme en rajoutant que « *les ontologies associent les concepts de base d'un domaine précis et les relations entre ces concepts, tout cela d'une manière compréhensible par les machines. Elles encodent la connaissance d'un domaine particulier ainsi que les connaissances qui recouvrent d'autres domaines, ce qui permet de rendre les connaissances réutilisables* »^[14].

Sur un plan technique, les ontologies sont des réseaux sémantiques à grandes envergures (dizaine de milliers de « concepts » inter reliés servant, entre autres, à des affinités de classification et de description (indexation) des produits ou des services d'information pour réaliser des opérations d'interprétation et d'inférence d'information pertinente en rapport avec des besoins de recherche. Fonctionnant sur la base d'une couche de métadonnées descriptives des ressources, les ontologies ont besoin d'un modèle de déclaration pour pouvoir normaliser la structure des dites ressources selon des schémas bien connus comme les schémas XML et les DTD (Document Type Definition).

^[13] Agence Universitaire de la Francophonie. Programmes d'action : Langue française, francophonie et diversité culturelle. [<http://www.auf.org/programmes/programme1/>]

^[14] Questions Fréquemment Posées sur le Langage des Ontologies Web (OWL). [<http://www.w3.org/2003/08/owlfaq.html.fr>]

Un modèle bien connu, situé au centre des travaux du Web sémantique, en l'occurrence le modèle d'assertion RDF (Ressource Description Framework), devient, depuis 1999, un modèle normalisé adapté pour représenter des métadonnées associées à des ressources disponibles sur le Web. C'est à ce juste titre que les ontologies vont intervenir pour définir les règles de base pour l'insertion de métadonnées sous forme de concepts liés à un terme approprié. Ce mécanisme servira ultérieurement à retrouver des ressources pertinentes même si elles ne contiennent pas le terme exact recherché (anti-silence), ou au contraire pour éliminer les ressources contenant le terme recherché, mais dont la signification est différente (anti-bruit). Le schéma RDF est pourtant jugé simple et inadapté pour obtenir une certaine interopérabilité entre des schémas nombreux et complexes, développés et gérés de manière autonome. Ceci est dû au fait que RDF n'applique aucune forme de normalisation sur la nomenclature des entités ayant le même sens empêchant de la sorte toute forme d'agrégation ou d'interopérabilité entre des données provenant de plusieurs sources différentes.

Les ontologies allaient donc prendre cette contrainte en charge en décrivant formellement les termes et les relations entre eux. Le langage OWL (Web Ontology Language) est ainsi conçu pour permettre de publier et partager plusieurs ontologies sur le Web en aidant les machines à gérer les informations, plutôt qu'à les rendre lisible à l'écran. Il s'agit d'une extension de RDF, avec un plus majeur, celui de créer des passerelles entre les désignations d'objets d'une ontologie à une autre.

Selon Rastier, chercheur au CNRS, cette migration des réseaux sémantiques vers les ontologies « *n'est pas simplement cosmétique : elle ancre ces formalismes de représentation sémantique par graphes dans la tradition philosophique du cognitivisme orthodoxe, qu'illustre notamment le psychologue George Miller...* » [RASTIER, 2004]. L'ouverture gigantesque et les potentialités de repérage que les ontologies ont apportée aux réseaux sémantiques ont profondément changé les modalités de recherche et d'accès à la connaissance et au savoir sur les réseaux. Ce sont désormais les modèles cartographiques des réseaux de concepts, les arbres de connaissances et les stratifications par clusters qui remplacent les formes documentaires classiques d'accès séquentiels par index. Ces nouvelles fonctions permettent, en effet, d'obtenir des vues synthétiques sur le contenu d'un agrégat de documents, réalisées via des algorithmes de *clusterisation* qui déterminent les rapports sémantiques entre toutes les entités extraites lors de l'opération de recherche. Cette technique, reconnue sous le concept de Topic Maps, est le résultat d'un long travail de recherche autour de la problématique des index, de l'organisation d'informations et de la schématisation des connaissances. Elle fera également objet de long processus de normalisation au sein de l'ISO et du groupe de travail TopicMaps.Org pour aboutir en 2000 à une norme internationale, ISO/IEC 13250, décrivant de façon non formelle les types d'objets qui composent une Topic Map, ainsi qu'une syntaxe XML de représentation et d'échange, en l'occurrence XTM 1.0 [PEPPER, 2001].

D'un point de vue linguistique, les ontologies ont aussi trouvé très vite leur chemin vers les réseaux ouverts et distribués pour résoudre des problèmes d'interprétations sémantiques multilingues. A partir de justificatifs linguistiques qui démontrent les niveaux de complexité de la compréhension du sens dans une langue particulière, et plus encore la tâche ardue de comprendre du sens lorsque plusieurs langues entrent en jeu, on peut facilement appréhender les vrais défis d'un Web sémantique ou d'une ontologie multilingues. Selon Lapalme, enseignant chercheur à l'Université de Montréal et Macklovitch, directeur du laboratoire RALI à Montréal, « *on pourrait imaginer que la compréhension est un processus qui se passe à un niveau indépendant de la langue mais cette question des subtiles interactions entre le langage et la compréhension est loin d'être réglée* » [LAPALME, 2005].

Déjà, dans le contexte documentaire classique, les thésaurus multilingues, parents proches des ontologies, ont soulevé des problèmes importants inhérents, surtout, à la traduction systématique d'un thésaurus unilingue pour produire un thésaurus multilingue. Ces cas de figure ont toujours engendré des complexités de non-correspondance de termes d'une langue à une autre, de traductions variées selon le sens, de pauvreté de vocabulaire dans une langue et de découpage différent d'un champ sémantique selon les langues. Fietzer [FIETZER, 2002], cité par Da Sylva [DA SYLVA, 2003],

synthétise l'état des relations entre des paires de termes répertoriés dans des thésaurus multilingues comme suit : équivalences exactes, inexactes, partielles, d'une à plusieurs, ou non-équivalence. Cette synthèse peut donner une idée des problèmes potentiels rencontrés au moment de l'élaboration d'ontologies multilingues bien qu'elles soient relativement différentes des thésaurus sur le plan de la structure hiérarchique. En effet, dans une ontologie, les concepts sont structurés conformément aux relations sémantiques qui relient les termes qu'ils représentent. Ils sont autonomes sur le plan linguistique mais restent rattachés à un ou plusieurs termes dans plusieurs langues à la fois. Il est toutefois possible que, lorsqu'une ontologie est construite à partir de termes extraits dans plusieurs langues, une des langues concernées ne dispose pas de termes adéquats pour désigner un concept.

La problématique fondamentale de ce document se situe peut être à ce niveau : dans tous les paliers du travail terminologique ou sémantique mondial en cours, où se situent les langues dites minoritaires dans leurs processus d'actualisation (i.e. modernisation sémantique, enrichissement lexical) et de normalisation de leurs langues respectives ? Les linguistes prétendent que les langues sont vivantes et dynamiques. On voit pourtant plusieurs langues minoritaires adopter les solutions faciles de l'emprunt et de la translittération pour combler des lacunes dans beaucoup de disciplines nouvelles. Le travail des académies de langue est crucial à ce niveau et les enjeux sont fondamentaux pour l'avenir des langues minoritaires et de leurs ressortissants ^[15]. Que dire alors de l'équilibre à envisager en termes d'autonomies propres et de l'urgence par rapport à des axes de recherche innovants comme les ontologies et les Topic Maps !

Ce n'est pas dans les objectifs de ce document de trouver la réponse à ces interrogations, autant qu'il est question d'explorer un domaine dans lequel les pays en voie de développement seraient obligés un jour de s'y aventurer. Alors que le problème terminologique se pose toujours dans ces pays dans une acuité paralysante à la culture technologique et à l'information numérique, les contextes linguistiques dominants évoluent déjà vers des phases avancées de recherche en terminologie et en vocabulaire spécialisés. Pourtant, la lutte contre la fracture numérique fait toujours rêver beaucoup de gens. Mythe ou réalité ? L'avenir nous le fera découvrir.

5. Les perspectives : des ancrages possibles pour les pays émergents

S'il est vrai que les statistiques de la présence des langues dans les sphères de la technologie numérique mondiale sont alarmantes ^[16], leur aménagement mondial par effectifs de locuteurs l'est beaucoup moins ^[17]. C'est déjà un indicateur de *bonne espérance* pour pouvoir remonter la pente et revenir dans la diversité du contexte linguistique mondial moyennant des solutions déjà apportées par la communauté internationale. L'avance technologique évoluera sans doute, et pour encore longtemps, au détriment des pays du Sud, mais des voies de contribution leurs sont déjà beaucoup plus ouvertes qu'avant pour s'ancrer aux amarres de la société globale de l'information. Deux pistes porteuses sont à explorer dans ce sens : la production des contenus numériques multilingues et la normalisation des ressources pédagogiques. Une troisième piste, beaucoup plus globale, coûteuse et à long terme, est en train de s'ouvrir au multilinguisme universel. Elle concerne l'internationalisation des zones systèmes, comme les noms des domaines (i-DNS) quasi exclusives à l'ASCII.

5.1. La production des documents numériques : le miracle Unicode

Les acquis du multilinguisme numérique depuis le miracle Unicode (recommandation ISO/IEC 10646 en 1993) et ses divers codages réduits (UTF-7, UTF-8), ne laissent plus prétendre aucun obstacle

[15] Une lourde charge est attribuée à l'académie des langues africaines (<http://www.acalan.org/index.html>) créée en septembre 2001 pour remettre les langues africaines dans le courant de l'ère numérique.

[16] Voir le site Global Search. Global Internet Statistics. (<http://www.greach.com/globstats/>) (Visité le 15 juin 2005)

[17] Voir Jacques Leclerc. L'aménagement linguistique dans le monde. (<http://www.tlfq.ulaval.ca/axl/index.shtml>) (Visité le 15 juin 2005)

technologique majeur pour le traitement d'une langue donnée. On n'en parle plus d'ailleurs qu'en termes d'adaptabilité et de cohérence des modes de codages des caractères codés. Et là, aucune comparaison à faire avec l'austérité des techniques de codage linguistique de la norme précédente ISO 8859. Dans un travail académique cumulateur, Da Sylva [DA SYLVA, 2003], de l'Ecole de Bibliothéconomie de Montréal, fait le point sur les différentes formes des technologies facilitatrices pour la dissémination de la communication scientifique multilingue. Son travail aborde les aspects touchant les modes de représentation des caractères, des langages de structuration et de balisage des documents numériques multilingues, des parseurs détecteurs automatiques des langues, des ressources lexicales et sémantiques multilingues disponibles, des outils de reconnaissance optiques des caractères, des outils de traitement automatisés des langues diverses, des outils de recherche translinguistiques, des outils d'aide à la traduction automatique etc. En somme, le multilinguisme numérique n'est plus du ressort de l'hypothèse. Tous les indicateurs potentiels l'affirment comme une réalité tangible. Pourtant les pays émergents ne parviennent toujours pas à remonter les pas de l'échelle de production des ressources numériques accessibles sur le Net. S'il est vrai que ceux-ci sont contraints par la carence des ressources financières, des performances des ressources matérielles et des compétences des ressources humaines nécessaires, les enjeux stratégiques et identitaires de la société de l'information et des ressources numériques sont tellement grands qu'aucune logique prioritaire ne devrait reléguer ce souci à un deuxième rang de préoccupation.

5.2. La normalisation des TICE : une chance de repositionnement

Dans l'attente de voir ces acquis technologiques multilingues optimisés par la communauté linguistique des pays émergents, il y a lieu d'investir en parallèle dans les activités de normalisation en cours, surtout celles qui viennent de commencer. A n'en citer ici que les projets de normalisation d'un nouveau domaine qui se construit ; le domaine de la formation ouverte et à distance et les axes subsidiaires qu'il embarque comme les métadonnées, le développement des ontologies et des réseaux sémantique. Les thèmes paraissent a priori porteurs d'une dimension qui dépasse les potentialités ou les priorités des pays émergents, mais la réalité est toute autre. Il est certes vrai que la technologie, comme pilier central de ces processus de normalisation, est déjà acquise pour les pays développés. Mais il est moins certain que ces pays développés aient déjà acquis des élans éloignés par rapport aux autres dans la maîtrise des spécificités de ces domaines. C'est maintenant où jamais, peut-on dire, pour prendre le train en marche et agir sans complexes au sein de la communauté internationale, non pas uniquement pour être informé en temps réel des réalisations en cours, mais aussi pour pouvoir défendre ses propres intérêts stratégiques dans un contexte normatif à retombées internationales évidentes comme celui de l'ISO.

L'enseignement à distance, avec toute les valeurs innovatrices et stratégiques d'éducation pour tous, de mutualisation de connaissances et d'appropriation de savoirs qu'on lui attribue ces derniers temps, est en train d'être structuré par des normes internationales qui pourraient redessiner de façon définitive, contraignante et unanime toutes les formes d'articulations entre ses outils, ses contenus et ses acteurs humains et institutionnels. Si l'on connaît que dans l'enceinte normative réduite et hermétique du sous comité 36 de l'ISO-CEI/JTC1, où les seules spécificités linguistiques et culturelles de référence prises en comptes sont celles des experts qui les analysent, et si l'on sait aussi qu'aucun pays émergent n'est officiellement présent dans ce club des grands, les prévisions sur les couleurs des normes à venir ne sont plus de l'ordre du secret. Comment peut-on envisager que dans les travaux du groupe de travail 7 (WG7), chargé de l'adaptabilité linguistique et culturelle des normes à venir, les spécificités culturelles et linguistiques des communautés arabophones ou lusophones ou africaines pourraient être prises en compte ou même signalées, si aucune représentation de ces communautés n'est là pour les défendre ? On a eu, à titre d'exemple, l'occasion d'observer dans un travail de recherche [BEN HENDA, 2003], que des particularités linguistiques et culturelles propres à la langue et à la culture arabe n'ont pas été prises en compte dans la formulation d'un projet de normes pour l'enseignement à distance, en l'occurrence le Learning Object Metadata (LOM IEEE 1484.12.1-2002). Quelle serait l'attitude de la communauté arabophone, si cette proposition a été validée comme norme internationale avec ces carences culturelles et linguistiques ? Ce serait certainement l'une des deux options : admettre ces carences avec tout l'impact que cela

engendrerait sur les usages où rejeter la solution et entrer dans le cercle de l'exclusion volontaire. Cet exemple pourrait s'appliquer aux 6700 langues dans le monde.

6. Conclusion

Il est peut être abusif dans ce stade de développement économique et social mondial d'exiger un niveau d'excellence très poussé des pays émergents en termes de TIC, de normalisation et de concurrence culturelle et linguistique alors que leurs enjeux sont d'ordres plutôt économiques et sociaux. Sauf que le contraire aussi ne doit pas verser dans un laxisme qui aurait des effets opposés à l'envie général de réduire la fracture numérique entre le nord et le sud. Information et technologies, sont désormais les ingrédients de base d'une potion magique vers le développement global tellement elles empreignent tous les mécanismes de fonctionnement des sociétés. La recherche, l'éducation et la formation bâtissent les fondements de la force motrice du savoir. La standardisation et la normalisation la canalisent vers une capitalisation de acquis et une rentabilisation des projets.

La potion magique est en train de bouillonner dans plusieurs chaudrons de la société de l'information. Les pays émergents n'ont qu'à choisir ceux où ils y rajouteront les ingrédients culturels et linguistiques, en fonction de leurs moyens et disponibilités, pour les assaisonner à leurs goûts et à leurs arômes. Des initiatives comme les archives ouvertes ^[18], avec tous leurs potentiels éditoriaux et diffusionnistes de l'information scientifique et technique, sont des alternatives jouables et immédiates qui peuvent rehausser la contribution des pays du Sud, dans toutes leurs diversités culturelles et linguistiques, pour diffuser un patrimoine scientifique local et l'intégrer aux réservoirs universels de la connaissance. Des actions de normalisation courantes, comme pour la formation ouverte et à distance autours du SC36, pour les archives numériques autours de la normes DTD-EAD d'encodage des instruments de recherche archivistiques [DHERENT, 2000], pour les archives ouvertes autours du protocoles de collecte de métadonnées OAI-PMH [FOULONNEAU, 2003], sont autant de chantiers en cours de développement qui n'enregistrent malheureusement pas de contribution notoire de la part des acteurs du Sud.

On connaît, certes, tous les problèmes inhérents à l'inégalité d'accès aux technologies, traduite par des inégalités importantes d'accès aux infrastructures de télécommunications, aux équipements informatiques dans les universités, dans les entreprises et pour les particuliers. On connaît bien aussi que l'inégalité des accès est souvent accentuée par des problèmes d'illettrisme informatique. Mais, on doit admettre aussi qu'on ne peut imposer aux acteurs réels de la société de l'information, généralement situés au nord, de veiller aux petits soins de l'adaptabilité culturelle et linguistique universelle de leurs projets et programmes impliquant des langues et des cultures différentes, et à plus forte raison dans les technologies de l'information. Une grande part de responsabilité reste endogène, car la dynamique ne peut être appréhendée autrement qu'à travers une volonté politique locale de pousser vers les choix les plus porteurs inscrits dans une vision de stratégies ou de politiques nationales, régionale et internationales d'information.

.../...

[18] Site officiel de l'Initiative des Archives Ouvertes. [<http://www.openarchives.org/>]

Bibliographie

1. ANDRE Jacques et GOOSSENS Michel. Codage des caractères et multilinguisme : de l'ASCII à Unicode et ISO/IEC-10646. In : Cahiers GUTenberg n°20-mai 1995.pp.1-53
2. BEN HENDA Mokhtar, ZGHIBI Rachid, Normalisation et EAD face aux contraintes linguistiques et culturelles de l'I18n dans le contexte arabophone. Symposium de Versailles, France. 18 mars 2003. [http://www.initiatives.refer.org/Initiatives-2003/_notes/_notes/benhendazghibi.htm]
3. BERLAND Sophie. *Constitution de corpus à partir du Web pour l'acquisition terminologique : une expérience*. Sous la direction de Benoît Habert. Mémoire de DESS en Ingénierie Multilingue, CRIM/INaLCO, année universitaire 1999 - 2000
4. BLANDIN Bernard. Normes : définitions et enjeux. FFFOD-CESI, juin 2003
5. BORDERIE Xavier. Le Web sémantique. Revue JDN Développeur. 18 novembre 2002
6. BOURIGAULT Didier., SLODZIAN Monique. Pour une terminologie textuelle. In *Terminologie et Intelligence Artificielle, actes du colloque de Nantes, 10-11 mai 1999*, Nantes. Tutoriel à TIA 1999, paru dans Terminologies nouvelles n°19
7. BRETON Philippe. Une histoire de l'informatique. La Découverte, 1990
8. CABRE Térésa. Terminologie et linguistique : la théorie des portes. Revue Terminologie et diversité culturelle. Rifal, Montréal. N°21, juin 2000. Pp. 10-15
9. CAUSSANEL Jean, CAHIER Jean-Pierre, ZACKLAD Manuel. Etude des interactions cognitives dans l'ingénierie du Web Sémantique. Contribution du laboratoire TEC-CICO à l'AS Web sémantique. Mai 2003. [<http://www.lalic.paris4.sorbonne.fr/stic/data/Caussanel.htm>]
10. DA SYLVA Lyne. Technologies facilitatrices pour la dissémination de la communication scientifique multilingue. EBSI, Université de Montréal, 2003
11. DHERENT Cathénine. La normalisation dans les archives. ARCHIVES, volume 31, N°3, 1999-2000
12. FIETZER William. Integrating Metadata Frameworks into Library Description In : Libraries, the Internet, and Scholarship. Tools and Trends Converging. New York ; Bâle : Marcel Dekker, 2002
13. FOULONNEAU Muriel. Le protocole OAI-PMH : une opportunité pour le patrimoine numérique. Relais-Culture-EuropeMission de la recherche et de la technologie. Janvier 2003
14. HEFLIN Jeff. Les cas et conditions d'utilisation du langage d'ontologie Web OWL. Recommandation du W3C du 10 février 2004
15. HUDRISIER Henri. E-Learning et normalisation : définitions, enjeux et contexte. In : *Enseignement ouvert et à distance : épistémologie et usages*. Lavoisier, 2004, pp. 203-225
16. JOBERT Gabrielle Kahina. Application des Réseaux et de la Télématic. Mémoire de DESS dirigé par Gilbert Sol. Université Denis Diderot, 2003
17. LAPALME Guy, MACKLOVITCH Elliott. Compréhension et multilinguisme. *Compréhension automatique des langues et interaction, séries. Cognition et traitement de l'information*, 3.6, 18 p. Hermes Science, Paris, décembre 2005
18. PEPPER Steve (éd.), MOORE Graham (éd). XML Topic Maps (XTM) 1.0. Topic Maps.Org. 2001. [<http://www.topicmaps.org/xtm/1.0/xtm1-20010806.html>]
19. RASTIER François. Ontologies. *Revue des sciences et technologies de l'information*, série : *Revue d'Intelligence artificielle*, 2004, vol. 18, n°1, p. 15-40
20. SAILLANT Jean Michel. L'impact de la normalisation sur les dispositifs d'enseignement. Rapport du groupe Normes et standards de GEMME. Février 2002
21. SALAM Ismail. Normalisation et standardisation dans les technologies de l'information : des enjeux linguistiques déterminants. La Tribune des industries de la langue et de l'information électronique, n° 20-21-22 : Du traitement à la création de documents.
22. YANG Jian. Comment construire une banque de terminologie véritablement multilingue. Actes du colloque : L'impact des nouvelles technologies sur la gestion terminologique. Collège universitaire Glendon, Université York. Toronto (Ontario) Canada. 18 août 2001.
23. ZWEIGENBAUM Pierre. Encoder l'information médicale : des terminologies aux systèmes de représentation des connaissances. *ISIS*, (2-3), 1999. Pp. 27-47