

Cartago, une base de donnée terminologique multilingue « large langues » dédiée à l'élaboration des normes de l'e-Learning.

Henri Hudrisier* ; Mokhtar Ben Henda**

* Maître de conférences, Liaison A de l'AUF à l'ISO
Laboratoire Paragraphe, Université de Paris 8

** Maître de conférences, Liaison A de l'AUF à l'ISO
CEMIC-GRESIC, Université de Bordeaux 3

Résumé :

Dans cet article, nous présentons le cadre de synergie qui s'est mis en œuvre pour définir les termes et les concepts des technologies de l'information et de la communication pour l'enseignement (TICE) avec pour objectif d'usage prioritaire les experts en normalisation. Nous décrivons le contexte particulier de réalisation de ces terminologies, les choix méthodologiques (méthode onomasiologique ou simple recueil de vocabulaire), deux ou trois langues de travail ou la volonté d'un multilinguisme extensif (large langues). Nous situons le rôle de l'Agence universitaire de la Francophonie (AUF), de sa Liaison à l'ISO et plus particulièrement sa participation à la définition des normes des TICE comme un enjeu primordial pour préserver la diversité culturelle et linguistique dans l'échange numérique du savoir. Nous insisterons particulièrement sur l'Alliance Cartago qui répond aux exigences les plus larges : multilinguisme extensif, méthode onomasiologique et normalisation des principes et méthodes terminologique et terminotique, choix d'un modèle applicatif du Terminological Markup Framework (TMF) adapté à nos besoins. Nous ouvrons aussi sur des questionnements prospectifs tant techniques que d'appropriation d'usage.

Mots clés : Terminologie ; normalisation ; e-Learning ; multilinguisme

Abstract:

In this paper, we present the synergy framework that has been implemented to define the terms and concepts of Information and Communication Technologies for Education (ICTE) with standards experts as main community of practices. We describe the particular context of achieving these terminologies, the methodological choices (Onomasiologic method or mere compilation of vocabulary), two or three working languages or the will of a multilingual extensive (broad language). We consider the role of the Agency of French Speaking Universities (AUF), its Liaison with ISO and particularly its participation in the definition of standards of ICT, as a leading issue for the preservation of cultural and linguistic diversity in the digital exchange of knowledge. We will insist particularly on the Cartago Alliance that meets the largest requirements: extensive multilingualism, onomasiologic method and standardization of terminology principles and methods, choosing the application model of Terminological Markup Framework (TMF) tailored to our needs. We address also prospective questions of both technical and usage considerations.

Keywords: Terminology; standardization; e-Learning; multilingualism

1. Introduction

Un grand nombre de pré-requis sont indispensables aujourd'hui pour comprendre le contexte de création des ressources terminologiques. De l'usage prévu, découlent des choix de méthode terminologique ⁽¹⁾ et des principes éthiques notamment pour ce qui concerne la préservation extensive de la diversité linguistique, ou large langues, et le principe d'éviter au maximum la marchandisation des concepts fondamentaux de l'éducation. Ces principes éthiques correspondent aux besoins de la Francophonie de préserver la diversité de ses langues et cultures partenaires mais ils sont aussi conformes au respect de la diversité culturelle tel qu'il a été défini par l'UNESCO. La non marchandisation des concepts de base de l'éducation, qui sont un bien public mondial inaliénable, dans le contexte numérique constitue aussi un enjeu primordial pour tous les pays notamment ceux du Sud confrontés aux risques de la fracture numérique.

Nous décrivons plus loin ce qu'est la normalisation des TICE, mais soulignons déjà que comme la langue d'Ésope, elle peut être la meilleure et la pire des choses. Elle pourrait être soit une cause d'exclusion de toute possibilité d'expression de la diversité d'offre ou d'usage des TICE, soit, si nous savons participer à la définition et à la construction de ces normes, elle pourrait être au contraire un moyen de défendre et de préserver la diversité dans ces technologies de la transmission du savoir. Soulignons le fait que la diversité a de nombreuses facettes. Elle peut être celles de la diversité des territoires et des institutions éducatives ou de formation, celles des cultures pédagogiques, des nations, des économies, des langues ou des écritures. C'est aussi la diversité des métiers, des disciplines scientifiques, des publics à former, des styles et des modalités pédagogiques.

Dans tous les cas de figures, la prise en compte de ces diversités ne sera pas définie comme contrainte normative pour être ensuite instrumentalisées dans des TICE effectivement développées que si nous savons construire un référentiel sémantique, en premier lieu, une terminologie largement multilingue de référence, associée à son tour au développement d'une ontologie, susceptible d'exprimer et de rendre communicable au monde entier la réalité de cette diversité.

L'Alliance Cartago que nous présentons dans ce papier, s'appuie sur un potentiel de synergie mondiale à travers l'ISO : plus précisément le sous comité 26 (SC36) et ses experts représentant 29 pays ⁽²⁾ et beaucoup plus quand on prend en compte une délégation comme

¹ A titre d'exemple, la méthode onomasiologique se base sur un strict respect des normes terminologiques du sous comité 37 (TC37) de l'ISO

² Pays membres (dont certains sont observateurs – *O Member*): Algérie (O); Allemagne; Australie; Autriche (O);

l'AUF qui s'appuie sur l'ensemble des pays francophones.

Si nous considérons les langues et les écritures, la diversité des communautés qui délèguent ces experts est aussi grande que celle des nations: allemand, anglais, chinois, coréen, danois, espagnol, finlandais, français, grec, italien, japonais, malais, néerlandais, norvégien, russe, suédois, tchèque, turc et ukrainien. On remarquera que cette liste prend en compte un nombre minimum de langues présentes dans les seuls états membres.

2. Le contexte numérique de la transmission du savoir : enjeux des normes pour défendre la diversité

La transmission du savoir est d'année en année toujours plus dépendante de la mondialisation numérique. Soit l'enseignement ou la formation sont intégralement médiatisés par des moyens numériques, soit, et c'est le cas le plus général, s'installe un mélange de présentiel et d'e-enseignement. Si l'on admet que l'éducation, l'enseignement et la formation sont des enjeux primordiaux pour le développement de toutes les sociétés, il est de ce fait très stratégique de s'assurer que nos sociétés respectives sauront préserver leur capacité d'accès dans leur langues et selon leurs critères culturels propres aux réseaux d'échange de savoir et aux TICE. Il faudrait s'assurer qu'elles sauront aussi préserver dans leur langue une complétude de tous les champs linguistiques se rapportant à ces technologies comme meilleur garant, on le sait, pour préserver un savoir faire industriel et éditorial de ces technologies et de ces ressources numériques. Remarquons aussi que ce que l'on appelle les TICE (Technologies de l'Information et de la Communication pour l'Enseignement), ne se limitent pas à l'e-enseignement. Les campus numérique et l'offre numérique en ligne d'enseignement deviennent aussi des composantes majeures de la transformation numérique de l'enseignement et de sa mondialisation.

L'exigence de multilinguisme systématique des offres d'enseignement peut sembler exorbitante au premier abord, mais quand on considère les efforts humains et économiques que toute communauté consacre à l'acquisition de savoir, les efforts de mise en visibilité multilingue mondiale sont comparativement peu importants. Encore faut-il que ce travail de localisation globale de l'offre soit réellement interopérable, donc normalisé et que la sémantique de cette offre globale soit conceptuellement cohérente ce qui implique un chantier terminologique et ontologique à la mesure des ambitions. C'est un des buts de l'Alliance Cartago.

Canada; Chine; République tchèque; Danemark; Finlande; France; Grèce; Irlande; Italie; Japon; Kenya; Malaisie; Hollande; Nouvelle Zélande; Norvège; Russie (O); Singapour (O); Espagne; Corée du Sud; Suède; Suisse; Turquie (O); Ukraine; Grande Bretagne; USA.

3. La terminologie comme outil et activité sémantiques

Les deux auteurs de cet article ne se prétendent aucunement terminologues ou terminoticiens. Cependant nous revendiquons une légitimité académique pour ce qui pourrait se définir comme l'étude de l'appropriation des méthodes de la terminologie et des outils terminotiques normalisés. En effet, nous entrons dans l'âge du Web sémantique, ce qui signifie que des volumes de ressources de plus en plus vastes du réseau Internet répondront par construction à une logique sémantique des concepts. Dans ces sites «labelisés sémantiques» les termes sur lesquels pourront s'effectuer des recherches seront garantis correspondre à des termes inclus dans des terminologies multilingues normalisées associés dans les ontologies elles même normalisées.

Quand une masse critique suffisante de sites constitueront un Web sémantique réellement opératoire, les terminologues et les terminoticiens seront alors confrontés à une évolution de leur devenir professionnel semblable à celle qu'ont connu 15 ans avant eux une activité voisine : les documentalistes et les bibliothécaires. A partir des années 90, grâce à l'interconnexion mondiale des réseaux puis à la généralisation du «web référentiellement cohérent»⁽³⁾, les documentalistes ont dû admettre que leurs professions avaient évoluées. Tout usager averti du Web pouvait dès lors prétendre à s'approprier les méthodes documentaires, tant pour mettre en ligne des ressources qu'il doit documentairement aménager dès le dépôt sur le Web, que pour consulter le Web avec son navigateur qui fait l'office d'un automate documentaire. La valeur ajoutée des documentalistes ne réside plus dans leur savoir faire traditionnel mais dans deux pistes d'évolution professionnelles : la transmission des savoir faire vers des producteurs ou des utilisateurs de ressources sur le Web ou la recherche-développement en ingénierie documentaire.

Quinze ans après, c'est au tour des terminologues d'occuper le devant de la scène. Leur activité était relativement confidentielle dans les années 90. Maintenant, c'est une activité de plus en plus demandée et il est vraisemblable que comme pour les professions documentaires, les terminologues et les spécialistes des ontologies vont connaître une évolution similaire. Leur savoir faire se banalisera. Ils devront aider les créateurs et les utilisateurs des ressources terminologiques ou ontologiques à s'approprier leurs méthodes et leurs outils. C'est à ce moment seulement que l'on pourra réellement parler du Web sémantique : quand un grand nombre de producteurs et d'utilisateurs de ressources de documents

³ Le Web des années 95 est formellement cohérent (on peut maîtriser la structure de sa mise en page) et référentiellement cohérent (on peut identifier une adresse, une ressource bibliographique de façon certaine), mais il n'est que très faiblement sémantiquement cohérent.

se seront appropriés les savoir-faire terminotiques et ontologiques de base et que cette culture terminologique et sémantique se sera banalisée dans la réalité sociale du système technique informationnel.

C'est la raison pour laquelle cet article fera référence en priorité à deux instances de normalisation : d'une part le SC36 ⁽⁴⁾, sous-comité d'experts dans lequel s'élabore les normes des TICE ⁽⁵⁾, et d'autre part l'ISO TC37, le Comité Technique (Technical Committee) dans lequel se développent les normes de la terminologie, de la terminotique et des ressources linguistiques.

Nous pensons aussi que le devenir de l'AUF est hautement dépendant de la maîtrise sémantique multilingue des réseaux du savoir et ce dans deux directions :

- Celle des langues partenaires de la Francophonie (arabe, langues de la péninsule indochinoise, langues africaines et océaniques) en direction desquelles il faut savoir développer des grandes terminologies multilingues sans se soucier de l'importance économique apparente de ces langues ;
- Participer aussi à des grands projets de développement de terminologies multilingues avec les autres langues du monde sans distinction d'écritures et sans tenir compte des anciennes notoriétés des langues du monde : terminologies qui seront fondamentales pour assurer l'interopérabilité et la réusabilité des TICE ainsi que la circulation des ressources pédagogiques dans l'environnement émergent du Web sémantique.

4. TICE et référentiel terminologique multilingue

L'exposé que nous développons ici exige quelques pré-requis minima. Beaucoup d'universitaires s'offusquent souvent que l'on veuille normaliser les TICE. Ils croient à tort que normaliser les TICE reviendrait à unifier les contenus, unifier les styles pédagogiques et les modes d'évaluation. C'est exactement le contraire qui est en cause : la normalisation est destinée à prévoir la grande diversité des plates-formes techniques, des logiciels, des contenus, des niveaux des apprenants... et à créer un espace d'interopérabilité, de réusabilité, de facilitation des localisations multilingues. On comprend dès lors que si nous ne participons pas à la normalisation des TICE, si nous n'imposons pas dans toutes les commissions de travail du SC36 la prise en compte de nos exigences dans les particularismes linguistiques,

⁴ Sigle court de l'ISO/IEC JTC1 SC36 : l'instance de normalisation des technologies de l'information et la communication pour l'enseignement et la formation. Les deux auteurs de la présente communication y représentent l'AUF (Liaison A) et sont aussi membre des délégations de l'AFNOR (Association Française de Normalisation)

⁵ TIC : Technologies de l'Information et de la Communication ; TICE : Technologies de l'Information et de la Communication pour l'Enseignement

culturels, institutionnels, scripturaux (la prise en compte d'écritures spécifiques), disciplinaire⁽⁶⁾ il est alors certain que la fracture numérique (notamment Nord-sud) s'aggravera notablement.

En effet, la mondialisation installe de fait une circulation des ressources d'enseignement, leur marchandisation, la mise en place aussi d'une offre d'enseignement en ligne hautement concurrentielle. Si les différentes communautés nationales ou linguistiques (par exemple l'AUF) ne savent pas répondre aux défis posés par les acteurs les plus avancés de cette compétition mondiale, alors les risques de domination unilingue anglophone deviendront une réalité très difficilement réversible. Ce risque de domination d'une communauté linguistique sur toutes les autres⁽⁷⁾ est la conséquence de ce que la totalité du système technique des TIC (qui inclus les TICE) est soumis à une dynamique de convergence et d'interopérabilité fantastique que nous appelons nous même de nos vœux. Ainsi nous nous étonnons quotidiennement, nous sommes même exaspérés, que certains objets techniques ou logiciels ne soient pas compatibles les uns avec les autres⁽⁸⁾.

Cette dynamique de convergence et d'interopérabilité qui se développe exponentiellement dans la bureautique, les traitements de texte ou l'audiovisuel se développe aussi naturellement pour les TICE. Il ne s'agit donc pas comme certains académiques frileux pourraient le croire de normaliser les contenus d'enseignement ou de formation ; ni de normaliser les processus d'apprentissage ou les modalités d'évaluation pour les ramener à des modalités simplifiées et univoques, mais au contraire de définir de façon normative un maximum de processus, de contraintes réseaux, de caractéristiques des composants, des logiciels. Il s'agit aussi de normaliser la structure modulaire des contenus pour pouvoir les recomposer en les adaptant à des publics disparates. Il est aussi question de normaliser les modes d'accès documentaires aux ressources pédagogiques pour en permettre l'interopérabilité, donc la réusabilité et la circulation facile. Ceci constitue la meilleure garantie pour préserver une diversité des offres, une diversité des ressources en toutes langues, correspondant aux spécificités pédagogiques qui, comme les spécificités culturelles, doivent être reconnues et instrumentées numériquement pour pouvoir être imposés dans les

⁶ Les institutions et les rôles académiques sont différents selon les pays et les cultures mais les disciplines (ainsi que les métiers) sont conceptuellement découpés de façon parfois très diversifiées.

⁷ Pour ce qui est de la production de savoir numérique ainsi d'ailleurs que les produits culturels (notamment audiovisuel. Voir Tardy (Jean) & Farchy (Joëlle). Les enjeux de la mondialisation culturelle. Paris : Ed. Hors Commerce, 2006

⁸ Pourquoi une photographie ou un film que nous venons de recevoir par courriel ne s'ouvre pas dans notre système de messagerie ? Pourquoi je ne peux pas imprimer une image ou un texte que je viens d'ouvrir et de copier-coller dans une présentation de cours électronique que je suis en train de concevoir ? Pourquoi mon tel mobile européen ne fonctionne pas en Corée ?

cahiers des charges des futures offres des TICE. C'est en nous mettant au niveau des pays les plus avancés pour les TICE et en imposant par les normes nos propres contraintes (notamment linguistiques) que nous saurons défendre nos communautés linguistiques contre tout risque de fracture numérique ⁽⁹⁾.

5. Une terminologie des TICE au SC36

Pour mieux comprendre le fonctionnement du SC36, il faut savoir que les grandes organisations de normalisation comme l'ISO, l'IEC ⁽¹⁰⁾ ou l'UIT ⁽¹¹⁾ se divisent en Comités techniques (TC : Technical Committee) chacun responsable d'un métier ou d'une activité. On trouve par exemple à l'ISO le TC46 pour la documentation ⁽¹²⁾ et le TC37 pour la terminologie ⁽¹³⁾.

Entre l'IEC, vieille institution internationale préexistante à l'ISO ⁽¹⁴⁾ pour la normalisation des technologies électriques et électroniques, et l'ISO, qui dans ce domaine s'intéressait plutôt aux processus et aux services logiciels, a été créé le *Joint Technical Committee N°1* ⁽¹⁵⁾. C'est dans ce JTC1 que s'est développée la normalisation de très nombreux domaines des TIC et du multimédia. Par symétrie structurelle avec les comités techniques de l'ISO et de l'IEC qui se divisent en Sous Comités (SC), le JTC1 s'est lui aussi découpé en SC, mais dans ce très vaste comité technique spécial que constitue le JTC1, se sont les SC qui constituent chacun l'instance de normalisation d'un métier ou d'une activité

Depuis 2000, l'ISO a mis en place un groupe d'experts, le SC36 ⁽¹⁶⁾ qui a pour objectif d'élaborer les normes des TICE. Comme la quasi-totalité des instances de normalisation, un des premiers soucis du SC36 a consisté à créer dès mars 2000 un sous groupe de travail chargé de rassembler une terminologie ou un vocabulaire ⁽¹⁷⁾ du domaine.

⁹ OILLO (Didier), MVE-ONDO (Bonaventure) Dir. : 2006, Fractures dans la société de connaissances. (Paris : Hermès n° 45, CNRS éditions).

¹⁰ IEC, International Electrotechnical Commission, en français : CEI, Commission Électronique Internationale

¹¹ Union Internationale des Télécommunications (en anglais ITU : International Telecommunications Union). Contrairement à l'ISO et l'IEC les TC se nomment TG (Technical Group) à l'UIT

¹² En sigle long ISO TC 46

¹³ En sigle long ISO TC 37

¹⁴ L'IEC a été fondée en 1906, l'ISO en 1947

¹⁵ Le JTC 1 est le seul JTC.

¹⁶ Le SC36 est structuré en groupes de travaux : WG1 : vocabulary ; WG2 : Collaborative technology; WG3 Learner information; WG4 : Management & delivery of learning education & training; WG5 : Quality assurance and descriptive frameworks; WG6 : International standardized profiles; WG7 : Culture, language & human-functioning activities; SWG (Special Working Group) : Business Planning; RG1: Marketing Rapporteur Group n°1; Privacy ad Hoc Group : non encore formalisé par un sigle.

¹⁷ La question n'est pas tranchée. D'évidence la spécificité de l'activité (ce n'est pas un secteur de normalisation strictement technique) plaiderait pour la constitution d'une terminologie.

Dans la famille de normes développées par cette instance⁽¹⁸⁾, une norme de vocabulaire du domaine, référencée ISO/IEC 2382-36, a été mise en chantier dès l'origine et nous en sommes à la 2^e édition. En fait nous le verrons plus loin, ISO/IEC 2382-36 est un sous ensemble d'une norme plus vaste ISO/IEC 2382 qui constitue le vocabulaire commun du JTC1. Le suffixe « -36 » précise bien qu'il s'agit du vocabulaire du SC36.

Ce groupe de travail tient compte de plusieurs synergies externes notamment : a) La mise en commun des vocabulaires de tous les Sous Comités du JTC1, b) La création de l'Alliance Cartago et le constat de son efficacité pour produire des ressources multilingues correspondant à un schéma XML normalisé TMF⁽¹⁹⁾ ; ce qui d'ailleurs a motivé la mise en place d'une liaison de catégorie C entre l'Alliance Cartago et le groupe de travail WG1 du SC36.

Partant du fait que le JTC1 a fondé successivement ses SCs, on comprend aisément que le SC36, l'un des derniers fondés, achoppe maintenant sur des besoins terminologiques beaucoup moins strictement techniques que la majorité des autres SCs. Cependant, la décision de regrouper tous les vocabulaires du JTC1 dans un seul et même ensemble commun pose des problèmes complexes de recoupement des mêmes concepts selon des « sémantiques métiers distinctes » qui ne seront pas forcément simples à gérer dans le futur. C'est la base terminologique *Termium*⁽²⁰⁾ qui a offert ses services pour répondre à ce besoin.

Le SC36 se conforme naturellement à cette logique de mise en commun de son vocabulaire normalisé. Ce vocabulaire, pour le moment assez réduit (une cinquantaine de termes), est établi sur deux listes (anglais et français) en principes co-établies en parallèles et organisés en 7 domaines⁽²¹⁾. On perçoit bien que la co-rédaction bilingue de ces listes ne pose pas d'énormes problèmes tant que le nombre de termes est réduit et tant que l'on travaille sur des langues uniques assez voisines comme l'anglais et le français. Notons aussi que la réalisation de vocabulaire fonctionne assez bien dans les domaines assez techniques comme les systèmes et outils, les processus et les ressources et contenus. Cela fonctionne

¹⁸ Une instance normative responsable d'un métier ou d'une discipline élabore toujours une famille de normes c'est-à-dire les normes référant à cette activité.

¹⁹ Un cadre de mise en œuvre normalisé pour la terminotique, lui-même lié au TML (*Terminological Markup Language*), un langage XML spécialisé (en fait une DTD) pour l'expression des données terminologique

²⁰ C'est la base de données terminologique *Termium* du Bureau fédéral de la traduction canadien. Celle-ci a offert ses services au JTC1 pour collecter et devenir le lieu de référence de la terminologie de tous les sous-comités qui en dépendent (y compris la terminologie e-Learning du SC36). La base *Termium*, qui gère un fonds très important de ressources terminologiques est cependant organisée prioritairement pour répondre aux besoins de traduction d'un pays bilingue. De ce fait, elle propose aujourd'hui des terminologies en français et en anglais. A moyen terme elle est prévue pour être opérationnelle en 4 ou 5 langues au plus.

²¹ Respectivement : 1 Termes généraux 2 Utilisateurs, organisations et rôles 3 Systèmes et outils 4 Processus 5 Ressources et contenus 6 Enseignement et apprentissage 7 Information sur l'apprenant

beaucoup moins bien pour les domaines des utilisateurs, organisations et rôles, des enseignements et apprentissages et de l'information sur l'apprenant qui sont des domaines très dépendants des institutions, des modes d'organisation des cultures et du point de vue pédagogiques voire éthique. Cela dysfonctionne totalement quand on s'attaque aux termes généraux (domaine 1) qui sont eux très dépendants des langues et des cultures.

6. Des méthodes terminologiques et terminotiques complexes

Sur un plan pratique, le SC36 a mis longtemps, à travers son premier groupe de travail (WG1) spécialisé dans le vocabulaire e-Learning, à préciser ses méthodes d'action. On a pu constater, en effet, un certain écart entre les besoins réels à terme, les habitudes des groupes de normalisation habituellement beaucoup plus strictement techniques alors que l'enseignement est *a priori* une activité hautement complexe, grandement dépendante des langues, des cultures, des institutions nationales, des disciplines ou des métiers concernés. Cet écart aurait pu être comblé par une capitalisation des acquis réels et semblables réalisés dans d'autres structures de normalisation si les mécanismes connus par le concept de liaison ont été plus efficaces. Dans le cadre de la terminologie e-Learning, paradoxalement, les conseils de la liaison technique TC37 auprès du SC36 n'ont pas été pris en compte.

Il faudrait peut être rappeler à ce sujet que toute instance de normalisation a pour objet de formaliser de façon très précise les caractéristiques principales de produits et de services ainsi que leur *process* de production, voire d'usage, pour assurer, grâce à la normalisation de ces caractéristiques l'interopérabilité, la modularité, la réusabilité, l'intercompatibilité entre des marques différentes, la sécurité d'emploi, le respect de normes écologiques, éthiques, etc... Les experts normalisateurs devraient être de ce fait des savants omniscients. Pour répondre à cela, toute instance de normalisation a le devoir absolu d'entretenir des liaisons techniques avec d'autres instances de normalisation qui recoupent leurs préoccupations ; ce qui leur évite, en principe, de réinventer et de re-normaliser ce que d'autres experts ont déjà normalisé. Pour ce, le SC36 entretient entre autres des liaisons avec le TC46 (documentation), avec MPEG (pour les ressources audiovisuelles), etc... Le SC36 active aussi une liaison avec le TC37 (Terminologie). Depuis 2008, se met en place une liaison technique de catégorie C entre l'Alliance Cartago et le WG1. Ceci est de nature à permettre d'imposer plus nettement les méthodes normalisées de la terminologie qu'avec le TC37 qui bien que « liaison » est de fait absent dans la quasi-totalité des plénières du SC36 et des séances de travail de son WG1. Les Liaisons peuvent être aussi non seulement techniques mais institutionnelle. C'est à cette catégorie de liaison qu'appartient la Liaison catégorie A de l'AUF avec le SC36.

Cependant à partir de 2005, notamment avec la création de l'Alliance Cartago, la quasi majorité des experts ont admis :

- 1- que leur groupe de travail avait souffert d'avoir réinventé ses propres méthodes terminologiques,
- 2- que les besoins terminologiques du SC36 dépassaient de beaucoup les besoins plus strictement techniques de la plupart des autres sous-comités du JTC1,
- 3- qu'il fallait réaliser une terminologie beaucoup plus largement multilingue. A l'évidence la normalisation des TICE impose une prise en compte des contextes institutionnels, linguistiques, culturels et pour ce faire on ne peut pas, comme dans d'autres activités plus industrielles, se contenter de listes de vocabulaires établies dans les deux premières langues officielles de l'ISO, à savoir l'anglais et le français⁽²²⁾.
- 4- que les démonstrations auprès du WG1 des premières maquettes de Cartago, réalisées grâce à Génétrix, ont été convaincantes. La majorité des experts sont maintenant persuadés que des terminologies très multilingues, devant atteindre à terme des masses assez conséquentes d'*entrées terminologiques* ne pouvaient pas se réaliser en alignant successivement des termes dans un simple fichier Word. Les experts ont été notamment impressionnés par le potentiel éditorial de ressources terminologiques stockées dans un schéma XML correspondant à une application TMF (*Terminological Markup Framework*). Nous leur avons ainsi démontré que l'Alliance Cartago dispose de ressources terminologiques organisées selon la méthode onomasiologique, présentant en parallèle et de façon cohérente des termes dans 5 ou 6 langues (anglais, français, coréen, arabe, espagnol, berbère, malgache...) et qui pourraient cependant générer à la demande des listes de vocabulaire répondant strictement à la norme ISO/IEC 2382-36.
- 5- que la réalisation d'un tel chantier ne pouvait pas, tout au moins dans un premier temps, être entreprise dans le cadre d'un groupe de travail ayant à l'origine des objectifs beaucoup plus limités.
- 6- qu'il pouvait y avoir une complémentarité entre la réalisation de listes parallèles multilingues de vocabulaires (des listes de termes) dans le cadre strict du SC36 WG1 et la réalisation d'une véritable terminologie du domaine, organisée selon des principes terminologique et terminotiques plus rigoureux dans un nombre extensif de langues dans le cadre de l'Alliance Cartago. Ce constat a été formalisé dans des résolutions du SC36 lors de sa dernière plénière de Septembre 2007 à Toronto pour instaurer un cadre institutionnel de

²² La Russie n'a que très tardivement envoyée une délégation au SC36.

collaboration entre le SC36 et l'Alliance Cartago.

7. L'Alliance Cartago

L'Alliance Cartago est une initiative fondée par des experts normalisateurs des TICE réunis à Tunis dans le cadre du SMSI⁽²³⁾ et plus particulièrement au cours d'un Open forum élaboré en commun par l'AUF⁽²⁴⁾ et le SC36. En février 2008, l'Alliance Cartago est devenue une Association Loi 1901⁽²⁵⁾. Cette Alliance Cartago a pour objectif de développer une base de données terminologiques multilingues normalisées « large langues »⁽²⁶⁾ car nous estimons que le TMF (*Terminological Markup Framework*) permet désormais la réalisation de bases terminologiques qui devraient pouvoir accepter un nombre extensif de langues à condition de se conformer strictement à un schéma terminotique normalisé. Par « large langues » nous voulons induire qu'à l'ère du Web sémantique les terminologies multilingues doivent résolument changer de nature et d'échelle. Elles ne peuvent plus être, comme c'est le plus souvent le cas aujourd'hui, des terminologies comportant une demi-douzaine de langues. Dans le cas qui nous concerne, celui de Cartago, nous pensons pouvoir être capables de gérer les concepts fondamentaux des TICE dans le plus grand nombre de langues possibles, notamment dans les langues partenaires de la Francophonie et dans de nombreuses langues du monde, car nous savons que c'est l'une des conditions préalables à la protection de la diversité culturelle, linguistique, scripturale, disciplinaire et académique confrontée à la mondialisation numérique.

L'Alliance Cartago s'est donc fondée pour répondre au désir de certains participants du groupe de travail SC36 WG1 de disposer d'un référentiel terminologique sur les TICE, largement multilingue et à même de documenter les différents domaines indispensables pour construire le cadre d'interopérabilité des TICE. Ce besoin exprimé par une minorité agissante de experts du WG1 particulièrement intéressée par le multilinguisme (mais aussi la pluralité des écritures) a rencontré la volonté personnelle des deux auteurs de cet article de s'impliquer et d'impliquer l'AUF et une équipe projet dans la réalisation d'un tel référentiel terminologique sur les TICE indispensable à la mise en synergie des différentes langues partenaires de la Francophonie avec le français. Les experts coréens du SC36 ont été, avec d'autres (Australie, Canada, Allemagne, Finlande), associés dès le départ dans la fondation du

²³ SMSI Sommet Mondial pour la Société d'information, Genève 2003, Tunis 2005.

²⁴ Colloque Initiatives 2005 : la norme comme instrument de réussite d'une Société de la connaissance partagée. Kram - Tunis, Tunisie. 14-15 novembre 2005 (en ligne sur le site de l'AUF).

²⁵ Alliance Cartago, MDA, 23, rue Vernet 75008 Paris, <http://www.cartago-alliance.org/>

²⁶ Nous dérivons ce mot valise du mot « large bande » qui s'emploie en électronique pour qualifier notamment des émetteurs ou des récepteurs hertziens capable de travailler sur de large bandes de fréquences.

noyau fondateur à Tunis pendant le SMSI de novembre 2005. L'Union Latine et bien sûr de nombreux chercheurs francophones ont aussi participé à la création de Cartago.

A ce jour, nous savons que, dans le cadre de Cartago, nous pouvons répondre efficacement à des questions de méthode tant dans le recueil que dans la mise en place d'un système terminotique parfaitement normalisé. La réalisation de nos premières maquettes de ressources et de serveur sur Génétrix le démontre parfaitement. Nous avons aussi éclairci les rapports de synergie entre l'ISO SC36 et l'Alliance Cartago. Nous sommes en phase de consolidation du statut institutionnel de cette Alliance qui ne saurait rester de la responsabilité exclusive de l'AUF et de la MSH Paris Nord qui lui ont servi de cadre de développement à ce jour : le contexte francophone est trop restrictif pour accueillir la pluralité mondiale des langues dans le contexte des TICE qui est le projet de l'Alliance Cartago. La multiplication des langues dans Cartago et la mise en chantier de nombreuses autres est pour nous un facteur de réussite mais nous confronte aussi à la difficile gestion de la croissance et de la capacité technique et humaine pour gérer des masses importantes de ressources terminologiques.

8. Objectifs, outils et méthodes pour une terminologie multilingue

Pour atteindre cet objectif de démultiplication linguistique de la terminologie e-Learning, des mesures méthodologiques sont à considérer comme facteurs essentiels de réussite. Il s'agit d'une part de clarifier un système de concepts essentiel pour s'assurer de la compréhension commune par les experts et les utilisateurs futurs des normes du système technique, du réseau social ou institutionnel et des contenus pédagogiques que l'on cherche à rendre interopérables au SC36. Il s'agit d'autre part de s'assurer une mise en parallèle multilingue de ces concepts pour des groupes d'experts travaillant dans un contexte multilingue international.

8.1 Un système de concept cohérent

A supposer que les experts rassemblés pour développer des normes pratiquent tous la même langue maternelle, ces experts ont néanmoins besoin d'être certains qu'ils travaillent bien sur les mêmes notions et sur les mêmes concepts. Dans une instance nationale de normalisation, à l'AFNOR par exemple, on partage une même langue ; mais cela cesse d'être le cas dans toutes les instances internationales comme l'ISO par exemple. Dans ce cas, même si les experts délégués parlent tous au minimum un anglais basique (le Globish), il reste que les experts anglophones, ou ceux des pays où la totalité de l'enseignement supérieur et de l'ingénierie se fait exclusivement en anglais, sont considérablement avantagés. Il faut noter aussi que dans ces discussions, la question de la langue n'est pas tant celle de la

compréhension des mots du langage courant. Le non anglophone bute très souvent sur des acronymes, des noms propres d'institution, de logiciels, de *process* qui sont nommés autrement dans nos langues et cultures. Pire, la non compréhension de concepts émergents, très souvent uniquement énoncés en anglais et n'existant encore que dans la littérature technique et scientifique anglophone, pose des questions de sous développement car c'est la communauté technique ou académique qui ne comprend tout simplement pas le ou les concepts en discussion. Ceci induit conséquemment un sous développement de ce domaine dans la communauté linguistique considérée. Ce sous développement peut être de plus ou moins grande ampleur : soit seuls les utilisateurs, donc les clients potentiels, ne comprennent pas et donc ignorent ces concepts innovants ; soit les chercheurs et les industriels ne les ont pas compris, ce qui est beaucoup plus rare. Cependant, dans les deux cas de figure, l'inexistence du marché dans le territoire national et/ou linguistique entraînera un sous développement et donc des pertes d'expérience et de compétence qui accentueront un effet de fracture techno numérique.

8.2 Des concepts multilingues parallèles

Pour tout ce qui relève des approches techno matérielles, la question peut apparaître assez triviale sur le plan des méthodes. Les objets techniques, notamment les composants électroniques sont développés comme produits semi-finis à une échelle industrielle mondiale et, sauf exception, leur dénomination et leur intégration dans des nomenclatures et des catalogues pose peu de problèmes. A un niveau plus prospectif, quand on décrit des objets et surtout des services qui ne sont pas encore développés (ou en cours de développement), la question est plus délicate. La classification des objets techniques peut être précisément remise en cause à partir du moment où les innovations bouleversent l'organisation d'un système technique préexistant et de son découpage générique et partitif en sous ensembles techniques. La plupart des révolutions techniques réorganisent, en effet, le système technique lui-même créant des intégrations et des regroupements inattendus. Par exemple, dans un véhicule automobile classique, le système de freinage sera totalement distinct du système moteur et de la distribution d'énergie. Dans une optique plus prospective, le moteur servira de frein et si c'est un moteur électrique, il deviendra dynamo et producteur d'énergie électrique qui, stockée dans des accumulateurs, pourra être réutilisée comme énergie du moteur. On voit bien à travers cet exemple qu'en rigidifiant le système conceptuel de la nomenclature, on risque de rendre moins visible les objets et concepts innovants. En fait, cet exemple est marginal quand aux méthodes des nomenclatures technique. En général, elles posent peu de problèmes quant à

l'ambiguïté multilingue des concepts et dans le découpage conceptuel et sa hiérarchie générique et partitive.

Mais la normalisation qui concernait au début surtout des objets ou des composants, concerne maintenant de plus en plus des services à forte composante sociale et humaine. En normalisant l'e-Learning (aussi bien l'e-enseignement que l'e-formation), on est typiquement dans cette nouvelle catégorie de familles de normes qui concerne non seulement des composants matériels (réseaux, supports mémoires et terminaux), voire des *process* industriels, mais qui touchent à des ressources d'enseignement, des styles d'apprentissages, des compétences, des objectifs d'apprentissages, des modes de coopération des descriptions d'acteurs ou d'institutions académiques. Bref, on touche toutes choses qui sont éminemment dépendantes des cultures, des langues, des métiers ou des disciplines et qui posent des problèmes terminologiques autrement plus complexe que la description d'un objet technique. Cartago est né d'un débat et d'une complémentarité de méthode entre l'ISO et des experts de l'e-learning

On a vu que la majorité des experts du SC36 avaient proposé que soit réalisée une liste de vocabulaire en anglais et en français découpée en un certain nombre de domaines. Pour des raisons que l'on comprend aisément, il ne pouvait être de la responsabilité d'un groupe ISO qu'un tel vocabulaire puisse être développé dans un nombre exhaustif de langues : par exemple au minimum les langues des différents pays participants au groupe de travail SC36 WG1.

Dans le cas particulier du SC36, le travail terminologique s'élabore donc à la fois dans le cadre même du groupe de terminologie obligatoirement mis en place dans une instance de normalisation (ici le SC36), mais aussi et de façon plus systématiquement multilingue et plus strictement conforme aux recommandations normatives terminotique (ISO TC37) dans le cadre de l'Alliance Cartago.

Nous sommes certains, par contre, que ces bases terminologiques « très large langues » confrontées à la disparité des écritures, des modes de classification alphabétique ou idéographique, confrontées aussi à la difficulté des requêtes orientées concepts poseront quantité de nouveaux problèmes terminologiques et terminotiques qui seront autant d'occasion d'entreprendre des recherches.

9. Historique d'une méthode

Le SC36 s'est ainsi conformé aux règles de la majorité des instances normatives et il a créé un groupe de travail : le SC36-WG1 : *terminologie* qui se restreint assez vite pour

devenir un groupe *vocabulaire*. Il s'avère que pour des raisons diverses la normalisation de l'e-Learning a surtout été amorcée par les « pionniers du domaines » essentiellement des industriels de l'*e-training* aéronautique, militaire et de sécurité auxquels se sont joint des enseignants européens des sciences exactes et expérimentales. Ces pionniers s'accommodaient incomparablement mieux d'un unilinguisme pragmatique (c'est le cas quasi à 100% dans le secteur aéronautique), d'une relative univocité des concepts des matières enseignées que les nouveaux entrants ultérieurs (surtout des universitaires).

Cette première prépondérance des pionniers s'estompe aujourd'hui même s'ils représentent toujours une part importante du marché global de l'e-enseignement. Les cultures linguistiques autres qu'anglophones et européennes (notamment Corée, Chine et Japon) affirment d'année en année un dynamisme et une croissance importante. Le Sud-est asiatique pourrait à lui seul multiplier le marché de façon considérable, et pourtant ce n'est que très récemment que cette partie du monde commence à affirmer son droit à la diversité linguistique et culturelle dans le processus normatif ⁽²⁷⁾.

L'évolution historique du SC36, qu'il serait hors sujet de relater en détail, fait que nous disposons maintenant d'un groupe de travail produisant un vocabulaire à ce jour bilingue (anglais/français), co-élaborée en consensus dans les deux langues. Ce vocabulaire n'a pas vocation à être très important ⁽²⁸⁾. Par décision de l'instance hiérarchiquement supérieure au SC36, en l'occurrence le JTC1, ce vocabulaire devrait en principe être mis en commun avec tous les autres vocabulaires ou terminologies du JTC1 dans une base terminologique commune gérée par Termium.

10. La pragmatique normative : un choix raisonné entre science et technique

Nous sommes rendus à un moment précis de notre exposé où nos précautions oratoires de non appartenance à la communauté scientifique des terminologues sont importantes et ne doivent pas être prises en compte comme une simple politesse académique. Le développement d'une activité industrielle ou sociale est toujours un choix raisonné entre science et technique. Ainsi, il existe une théorie des sciences de l'éducation ainsi qu'une science de l'informatique et une théorie des technologies numériques et multimédia ; mais il existe aussi une pragmatique du développement de l'e-Learning et des TICE et de leur appropriation par les

²⁷ Voir HUDRISIER (Henri), « Société de connaissance, le paradigme de l'appropriation », et BEN HENDA (Mokhtar), « Les contradictions d'une politique de diversité culturelle », in OILLO (Didier), MVE-ONDO (Bonaventure) Dir. : 2006, *Fractures dans la société de connaissances*. (Paris : Hermès n° 45, CNRS éditions).

²⁸ Quelques centaines de termes au maximum.

apprenants, les enseignants, les créateurs de contenus ainsi que les décideurs et les gestionnaires des institutions et des grands réseaux mondiaux d'e-enseignement. En bonne gouvernance mondiale, il est de la responsabilité de tous ces acteurs de s'impliquer dans la préservation pragmatique de leurs intérêts quelquefois contradictoires pour établir en consensus des normes (soit nationales, soit européennes ou mondiales) à même de préserver au maximum leurs besoins d'expression de leur diversité : c'est fondamentalement notre rôle au SC36.

Du côté de la terminologie et des ressources linguistiques, notre pratique de normalisateurs ⁽²⁹⁾ nous autorise, croyons nous, à proposer comme pour l'éducation une pragmatique du développement linguistique, terminologique et terminotique, tout au moins pour ce qui concerne une gestion sémantique de la transmission numérique du savoir. Nous examinerons successivement la question d'un point de vue global puis du point de vue pragmatique de nos choix de méthodologie terminologique dans le cadre de Cartago.

10.1 Pragmatique normative d'un point de vue global.

La normalisation linguistique ou terminologique achoppe en tout premier lieu sur l'importante question de normalisation linguistique, distincte mais évidemment liée à la normalisation des produits et services telle que l'entendent l'ISO, l'AFNOR ou même les diverses commissions de normalisation terminologique telles que nous les connaissons par exemple en France, au Québec, en Catalogne ou dans le monde arabe. Nous savons bien sûr que la plupart des grands auteurs fondateurs de la linguistique définissent les langues par leurs pratiques (notamment orales). Dès lors, toute volonté de normalisation ou de gestion linguistique serait illégitime. Nous savons cependant que les grandes langues (le français, l'arabe, le chinois en sont des exemples), se sont construites par la volonté politique et maintenant industrielle et économique de stabiliser des états normatifs de ces langues et d'en assumer la gestion. Les codes de langues de l'ISO ⁽³⁰⁾ reflètent aussi au niveau du repérage de

²⁹ Les deux auteurs de cet articles ont été à des degrés divers de responsabilité (national, européen, ISO) impliqués dans plusieurs commissions de normalisation : TC46, Documentation ; JTC1 SC2, codification des caractères ; JTC1 SC29, codage de l'audiovisuel (MPEG, JPEG) et bien sûr SC36. Ils ont aussi assisté en tant qu'observateur à plusieurs réunions du TC37 (terminologie et autres ressources linguistiques). Ils assurent une veille active des normes et travaux en cours dans ce TC37.

³⁰ L'ISO 639 est une norme internationale qui définit des codes pour la représentation des noms de langues : ISO 639-1 (alpha-2) utilise des codes sur 2 caractères, et les associe avec les noms en français, en anglais et dans la langue elle-même. ISO 639-2 (alpha-3) utilise des codes sur 3 caractères et a deux codages possibles : ISO 639-2/B (bibliographiques) et ISO 639-2/T (terminologiques) ; les codes sont associés avec des noms en français et en anglais. ISO 639-3 complète l'ISO 639-2 avec de très nombreuses autres langues encore manquantes.

ces langues cette réalité pragmatique.

10.2 Nos choix pragmatiques de méthodologie terminologique dans le cadre de Cartago.

Dans le cadre de Cartago, nous mettons en œuvre un dispositif de gestion terminologique pour lequel nous avons décidé de nous conformer strictement à différentes normes élaborées par le TC37 (ISO/IEC 704) principes et méthodes en terminologie, ISO/IEC12620, catégories de données en terminologie, ISO/IEC 16642 *Terminological Markup Framework*). Nous avons bien sûr calqué notre méthode sur le TMF tel qu'il a été défini par le Loria. Ayant déjà collaboré à de nombreuses occasions avec Laurent Romary, le recours au Loria a été facile. C'est directement Laurent Romary et ses collaborateurs qui ont été les artisans premiers de Cartago⁽³¹⁾. C'est uniquement parce qu'il nous était impossible de mettre en place un financement que nous avons instrumenté ces premiers travaux dans le cadre de Génétrix. Nous avons, à cette occasion, complété, consolidé et validé un premier état du modèle d'application XML-TMF pour Cartago. Il est à noter que la migration des ressources (limitées il est vrai) s'est effectuée très facilement, ce qui démontre la réelle interopérabilité du modèle.

Résumons nos besoins terminologiques. Nous devons assurer certaines fonctionnalités informationnelles de ces ressources terminologiques qui nous ont paru importantes dans l'équipe projet⁽³²⁾ Cartago. Nous avons besoin de disposer d'un cadre global de recueil, description, administration et gestion terminotique de nos ressources terminologiques. Il nous fallait aussi assurer une interopérabilité avec les choix de recueil de vocabulaires en parallèle multilingue normalisé par le SC36. TMF, et plus particulièrement le système Génétrix qui lui est conforme, répondent actuellement à ces premières contraintes globales. Ce cadre structurel du TMF est connu par un grand nombre de terminologues intéressés par la terminotique normalisée. Nous en présentons donc très rapidement les grandes lignes.

11. Les bases d'une méthodologie pragmatique

11.1 Deux logiques d'approches méthodologiques

La collecte des termes, qui est une activité intellectuelle très ancienne, s'organise naturellement selon deux démarches de directions inverses mais synergiques : *sémasiologique*

³¹ Ce sont également eux qui nous ont initiés au TMF et au TML.

³² Nous distinguons évidemment l'équipe projet (AUF, Loria Nancy, André Le Meur, MSH Paris Nord), de l'Alliance Cartago qui s'approprie progressivement Cartago au fil de l'intégration de nouvelles langues dans Cartago.

et *onomasiologique*.

La démarche *Sémasiologique* : Pour des rédacteurs monolingues, des traducteurs, dans une certaine mesure pour les lexicographes, notamment le rédacteur de dictionnaires de langue ou des dictionnaires bilingues, il est fondamental de partir des termes, des mots ou des expressions (syntagmes) pour pointer sur leur signification (des concepts et leur définition).

La démarche Onomasiologique : pour des documentalistes, pour des spécialistes du knowledge management, des ingénieurs ou des chercheurs analysant des processus techniques ou un phénomène scientifique, il est nécessaire d'avoir une démarche de direction opposée. Il faut partir des concepts, que l'on organise en système de concepts, avec des graphes de relations (génériques, partitives, relationnelles), pour pointer sur un deuxième niveau sur les termes eux-mêmes dans des langues naturelles, voire des langues de spécialité.

Pour ce qui est de l'aspect humain de l'activité terminologique ou lexicographique, la synergie entre les deux directions de méthode est la règle. Toutes les terminologies se réalisent en liant les deux méthodes en synergie dialectique. Aujourd'hui la communauté des terminologues mais aussi celle des lexicographes se sont entendus pour normaliser une démarche unique. Celle-ci est onomasiologique et ce pour plusieurs raisons.

Une banque terminologique sémasiologique connaît une explosion des liens de relations, surtout si elle est multilingue. En effet, si on construit des banques multilingues en partant des termes, on rencontre les plus grandes difficultés à apparier *in fine* les structures conceptuelles, car leur nombre croît exponentiellement par rapport au nombre des termes. De ce fait, cette méthode est quasi-incompatible avec la logique informatique. Notons aussi que l'univers des relations conceptuelles ainsi créé perd toute utilité du fait de son immensité et de son extrême complexité.

A contrario, la démarche onomasiologique, découle hiérarchiquement des concepts, ce qui s'accorde parfaitement avec la logique du calcul informatique. Sa relative simplicité d'organisation hiérarchique des concepts (génériques et partitifs), lui permet, au contraire de la démarche sémasiologique, d'être parfaitement utilisable pour modéliser des référentiels multilingues et multiculturels. C'est la raison pour laquelle la communauté des normalisateurs terminologues et lexicographes a normalisé comme seule valide les principes et méthodes correspondant à la démarche onomasiologique (ISO704, Terminologie : principes et méthodes). Ils ont ensuite normalisé un catalogue ouvert de catégories de données apte à définir des données terminologiques ou lexicographiques (ISO/IEC12620). Puis ils ont enfin

normalisé un cadre commun de mise en œuvre terminotique à même d'assurer l'interopérabilité et la réusabilité des ressources terminologiques indépendamment des diverses banques de données terminologiques. Ce cadre commun, le TMF (ISO/IEC 16642) nécessite bien sûr que ces différentes bases respectent le métamodèle XML TMF, ou exige que les ressources terminologiques soient reformatées selon ce même modèle.

Nous savons bien sûr que les bases terminologiques développées en TMF peuvent in fine générer des ressources présentées aussi bien selon la logique onomasiologique que sémasiologique (fonction disponible sur la base Dhydro du Loria et que Termbridge permet aussi dans Génétrix).

Nous présentons ci-dessous quelques copies d'écrans de nos travaux actuels mais qui demandent à l'évidence à être instrumenté dans le cadre concret d'une base terminologique intégrée qui dépasserait la seule saisie des données dans le cadre d'un contexte XML amélioré. C'est ce chantier de constitution d'une équipe élargie et d'un environnement matériel et logiciel efficace qui constitue le chantier ouvert par l'Alliance Cartago pour les prochains mois.

11.2 Le cadre d'interopérabilité entre les bases de données terminologiques

L'intérêt de ces normes du TC37 c'est qu'elles ouvrent un cadre d'interopérabilité entre différentes bases de données terminologiques. Cette interopérabilité est d'abord assurée par l'identité de méthode et du mode de description. Plusieurs bases de données terminologiques qui s'appuient sur la norme ISO 704 (principes et méthodes) et ISO/IEC 12620 (Catégories de données en terminologie), seront déjà assurées d'avoir constitué des ressources terminologiques qui peuvent, au prix d'efforts informatiques importants mais possibles, être récupérées pour être interopérables entre elles.

Dès lors, il devient clair que ces dix dernières années, les efforts terminotiques ont été très importants et se sont focalisés sur le TMF et l'adaptation des nouvelles versions de l'ISO/IEC 12620 à ces nouvelles méthodes. La terminotique n'a réellement bouclé ses méthodes que grâce au progrès de l'information structurée (SGML et surtout XML) qui ont permis d'organiser toutes les banques de données terminotiques (notamment les terminologies multilingues) aussi bien en permettant de concilier la logique lexicographique et la logique terminologique mais en permettant surtout de concevoir un schéma terminotique unique basé sur XML-TMF qui permet d'assurer l'interopérabilité de toutes les banques terminologiques. L'interopérabilité des terminologies dans le futur web sémantique exigeait notamment un tel

choix normatif. Les mécanismes du TMF et de la normalisation des catégories de données permettent précisément de servir de base à la réalisation modulaire d'ontologies elles mêmes en cours de standardisation par le W3C par son langage OWL (Web Ontology Language).

11.3 La logique du schéma TMF

Le TMF est un métamodèle ce qui signifie en pratique qu'il pourra s'instancier en un grand nombre de schémas applicatifs intercompatibles entre eux et répondant tous à une même logique XML. Le TMF est ainsi défini selon une logique hiérarchique par niveaux TDC (*Terminological Data Collection*). Ce niveau est celui de la base elle même. GI (*Global Information*) et CI (*Complementary information*) sont deux niveaux (en fait deux registres) dans lesquels sont stockées les données référentielles importantes pour administrer ou faire fonctionner la base. Ces données n'appartiennent pas directement à la collection terminologique.

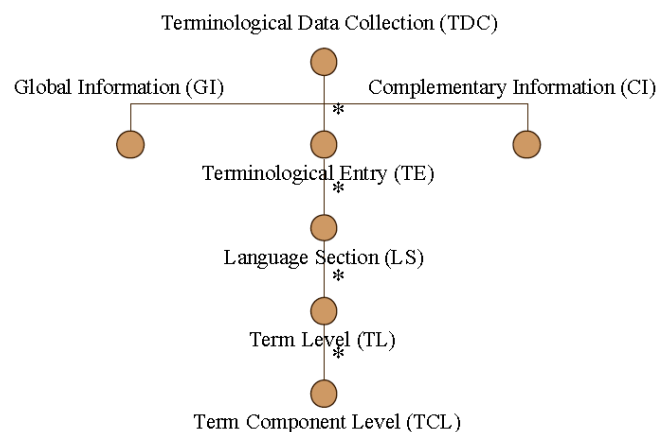


Fig.1 : Méta-modèle d'organisation d'une base de données terminologique

Le schéma logique structurel qui permet d'optimiser la démarche onomasiologique s'organise dans les niveaux TE (*Terminological Entry*), LS (*Language Section*), TL (*Term Level*) et TCL (*Term Component Level*) comme présenté dans la figure 1 ⁽³³⁾.

- **TE (*Terminological Entry*)** ⁽³⁴⁾ : malgré son nom, c'est exclusivement le niveau du concept. Dans le schéma Cartago ce niveau est celui d'un identifiant unique par concept ⁽³⁵⁾. C'est en

³³ Notons que pour exemplifier l'exposé nous mélangeons volontairement l'option du modèle applicatif Cartago et l'exposé fonctionnel du métamodèle TMF.

³⁴ On ne peut changer (sauf révision de la norme) ce qui a été normalisé, mais certains experts s'accordent pour convenir qu'à ce niveau l'adjectif « *conceptual* » aurait été plus clair que « *terminological* ».

³⁵ La logique qui consiste à identifier chaque entrée terminologique (chaque concept) uniquement par un identifiant (numérique ou alphanumérique) est méthodologiquement indispensable. Il s'agit clairement d'affirmer que le concept n'équivaut pas aux mots pour le dire. Dans la pratique l'utilisateur peut toujours recourir à une facilité de rappel automatique du terme vedette le plus proche du concept (soit dans la langue

ce sens que cela justifie la dénomination *Terminological Entry*. Tous les termes en toutes langues sont traités à un niveau hiérarchique inférieur lié à ce niveau de « l'entrée » et réfèrent tous à ce même identifiant. C'est aussi à ce niveau que l'on peut décrire les caractéristiques du système (ou graphe) de concept : générique ou partitif⁽³⁶⁾. Nous devons permettre la réalisation d'un système de concepts organisé selon une méthode onomasiologique bien adaptée à la maîtrise primordiale des concepts plutôt que celle des termes. En effet, des experts en normalisation ont besoin de s'entendre collectivement et au-delà du (ou des) nom(s) des choses (composants, services, institutions, rôles, contenus...) dans leurs langues respectives. Ils peuvent ainsi organiser ces concepts (qui ne sont à ce stade que des identifiants) les situer en relation de génériques ou partitives. L'identifiant, la qualité de généricité et partitivité sont des informations qui relèvent du niveau TE (Terminological Entry, niveau du concept),

La définition d'autres types de relations entre les concepts au-delà de ces deux relations (génériques ou partitives), est du ressort de la réalisation d'ontologie. A terme, nous prévoyons que l'Alliance Cartago devra réaliser des ontologies du domaine. Les terminologies seront d'autant plus faciles à chaîner ensemble (par exemple avec le langage OWL) qu'elles se limiteront aux seules relations génériques et partitives. La notation provisionnelle de tels renseignements, quand ils existent, permet cependant d'avancer le chantier ontologique qui n'est pas notre première priorité. La description des relations ontologiques est aussi stockée au niveau TE,

- **LS (*Language Section*)** : comme son nom l'indique, c'est plus une section qu'un niveau. Le LS intégrera tout ce qui est dépendant des langages, et il s'oppose clairement au niveau des concepts. C'est à ce niveau hiérarchique que l'on ouvrira des langues. Dans Cartago, nous ouvrons en LS des descripteurs « *definition* » qui permettent ici de définir en n langues un même concept situé lui au niveau TE. Cette architecture du schéma est clairement conceptuelle (c'est notre choix Cartago). Elle nous permet de définir ces concepts en consensus dans autant de langues de travail ou de langues objet que l'on souhaite. Dans le modèle structurel applicatif TMF que nous avons arrêté pour Cartago, nous avons choisi de ne pas attacher les définitions aux termes eux-mêmes mais aux concepts. Nous sommes donc

de référence, soit dans sa propre langue). Mais cette commodité d'interface ne doit jamais faire oublier que le concept n'est pas un terme. On peut même envisager de proposer une image fixe ou une vidéo, mais en gardant présent à l'esprit, selon le mot de Magritte : « ceci n'est pas une pomme ! » ; autrement dit le concept n'est pas le mot.

³⁶ On peut envisager de décrire aussi des relations d'autre nature, dépendance, inclusion, relation... mais hors des 2 premières (générique ou partitive), on quitte le domaine des terminologies pour être dans celui des ontologies. Le modèle TMF autorise cependant ce type de catégorie de description.

bien dans la posture de définition conceptuelle collaborative entre langues distinctes. Ces définitions sont stockées dans le niveau LS (Language Section et non pas comme c'est très souvent le cas dans d'autres modèles applicatifs à l'intérieur du TL : Term Level) (³⁷).

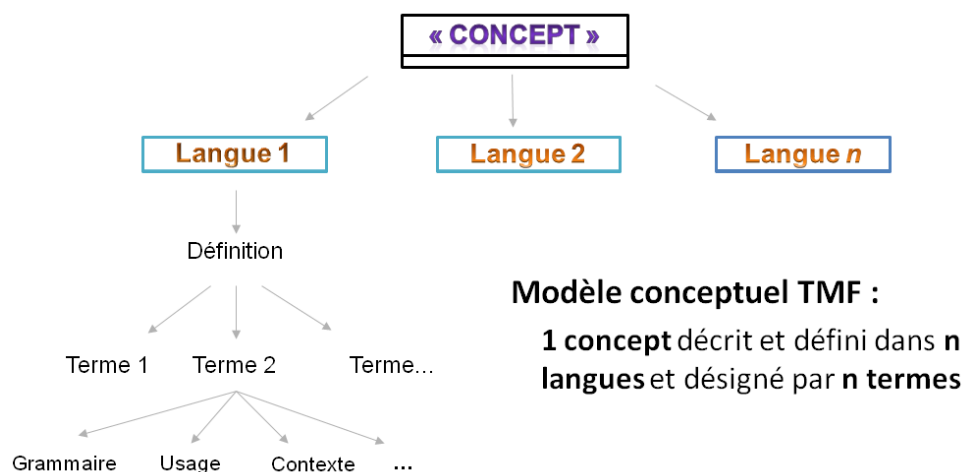


Fig. 2 : La section Langues dans le modèle TMF

- **TL (Term Level)** : Nous avons besoin bien sûr d'apparier chaque concept (chaque terminological entry) avec des termes dans les langues traitées par Cartago. Ces termes sont réputés synonymes. Des deltas entre termes sont inévitables : par exemple EAD (Enseignement à Distance) est fréquemment donné comme équivalent du terme anglais e-Learning. C'est vrai pour la stricte acception du terme dans le cadre numérique, c'est faux quand l'EAD signifie enseignement par correspondance postale. Des catégories de données peuvent à ce TL décrire l'acception particulière qu'a tel ou tel terme (ici, par exemple EAD strictement numérique à l'exclusion du sens EAD par correspondance). Des termes différents d'une même langue peuvent d'ailleurs exprimer en stricte synonymie le même concept. Ainsi, par exemple, le premier niveau de l'enseignement secondaire français, sera désigné par le terme 6^{ème} en France, à Lausanne et dans tous les lycées français du monde. Il s'appellera 1^{ère} à Genève, en Tunisie, au Maroc et dans bien d'autres endroits francophones. Dans d'autres langues et dans d'autres cultures, l'institution éducative ne sera pas forcément découpée selon la même logique (³⁸). C'est à ce niveau que s'ouvrent un ou N termes supposés

³⁷ Nous verrons que nous réservons pour ce TL des catégories de données permettant de préciser ce qui distingue les termes donnés dans chaque langue pour équivalents synonymes du concept de référence.

³⁸ On entrevoit sur cet exemple apparemment trivial comme celui du premier niveau d'enseignement secondaire la difficulté à construire réellement un référentiel sémantique commun indépendant des langues et des cultures. La succession des phases éducatives est loin d'être universelle ; école maternelle, école primaire, enseignement secondaire (découpé ou non en collège puis lycée), l'enseignement universitaire ou professionnel correspondent dans d'autres cultures à des regroupements qui peuvent être très différents. On voit bien de ce fait que la terminologie interculturelle et interlinguistique est bien autre chose qu'une

être tous plus ou moins synonymes dans autant de LS qu'il y a de langues dans la base. C'est à ce niveau que peut se faire une description morphosyntaxique⁽³⁹⁾ des termes. Dans Cartago nous préférons une description assez sommaire car la description linguistique n'est pas notre propos : genre, nombre, catégorie du discours (nom, adjectif, verbe, syntagme, mots valise..), abréviation, acronyme⁽⁴⁰⁾. Nous décrivons aussi pour les langues non latines la translittération voire la transcription phonétique qui permet à des experts non pratiquants de ces langues d'en parler avec leurs collègues concernés.

- **TCL (*Term Component Level*)** : C'est ce niveau que le modèle TMF permet de décrire (comme son nom l'indique) les composants d'un terme. Dans le cas concret du SC36 et dans beaucoup de terminologies d'instances normatives, nous devons gérer une proportion importante de mots valises ou de syntagmes⁽⁴¹⁾. Après longue discussion, Cartago a fait le choix d'activer ce niveau du TMF, bien qu'il ne soit pas obligatoire et que beaucoup de terminoticiens le considèrent comme un niveau de détail inutile. Formellement chaque élément d'un terme composé, dans chaque langue permet d'ouvrir un TCL (*Term Component Level*). On ouvrira donc autant d'item pour un terme qu'il y a de composants dans le terme. En effet l'information terminologique partagée par des experts en normalisation est assez spécifique. Elle comporte d'une part de très nombreux acronymes, sigles et autres formes courtes ou développée. Cela rend bien sûr difficile la compréhension de l'information par des non initiés et des experts étrangers à la langue dominante de rédaction de la norme. Elle reste difficile d'accès aussi à des traducteurs ou des interprètes non spécialisés. C'est la raison pour laquelle nous avons décidé de mettre en place ce TCL.

Paradoxalement en nourrissant la base Cartago en anglais, français et arabe, les deux auteurs ont constaté l'intérêt de ce niveau même en termes d'économie descriptive. Par exemple, entre « système de gestion de l'apprentissage », « système de gestion de contenus d'apprentissage », « système d'apprentissage basé sur la technologie », « système d'apprentissage distribué basé sur la technologie », « système de gestion de contenu »,

correspondance terme à terme.

³⁹ Bien sûr l'arrivée de langues syntaxiquement très différentes (malgache, coréen et même le berbère ouvre des discussions pour ajouter des descripteurs morphosyntaxiques non prévus lors de la mise au point du modèle applicatif. Ces descripteurs étant par nature spécifiques ils ne posent pas de grands problèmes quand au respect de la cohérence historique des anciennes ressources.

⁴⁰ La maîtrise des sigles et acronymes est très importante pour la normalisation. Nous devons pouvoir ainsi gérer comme « termes synonymes » : ISO/IEC JTC1 SC36, sa forme courte (*short form*) SC36 ou même (ce qui est en principe incorrect) ISO SC36, voire ISO/CEI JTC1 SC36 ainsi bien sûr que les formes développées de ces sigles.

⁴¹ Enseignement à distance, style d'apprentissage, ingénierie pédagogique, système d'apprentissage distribué basé sur la technologie, e-apprentissage, *m-learning*, etc...

l'économie de description est évidente. Des secteurs entiers de la terminologie sont ainsi générés à partir de module de base⁽⁴²⁾ comme une sorte de mécano linguistique. L'intercompréhension multilingue de ce mécano et des différentes options de *patterns* linguistiques est très éclairante pour les experts qui ont expérimenté la base. Les catégories descriptives reprennent à ce niveau la description du terme mais paradoxalement en l'élargissant sur les catégories du discours⁽⁴³⁾. Nous avons aussi jugé utile à ce niveau de reprendre la translittération (pour rendre le *pattern* plus facilement accessible à des experts étrangers à la langue considérée) ainsi qu'une traduction dans les langues de travail de Cartago⁽⁴⁴⁾ pour la même raison que précédemment. Nous pensons ainsi que ce niveau de Cartago sera très utile pour proposer la création de nouveaux termes (le plus souvent des termes composés) pour élaborer des terminologies techniques dans des langues qui ont subi des pertes de domaines linguistiques⁽⁴⁵⁾ importantes. Cela permet, en effet, aux experts terminologiques de ces langues de comparer des patrons types de mots composés réutilisables éventuellement dans leur communauté linguistique.

Notons enfin que à tous les niveaux du schéma sont prévu des descriptifs spécialisés pour la gestion administrative des données : les contributeurs, leur rôles, les dates de contribution, de vérification, de révision, etc. De façon systématique chaque terminologie est obligée de recourir à ces catégories de gestion administratives. Mais dans le cas des terminologies dédiées à la normalisation (ce qui est le cas de Cartago), ces questions doivent être aménagées avec le plus grand soin pour pouvoir qualifier les différents niveaux du *process* normatif (CD, FCD, DIS, FDIS, IS)⁽⁴⁶⁾, mais aussi qualifier plus trivialement les diverses contributions terminologiques en n langues et leur validation dans le *process* Cartago interne.

⁴² Les composants isolés ou même des blocs entiers de 2 ou 3 composants unitaires qui sont eux-mêmes composants d'un syntagme plus complexe.

⁴³ Au-delà de nom, verbe, adjectif, nous rencontrons adverbe, préposition, article et même des sous parties (préfixe, suffixe... et nous n'avons pas encore expérimenté l'allemand qui posera évidemment quantité de problèmes qui sont l'objet d'une littérature théorique pour laquelle nous demanderons aide et assistance.

⁴⁴ Dans un premier temps le français et l'anglais partagés par un maximum d'utilisateur à ce stade de développement. On sait par ailleurs que la logique de déclinaison fédérale de Cartago, entraînera la mise en place de pôle Cartago linguistiques. Cartago coréen aura sans doute comme langue de travail le coréen et l'anglais.

⁴⁵ NILSSON (Henrik) : 2006, « Perte de domaine : indicateurs et enjeux », in AUF & AILF éd. *La langue française dans l'aventure informatique*, (Paris : colloques Lexipraxis 2005 & 2006), pp. 149-158

⁴⁶ CD : Committee Draft, FCD : Final Committee Draft, DIS : Draft International Standard, FDIS : Final Draft International Standard, IS : International Standard



Fig. 3 : Structure d'une fiche terminologique conforme TMF sous Genetrix

12. Facilités éditoriales avec XML

Le TMF repose intégralement sur la logique XML. Il est maintenant très connu que les informations disponibles sur un schéma XML peuvent être facilement exploitées grâce à des « feuilles de styles » ad hoc pour générer diverses applications éditoriales.

C'est ainsi que, même si nous avons dans Cartago, beaucoup plus d'informations sur chacune des entrées de la norme terminologique du SC36 (ISO/IEC2382-36), nous pouvons très facilement éditer un document (numérique ou papier) absolument conforme au format prévu par le SC36WG1 (liste bilingues en parallèle organisées en 7 sections thématiques ne comportant que les numéros identifiants, les termes et les définitions, suivit d'index alphabétiques). Nous avons pu ainsi montrer au SC36 qu'il nous était facile de mettre en parallèle d'autres langues. La compatibilité de la base la plus sophistiquée (Cartago) vers celle qui l'est moins (le vocabulaire du SC36) est bien sûr évidente.

36 Information Technology for Learning, Education, and Training.	36 Technologies de l'information pour l'apprentissage, l'éducation et la formation	36 تكنولوجيا المعلومات للتعليم والتربية والتكوين	36 학습, 교육, 그리고 훈련을 위한 정보 기술
36.01 36.01 General terms	36.01 36.01 Termes généraux	36.01 مصطلحات عامة	36.01 36.01 일반개념
36.01.01 learning (preferred term) acquisition of understanding, knowledge, skills and/or Attitudes	36.01.01 apprentissage (terme préféré) acquisition de connaissances, développement d'habiletés et d'attitudes	36.01.01 تَعَلُّم (preferred term) التعلم (terme admis) عملية اكتساب الفرد لمعلومات أو طرائق جديدة للسلوك، أو مهارات جديدة في العمل المعيني	36.01.01 학습 이해, 지식, 기술, 및/또는 태도의 습득
36.01.02 education (preferred term)	36.01.02 éducation (terme préféré)	36.01.02 n, (preferred term)	36.01.02 تربية

Fig. 4 : Édition multilingue de la liste ISO 2382-36

Évidemment, les concepts formalisés dans ce référentiel commun peuvent être spécifiques à une langue et à une culture. Cela n'en est que plus bénéfique à la communauté des experts. Il s'agit là d'une notion bien banale pour des linguistes. Chaque langue « réinvente le monde à sa façon ». Cela est particulièrement vrai dans le secteur de l'éducation puisque l'apprentissage des langues (un des premiers objets que l'enfant puis l'adulte ont à apprendre puis à approfondir) nécessite des modalités d'acquisition spécifiques à telle ou telle langue. Par exemple distinguer en arabe le niveau de maîtrise de la lecture accentuée du niveau non vocalisé perd tout sens dans une langue alphabétique (et non plus consonantique). L'apprentissage d'une langue idéographique (intimement lié à l'apprentissage du vocabulaire et de sa notation idéographique) est conceptuellement très éloigné de l'apprentissage initial de l'écriture et la lecture quasi indépendant de l'acquisition du vocabulaire. Les institutions éducatives de toutes les nations sont autant de concepts relativement distincts les uns des autres mêmes s'ils peuvent être appariés. De nombreuses nations n'ont pas un ministère spécifique à l'éducation. En France nous partageons la transmission du savoir entre plusieurs

ministères (Travail et affaires sociales, santé, recherche et enseignement supérieur, culture, armées, agriculture...). On voit bien que le pragmatisme nous obligera à arbitrer entre deux scénarios extrêmes : a) définir des concepts sommairement définis et associer des quantités de termes dans chaque langues, b) multiplier à l'infini le nombre des concepts pour préserver dans le référentiel tous les particularismes culturels, institutionnels, linguistiques et noyer les experts sous trop d'information. On est là au cœur du débat de la normalisation respectueuse de la diversité.

13. Conclusion

On ne saurait conclure un projet qui en est en phase de mise en œuvre opérationnelle. De nombreux problèmes restent ouverts. L'équipe projet est trop réduite pour les résoudre et n'en a pas de toute façon l'entière compétence.

L'équipe Cartago a besoin de s'ouvrir à des réseaux comme le réseau thématique de l'AUF mais aussi à d'autres réseaux extra-francophones tant pour ce qui est du recueil et de la gestion des ressources terminologiques que pour ce qui est du conseil et de l'aide au développement terminotique.

En première approche nous avons repéré un certain nombre de questions ouvertes. Comment réaliser concrètement en travail collaboratif à distance entre un grand nombre de langue un système de concept cohérent ? Comment concevoir des modes d'accès à Cartago respectueux à la fois d'interfaces alphabétique et de classement par clefs idéographique ? Quels environnements techniques seront capables de gérer concurremment plus d'une dizaine d'écriture dans des applications dépassant la complexité d'un traitement de texte ? Nous avons surtout conscience qu'il est essentiel que l'Alliance Cartago réussisse à très court terme la définition d'un contexte de travail tant technique que humain. Il lui faut pour cela réussir aussi son modèle de financement.

Bibliographie :

BEN HENDA (Mokhtar) : 2006, « Les contradictions d'une politique de diversité culturelle », in OILLO (Didier), MVE-ONDO (Bonaventure) Dir., *Fractures dans la société de connaissances*. (Paris : Hermès n° 45, CNRS éditions), pp. 41-48.

BEN HENDA (Mokhtar), HUDRISIER (Henri) : 2008, « Actions francophones autour des normes e-Learning à l'ISO Pour un accès multilingue et multiculturel égalitaire à l'éducation » *Distance et Savoirs*, Hors série, pp. 1-18.

CABRÉ (Maria Theresa) : 1998, *La terminologie, théorie, méthodes et applications*, (Paris : Armand Colin, Les Presses de l'Université d'Ottawa).

HUDRISIER (Henri) : 2006, « Société de la connaissance, le paradigme de l'appropriation », in OILLO (Didier), MVE-ONDO (Bonaventure) Dir., *Fractures dans la société de connaissances*. (Paris : Hermès n° 45, CNRS éditions), pp. 153-164.

HUDRISIER (Henri) : 2005, « Cartago : un outil et un réseau pour la terminologie notamment berbère, une alliance multidisciplinaire et internationale de chercheurs », in *Études et documents berbères*, n° 23, éd. La boîte à documents, MSH Paris Nord, 2005.

HUDRISIER (Henri) : 2005, « La normalisation comme cadre de développement d'un enseignement-recherche ouvert sur le web-sémantique », in Uni-Mail Genève, éd. 22^e congrès de l'AIPU (Genève : AIPU, 12-14 septembre).

HUDRISIER (Henri) : 2006, « Vers une société de connaissances partagées : initiatives universitaires francophones dans la formation ouverte à distance : l'enjeu des normes et standards », in AUF & AILF éd. *La langue française dans l'aventure informatique*, (Paris : colloques Lexipraxis 2005 & 2006).

HUSSON (Jean-Luc), RODHE (Hans), ROMARY (Laurent), VISCOGLIOSI (Nadia), VAN CAMPENHOUDT (Marc) : 1999, « Normes expérimentales pour la définition d'un forum terminologique : le projet européen Dhydro », in *Conférence « AET » sur la coopération dans le domaine de la terminologie en Europe*, (AET, 17-19 mai)

KRAMER (Isabelle) : 2004, *Easy Time TMF, livre blanc du TMF*, (INRIA/LORIA)

NILSSON (Henrik) : 2006, « Perte de domaine : indicateurs et enjeux », in AUF & AILF éd. *La langue française dans l'aventure informatique*, (Paris : colloques Lexipraxis 2005 & 2006), pp. 149-158

OILLO (Didier), BEN HENDA (Mokhtar) et HUDRISIER (Henri) : 2007, « Les

disparités culturelle et linguistique de la Francophonie comme atout de son développement technologique en communication et en éducation », in Université de Balamand éd. *Francophonie : Conflit ou complémentarité identitaire ?*, (Tripoli, Liban : Université de Balamand, 16–20 avril).

OILLO (Didier), MVE-ONDO (Bonaventure) Dir. : 2006, *Fractures dans la société de connaissances*. (Paris : Hermès n° 45, CNRS éditions).

ROMARY (Laurent), HUDRISIER (Henri) : 2003, « Le balisage normalisé des concepts et documents en liaison avec les normes de l'EAD », in *Initiatives 2003 : open forum AUF/SC36* (Versailles : AUF/SC36, 19 mars).

ROMARY (Laurent), VAN CAMPENHOUDT (Marc) : 2001, « Normalisation des échanges de données en terminologie : le cas des relations dites conceptuelles », in *Conférences TIA*, (Nancy ; TIA, mai).

WÜSTER (E.) : 1976, « La théorie générale de la terminologie, un domaine interdisciplinaire impliquant la linguistique, la logique, l'ontologie, l'informatique et les sciences des objets », in DUPUIS (H.), éd. *Essai de définition de la terminologie. Actes du colloque international de terminologie* (Québec, Manoir du lac Delage, 5-8 octobre 1975), Québec, Régie de la langue française, pp. 49-57.