

15 ans de LexiPraxi

Florilège
FR



En partenariat avec la Délégation générale à la langue française et aux langues de France

LexiPraxi

Auteurs

Jacques André – Sylvie Baste – Pierre Baylet – Jean-Michel Borde –
Karl Dubost – Henri Hudrisier – Jean Herr – Jean Hyenne – Patrick Leboeuf –
Bernard Marti – Didier Oillo – Laurent Romary – Rachid Zghibi

Relecteurs

Jean-Alain Hernandez – Alain Vaucelle

**Publié sous la direction du comité scientifique de l'AILF
Collection LexiPraxi**

© AILF - Informatique & langue française

Éditeur

**19 rue Bergère
75009 Paris
France
(+33) 1 42 93 11 07
www.aifl.asso.fr
lexipraxi@aifl.asso.fr**

Paris décembre 2008

collection LexiPraxi 2008

15 ans de LexiPraxi

Florilège

Écritures, enseignement, normalisation

*Sous la direction de Jacques André, Jean-Michel Borde,
Jean-Alain Hernandez, Henri Hudrisier, Alain Vaucelle*



Association des Informaticiens de Langue Française

*En partenariat avec la
Délégation générale à la langue française
et aux langues de France*

LexiPraxi 2008

Remerciements

L'AILF remercie la Délégation générale à la langue française et aux langues de France pour son soutien et sa contribution à la réalisation de cet ouvrage. Ces remerciements vont plus particulièrement à Monsieur Xavier North, délégué général, pour sa bienveillante attention aux thématiques chères à notre association et à ses membres.

6, rue des Pyramides 75001 PARIS
Métro : Pyramides ou Tuileries
Téléphone : 33 (0)1 40 15 73 00
Télécopie : 33 (0)1 40 15 36 76
Courriel : dglf@culture.gouv.fr
Site : <http://www.dglf.culture.gouv.fr>

Délégué général
Xavier North

Remerciements aux auteurs des articles parus dans les actes LexiPraxi de 1990, 1992, 1994, 1996, 1998, 2000, 2002, 2006.

Remerciements aux relecteurs et contributeurs de la présente édition Messieurs Jacques André, Jean-Michel Borde, Jean-Alain Hernandez, Henri Hudrisier, Alain Vaucelle.

Avertissement

Les éditeurs de ce florilège ont opté pour une mise en page aussi intuitive que possible afin de rendre compte des différents niveaux de texte : extraits originaux des contributions des intervenants aux différents LexiPraxi, sur fond blanc ; commentaires du collectif d'éditeurs et résumés minimum pour rendre compte des coupes entre les extraits imprimés sur fond tramé. Nous pensons que nos lecteurs s'approprient facilement ce code graphique des différents niveaux de texte. Deux niveaux de notes ont été traité en parallèle (notes originales des auteurs et notes des éditeurs). Le même code graphique fond blanc et tramé, permettra au lecteur d'appréhender facilement les niveaux de notes respectifs.

15 ans de Lexipraxi – Florilège

Table des matières

Remerciements	6
Introduction	8
La culture générale du 21 ^e siècle	14
Avant propos	24
De l'enseignement à distance à l'université virtuelle	25
Le concept d'Université virtuelle	27
En conclusion	29
Codage des caractères	32
LP1990 - Évolution des méthodes de codage de caractères pour la prise en compte du multilinguisme	34
LP2000 - Exigences culturelles et internationalisation	37
LP2000 - De Babel Web à Orléans : pluri-linguisme, Internet et recherche universitaire	43
LP2006 - UNICODE et ISO/CEI 10646 – le codage informatique des langues minoritaires : le berbère	46
Bibliographie commune aux quatre articles	52
Normalisation	54
LP1990 - Exposé pour le colloque Lexipraxi 1990	55
LP1992 - Le français, langue officielle de la normalisation internationale	59
LP1992 - Normalisation des différents média et problèmes de coordination	64
LP2000 - Être strict pour partager : les standards du W3C pour garantir le multilinguisme sur Internet	70
Enseignement en réseau	74
LP2000 - Normalisation des systèmes de médiation pédagogiques : vers l'évolution de l'édition sous forme d'information structurée	74
LP2002 - Des modèles et des identifiants : les travaux de normalisation (ISO/TC46, AFNOR/CG46) potentiellement profitables au Web sémantique	81
LP2002 - Métadonnées et standards en ingénierie des langues	88

L'AILF (Association des Informaticiens de Langue Française) organise, depuis presque 20 ans, un colloque biennal, Lexipraxi. Celui-ci traite des TIC, notamment, de leur appropriation socioprofessionnelle, dans un contexte francophone et, plus largement, dans un environnement multilingue. C'est la préoccupation de la Francophonie et des langues partenaires (Langue de France pour la DGLF-LF).

Au cours de ces 20 ans, la somme cumulée de ces textes représente un patrimoine savant qu'il nous a paru intéressant de solliciter, pour en éprouver les visions et restituer les tendances, parmi les plus fondamentales.

En une quinzaine d'années les TIC ont énormément évolué et le passage au « tout numérique » de l'information induit d'encore plus fortes évolutions, dans les décennies à venir.

Pour éviter la désorientation liée à ces transformations radicales de l'environnement des TIC, il nous a semblé utile de distinguer, au long de nos communications Lexipraxi, de ces quinze dernières années, les contributions qui ont démontré une anticipation remarquable. Celles qui ont su établir la

différence entre la présentation de techniques émergentes et la véritable vision actuelle mais aussi justement prospective, pour être encore intéressante aujourd'hui.

Établir un florilège, aller chercher dans des textes considérés a priori comme originaux, puisqu'ils ont été sélectionnés en leur temps par un comité scientifique, n'est pas une tâche aisée. Cela voudrait dire que l'on imposerait un pouvoir de sélection dans la sélection. C'est pourtant l'exercice auquel nous nous sommes livrés, avec une vision qui est obligatoirement subjective et donc focalisée. Nous avons déterminé 3 axes de sélection partant de l'observation de plus de 15 ans de Lexipraxi. Cela nous a permis de détecter les sujets d'étude de fond qui, ayant eu vocation à s'épanouir, sont devenus des champs d'applications, investis par la suite par le marché et les utilisateurs.

Les « éditeurs de ce florilège » ont, de cette façon, dégagé trois axes thématiques, qui correspondent à la majorité¹ des champs d'application traités dans les différents colloques Lexipraxi.

Après l'introduction, un premier thème traite de l'écriture électronique par la représentation codée de caractères.

Il nous a paru être le fondement commun des autres thématiques abordées dans Lexipraxi. L'écriture des nombres et du calcul, puis très vite l'écriture de texte en langage naturel auront été,

¹ On notera dans ce choix des éditeurs au moins 2 ou 3 axes que nous n'avons pas sélectionné pour ce florilège mais qui aurait pu l'être si le nombre des textes traitant de ces thématiques avait été plus important : les technologies des jeux et la téléphonie (notamment mobile) ainsi aussi que le « broadcast numérique » (radio ou TV).

historiquement, le premier mode de médiation de l'information qui ait été informatisé. Comme le précisent les textes, qui ont été choisis dans le florilège, c'est au cours de ces 15 dernières années que la technologie numérique de la codification des écritures est devenue universelle et compatible pour la quasi-totalité des écritures de la planète (y compris les écritures des époques historiques).

Deuxième thème : normalisation

La standardisation, comme la normalisation, sont très liées aux deux autres thématiques. Il n'est pas possible que tous les documents écrits du monde, que toutes les ressources d'enseignement conçues dans des disciplines, des langues différentes et dans des contextes pédagogiques distincts, puissent circuler et être réutilisés sur une grande échelle, s'ils n'étaient pas normalisés. La poursuite de l'évolution des technologies du numérique, dans les 15 à 20 années à venir pourrait remettre en cause la primauté de l'écriture sur les autres modes de médiation. Le multimédia et l'hypermédia en réseau se sont imposés comme réalités, ce qui nous conduit à la prudence et cette réflexion pourrait, peut-être, alimenter les thématiques abordées par les intervenants de Lexipraxis dans les 15 années qui viennent.

Cependant dans le domaine des TIC, et en dehors des deux autres thématiques déjà citées, la normalisation est elle-même devenue, également pendant ces vingt dernières années un champ économique et social particulier.

De la convergence des médias, des réseaux, des langues, des méthodes d'intelligence et de requête sémantique émerge une forme de communication interpersonnelle multimédia.

La normalisation n'est plus dès lors un simple processus d'harmonisation et d'assurance qualité ou sécurité, elle devient, dans ce champ particulier des TIC, le cadre d'innovation principal des technologies et des applications à venir. Elle devient aussi la structure du cadre prospectif du co-développement futur des technologies des prochaines décennies. Quand on observe avec un certain recul les papiers qui traitent de l'enseignement électronique on observe qu'ils abordent ce thème en mettant en avant toutes sortes de facettes innovantes pour l'époque. Dans ce type d'approches, le temps a fait son œuvre pour certaines d'entre elles ; tel ou tel objet qui retenait l'attention de la communauté scientifique de l'époque ne l'intéresse plus aujourd'hui. Cependant, en prenant du recul vis-à-vis des textes produit dans les différents Lexipraxis, on se rend compte que ce qui les caractérise, c'est la question, aujourd'hui très à l'ordre du jour, de la prééminence du concept sur le terme. La question récurrente d'une approche conceptuelle pour la transmission du savoir à l'ère de la mondialisation numérique en découle : S'agit-il d'accéder à l'essence, aux concepts des savoirs ou aux savoir faire (savoir être), que l'on enseigne et non aux mots des textes décrivant un savoir ? Dès le Lexipraxis 1990 Pierre Baylet le met en évidence, dans une session du Lexipraxis

1990 au titre édifiant: «Enseigner les concepts, voilà le problème».

Troisième thème: TICE et de l'enseignement en réseau

Cette thématique vient naturellement clore et compléter les deux premières. Elle s'enchaîne et complète logiquement avec les précédents car écriture et normalisation sont indispensables à la transmission des savoirs ou d'élaboration des concepts.

Aujourd'hui un regard rétrospectif nous montre deux, ou plutôt trois thématiques successives.

Un intérêt pour une approche conceptuelle de l'enseignement, qui constitue évidemment un préalable au multilinguisme, à la mondialisation des contenus et à leur réutilisation dans des contextes nouveaux. À cette vision «conceptuelle» de l'enseignement électronique à distance succède d'autres textes traitant de l'ingénierie de l'EAD. Ici encore des interventions montrent combien l'approche conceptuelle doit être prolongée par une vision multilingue, mondiale et normalisée. Celle-ci implique bien sûr une approche plus conceptuelle et non pas seulement anecdotique, ou factuelle. Cela nous incite à mettre en sous-titre à cette partie: **ENSEIGNER LES CONCEPTS ET CONCEVOIR LA TRANSMISSION DU SAVOIR A L'ÈRE DE LA MONDIALISATION NUMÉRIQUE.**

LP1990 - Le multilinguisme à télécom Paris

Pierre Baylet

...[... La pratique de l'enseignement et de la formation, surtout à un certain niveau et dans des domaines fortement techniques est une école de rigueur et de précision. Faire un cours nécessite une clarification des idées et une toute autre maîtrise de son sujet qu'une simple conversation. Un des projets du Directeur de la Formation de TELECOMS Paris est d'ailleurs de prévoir dans le cursus normal des élèves une séquence où ils enseigneraient à leurs camarades.

En formation continue, le public a perdu la pratique quotidienne de l'assimilation d'informations nouvelles. Il est donc essentiel de faire passer des messages clairs, des concepts dépourvus d'ambiguïté. Ceux qui pratiquent la formation continue savent que c'est un problème quotidien, le fait d'être expert d'un sujet ne garantissant pas le sens pédagogique nécessaire.

Naturellement la situation se complique lorsqu'à cette opération de transfert de connaissances s'ajoute un processus de traduction dans une autre langue.

TELECOMS Paris organise depuis trois ans un cycle de deux mois, en anglais, pour un public anglophone (anglophone non natif). Ceci constitue une expérience intéressante et donne quelques repères.

Intitulé “New Telecommunications Techniques” et réalisé avec SUPELEC, il a accueilli au total 87 stagiaires (de niveau technique et linguistique assez hétérogène) provenant de 29 pays différents. Les cours étaient exclusivement réalisés en anglais, par des intervenants français. Le degré de satisfaction global des participants a été élevé, comme en témoigne le nombre croissant de stagiaires des mêmes pays.

Quelques remarques instructives peuvent en être tirées :

- Les conférenciers pressentis, déjà connus de nous, (car étant intervenus en Français) ont généralement accepté très vite d’effectuer leur prestation en anglais (seule la réalisation de transparents et de documentation n’a pas toujours été à la hauteur, peut-être plus pour des raisons de temps que de compétence) ;
- Les stagiaires n’ont pas manifesté de critiques vis-à-vis du niveau et de la qualité d’expression en anglais des animateurs.

Il n’est donc pas apparu de problèmes majeurs de traduction proprement dite.

Toutefois si l’on se place sur le terrain de la compréhension et de l’interprétation, les choses sont moins simples.

1) Certaines notions n’ont pas le même sens selon la population à laquelle on s’adresse, même dans un pays donné. Le concept «réseau local» à ainsi longtemps eu en France deux sens, l’un

pour les télécommunicants (la partie finale de la distribution téléphonique) l’autre pour les informaticiens (le sens aujourd’hui reconnu). L’expression de «gestion des réseaux» est actuellement dans la même incertitude.

2) La traduction directe de notions complexes, même en utilisant des formules apparemment standardisées au plan international a révélé l’ampleur de ces écarts. Il n’y a pas toujours de vocabulaire international universel, même pour des sujets apparemment classiques.

3) Cela est particulièrement vrai dans le cas de domaines nouveaux. Il en a été ainsi par exemple pour le concept de «réseaux intelligents» (traduction : intelligent network) qui n’est pas encore une appellation contrôlée, et qui était même une formule totalement nouvelle pour certains des participants. Or c’est dans la phase d’émergence d’un nouveau concept que les besoins d’information sont les plus importants et l’exigence de précision la plus absolue.

4) Chacun s’impose naturellement un effort d’explicitation d’une notion complexe lorsqu’il intervient dans sa langue maternelle. Cet effort semble être allégé, voire gommé quand on renvoie à une langue étrangère. Elle est supposée, à tort, être la même pour tous. On se donne donc rarement la peine de donner une définition précise. On se contente de dire «ce que les anglosaxons appellent». Il y a là une sorte

de paradoxe car on utilise mal un outil moins performant (une langue qui n'est pas sa langue maternelle).

5) S'exprimer dans une langue compréhensible est aussi un souci de correction et d'effort vis-à-vis de l'autre. Au contraire, l'utilisation abusive de vocabulaires étrangers est, soit une solution de facilité, soit une volonté délibérée de faire sentir à son interlocuteur son incompetence sur le sujet.

6) Tous les termes n'ont pas la même valeur.

- Certains représentatifs de données physiques ou descriptifs de matériels me paraissent pouvoir être transportés sans problème d'une langue dans une autre. La langue française est particulièrement riche de ces apports extérieurs.
- En revanche, les concepts sont l'expression d'une pensée. Ils doivent donc être les plus clairs et plus purs possibles. Le souci du mot juste et de définitions non ambiguës ne s'accomode pas d'une langue que l'on maîtrise mal.

C'est donc sur ce critère que pourrait se faire un clivage entre ce qui est du domaine ouvert, libre et accueillant aux termes étrangers les plus commodes, et

ce qui est du domaine réservé à la réflexion et au raisonnement et qui doit s'appuyer sur une langue forte². ...]

Cette réalité que les concepts prédominent, sinon précèdent les mots pour le dire (dans les différentes langues naturelles, mais aussi langue de métier ou de spécialité disciplinaire) seront sous-jacents, sinon au cœur des interrogations de beaucoup d'auteurs de Lexipraxi. La notion de concept (proche parente de celle d'ontologie, de web sémantique, d'ingénierie multilingue) est ainsi au cœur des interrogations qui traverseront tant les thématiques de nombreux Lexipraxi et nécessairement de ceux qui traitent des TICE³ dès le début prise en compte par les auteurs de Lexipraxi mais aussi les organisateurs (AILF) et les conseils scientifiques des différents Lexipraxi.

Ainsi, dans cette même livraison 1990 Jean Herr dans sa courte introduction « Le multilinguisme à Telecom Paris » insiste sur la nécessaire dialectique entre l'enseignement, l'informatique, la maîtrise multilingue (tant à titre personnel qu'en tant que compétence d'ingénierie) mais surtout (et c'est un discours novateur à l'époque) la nécessaire attention à la langue française au cœur de ces enjeux.

Jean Herr ...[...]

Avant de m'exprimer en tant que directeur de l'École des Télécommunications, je dois vous avouer que j'ai une position radicale et affective en faveur

³ Terme utilisé ici avant la lettre car il ne sera introduit que beaucoup plus tard.

² Pierre Baylet, Apprendre et enseigner l'informatique, dans quelles langues ?

du multilinguisme. En effet, ma langue maternelle est l'alsacien et je n'ai appris le français qu'à l'âge de 7 ans...[...].

L'informatique me concerne également de fort près, puisque je dirige l'une des Écoles leader de ce domaine en France. À TELECOM Paris, faut-il le rappeler, l'informatique constitue l'un des pôles d'excellence, tant il est vrai qu'on ne conçoit plus d'informatique sans télécommunication, ni de télécommunication sans informatique.

De quelle manière TELECOM Paris prend-elle part au combat pour le multilinguisme et au-delà, pour la défense de la francophonie? : Par ses enseignants-chercheurs et leur rayonnement mondial - J.-A. Hernandez, Président de l'AILF n'est-il pas l'un d'entre eux;

Dans les Lexipraxis 98 et 2000 deux auteurs font particulièrement rebondir la question du « concevoir la transmission du savoir à l'ère numérique » : J.-P. Archambault et Didier Oillo. C'est avec des extraits de ces 2 auteurs que nous avons composé la très courte anthologie d'extraits sur ce thème. J.-P. Archambault ouvre la thématique en 1998⁴ en proposant une approche philosophique, culturelle mais aussi citoyenne de l'introduction des TICE à l'école. Son ancrage est résolument contemporain (Derrida, Vygotski) mais il aurait pu être celui de la similitude de démarche avec une approche pédagogique Freinet où l'informatique remplace l'imprimerie ou encore celle de Platon pour lequel l'innovation d'écriture est au cœur de l'interrogation citoyenne⁵.

ses élèves pour lesquels le multilinguisme est une réalité; sa recherche sur les langages naturels et la communication homme-machine. [...].

LP1998 - Lire, écrire, compter : les différentes influences des Tic

Jean-Pierre Archambault

...[... Les missions fondamentales de l'École demeurent : former l'homme, le travailleur et le citoyen. Mais elles s'inscrivent dorénavant dans une société profondément transformée par les TIC (technologies de l'information et de la communication). Les enjeux éducatifs de ces dernières sont de natures différentes et il faut bien les distinguer. Ils portent à la fois sur des contenus nouveaux à enseigner, sur le métier de l'École (au sens où l'on parle des métiers de la chimie ou de la métallurgie), c'est-à-dire la pédagogie, et sur les mutations des « lire, écrire, compter » eux-mêmes.

⁴ « Lire, écrire, compter : les différentes influences des TIC ».

⁵ La référence à la Grèce antique et plus particulièrement à Athènes comme ayant insisté sur la différence entre terme et concept est clairement affirmé dans le sigle OWL (à la fois la chouette en anglais mais aussi attribut tutélaire d'Athéna et symbole de la sagesse et du savoir et OWL, Ontological Web Language.

La culture générale du 21^e siècle

Le citoyen

«Les autoroutes de la communication aideront à créer des emplois à hauts revenus, à résoudre la crise du système de santé, à remettre en état le système éducatif et scolaire, et renforceront la domination technologique américaine» affirme en 1993 The United States Department of Commerce National Telecommunications and Information Administration (Direction des Télécommunications et de l'Information au Ministère du Commerce).

Renforcer la domination technologique américaine, on aurait plutôt tendance à le croire. Par contre, on serait plus dubitatif quant à l'aide annoncée à la résolution de la crise du système de santé. Mais quid du jugement et de l'intervention dans le débat citoyen de celui qui ignore tout des autoroutes de la communication, qui n'en a pas les connaissances et les représentations nécessaires.

Joël de Rosnay nous dit que «le contrat de travail enferme aujourd'hui la logique de la croissance dans une matrice à une seule dimension: du temps contre du salaire, que les règles traditionnelles d'unité de lieu, de temps et de fonction bloquent l'essor de l'éco-

nomie informationnelle, que l'on peut travailler à distance, que de nombreux salariés préféreraient une amélioration de la qualité de la vie à une augmentation de salaire...»⁶.

Si l'amélioration est réelle, si les liens entre les individus ne se distendent pas... mais il ne faudrait pas qu'en la circonstance Internet ait bon dos. Des questions sérieuses en tout cas. Quid encore de celui qui n'a pas les connaissances pour argumenter sur l'essor de l'économie informationnelle.

Un dernier exemple. Les réseaux favorisent des modalités de travail coopératif. Infrastructure pour une intelligence collective, sont-ils pour autant, comme le suggère Pierre Lévy, le fondement d'«un projet de civilisation synonyme de recréation du lien social par les échanges de savoir, de reconnaissance, d'écoute et de valorisation des singularités, de démocratie plus ouverte, plus directe, plus participative.»⁷.

L'observation du fonctionnement des groupes utilisant la messagerie électronique montre la nécessité de s'appuyer sur des réseaux humains préexistants. Le problème reste donc entier.

⁶ numéro spécial de *Manière de voir* (Le Monde Diplomatique) : Internet, l'extase et l'effroi.

⁷ éviter, par exemple, que le citoyen ne se retrouve pas dans la situation d'un interlocuteur de J. Lacan ignorant les nombres complexes ou le tore topologique ; voir *Impostures intellectuelles* de A. Sokal et J. Bricmont.

Ces questions, et bien d'autres, font partie intégrante du débat citoyen. Mais imagine-t-on un instant que l'on puisse débattre sereinement, sur un pied d'égalité si l'on n'a pas une solide culture en matière de TIC. Bien évidemment non. Il y a un risque bien réel d'avoir des citoyens de seconde zone car, citoyens de seconde zone en matière de TIC, (au même titre que l'on est citoyen de seconde zone lorsque l'on ne maîtrise pas la lecture, l'écriture ou le calcul). Pour conjurer ce danger pour la démocratie, le rôle de l'École est fondamental. Aux « lire, écrire, compter », s'ajoute désormais le « connaître les TIC ».

Le travailleur et l'homme

Les TIC font aujourd'hui partie intégrante de la culture générale de tous.

En tant que citoyen avons-nous vu, mais aussi en tant que travailleur. Les entreprises, industrielles et tertiaires, se sont profondément transformées, les métiers, les qualifications et les diplômes également. On est passé de la machine à écrire au TTX, du fichier carton à la BDD relationnelle, de la planche à dessin au DAO, de la lime à la FAO...

En tant qu'homme également, car tous les champs d'activité et la vie quotidienne sont, peu ou prou, modifiés par les TIC, de la réservation d'un billet de train à de nouvelles formes d'expression artistique.

Des contenus et des méthodes

Au même titre qu'on explicite le bagage mathématique du citoyen de l'an 2000 (calculs, fonctions, probabilités...), on explicite le bagage TIC du citoyen de la société informationnelle, en n'oubliant pas que l'ordinateur est avant tout une machine conceptuelle, dont l'utilisation n'est pas aussi simple que l'on veut bien le laisser entendre parfois (interroger une base de données par exemple), et qu'il est illusoire de penser pouvoir se dispenser d'un apprentissage structuré sur la durée.

Il faut énoncer (et distinguer) les objectifs (former l'homme, le travailleur, le citoyen de la société informationnelle), les contenus permettant de les réaliser (en termes de structures de données, de recherche documentaire, de configuration des réseaux, de dialogue des ordinateurs, etc.), les progressions didactiques (telle notion, abordée au collège, sera approfondie en classe de seconde puis en première) et les méthodes pédagogiques qui tiennent compte des élèves concrets que l'on a devant soi (leur niveau, leur motivation, leur section...).

Un outil pour apprendre à écrire

Les TIC ne constituent pas un outil miracle permettant à la société et à son système éducatif de répondre à l'exigence sociale croissante de formation et au défi d'un enseignement de qualité pour tous. Néanmoins, il serait dommageable de se priver de leurs apports réels et constatés pour aider les élèves (et leurs

enseignants) dans les apprentissages « du lire, de l'écrire et du compter ». Un exemple⁸.

L'enseignement de l'écriture au collègue

Les vertus vantées du traitement de texte - qui vont suivre - ne valent pas que pour les publics scolaires en difficulté mais pour tous les élèves. Pour s'en convaincre il suffit de penser à soi-même en train d'écrire un texte à l'ordinateur.

1) Écrire c'est réécrire, une banalité mais une lourde tâche quand on veut que les élèves « revoient leur copie ». Réécriture suppose relecture. Mais les élèves rechignent à le faire. Quelques annotations de l'enseignant ne suffisent pas. On obtient souvent au mieux quelques corrections orthographiques et de ponctuation.

Or il arrive que les élèves doivent se persuader qu'ils n'ont pas maintenu le dialogue implicite avec un lecteur (ils ont tu des données...), qu'ils ont insuffisamment fait la différence entre ce qu'ils voulaient dire et ce qu'ils ont réellement écrit, qu'ils ont mal perçu les registres de langue...

Des enseignants disent : lors d'un travail de réécriture, différents outils ont leur place (ordinateur, dictionnaire, stylo, grammaire...) et l'ordinateur a un apport réel dans la maîtrise des procédés de réécriture.

Déplacer un mot, une phrase, un paragraphe, corriger quelques fautes, recopier une nouvelle version issue d'un brouillon vite devenu illisible de par la multiplicité des modifications... tout cela est fastidieux et rédhibitoire s'il n'y a pas une forte motivation. Mais avec un traitement de texte, s'il faut repérer des répétitions ou mettre en évidence ce qui relève du langage parlé, l'enseignant peut demander de mettre les mots en caractères italiques. Erreurs, ratures, ajouts ne sont plus insupportables. La reprise est facile. On échappe à la lourdeur de la réécriture à la main. Une mauvaise graphie ne s'oppose plus à la lecture par les autres, une écriture illisible de par des troubles de motricité fine n'est plus un obstacle.

L'ordinateur se révèle être une condition (nécessaire?) d'existence d'opérations intellectuelles en ce sens qu'il en permet la réalisation en la rendant infiniment plus aisée, en en supprimant les contraintes « bassement matérielles ». Comme si la portée de l'outil était d'autant plus grande que son effet anodin.

2) Cliché certes, mais réalité soulignée avec force, le statut de l'erreur se transforme : la conservation aisée des états antérieurs permet à l'élève de voir les effets produits par ses tentatives successives, son travail et les résultats de ses efforts, et non de voir des échecs ! Une sortie propre à l'imprimante accroît

⁸ Médialog n°31, dossier Un ordinateur dans la classe de français.

naturellement les exigences (les fautes d'orthographe choquent davantage). La distance objective que procure l'ordinateur entraîne une lecture différente de ce que l'on a écrit. Un texte propre aide à donner une représentation positive et un modèle de l'écrit (voir le miroir de la danseuse). Le statut de l'erreur, que l'on fait disparaître comme par enchantement, change du tout au tout. L'ordinateur n'est pas perçu comme celui qui juge ou sanctionne. L'élève se trouve dans d'autres dispositions par rapport au nécessaire effort scolaire.

3) Autre fondation pour l'utilisation de l'ordinateur dans des activités d'écriture, le fait que l'on écrit pour quelqu'un. Or les élèves ont du mal à imaginer leur lecteur. Cette difficulté va de pair avec l'absence de recul par rapport à leurs propres écrits, l'absence de conscience de soi. Or, l'ordinateur devient même lecteur de ce qu'on lui fait écrire, instaurant l'image mentale du destinataire⁹.

La facilité de réécriture permet de multiplier les destinataires, d'aller au-delà des écrits scolaires, de sensibiliser aux registres de langue. On écrit à la manière de et alors la méconnaissance des codes devient transgression volontaire cautionnée par des auteurs reconnus.

Et les enseignants peuvent d'autant plus exiger des élèves un travail sur leurs textes que la production deman-

dée est socialisée dans la classe ou à l'extérieur. L'inscription des textes d'élèves dans une situation de communication claire (explicitant le but, le ou les destinataires, les enjeux de l'écriture) justifie les contraintes linguistiques.

4) On pourrait multiplier les arguments. Terminons par le cas du journal scolaire. La discussion dans la classe sur les articles des uns et des autres amène à considérer sa propre écriture comme un « texte littéraire », à comprendre que le scripteur peut agir sur le lecteur, à s'impliquer sur le fond et sur la forme, à accepter les contraintes, à s'intéresser à la mise en page, à la structure (titre, sous-titre, paragraphe, encadré), pour mieux distinguer l'essentiel de l'anecdotique.

Sociabilité et travail d'écriture (préciser sa pensée, l'organiser permet de l'approfondir) favorisent les apprentissages

Le bonheur pédagogique existe donc ! Il faut néanmoins attirer l'attention sur la nécessité d'aller au-delà de situations de communication qui s'usent, d'ancrer le plaisir d'écrire pour écrire, y compris des textes scolaires, de ne pas leurrer les élèves en leur laissant croire qu'ils auraient produit des chef-d'œuvres. En quelque sorte revenir à des activités scolaires après s'en être éloigné. ...]...

⁹ J. Derrida : « l'ordinateur entretient l'hallucination d'un destinataire ».

Suit un développement similaire à celui sur l'écriture qui analyse l'ordinateur comme un outil pour apprendre à compter autrement. En 1998, le concept de TICE est aussi au cœur d'un colloque qui porte comme titre: Former des citoyens pour maîtriser la société de l'information.

On y trouve une contribution de Claude Riso-Lévi qui fait état de l'importante contribution de Seymour Papert au domaine de l'éducation par l'électronique, mais aussi la contribution de J.-P. Archambault dans une session «Comment éviter la fracture

culturelle?» On relève aussi dans ce Lexipraxis de 1998, un article de Jean Mesnager: «Multimédias: de nouveaux modes de lecture?»

C'est surtout dans le Lexipraxis 2000 que la question des TICE fait l'objet d'une session complète¹⁰. L'utopie sociale¹¹, les enjeux sociétaux¹², la mondialisation¹³, le multilinguisme¹⁴, les normes¹⁵ des TICE y sont spécifiquement traitées. Encore avec J.-P. Archambault¹⁶ puis Didier Oillo¹⁷ que se déploient véritablement JPA d'abord avec cet article:

LP2000 - L'école publique, plus que jamais, pour de nouveaux savoirs

Jean-Pierre Archambault

...[...]

Professeur - mission veille technologique et industrielle.

L'intégration des Technologies de l'Information et de la Communication (TIC) dans les établissements scolaires se poursuit. On pense beaucoup aux TIC pour faire face aux défis nés de la massification de l'enseignement. Ce déploiement des nouvelles technologies

à l'École, entamé il y a une vingtaine d'années, interfère avec le processus plus récent de marchandisation de certains secteurs de l'éducation, à base de réseaux et de ressources électroniques. Dans l'un et l'autre cas, il est beaucoup question de pédagogie. Où technologies et pédagogie, à la fois, posent de vraies questions, constituent de réels apports et servent de chevaux de Troie.

On compte de l'ordre du million d'ordinateurs dans les établissements scolaires, avec des disparités selon les ordres d'enseignement, davantage de machines dans les lycées que dans les collèges ou les écoles, et les secteurs, avec une quasi-banalisation dans les séries techniques et professionnelles et un moindre développement dans l'en-

¹⁰ Les savoirs sur la toile.

¹¹ Les utopies actuelles sur l'acquisition de savoir sur la Toile, par Richard Walter.

¹² Enjeux sociétaux d'une économie du savoir sur la toile, par Michel Gensollen.

¹³ Quelle réalité à la mondialisation de l'apprentissage et du savoir par la toile? Par Louis Weber.

¹⁴ Une machine plurilingue pour l'acquisition des savoirs sur la toile, par Winfried Schmitz-Esser.

¹⁵ Normalisation des systèmes de médiation pédagogiques: vers l'évolution de l'édition sous forme d'information structurée, par Henri Hudrisier.

¹⁶ L'école publique, plus que jamais, pour de nouveaux savoirs, par Jean-Pierre Archambault.

¹⁷ Universités virtuelles, universités plurielles, par Didier Oillo.

seignement général. Il existe également, dans le cadre des lois de décentralisation, des différences géographiques dues aux politiques diverses de dotation des collectivités locales, même si grandit partout la conviction qu'il faut aller progressivement vers une incontournable généralisation.

Un processus d'intégration

Les opérations d'équipement s'accompagnent d'actions de formation diversifiées, de la mise à disposition de ressources électroniques et de la constitution de centres dont les personnels ont pour mission de favoriser l'intégration des TICE dans l'ensemble du système éducatif. Au fil des années, des usages se sont imposés, dans une espèce de darwinisme pédagogique. Ainsi, le traitement de texte fournit-il une aide appréciable à l'apprentissage de l'écriture¹⁸ et Internet à la motivation des élèves¹⁹. Les acquis sont indéniables. L'ordinateur fait désormais partie de la panoplie des outils à la disposition des enseignants pour exercer leur métier de pédagogue. À l'EAO des pionniers, a succédé une phase caractérisée par l'utilisation des progiciels (traitement de texte, tableur, gestionnaire de fichiers et de bases de données...) et du réseau pour l'accès à des ressources et des activités coopératives, avec un reflux de l'utilisation de logiciels disciplinaires spécifiques. Aujourd'hui, une diversité et une variété plus grandes des res-

ces électroniques disponibles, une qualité supérieure liée aux performances des machines font qu'elles se prêtent à tous les styles pédagogiques, aussi bien à des méthodes actives qu'à des démarches comportementalistes, à des utilisations individuelles ou collectives.

Mais, l'omniprésence des nouvelles technologies dans la société fait qu'il devient difficile de ne pas les considérer également comme un élément à part entière de la culture générale. À ce titre, elles constituent un objet d'enseignement, comme dans le cours de technologie au collège. Les TICE sont aussi un facteur d'évolution des disciplines générales et techniques que l'on enseigne à l'École, de leurs objets et de leurs méthodes. On les retrouve alors dans les sciences expérimentales (simulation, expérimentation assistée par ordinateur), les mathématiques (outil pour établir des conjectures), la géographie (cartographie), l'histoire (banques de données, traitements statistiques), les matières tertiaires (traitement de texte, bases de données), industrielles (DAO, CFAO...). Pour les disciplines techniques et professionnelles, il s'agit de former à des qualifications nouvelles et à des environnements de travail coopératif, avec des ressources électroniques à disposition qu'il faut savoir utiliser et mobiliser. Enfin, les enseignants se servent de l'ordinateur pour préparer leurs cours et, à l'instar des entreprises et des autres administrations,

¹⁸ Dossier « L'ordinateur dans la classe de français », Médialog n°31.

¹⁹ « Internet à l'École », Manière de voir « Internet, l'extase et l'effroi ».

l'Éducation nationale s'informatise pour son propre fonctionnement. Les TIC sont alors outil personnel de travail de l'enseignant et de la communauté scolaire dans son ensemble.

Scolarisation et coûts

Les statuts des TIC à l'École sont donc multiples. Ils engendrent des attentes diverses qui ne se démentent pas au fil des années. Ainsi, les travaux européens sur la « Société de l'information » et sur la « Société cognitive » (1995) concluaient-ils que *« pour préparer les Européens à relever les défis de la société de l'information, les systèmes d'éducation et de formation doivent se fixer un objectif ambitieux : offrir un enseignement de qualité tout au long de la vie, et satisfaire des besoins croissants et toujours diversifiés. Dans un contexte de restrictions budgétaires et de concurrence accrue, les institutions d'enseignement comme les entreprises doivent remplir cette mission à moindre coût. Une telle contrainte les oblige à rechercher de nouveaux outils éducatifs et à mettre en œuvre des méthodes pédagogiques adaptées. Le multimédia éducatif peut apporter des*

éléments de réponse à ce double défi par l'individualisation et la flexibilité de l'apprentissage qu'il permet ». Des objectifs éducatifs classiques, former les élèves à la société telle qu'elle évolue ou utiliser de nouveaux outils pédagogiques, en côtoient d'autres plus terre à terre comme celui de « l'impérieuse » nécessité de réduire les coûts de l'enseignement²⁰. Les TIC sont alors positionnées, d'une manière subreptice mais de plain-pied, dans le champ politique.

En effet, dans le monde entier, l'heure est à la prolongation de la scolarité, sous l'effet conjugué de la complexification croissante des sociétés et de la demande sociale. Tous les pays se trouvent ainsi confrontés à des défis nés de la massification de l'éducation, qui ne se confond pas complètement avec la démocratisation porteuse, elle, de l'idée de qualité. Pour sortir l'École de la « crise »²¹ qu'elle connaît, pour améliorer ses performances on se tourne vers les nouvelles technologies, non sans certaines ambiguïtés. Par exemple quand on en parle en termes d'efficacité sans que l'on discerne toujours très bien s'il s'agit de

²⁰ La période 1945-1975 a vu la part du financement public des dépenses d'éducation passer de 2,5 à 5 % du PIB (en France 5,8 %). Depuis 1975 on assiste à une stabilisation avec une convergence, à l'exception de la Chine et du groupe des 45 pays les plus pauvres ; avec le financement des familles et des entreprises, on arrive à 6,1 % du PIB, François Orivel, directeur de recherche IREDU-CNRS, Université de tous les Savoirs (30 août 2000).

²¹ L'école est toujours en crise, de par l'irréductibilité de l'écart entre l'idéal et la réalité. Ses hautes ambitions, émancipation, citoyenneté, égalité, solidarité... ne peuvent qu'être en décalage avec son fonctionnement et ses résultats. Autre chose est la crise d'un appareil de formation, sous l'effet du nombre, des tendances individualistes à l'œuvre dans la société alors que l'éducation reste fondamentalement un acte collectif, et dans la mesure où l'effort que consent la nation, les choix qu'elles opèrent en définitive, inscrivent l'investissement éducatif dans certaines limites.

qualité de l'enseignement ou d'efficacité économique. Les apports pédagogiques et didactiques sont incontestables. Mais, bien que les études réalisées sur les coûts économiques des TIC ne soient pas légion²², on sait que dans la classe un surcroît d'efficacité sur le plan pédagogique revient relativement plus cher, d'une manière non linéaire. Dans un premier temps, l'introduction des TIC représente des coûts fixes importants. Elle ne devient rentable que lorsque des seuils d'élèves ou d'étudiants sont franchis. Le cas de figure de l'enseignement à distance illustre à sa façon l'existence de ces deux notions distinctes d'efficacité qui relèvent de logiques différentes. Les dépenses qu'il occasionne sont moindres que celles de l'enseignement présentiel. Alors, même lorsqu'il donne des résultats scolaires inférieurs à ceux de l'enseignement traditionnel, il peut se justifier, par exemple si l'on veut toucher des publics larges, difficiles à regrouper, comme dans de nombreux pays d'Afrique, d'Amérique du Sud ou d'ailleurs. Le discours gagne en clarté lorsque l'on distingue clairement qualité pédagogique et hauteur de l'investissement éducatif, qui

appartiennent à des registres distincts de l'argumentation. Mais, supports de nouveaux services, les TIC donnent lieu à d'autres glissements « sémantiques ».

Des environnements éducatifs nouveaux

Fait relativement nouveau depuis quelques mois, il ne se passe de jour sans l'annonce de formes nouvelles d'organisation institutionnelle de l'enseignement permises par les technologies de l'information et de la communication²³. Des environnements pédagogiques apparaissent dans l'enseignement supérieur et d'autres modalités de formation se mettent en place avec les TIC pour les entreprises, les adultes, le grand public autodidacte, en présentiel ou à distance, à base de centres ressources, de services éducatifs, de logiciels et de cédéroms, de systèmes en réseau, d'autonomie et de tutorat²⁴. La liste est longue : plate-forme d'échanges en ligne « *afin d'optimiser les relations entre apprenants, formateurs et tuteurs en entreprise* » pour des promotions d'ingénieurs en alternance ou de techniciens supérieurs en maintenance²⁵ ;

²² François Orivel, Patrick Chevallier, Cnam, atelier du Préau, octobre 98 (la situation n'a pas radicalement évolué depuis cette date).

²³ Le premier Marché mondial de l'Éducation de Vancouver (mai 2000) a été de ce point de vue un moment privilégié.

²⁴ L'OCDE chiffre la part du financement privé des dépenses éducatives à 1,2 % du PIB ; on n'observe pas de convergence ; pour un premier groupe (USA, Japon, Corée du Sud...), les chiffres varient de 1,5 à 3 % ; pour le deuxième groupe, elle est de 0,5 % (la France se situe dans la moyenne) ; concernant le premier groupe, on observe une concentration sur le post-obligatoire et pas de financement significatif pour le primaire et le secondaire, François Orivel, Université de tous les Savoirs (30 août 2000).

²⁵ Outil développé par le CESI Normandie.

offre en e-formation destinée aux entreprises, avec l'objectif de « *leur fournir les moyens de maîtriser les coûts et les délais de développement et de déploiement de sessions de formation pour des utilisateurs à l'échelle mondiale* »²⁶; site Internet consacré « au soutien scolaire synchrone pour les élèves de la 6^e à la terminale » avec accès à des classes virtuelles ou à des cours particuliers²⁷; premières formations et conférences en français utilisant IeL (Interactive e-learning), « *technique interactive qui permet aux stagiaires de suivre une formation depuis le lieu de leur choix, et à l'heure de leur choix* »²⁸, programme de téléformation par satellite « Les paraboles du savoir » et campus électronique du CNED (créé en 1995); formation des chirurgiens à l'Institut de Téléchirurgie de Strasbourg où « *les 17 tables d'opération, reliées par un réseau à large bande, sont supervisables en temps réel par les enseignants qui peuvent intervenir soit oralement par le micro, soit au moyen d'un crayon optique* »²⁹, universités virtuelles aux USA ou au Sénégal...

Ces évolutions vont au-delà de la simple médiatisation technique de l'acte pédagogique. Elles relèvent d'une réorganisation des appareils de formation, privés et publics. Les nouveaux environnements concernent surtout la formation des entreprises, des adultes ou le monde universitaire, mais ils commencent à être proposés au grand public et à l'enseignement scolaire. Mais, « *le volume d'affaires réalisé par la formation sur Internet ou Intranet pour les entreprises s'élevait à un milliard de dollars en 1999, soit un tiers du marché total des formations utilisant les technologies, et devrait atteindre deux milliards en 2000 pour progressivement placer au second plan les autres technologies de formation* »³⁰. Dans les faits, les TIC sont devenues plus qu'un Cheval de Troie dans les processus de marchandisation de biens et de services éducatifs.

La fin de l'article détaille les questions de la marchandisation de l'Éducation et J.-P. Archambaut insiste sur la notion d'exception éducative.

³⁰ Étude 2000 « Choisir une solution de téléformation » réalisée par Le Préau, Aska et Klir.fr sur l'offre de plates-formes et portails de téléformation.

²⁶ Par Hewlett-Packard et Docent.

²⁷ Monprof.com ouvrira en septembre ou en octobre 2000; le contenu pédagogique se calquera sur le programme de l'Éducation nationale; l'objectif n'est pas de se substituer à l'enseignement classique mais d'offrir un complément.

²⁸ Par Global Knowledge; IeL utilise l'interactivité sur support vocal en temps réel entre formateur et apprenants, un tableau blanc grâce auquel le formateur peut projeter des diapositives, des schémas, les modifier instantanément ou annoter le texte; un chat-box permet aux apprenants de dialoguer entre eux et avec le formateur, par écrit; les formations sont enregistrées, ce qui permet d'accéder aux archives.

²⁹ voir dossiers de la Mission TIC du Premier Ministre (<http://www.mtic.pm.gouv.fr/dossiers>).

...[...Les technologues, dans la tradition du modèle industriel, et certains parmi les décideurs s'expriment volontiers en termes d'efficacité, les pédagogues, eux, d'utilité sur le long terme pour les élèves. L'éducation vise à construire des personnalités, à créer des œuvres qui durent, et non à produire des biens de consommation éphémères. Des institutions lui sont indispensables pour sa pérennité. Pour l'intégration des technologies de l'information et de la communication favorisant la massification et la démocratisation de l'enseignement, pour l'exception éducative, le rôle du service public reste irremplaçable et il serait périlleux de l'affaiblir...]

Dans le même Lexipraxis, Didier Oillo nous propose aussi un article qui conceptualise les TICE à l'ère de la mondialisation. Il intervient forcément avec l'expérience qui est la sienne, celle d'un responsable d'un réseau de campus numériques (une cinquantaine) réparti dans des établissements d'enseignement supérieur de la Francophonie : plus de 600 établissements membres répartis dans plus de 70 pays. Ce type de responsabilités intériorisées au quotidien par Didier Oillo le contraint obligatoirement à réfléchir à la question des TICE en terme de pluralité mais aussi en terme de virtualité numérique. Pour être utile l'AUF est condamnée à réussir un certain nombre d'enjeux prospectifs de l'EAD potentiellement développables grâce au numérique en réseau. Mais réussir cela nécessite dès lors de le conceptualiser, de le développer d'en donner les scénarios possibles de façon à préserver la pluralité des langues, des cultures, des institutions, des disciplines, des économies ou des disparités géographiques des différents territoires où sont situées les universités membres de l'AUF.

LP2000 - Universités virtuelles, universités plurielles

Didier Oillo Agence université de la Francophonie

Avant propos

Lorsqu'en 1993, le vice-président des États-Unis, Al Gore, annonce l'ouverture d'un nouveau chantier, les autoroutes de l'information, pouvions-nous imaginer que nous aborderions une révolution technologique touchant tous les secteurs de nos sociétés. Les universités, en particulier celles des pays les plus avancés ont apporté une forte contribution, en terme de recherche, à l'émergence des nouvelles technologies de l'information. Depuis l'invention du bouquet de protocole IP par Vinton Cerf en 1974, 15 années ont suffi pour créer les fondements d'une intelligence collective. Cependant, l'effort du secteur universitaire a surtout porté sur la recherche et le prototypage de nouveaux outils, laissant de côté leur socialisation. Les usages ont insuffisamment été étudiés. C'est le domaine de l'éducation et de la formation, l'une des activités essentielles des sociétés modernes, qui avait le moins bénéficié de l'apport des nouvelles technologies de l'information et de la communication. Aujourd'hui, le vocable Société du savoir remplace peu à peu celui de Société de l'information.

Ce secteur de la connaissance est aussi celui qui profitera ou subira, suivant l'orientation des mutations imposées par la mondialisation et la compétition internationale, des évolutions en contenus, en méthodes et en moyens.

Les années 1985 à 1995 ont vu croître la place des didacticiels dans le télé-enseignement, se substituant peu à peu à l'enseignement auto-dirigé ou à l'enseignement programmé ayant recours aux supports de cours traditionnels ou audiovisuels.

Les années 1995-2000 ont transformé et élargi ce processus d'évolution, en délocalisant apprenants et formateurs, en reconstituant les supports de cours naturels d'enseignement, en faisant transiter sur les réseaux de transmission de données les contenus stables et enrichis, en associant aux contenus magistraux la documentation requise, en permettant enfin, par correspondance électronique, par forum, les échanges entre enseignants, chercheurs et étudiants, en mondialisant le savoir.

Le contexte actuel en matière d'information est caractérisé par l'arrivée massive des solutions numériques qui bouleversent les réalités installées, par un risque majeur d'hégémonie d'une seule langue aux dépens du multilinguisme, d'une seule culture aux dépens de la pluralité, notamment à travers le développement très rapide de l'Internet, par un développement atomisé et à plusieurs vitesses, suivant les pays et les régions, des nouvelles technologies, et enfin, par une moindre visibilité des

contenus non numérisés pour des communautés de plus en plus larges. Parallèlement, les besoins vont croissant en informations, en méthodes et outils pour les produire, les organiser, les diffuser et les consulter.

La communauté scientifique est, rappelons-le, à l'origine de la plupart des solutions techniques et méthodologiques nécessaires à cette circulation de l'information, comme l'Internet et les standards associés, elle se doit aujourd'hui de réfléchir à ses usages et à sa pénétration inéluctable dans le monde de l'enseignement et de la recherche. Les expériences passées devront être analysées au regard de l'impact du numérique en particulier dans la circulation et la transmission du savoir. Le risque est grand de laisser le champ de la transmission des connaissances au marché.

Certains pensent que le concept d'industrialisation de l'enseignement est en cours de mutation profonde. Les bâtiments, les salles de cours, les amphithéâtres sont appelés à disparaître au profit d'espaces numériques, de lieux de savoir virtuel. Les enseignants seront-ils remplacés par des avatars numériques ou seront-ils happés dans le tourbillon du changement? Auront-ils tous la capacité de redéfinir leur rôle ou subiront-ils les mutations imposées par l'économie mondiale?

On doit considérer aujourd'hui la notion d'acte pédagogique, à l'instar d'un acte médical. Le pédagogue devient un médiateur du savoir.

Autant de questions auxquelles nous ne prétendons pas apporter des réponses. La reconfiguration de l'espace universitaire vers le concept d'université virtuelle mérite quelques précisions et amorces de réflexion.

De l'enseignement à distance à l'université virtuelle

L'éducation à distance a connu une histoire déjà longue de recherche et de réalisations, partant des études par correspondance jusqu'aux systèmes soutenus par l'audiovisuel à distance (en général basés sur satellite). Les NTIC sont en train d'apporter une révolution à cette discipline et devraient lui permettre de sortir des débats d'initiés et du scepticisme des pédagogues pour la transformer en industrie globale. Tandis

que la question de la qualité pédagogique de l'éducation à distance reste l'objet de nombreux débats contradictoires entre «traditionalistes» et «spécialistes», les NTIC commencent à apporter tous les éléments d'une véritable *révolution pédagogique* où les rapports enseignants-enseignés et enseignés-enseignés changent radicalement. Les concepts de «collaboration» et «d'enseignement asynchrone» commencent à s'imposer,

plus parce qu'ils sont le reflet des nécessités de l'évolution de la société, que pour des raisons purement pédagogiques. Cependant, ce changement porte les germes d'une véritable révolution pédagogique où les structures traditionnelles figées d'espace-temps-hiérarchie explosent. L'argument principal qui plaide pour ce changement est bien entendu économique; il faut être conscient toutefois que les tarifs de l'enseignement à distance restent encore aujourd'hui à la hauteur de ceux de l'enseignement traditionnel, les économies de masse permettant de rentabiliser un investissement initial supérieur, à tous points de vue, à celui de l'enseignement traditionnel, n'étant pas encore atteintes. Cela changera radicalement dans la phase de massification de cette technologie, qui pourrait intervenir après une période de transition de quelques années.

Il reste bien entendu de nombreux défis à relever:

- l'utilisation des différentes technologies dans les proportions et usages appropriés;
- la question des méthodes pédagogiques adaptées;
- la question de la production des cursus;
- la question de la propriété intellectuelle des contenus;
- la question de l'évaluation des enseignants et des enseignés;
- et, le plus important goulot d'étranglement: la formation (en particulier celle des enseignants) à une utilisation maîtrisée de ces outils nouveaux.

Le nombre de termes et expressions différents qui sont utilisés autour des concepts d'éducation à distance ou d'enseignement virtuel justifie un travail sémantique pour classifier, éviter les malentendus et distinguer les grands axes historiques et prospectifs.

- Les cours par correspondance: caractéristique des premières applications qui se contentaient d'utiliser la poste pour combler la distance.
- L'éducation à distance: la plus utilisée par les experts en pédagogie, existait avant l'essor des NTIC et semble résister aux changements technologiques. Le terme pourrait souffrir du retard technologique d'un certain nombre d'experts en pédagogie, mais il devrait maintenir son statut de terme de référence.
- Le télé-enseignement / télé-formation / télé-université: caractéristique de l'étape audiovisuelle (télévisuelle) de l'éducation à distance, celle-ci est fortement menacée par l'apport original des NTIC.
- L'université ouverte: met l'accent sur collaboration et globalité.
- L'université collaborative ou coopérative: met l'accent sur l'aspect collaboration, aussi bien entre enseignants et enseignés qu'entre partenaires pédagogiques.
- La formation asynchrone: met l'accent sur un des éléments fondamentaux du nouveau style pédagogique qui est rendu possible par la technologie, c'est-à-dire étudier à son propre rythme, sans impératif de temps.

- L'université globale : met en avant la vision géopolitique.
- L'éducation assistée par ordinateur (Computer mediated education) et la formation distribuée (Distributive learning) sont utilisées par les spécialistes de la technologie.
- L'université Internet : technologique et commerciale. Souvent une « méta université » qui vend un appareil logistique (logiciels et « www ») pour pouvoir produire des contenus pédagogiques destinés à être distribués à distance.
- L'université virtuelle : caractéristique de l'approche technologique avancée (Internet). Se veut un concept intégrateur du nouveau paradigme où l'espace physique et le besoin de synchronisme disparaissent. Toutefois, on peut apporter une restriction sémantique vers l'enseignement supérieur, alors que dans le nouveau paradigme, le terrain de l'enseignement a tendance à réduire les frontières entre les types d'études (primaire, supérieure, professionnelle).
- Le campus virtuel est une variante du précédent, peut-être avec une moindre ouverture.

Le concept d'université virtuelle est aujourd'hui le plus répandu. Il répond au besoin des universités traditionnelles d'offrir de nouveaux services, de trouver de nouveaux modes de relation avec les apprenants.

Trois termes souvent utilisés dans le même champ sémantique, n'ont pas encore trouvé de définition stable. Nous nous essayons ci-après à en cerner le sens dans le contexte de l'enseignement.

Électronique : qui fait appel aux couches basses de la technologie, plus proche du matériel physique et du transport de l'information.

Numérique : qui ressort du codage de l'information, qui est proche du message et donc du contenu.

Virtuel : vertueux étymologiquement, c'est le terme qui se rapproche le plus de l'individu, des groupes et de leur mode de représentation. On pourrait parler de couche haute.

Le concept d'Université virtuelle

- Les ingrédients de la réussite :
- l'utilisation des NTIC et une combinaison en « juste proportion » des différents outils technologiques dans la perspective d'un changement radical de l'équation du coût de l'enseignement ;
 - une pédagogie qui accompagne le nouveau paradigme technologique et permet :
 - un apprentissage qui peut être asynchrone,
 - une nouvelle relation entre les acteurs,

- une formation continuée tout au long de la vie;
- une vision participative de la formation; l'université virtuelle se trouve à l'intersection de trois espaces:
 - l'espace des possibilités,
 - l'espace des technologies,
 - l'espace de la pédagogie.

Dans l'espace des possibilités, nous pouvons distinguer une révolution économique démocratisante de l'éducation et la tendance à la disparition de la segmentation des formes d'éducation (primaire, secondaire, universitaire, pour adultes, professionnelle, finalement continue). Mais cette forme d'éducation a ses dangers, elle est porteuse de sa contradiction en ce que sa source émettrice n'est pas toujours bien identifiée, qu'elle ne respecte pas toujours des critères académiques et qu'elle tend vers une mercantilisation accrue des échanges.

Dans l'espace des technologies, les NTIC occupent une place au sommet de la hiérarchie, mais il est bon de comprendre que si ces technologies ne sont pas appliquées avec les pédagogies appropriées, elles n'apporteront rien d'autre à l'éducation que confusion et erreur d'objectifs. Par contre, si elles sont maîtrisées, le changement de paradigme qui les accompagne permettra une véritable révolution pédagogique. En premier lieu, elle sera marquée par l'abandon de la forme verticale de formation (transfert vertical de connaissances) au profit d'une forme en an-

neau où l'enseignant se transforme en facilitateur d'un processus centré sur l'enseigné et sa capacité de découvrir les connaissances à son propre rythme et en collaboration avec les autres enseignants et les facilitateurs. En deuxième lieu, cette nouvelle forme de pédagogie libère des contraintes d'espace et de temps, grâce à un mode de fonctionnement asynchrone. En perdant sa verticalité, l'apprentissage devient un processus de collaboration et de partenariats entre différents groupes.

L'espace des pédagogies est en pleine transformation. Certains apports récents ont été disqualifiés, d'une certaine manière, par les changements technologiques et une révision des concepts entre moyens et objectifs est en cours. Le grand défi est de concevoir une nouvelle pédagogie qui s'appuie sur les moyens technologiques et les transcende, sans pour autant prendre un point de vue de technologue ou de mythification des technologies. Le défi apparaît moins contradictoire quand les aspects sociologiques qui accompagnent la révolution technologique sont mis en avant.

Cette discipline, après plusieurs années de progrès lents et difficiles, est rentrée dans une phase de «poussée technologique» (technology push) qui arrive à maturité et permet une nouvelle phase de «conduite de marché global» (global market driven) qui voit accroître significativement ses parts de marché.

Les trois éléments clefs de la « poussée technologique » sont :

- ① Le changement pédagogique nécessaire, en particulier dans les relations étudiants-professeurs, étudiants-étudiants.
- ② Les ingrédients clefs pour le changement pédagogique: la recherche de l'interactivité et la proactivité du groupe (collaboration).
- ③ La détermination de la proportion judicieuse des moyens technologiques.

Ensuite, sur les bases obtenues, le marché commencera à orienter les choix dans plusieurs directions non nécessairement cohérentes :

- l'apparition de propositions académiques globales et à distance dans les principales langues (anglais, espagnol, français, suivis par les langues orientales et asiatiques), en

compétition sur le marché global de l'éducation avec une réduction progressive et sensible des coûts par la massification ;

- l'apparition d'un « bazar » de l'éducation globale au tout-venant où la démarche purement mercantile et la question de la propriété intellectuelle devienne critiques, à moins que l'évolution de l'Internet dans les cinq prochaines années commence à apporter des lignes de solutions.

...[... Didier Oillo examine alors les conditions concrètes de la mise en place des scénarios prospectifs envisagés et les méthodes d'expérimentation et d'observation et les conditions d'inventaire scientifique du domaine. Il établit des typologies qui à presque 10 ans de distance sont totalement opératoires...]...

En conclusion

L'imaginaire se nourrit des nouvelles métaphores produites par une communauté dont le lien socioprofessionnel n'apparaît pas en tant que tel, dont les frontières ne sont pas définies, porteuses de concepts peu solidifiés. C'est cependant dans ce nouvel espace que nous devons définir notre devenir d'enseignants, n'en comprenant pas toujours les règles. Les tentatives de normalisation de ces nouvelles méthodes de transmission de la connaissance par l'ISO sont également des facteurs

de troubles. L'inaction politique, face à ces enjeux, nous font craindre de nous retrouver exclus du débat et de devenir les instruments de nouvelles pratiques engendrées par le marché. L'université n'a jamais plus été en danger et pourtant elle doit saisir cette chance de devenir le nouvel acteur du développement des savoirs tout au long de la vie. L'université virtuelle doit être à même d'aider l'université classique à surmonter les obstacles qu'elle rencontre et à développer ses potentialités...]...

Le fil rouge de la conceptualisation des TICE rebondira ensuite dans les livraisons Lexipraxi 2005 & 2006 éditées sous la forme d'un ouvrage unique³¹ organisé sous la forme d'un «livre blanc»: TICE et téléenseignement francophone face aux défis stratégiques de la société de la connaissance. Nous ne retrouverons dans ce florilège que des articles confrontant TICE et Normalisation, ceux-ci se retrouvant dans l'axe développé ci-après (cf. Borde). Soulignons que Didier Oillo³² avait cité les normes et standards comme un des axes stratégiques du déploiement des universités virtuelles et plurielles. Il y revient avec plus de force dans sa contribution, comme Arbia Ouerqli³³ et Henri Hudrisier³⁴. Par contre si nombre d'autres auteurs de cette publication de 2006 rassemblant les Lexipraxi de 2005 & 2006 insistent sur la diversité linguistique (l'arabe³⁵, le berbère) un autre article³⁶ développe la

nécessité d'élaborer des terminologies normalisées comme base du développement des normes des TICE.

Avec maintenant deux ans de recul par rapport à cette dernière livraison de Lexipraxi nous ne pouvons que confirmer que l'avenir des TICE est vraiment étroitement lié à leur prise en compte conceptuelle et sémantique. En moins de 20 ans l'EAD et les TICE en général se sont complètement mondialisés et sont devenu un marché global dans lequel il est primordial de défendre l'offre francophone et celle des langues partenaires de la Francophonie.

La prospérité de la formation, tant académique que professionnelle, sera de ce fait étroitement dépendante des efforts faits par les chercheurs et les enseignants francophones pour défendre les concepts que sous-tendent les visions du monde tel que nous l'imaginons.

³¹ La langue française dans l'aventure informatique. TICE et téléenseignement francophone face aux défis stratégiques de la société de la connaissance.

³² L'AUF, un engagement dans la formation en ligne.

³³ Normalisation des métadonnées pédagogiques dans un contexte multilingue.

³⁴ Vers une société de connaissance partagée: «initiatives universitaires francophones dans la formation ouverte à distance: l'enjeu des normes et standards».

³⁵ Hocine Sadi: l'enseignement de la langue berbère en France et les TICE; mais aussi, Ahmed Boukouss, Francophonie, TICE et culture du partage - Le cas du Maghreb.

³⁶ Joseph Tientcheu & Bechir Boudir: Approche de la terminologie normalisée de la formation ouverte à distance (FOAD).

Conclusion et liaison avec le deuxième thème, normalisation :

Cet axe du florilège répond à une logique historique d'apparition des intérêts des chercheurs sélectionnés par Lexipraxi.

De 1990 c'est-à-dire avant le Web. On peut cependant supputer que les chercheurs des télécoms même si ils ne pouvaient pas raisonnablement supposer une telle rapidité de mondialisation du multimédia pouvaient d'ores et déjà anticiper l'importance à moyen terme de la généralisation du multilinguisme; anticiper aussi l'émergence probable d'une future gestion sémantique (donc conceptuelle des documents)

En 1998, puis 2000, soit six à huit ans après de démarrage du Web. Les scénarios prospectifs de 1990 sont devenus des réalités, qui reposent d'une façon non théorique la question de l'agora du village monde. Proposer de repenser en profondeur les fondements de l'apprentissage dans la mesure où il est devenu mondial, multilingue, en réseau et multidisciplinaire est devenu une nécessité. Les auteurs Lexipraxi l'ont bien compris et ils sont déjà pour quelques-uns complètement impliqués dans l'importante question de la normalisation des TICE, qui renvoie naturellement au troisième axe de notre Florilège.

Codage des caractères

Articles rassemblés et commentés par Jacques André.

La normalisation des caractères se réfère à la représentation codée de caractères, aussi n'est-il pas étonnant que nombres des articles des conférences Lexipraxis traitent du sujet du codage. Avant de les présenter, peut-être n'est-il pas inutile de rappeler succinctement ce qu'est un codage (voir [Marti 1990, André 2002]).

Les ordinateurs ne connaissent que les chiffres (ou plutôt les bits, chiffres binaires). Pour traiter des suites de lettres, il suffit de leur donner des numéros, appelés « codes », par exemple 1 pour A, 2 pour B, ... 12 pour L, etc. Mais il faut alors décider si 12 est la lettre L ou la suite des lettres AB. Pour cela les codages utilisent un nombre constant de chiffres. Par exemple avec 2 chiffres, on pourra coder ABL par 010212, mais on ne pourra coder que 100 caractères (de 00 à 99). C'est ainsi qu'on a eu des codages (les nombres comme 128 sont en fait liés aux nombres de bits) de 128 caractères (7 bits), 256 caractères (8 bits), etc. Mais alors, « comment coder plus de 128 caractères, ou plus de 256 caractères, etc. ? ». La réponse naïve, « prendre plus de chiffres », n'a pas été possible, pour des raisons historiques de taille des mémoires et

de temps de calcul, avant que ne naissent Unicode et ISO-10646 à la fin du XX^e siècle ! Toute l'histoire des codages est donc, principalement, la recherche d'une solution à cette question. Prenons le cas du « é » ; pas prioritaire pour entrer dans un codage (comme ASCII) à 128 caractères. On a imaginé alors plusieurs solutions : coder « é » comme étant la succession de « e » et de l'accent aigu « ´ », utiliser des tables annexes (comme avec les machines à écrire où à une touche peut correspondre 2 positions selon que le chariot est en position haute ou basse, par exemple « a / A », mais aussi « é / 7 ») ou des codes d'échappement (pour taper « Ê », on tape d'abord le circonflexe « ^ » puis le « E »), etc. Mais souvent, on utilise des codages de niveau supérieur, c'est-à-dire définis par une combinaison de caractères d'un codage plus simple. Ainsi, le langage HTML peut-il fonctionner pratiquement avec le seul codage ASCII, les caractères n'étant pas dans le jeu ASCII utilisant des combinaisons de ces caractères, par exemple « ´ » signifiant « é ». Mais d'autres codages utilisent pour « é » soit « =E8 » (Mime) soit « \'e » (TeX), soit... etc. Il est donc important de savoir de quel type de codage on parle !

Un codage est alors simplement une liste limitée de nombres, chacun correspondant implicitement à un caractère. Pour être efficace, il faut que ces codages soient reconnus internationalement par le maximum possible d'ordinateurs et correspondent aux besoins des utilisateurs et aux possibilités des ordinateurs (d'où pendant longtemps des nombres pas trop grands!)

Un peu de terminologie: on appelle *répertoire de caractères* un jeu complet de caractères abstraits qu'un système supporte, et *jeu de caractères codés* ou *codage de caractères* un ensemble de couples (*nom*, *valeur*) qui spécifie comment représenter un caractère de nom *nom* en utilisant un entier de valeur *valeur*.

Les conférences Lexipraxi ont donc eu la préoccupation constante de s'intéresser aux codages. Quatre articles sont repris ici:

- Dans le premier, Bernard Marti présente l'histoire de ces méthodes de codages montrant que des solutions souvent élaborées ont été mise en œuvre pour résoudre ces problèmes d'accès à des jeux

de caractères de plus en plus riches tout en minimisant la taille des répertoires.

- Dans le second, Sylvie Baste constate que les outils informatiques permettent de résoudre les problèmes de codage, mais ne suffisent pas. Elle répond alors à la question « Quelles exigences culturelles doit-on avancer pour rendre effective cette internationalisation ? »
- Le troisième article montre le pessimisme de Jacques André qui craint que les informaticiens, sous prétexte de compatibilité avec des codages antérieurs, ne freinent la diffusion d'Unicode.
- Enfin, Rachid Zghibi, s'intéresse à un cas très particulier, le codage des caractères berbères, et montre ainsi sur un cas très simple la difficulté de coder des langues dites minoritaires. Mais, avoir réussi à inclure les caractères tfinaghnes montre qu'Unicode s'intéresse bien au seul codage des caractères des langues sans la moindre connotation politique.

LP1990 - Évolution des méthodes de codage de caractères pour la prise en compte du multilinguisme

Bernard Marti, CCETT, Rennes (Lexi-praxis, 1990)

Au commencement, partout était l'anglais. Et encore, fallait-il l'écrire tout en majuscule. La capacité de reproduction des caractères des imprimantes était limitée à un répertoire d'une cinquantaine de caractères et tous les langages informatiques étaient limités à l'utilisation de ces quelques caractères.

Le premier code de caractères à usage de traitement électronique, l'alphabet international n° 2, ou code télex, avec 5 bits et un astucieux système d'échappement ne permettait effectivement que la traduction codée d'une soixantaine de caractères.

En 1963, le premier code normalisé pour le traitement de l'information est adopté par l'ISO après de laborieuses discussions qui ont duré 4 ans. En effet, ce code à 7 éléments ne permet que 128 significations et regroupe des fonctions de commandes et des caractères. 95 valeurs ont été conservées pour les caractères et cela est encore insuffisant pour faire autre chose que de l'anglais. Malgré cela, sur les signes auxiliaires, de nombreuses options possibles sont en débat. Un compromis est finalement trouvé. Le code international, dit version internationale de référence (IRV) de l'alphabet international n° 5 (IA5) comporte donc l'ensemble de 26 majuscules et

minuscules de l'alphabet latin simple, des chiffres, des signes de ponctuation et des caractères spéciaux. Parmi ceux-ci, un signe inconnu, dit caractère monétaire international, a été introduit à la demande de pays qui refusaient de voir le dollar considéré avec cet usage. À ce signe près, l'IRV, c'est la norme nationale américaine ASCII. À côté de cet IRV, la norme ISO 646 prévoit une table de base comportant des cases vides réservées aux options nationales : ce geste est la première reconnaissance officielle des nécessités de localisation des applications et c'est aussi l'origine de bien des ennuis. Chacun appelle donc ASCII sa version nationale d'une norme devenue ainsi floue et cause de multiples incompatibilités. Il y a cependant loin de ces versions nationales à un codage multilingue.

Les pionniers en la matière ont été les bibliographes du TC46 de l'ISO. Entre 1972 et 1980, ils définissent une série de 3 normes pour l'échange d'informations bibliographiques respectivement en alphabet latin, grec ou cyrillique. Ces normes comportent plus d'une idée originale.

D'abord ils utilisent la capacité d'échappement vaguement évoquée dans l'ISO 646 pour utiliser simultanément deux tables de code. Une table primaire qui, pour le latin est l'ASCII comporte l'alphabet de base, une ta-

ble secondaire comporte des symboles complémentaires et, seconde innovation, parmi ceux-ci, les signes diacritiques. Ainsi un caractère diacrité est codé comme la succession du code du signe diacritique et de celui du caractère porteur. Cette méthode est étendue aux caractères grecs et cyrilliques par la création des jeux de base nécessaires.

Ces idées originales donnent lieu à trois types de développements normatifs. Le premier c'est la norme des techniques d'extension qui permet d'utiliser simultanément jusqu'à 4 jeux de caractère, avec une diversification intéressante des techniques d'appel. Dans le système original pour passer d'un jeu à l'autre, il faut commuter les jeux par le code de sortie (SO) puis revenir au jeu d'origine par le code SI et enfin imposer la superposition de l'accent et du caractère en interposant le code BS (Retour arrière). Ainsi un caractère accentué réclame-t-il 5 octets de code. Dans le système d'extension nouveau de la norme ISO 2022, la technique du simple appel permet de n'utiliser qu'un seul code pour chercher un caractère unique dans un jeu accessible sans avoir à commander le retour. On réduit ainsi le code d'un caractère accentué à 4 octets.

On améliore encore la situation en introduisant enfin un codage à 8 éléments: l'élément supplémentaire permet de distinguer entre deux jeux et le codage d'un caractère diacrité tombe à trois octets.

Enfin, en 1977, les télécommunicants du CCITT décident d'adopter la méthode par composition pour le jeu de caractères codés pour la communication de texte. Ce travail est précédé d'une étude comparative de 39 langues européennes utilisant l'alphabet latin, permettant un inventaire exhaustif des caractères nécessaires. La norme 6937 est adoptée simultanément par l'ISO et le CCITT (sous le nom de T.51). En 1980, est constituée la base de codage de tous les services télématiques. La dernière amélioration apportée par cette norme est que l'accent flottant est considéré comme caractère sans chasse, c'est-à-dire de largeur nulle ce qui évite d'avoir à utiliser la fonction de retour arrière: un caractère *diacrité* est donc représenté en deux octets dans la version 8 bits de ce code.

Immédiatement, l'ISO essaie d'étendre ce mécanisme à des langues non latines mais des événements internes interrompent le processus. Celui-ci est repris dans un tout autre esprit en 1984. À ce moment le SC2 entreprend une norme multilingue concurrente de la 6937. Cette norme reprend l'idée de versions nationales en codage à 8 éléments. Un jeu de base, l'ASCII et un jeu de caractères complémentaires composés forment chaque partie de la norme. Chaque partie est affectée à une zone géographique. Six alphabets latins sont ainsi définis puis un alphabet latin-cyrillique, un latin arabe, un latin grec, un latin hébreu, tous basés sur le même principe: chaque caractère est codé sur

un seul octet. Bien adapté au traitement local, cette méthode ne permet pas la communication multilingue stricte-sensu. C'est ainsi, pour la petite histoire, que le projet initial comportait 4 alphabets latin, un pour l'Europe Occidentale, un autre pour l'Europe Orientale, un pour l'Europe du Sud, un autre pour l'Europe du Nord. Après cela, les Turcs, classés géographiquement à l'Est, ont considéré que leurs besoins de communication étaient plus forts avec français, italiens ou allemands qu'avec les lettons ou les polonais. Pour ce faire, un alphabet latin n°5 a donc été établi. Enfin, un alphabet complémentaire comportant tous les caractères oubliés (dont le couple œ - Œ du français) a été publié récemment (Latin-9; sur l'oubli du œ, voir [André 1996]). Comme on le voit, cette méthode d'ailleurs non approuvée par le France, ne permet pas réellement la communication multilingue. C'est cependant celle à qui l'actuel rapport de forces à l'ISO donne la faveur et les travaux sur les versions non latines du 637 ont été stoppés.

À l'autre extrémité de la gamme des travaux de l'ISO, a été entrepris en 1987 un véritable travail de titans : établir un inventaire exhaustif de tous les caractères de toutes les écritures du Monde ainsi qu'un codage sur plusieurs octets de ces caractères. Déjà le codage systématique sur deux octets est utilisé par chinois et japonais pour le codage de 6 à 7 mille idéogrammes. Chaque idéogramme est représenté par deux codes ASCII et la capacité du système

multi-octet est donc de $95 \times 95 = 9025$ caractères par table. L'extension à plus de deux octets ouvre donc des possibilités fabuleuses mais..., à quel prix ! Cependant, la baisse du prix des mémoires rend cette méthode praticable et, devant les lenteurs de l'ISO à sortir sa norme 10646, des standards de fait (Unicode) basés sur les mêmes principes commencent à apparaître.

Depuis, Unicode et ISO-10646 sont devenus deux répertoires « compatibles ». Leur méthode de codage reste simple et, grâce à quelques principes (unification, composition dynamique, etc.), rigoureux et ouvert (voir [Andries 2002] pour une histoire des principes de codage d'Unicode et [Andries 2008] pour une introduction à ce codage).

En conclusion : le multilinguisme dans les applications informatiques est de plus en plus à l'ordre du jour. Les usagers attendent des normes et lorsqu'elles tardent à venir, ils font des standards. Au niveau français, il faut mettre en avant la faiblesse du dispositif : le comité français correspondant ne comporte que 4 personnes, ce qui est insuffisant pour faire face à la fois au codage de caractères et à celui des images et des sons, ainsi qu'à assurer les contacts internationaux nécessaires. Je terminerai donc ce rappel de travaux en faisant un appel au peuple : engagez-vous, reengagez-vous, le comité national n°2, chargé de ces travaux au plan national manque de bras et de cerveaux, ce qui a, entre autres conséquences,

celle d'affaiblir le statut de la langue française dans ce mouvement accéléré de recherche des solutions aux besoins de la communication multilingue.

Hélas, l'appel au peuple de Bernard Marti n'a pas été entendu et nombre de particularités de l'écriture de la langue française ont quelque peu été oubliés dans les codages comme Unicode [André-Hudrisier, 2002].

LP2000 - Exigences culturelles et internationalisation

Sylvie Baste, Doctorante – Conservatoire National des Arts et Métiers (Lexipraxis 2000)

La représentation des langues dans les technologies de l'information et de la communication (TIC) passe par la présence des systèmes d'écriture qui transcrivent ces langues.

L'existence d'une norme et d'un standard tous deux internationaux, et proposant une architecture de table multilingue destinée à contenir les caractères du monde entier, atteste une prise de conscience. ISO/IEC 10646-1 et le standard Unicode sont des systèmes logiques où chaque caractère se voit attribuer un code qui lui-même ne correspond qu'à un caractère. Les caractères sont clairement et sémantiquement définis. La convertibilité est un autre principe du Jeu Universel des Caractères (JUC) qui contient tous les standards nationaux en vigueur.

Toutes ces caractéristiques concourent à l'internationalisation des pratiques et outils informatiques mais ne suffisent pas. Quelles exigences culturelles

doit-on avancer pour rendre effective cette internationalisation? Comment ces attentes sont-elles perçues par les instances de normalisation?

Après une définition sommaire de l'internationalisation et un exposé sur la rencontre entre l'écriture et les TIC, nous répondrons à ces questions.

Principes de l'internationalisation

Le concept d'internationalisation a pris son essor, après la Deuxième Guerre mondiale, au moment où les échanges commerciaux se sont développés grâce à l'ouverture des marchés.

Une nouvelle étape est aujourd'hui franchie. Les modifications des modes de gestion de production induites par les TIC entraînent des changements qui ne touchent pas seulement les habitudes de consommation. L'internationalisation des échanges s'accompagne et est dépendante d'une internationalisation des méthodes et produits informationnels. À tel point qu'il existe maintenant dans les comités techniques de normalisation ou de standardisation prenant en charge les technologies de l'infor-

mation, des groupes de travail voués à l'internationalisation³⁷.

Ces efforts visent à rendre accessibles les éléments culturels présents non seulement dans les données textuelles numérisées mais aussi dans les processus de codage informatique.

Il convient d'insister sur le rôle d'Unicode et de la norme ISO/IEC 10646. L'internationalisation est basée principalement sur cette norme et sur ce standard qui sont fondamentaux.

Les TIC à l'épreuve de l'efficacité scripturale

Aucune écriture n'a plus de valeur qu'une autre. Cette évidence est presque difficile à énoncer. Il ne nous viendrait pas à l'esprit de déprécier ou mépriser un tant soit peu un des aspects les plus marquants de la culture, le fait qu'elle s'exprime. La langue et l'écriture sont deux instruments de cette expression.

En ce qui concerne les systèmes d'écriture, que devient cette évidence quand on sait que seul l'alphabet latin (à travers l'anglais surtout) est vraiment représenté ?

Une des réponses à cette interrogation est présente dans le mot instrument. L'écriture est contenu mais elle est aussi contenant. Elle est un moyen. Sur ce plan, on lui demande d'être efficace, on juge sa performance en tant que mode de communication particulier.

Le code ASCII capable de donner une représentation et un traitement de l'alphabet latin non accentué était efficace au début de l'informatique. Les données textuelles étaient en anglais et d'autres langues pratiquées dans des pays industriels pouvaient se satisfaire de ce système.

Depuis, les choses ont évolué. Mais, contrairement à ce que l'on aurait pu croire, pas dans le sens de l'adoption unanime de l'anglais comme langue universelle de communication dans les technologies de l'information. Au contraire, un jeu universel des caractères est né d'un consensus inespéré entre les différents acteurs de la normalisation internationale (ISO, IEC, ITU, IETF, W3C...) d'une part, et, plus surprenant, entre les principaux constructeurs informatiques (IBM, Xerox, Apple), d'autre part.

On peut s'étonner en effet de l'existence du standard Unicode ou de la norme ISO/IEC 10646-1, quand on comptabilise le nombre de normes nationales ou internationales et de standards privés fabriqués dans le domaine du codage des caractères graphiques. Les industriels, la plupart du temps, cherchent à capturer des marchés en proposant des produits dont ils ont le monopole alors que les usagers et les exploitants encouragent la concurrence entre les fournisseurs [Marti 1990].

³⁷ Le groupe de travail WG20 du sous-comité SC22 est en charge de l'internationalisation dans le comité technique JTC1 (*Joint Technical Committee 1*) commun à l'ISO et à l'IEC. Le W3C tient compte de l'internationalisation dans toutes ses recommandations.

On peut comprendre qu'un niveau si fondamental de l'informatisation des écritures n'ait pas suivi le même processus que d'autres standards informatiques. La restitution de l'écrit est une nécessité qui subordonne l'utilisation de l'outil informatique lui-même. La table multilingue s'est imposée comme un impératif, un cadre pour codifier les écritures. Sans ce cadre consensuel, rien n'était possible en aval. De plus, la mondialisation des échanges informatiques est une réalité qui aboutit à la circulation des écritures du monde entier sur réseaux.

Mais force est de constater que le codage informatique des caractères n'est pas à l'abri de la tumultueuse standardisation. La liberté nécessaire à la conquête de nouveaux marchés ne s'exerce plus dans la problématique possibilité de créer des standards de jeux de caractères et de méthodes de codage mais dans la possibilité de proposer des produits à de nouvelles demandes : les attentes multilingues et celles concernant la localisation³⁸ de produits. Or, logiquement, ces requêtes n'aboutissent sur les tables de travail des industriels que si elles émanent de personnes ou de groupes capables de les financer ou de les acheter.

Le standard Unicode a sa propre vie à côté de la norme ISO/IEC 10646, alors qu'ils se définissent de la même façon. C'est un standard qui a pour voca-

tion de rassembler et coordonner plus que de monopoliser une technique. Cependant, il faut comprendre que malgré l'avantage évident que cela implique, il peut subsister un inconvénient majeur : ne pas prendre en compte tous les besoins culturels.

Exigences culturelles vis-à-vis de l'internationalisation

Les attentes culturelles vont au-delà de l'efficacité quasiment technique de l'écriture. On peut, dans une approche scientifique, considérer l'écriture comme une invention, même si ce terme n'est pas admis par tous les spécialistes. On constate alors que les systèmes d'écriture sont le résultat de bien des tâtonnements pour arriver à des systèmes alphabétiques simples et économiques après un passage par des systèmes pictographiques, idéographiques et syllabiques. Notons que toutes les écritures n'ont pas suivi cette évolution et pas de la même façon lorsque cela a été le cas. « L'écriture n'a pas été inventée en une seule fois, en un seul lieu. » [Calvet, 1996].

La diversité des systèmes d'écritures dans le temps et dans l'espace, révèle une autre propriété des mécanismes qui ont fait surgir l'écriture. « La grande richesse de la créativité humaine se manifeste aussi dans cette pluralité de solutions apportées au défi constitué par la fugacité de la parole. » [Calvet, 1996].

³⁸ La localisation est le procédé qui prend en charge la transformation d'un produit afin de l'adapter à des consommateurs (à leur langue, mode de vie, goût ...) d'un pays donné.

Considérer une évolution linéaire de l'écriture est une option qui comporte un écueil : se concentrer uniquement sur le système d'écriture occidental constitué de l'alphabet latin. Obnubilé par ce système, on risque de mépriser les autres, surtout ceux qui en sont très éloignés. Le fait de considérer l'écriture comme le support graphique du langage impose immédiatement une hiérarchie selon laquelle l'écriture doit être au service de celui-ci, de la manière la plus économique qui soit. On en arrive à considérer que la forme la plus évoluée de l'écriture est l'alphabet [Gaur, 1987]. Or, en représentant le langage parlé, l'écriture alphabétique possède ce que l'on pourrait considérer comme un inconvénient. Elle ne permet pas l'internationalisation aisée de l'écriture puisque celle-ci n'est alors pas indépendante de la langue.

L'histoire de l'écriture apprend que l'écriture est composite, qu'elle résulte de notions de quantité, de pratiques de calcul, d'échanges inter-ethniques, d'activités religieuses, de quêtes spirituelles etc., et que les postures, les instruments et les matériaux supports d'écriture ont joué un rôle dans la graphie.

L'éthnocentrisme appliqué à l'alphabet n'est pas récent. Il représente un risque que l'on peut difficilement courir quand on veut parler d'internationalisation. La première exigence culturelle à soutenir est que toutes les écritures soient représentables. Finalement, il s'agit là, peut-être, de la seule véritable exigence à formuler parce qu'on ne peut pas faire de choix préalable. Cependant,

cette attente se décline sous des formes multiples à l'image de la culture dans sa pluralité.

Les implications culturelles de l'internationalisation

Les énormes possibilités offertes par les TIC attirent tout le monde, et c'est légitime. Par exemple, il y aurait un paradoxe à faire vivre un Internet où les systèmes d'écriture et les langues ne seraient présents qu'en nombre limité. Cela poserait la question de la nature même du « grand réseau planétaire » sur lequel reposent tant d'espoirs.

Les écritures véhiculent des langues. Si on ne peut pas dire que les écritures sont le reflet d'un environnement culturel donné, puisqu'il n'y a environ que 25 systèmes d'écriture dans le monde [Malherbe, 1983], les innombrables langues sont souvent revendiquées comme telles. Il s'agit ici de laisser vivre la diversité culturelle, à l'abri des études comparatives. La comparaison des écritures ou des langues permet de bien comprendre les mécanismes scripturaux ou linguistiques. Elle comporte pourtant un danger : valoriser une écriture ou un groupe d'écritures et avec elles, valoriser peut-être une manière de penser. L'enjeu est que les hommes puissent s'approprier et poursuivre l'invention de l'écriture à l'ère de sa numérisation. Cette invention se nourrit entre autres de l'interculturel.

Il est important d'encourager les revendications de la part de ceux qui ne sont pas encore en position de revendiquer l'informatisation de leur langue.

Un retard par rapport à ces questions creusera davantage les inégalités économiques et sociales.

La plupart du temps, nous utilisons le réseau Internet pour chercher une

réponse à une question ou au moins tentons-nous de trouver des solutions. Dans cette démarche nous avons tout à gagner de la présence d'interlocuteurs nombreux et variés.

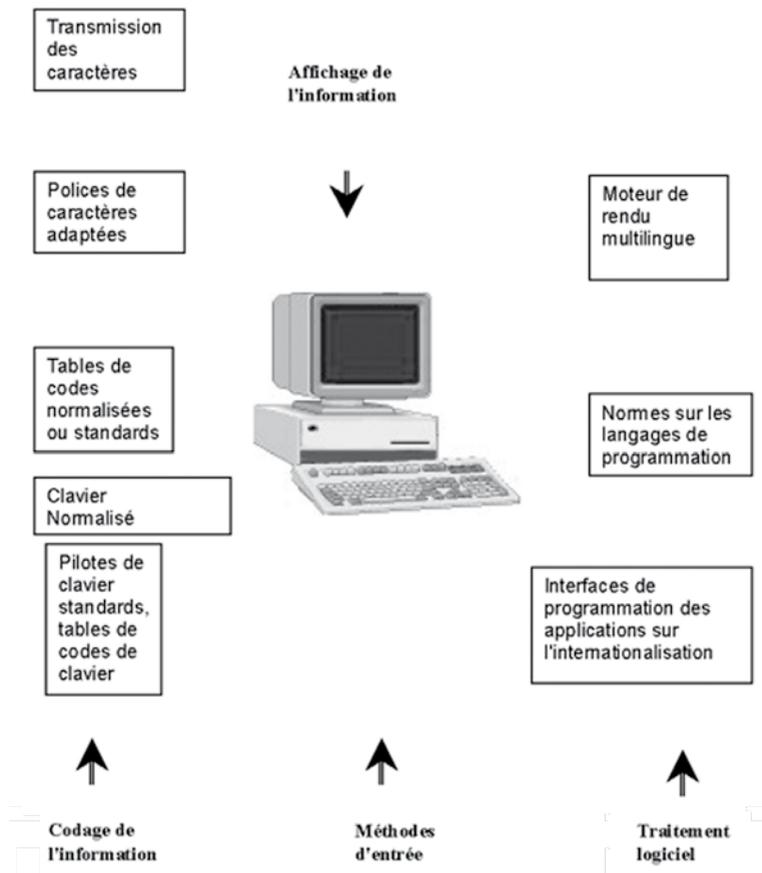


Figure 1 : niveaux d'intervention pour la définition d'un poste multilingue.

Traduction des exigences culturelles en exigences techniques

L'exigence culturelle primordiale selon laquelle toutes les écritures et les langues doivent être informatisables peut se décomposer en une multitude d'exigences qui touchent aussi bien le codage de l'information, que l'affichage de cette information, que les méthodes d'entrée et que le traitement logiciel. La figure suivante met en évidence les éléments qui définissent un poste multilingue.

Le codage des caractères correspond à un niveau élémentaire d'intervention. Grâce à Unicode et ISO/IEC 10646, les caractères et les méthodes de codage sont définis. Les problèmes de prise en charge des règles d'écriture peuvent être gérés par des programmes (moteurs de rendu³⁹) qui, s'ils sont basés sur la table multilingue, répondent au critère d'interopérabilité. Ce critère est nécessaire aux échanges de données multilingues. En revanche, il n'existe pas encore un réel consensus autour des polices de caractères.

Les groupes de normalisation et standardisation chargés de l'internationalisation proposent, par le biais de l'élaboration de convention ou l'adoption de résolutions, des solutions pour mettre en œuvre le jeu universel des caractères. Citons pour exemples quelques documents qui émanent de l'ISO [ISO/IEC, WG20] :

0. ISO/CEI TR 10176: 1998 - Guide pour la préparation (disponible actuellement en anglais seulement) de normes sur les langages de programmation.

1. ISO/CEI TR 11017: 1998 - Cadre pour l'internationalisation.
2. ISO/CEI 15897: 1999 - Procédures pour l'enregistrement des éléments culturels.
3. ISO/CEI 14651: Classement international de chaînes de caractères.

Ces normes sont destinées à des applications qui vont du traitement des jeux de caractères au formatage de la date, en passant par le tri et la recherche de données.

L'exigence culturelle première est de respecter tous les systèmes d'écritures et de considérer qu'ils ont la même valeur. D'un point de vue technologique, ceci suppose de considérer plusieurs niveaux de mise en application sur les plates-formes. Le travail collaboratif international doit permettre de partager un inventaire des problèmes posés et d'y apporter des solutions aussi universelles que l'ISO/IEC 10646-1 sur la base des propositions Unicode. Les producteurs de logiciels libres constituent également une sorte de groupe de pression. Ils préconisent l'adaptation multilingue, puisqu'il est dans l'essence même de leurs produits d'être utilisables par le plus grand nombre, à court et à long terme.

³⁹ Le moteur de rendu est capable de traiter différents aspects de l'écriture tels que la directionnalité, la combinaison de caractères, des changements de formes selon le contexte, le tri... [Bottin, 2000].

LP2000 - De Babel Web à Orléans : pluri-linguisme, Internet et recherche universitaire

Jacques André, Directeur de recherche – IRISA / INRIA Rennes (Lexipraxis, 2000)

L'étude des divers systèmes d'écriture, de leurs origines à nos jours, révèle qu'il n'y a pas d'écriture achevée. Le propre de l'écriture est de s'inscrire dans un processus de transformation par les supports, et par la main de l'homme. L'écriture est contrainte par les besoins d'échanges, libérée par la littérature, soumise à des changements qui semblent parfois aléatoires. En entrant dans les nouvelles technologies, elle suit toujours ce processus.

Ce savoir, à l'ère de l'informatisation, est une base précieuse pour forger la maîtrise du codage électronique des écritures. Il nous permet d'envisager les enjeux et la mise en pratique de la standardisation du codage des caractères. En somme, les enjeux sont ceux posés par l'écriture, notamment ceux qui concernent l'accès au savoir et l'expression de la diversité culturelle. Des langues meurent chaque jour et avec elles disparaissent des dispositions à la connaissance et, indirectement, à la vie.

Les progrès technologiques conduisent à reconsidérer le concept d'écriture. On ne demande plus à l'écriture de représenter simplement le langage mais aussi de faciliter l'accès à l'information et l'appropriation de celle-ci.

Les supports informatiques et le codage qui les accompagne sont les

nouveaux supports de l'écriture. Ils apportent des changements et imposent des choix dans les manières d'écrire et d'éditer les textes qu'il faut comprendre et gérer d'une façon universelle en espérant que les efforts de normalisation tendront vers une efficacité de l'information, plutôt que vers une homogénéisation des contenus.

Quelles recherches faut-il encore faire pour transmettre de l'information plurilingue sur Internet? On a envie de répondre: « plus aucune! »:

- Grâce aux contributions Unicode, officialisées par la norme ISO/CEI 10646, on sait manipuler tous les caractères usuels du monde (présents ou passés comme les hiéroglyphes, et il y a encore de la place pour de futurs caractères!) et même les mélanger (avec les diverses règles d'écriture de droite à gauche ou de gauche à droite, voire de haut en bas). Des travaux ayant évalué le nombre de caractères discernables créés par l'homme à environ 2 millions, on parlera bien ici des caractères usités de la vie courante et du commerce international.
- Grâce à la galaxie XML (qui, de façon normative, a choisi Unicode comme codage des caractères), il n'y aura bientôt plus de problèmes de transmissions de ces caractères.
- Grâce à divers outils en cours de développement (par exemple le concept de « textes parallèles »), voire de dif-

fusion (Omega, OpenType, etc.), on sait éditer, afficher, comparer, indexer, etc. des documents multilingues.

Certes, il y a encore à faire, notamment en ce qui concerne l'utilisation de ces textes, c'est-à-dire la compréhension des textes ou le passage de l'un à l'autre mais de nombreuses équipes de recherche linguistique « computationnelle » travaillent sur ces problèmes et on est en droit de penser que des progrès significatifs seront publiés prochainement. Mais savoir les transmettre c'est déjà beaucoup !

Tout va pour le mieux et l'on va vraiment vivre d'ici peu dans un village plannétaire multilingue.

Hélas, cette Tour de Babel sur le Web où chacun se comprendra est victime d'une rumeur d'Orléans et le plus curieux est que celle-ci se propage dans un milieu, celui des scientifiques (et pas seulement français), où on s'attendrait à un peu plus d'esprit critique et de rigueur. Le fait n'est pas nouveau : il y a une vingtaine d'années, la rumeur a circulé dans les comités de rédaction des grandes revues scientifiques qu'il fallait dire dans chaque article où on citait, par exemple, Unix que ce système était une marque déposée de la Bell, ou qu'il fallait écrire « le langage Ada[®] ». Au lieu d'en vérifier le bien fondé (qui n'existe pas), les scientifiques du monde entier se sont mis à propager, à informer leurs collègues de cette obligation. La baudruche a mis très longtemps à se dégonfler, sans doute grâce, entre autres, à un article en montrant le ridicule [Nedgin

and Bworn, 1984], mais elle subsiste encore. Autre exemple : dès qu'un virus est annoncé sur Internet, tout le monde, y compris nombre de scientifiques, se croit obligés de propager cette alerte, sans la vérifier, polluant plus les réseaux que le virus fantôme.

Quelle est donc cette « rumeur d'Orléans » liée au plurilinguisme ? C'est de dire que

- *Sur les réseaux, hors de l'Asciipasdesalut.* Bien sûr, aucun scientifique n'est concerné, mais on a le cœur sur la main : « tu comprends, il y a des gens qui n'ont pas les mêmes matériels que nous ; tu sais, les Américains ils n'ont pas toujours les lettres accentuées, etc. » Alors, on nivelle par la base. Et on ajoute deux axiomes :
 - *Pourquoi accentuer des mots français puisque de toutes façons on peut être compris ?*
 - *Pourquoi ne pas écrire en anglais puisque c'est la langue moderne de l'Internet ?*

Il suffit de regarder n'importe quelle liste de discussions, forum, groupe de nouvelles, etc. pour trouver de très nombreux exemples où ces trois principes sont explicitement énoncés, recommandés, etc. Certes, certains groupes tentent de montrer que oui on peut utiliser Latin 1 (par exemple fr.usenet.8 bits), d'autres imposent Latin-1 et corrigent les fautes d'accentuation (bien sûr les listes sur la typographie ou la langue française, mais il y en a d'autres) voire imposent Unicode (comme ce groupe

de discussion sur le grec ancien). Mais hélas, pour la majorité des forums scientifiques, le français non accentué est de règle. Dans le tout récent groupe francophone sur XML (et qui, je le répète, utilise Unicode comme codage normal), il est très fréquent de voir des phrases comme « Ca a l'air OK. Reste à savoir si... ».

En mathématiques, on apprend qu'il y a deux choses compliquées: passer de 1 à 2 et de n à l'infini. Je crains qu'il en soit de même pour le transfert plurilingue par Internet: passer de 1 (l'anglo-américain) à 2 (le français) est très difficile. Alors avant de passer à n ... De nombreuses études, tant sociologiques, que linguistiques ou politiques ont été publiées (une des meilleures synthèses regroupant ces divers aspects se trouve dans [Anis, 1999]) montrant que l'emploi du français sur Internet n'a ni la place ni la qualité que son statut de langue internationale pourrait laisser espérer.

Or, depuis longtemps (depuis ISO Latin 1), on sait mélanger français et anglais sur la toile, mais beaucoup de chercheurs pensent que non! Alors on imagine la difficulté de mélanger arabe et français, voire hébreu, ogam et hiragana. Non pas la difficulté technique (je le répète, grâce à XML et Unicode on sait faire!) mais la mauvaise volonté, le refus aveugle de croire à une autre langue internationale que le seul anglais. Alors oui, il y a des recherches à faire encore pour le plurilinguisme sur Internet: Jacques Anis dit « C'est le rôle des chercheurs en sciences humaines et en sciences cognitives, linguistes, psychologues, sociologues, pédagogues, philosophes de développer des recherches qui contrebalancent des publicités mensongères » [Anis, 1999, p. 182]. Et j'ajouterai « et de comprendre et faire comprendre à leurs pairs et confrères que l'ont peut et que l'on doit parler d'autres langues sur Internet que l'anglais et l'ASCII »!

LP2006 - UNICODE et ISO/CEI 10646 – le codage informatique des langues minoritaires : le berbère

Rachid Zghibi, Université de Tunis (Lexipraxis 2006)

L'article de Rachid Zghibi commençait par une présentation du concept de norme et faisait l'historique des normes ayant conduit à ISO 10646 et au standard Unicode. On retrouvera l'essentiel de cette information dans sa thèse [Zghibi, 2004] et dans [Andries, 2002].

Le berbère appartient aux groupes des langues chamito-sémitiques [Cohen, 1981]. « *Qualifié de langue essentiellement orale, le berbère est surtout une langue à alphabets multiples* ». En effet, Sur le plan des systèmes d'écritures, trois graphies sont en usage actuellement pour noter les langues berbères : les caractères arabes, les caractères latins et les tfinagh.

Les Tfinagh

« *Tfinagh (féminin pluriel) a pour singulier Tafineq « caractère, signe, lettre d'écriture », mot peu usité. Par tfinagh, les Touaregs désignent l'ensemble des caractères de leur écriture* ». Cependant, il convient de distinguer les *tfinagh authentiques*, originelles (en usage chez les Touaregs sahariens d'Algérie, en Libye, au Mali, au Niger, au Burkina-Faso et dans la diaspora touareg vivant au Nigeria, au Ghana, au Côte d'Ivoire, au Soudan, au Tchad) des *néo-tfinagh* (en usage en Occident et en régions berbérophones).

Les tfinagh originelles sont héritées des Anciens. D'après Mohamed Aghali-Zakara [1994], il existe une douzaine voire plus d'alphabets tfinagh selon les parlers. Les signes constituant ces alphabets comportent des variants et des invariants. Depuis les années 60, les tfinagh d'origine sont employées dans les journaux locaux et dans les centres d'alphabétisation des adultes au Mali et au Niger.

Les néo-tfinagh sont une création artificielle de l'« Académie Berbère ». L'objectif de l'Académie était d'adapter les tfinagh traditionnelles à la notation du kabyle, dialecte qui représente de nombreuses particularités phonétiques. Le nouvel alphabet de l'Académie Berbère comporte 37 signes dont 21 signes sont nouveaux par rapport à l'alphabet de l'Al hagggar.

Les caractères arabes

Le contact du monde berbère avec la civilisation musulmane est culturellement marqué par l'introduction des caractères arabes pour noter les parlers berbères. « *L'écriture arabe correspond à une deuxième période civilisationnelle, celle de la conquête arabe du Maghreb : les Arabes ont eu la tentation d'arabiser l'écriture du berbère pour faciliter la pénétration de la religion et de l'administration* ». En effet, à partir de XVI^e siècle, certains lettrés ont utilisé les lettres arabes pour noter des textes berbères notamment au Maroc. De même que

Alphabet arabe de base		Alphabet berbère (Académie Berbère)		Alphabet berbère (Système TEX berbère)		Alphabet berbère (IRCAM)	
1	ا	1	ا	1	ا	1	ا
2	ب	2	ب	2	ب	2	ب
3	ت	3	ت	3	ت	3	ت
4	ث	4	ث	4	ث	-	-
5	ج	5	ج	5	ج	4	ج
		-	-	6	چ	-	-
6	ح	6	ح	7	ح	5	ح
7	خ	7	خ	8	خ	6	خ
8	د	8	د	9	د	7	د
9	ذ	9	ذ	10	ذ	-	-
10	ر	10	ر	11	ر	8	ر
11	ز	11	ز	12	ز	9	ز
		12	ژ	13	ژ	10	ژ
		-	-	-	-	11	ر
12	س	13	س	14	س	12	س
13	ش	14	ش	15	ش	13	ش
14	ص	15	ص	16	ص	14	ص
15	ض	16	ض	17	ض	15	ض
16	ط	17	ط	18	ط	16	ط
17	ظ		-	19	ظ	-	-
18	ع		-	20	ع	17	ع
19	غ	18	غ	21	غ	18	غ
20	ف	19	ف	22	ف	19	ف
21	ق	20	ق	23	ق	20	ق
		21	ڤ	-	-	-	-
22	ك		-	24	ك	21	ك
			-	-	-	22	ك'
			-	25	گ	23	گ
			-	-	-	24	گ'
23	ل	22	ل	26	ل	25	ل
24	م	23	م	27	م	26	م
25	ن	24	ن	28	ن	27	ن
26	ه	25	ه	29	ه	28	ه
27	و	26	و	30	و	29	و
		-	-	-	-	30	و (avec accent circonflexe)
28	ي	27	ي	31	ي	31	ي
						32	ي (avec accent circonflexe)

Tableau n°2 : Les alphabets berbères

certaines tribus touaregs utilisant l'alphabet de leur propre culture, tels les Kel Antessar du Mali, ont tenté de vocaliser les tfinagh à l'aide de signes notant les voyelles brèves arabes «a», «i» et «u».

Pour rendre certains phonèmes propres à la langue berbère mais qui sont absents dans la langue arabe, de nouvelles lettres ont été créées à partir des lettres arabes de base modifiées par l'adjonction des signes diacritiques. Ces nouvelles lettres varient d'un système à un autre. À des fins de comparaison, nous présentons dans le tableau suivant le système de l'Académie berbère, le système TEX berbère [Harlambous, 1994] et le système de l'Institut Royal de la Culture Amazighe du Maroc (IRCAM).

Les caractères latins

Les premières tentatives de notation du berbère avec l'alphabet latin ont été réalisées depuis le XIX^e siècle, notamment par les missionnaires, les militaires et divers chercheurs occidentaux. Plusieurs alphabets ont été définis avec beaucoup de divergence. Un travail d'harmonisation a commencé avec la conférence internationale de Bamako en 1966 et s'est poursuivi par celle de 1984. Les alphabets élaborés, puis améliorés pendant ces deux conférences, ont été officiellement adoptés au Mali et au Niger dans les services d'alphabétisation des adultes et dans les écoles primaires expérimentales en Touareg. L'alphabet latin adopté lors de la conférence de Bamako en juin 1984 est le suivant :

a, ä, b, d, ð, e, ð̄, f, g, Ÿ, h, ħ, ç, i, j, k, l, l̄, m, n, ñ, n̄, o, q, r, °, s, s̄, š, t, t̄, u, w, x, y, z, z̄, ž.

Cette liste reste ouverte pour noter tout autre phonème découvert ultérieurement.

Le codage informatique des alphabets berbères par Unicode et ISO/CEI 10646

Dans cette section nous identifierons les codes réservés par le standard Unicode et la norme ISO/CEI 10646 pour coder les caractères latins et arabes utilisés pour transcrire le berbère ainsi que les caractères tfinagh

Caractères latins

Les caractères latins définis pendant les deux conférences internationales de Bamako en 1966 et en 1984 pour transcrire le berbère sont codés :

- au niveau des blocs de caractères *Latins de base* (Rangée 0, positions: U+0020-U+007E) et *Latin-1 Supplément* (rangée 0, positions: U+00A0- U+00FF);
- au niveau des blocs de caractères *Latins étendus A* (Rangée 1, positions 0100-017F) et *Latins étendus B* (rangée 1, positions 0100-017F);
- au niveau du bloc de caractères *IPA Extensions* (Rangée 2, positions 0250-02AF);
- au niveau du bloc de caractères *Latins étendus additionnels* (Rangée 1E, positions 1E00-1EFF);

Le tableau suivant identifie l'ensemble de caractères latins utilisés pour transcrire le berbère avec leurs positions exactes dans les différents blocs de caractères Unicode et ISO/CEI 10646.

<i>Caractère berbère</i>	<i>Code Unicode</i>	<i>Caractère berbère</i>	<i>Code Unicode</i>
a	U+0061	ɲ	U+0272 (IPA)
ă	U+0103 (latin A)	ŋ	U+019E (latin B)
b	U+0062	n	U+006E
d	U+0064	o	U+006F
ð	U+00F0 (Lat1 suppl.)	q	U+0071
e	U+0065	r	U+0072
ø	U+018F (latin B)	°	U+00B0 (lat1 suppl.)
f	U+0065	s	U+0073
g	U+0067	ş	U+1E63 (latin add.)
γ	U+0194 (latin B)	š	U+0161 (latin A)
h	U+0068	t	U+0074
ħ	U+1E25 (latin addit.)	ţ	U+1E6D (latin add.)
ç	U+00E7 (latin1 sup.)	u	U+0075
i	U+0069	w	U+0077
j	U+0069	x	U+0078
k	U+006B	y	U+0079
l	U+006C	z	U+007A
ḷ	U+1E37 (latin add.)	ẓ	U+1E93 (latin add.)
m	U+006D	ž	U+017E (latin A)

Tableau n°3 : Les caractères berbères latins dans Unicode

Caractères arabes

Dans la version 4.0.0, le standard Unicode et la norme ISO/CEI 10646 réserve 1026 codes pour coder l'écriture arabe de base ainsi que les différentes écritures dérivées de l'arabe de base. Les caractères sont regroupés en blocs de caractères au niveau de la rangée 6, FB, FC, FD et une partie de la rangée FE du Plan Multilingue de Base (BMP). Les rangées FB, FC et FD (U+FB50-U+FDFF) sont appelées *Formes de présentation arabes A* et la partie arabe de la rangée FE (U+FE70-FEFF) est appelée *Formes de présentation arabes B*.

Toutes les lettres arabes modifiées pour noter les phonèmes berbères qui n'ont pas d'équivalents dans la langue arabe sont codés dans Unicode et ISO/CEI 10646 excepté les lettres n° 22, 24 et 32 de l'alphabet berbère de l'IRCAM.

Le tableau n°4 identifie les lettres arabes de base avec leurs valeurs hexadécimales dans Unicode et ISO 10646 et le tableau n°5 identifie les lettres arabes modifiées :

N ^o	Caractères arabes	Valeurs Unicode
1	ء	U+0621
2	ا	U+0627
3	ب	U+0628
4	ت	U+062A
5	ث	U+062B
6	ج	U+062C
7	ح	U+062D
8	خ	U+062E
9	د	U+062F
10	ذ	U+0630
11	ر	U+0631
12	ز	U+0632
13	س	U+0633
14	ش	U+0634
15	ص	U+0635
16	ض	U+0636
17	ط	U+0637
18	ظ	U+0638
19	ع	U+0639
20	غ	U+063A
21	ف	U+0641
22	ق	U+0642
23	ك	U+0643
24	ل	U+0644
25	م	U+0645
26	ن	U+0646
27	ه	U+0647
28	و	U+0648
29	ي	U+064A

Tableau n° 4 : les lettres arabes

Les tifinagh : addition de l'écriture tifinagh au répertoire de l'ISO/CEI 10646 et le standard Unicode

Le 26 janvier 2004, une proposition d'ajout de l'écriture tifinagh au répertoire de la norme ISO/CEI 10646 et du standard Unicode a été formulée par Patrick Andries, Français Yergeau et Alain La Bonté auprès de l'ISO/CEI/JTC1/SC2/GT2. La proposition reprend les symboles tifinagh retenus par l'Institut Royal de la Culture Amazighe du Maroc (IRCAM), qui sont au nombre de 33. Cette liste n'est pas définitive et elle pourrait être complétée par de nouveaux caractères. Ceci a été fait en juin 2004 [Andries 2004]. C'est une écriture non cursive et non contextuelle. Elle s'écrit de gauche à droite et elle emploie les signes de ponctuation conventionnels qu'on retrouve dans les écritures latines. Il est préconisé d'utiliser les chiffres arabes occidentaux.

Tableau n° 5 : Les lettres arabes modifiées pour noter le berbère

N ^o	Caractères arabes	Valeurs Unicode
1	ژ	U+0698
2	ڤ	U+06A4
3	چ	U+0686
4	گ	U+06AF
5	پ	U+0695
6	ڤ	U+06C9

Rangée 2D : TIFINAGHE

	2D3x	2D4x	2D5x	2D6x	2D7x
0	◦	⊕	≠	Δ	
1	⊖	∅	!	⊔	
2	⊕	⋮	∩	∫	
3	⋈	∧	∴	⋈	
4	⋈	∩	○	↑	
5	⋈	⋈	⊙	⋈	
6	⊥	∴	∩		
7	∧	∩	∴		
8	∩	⋮	∴		
9	∩	∩	⊙		
A	∩	∩	⊙		
B	∩	⋈	⊙		
C	∩	≠	∩		
D	∩	∩	⋈		
E	∴	∩	⊙		
F	∩	∩	∩	∩	

Clé	
	Tifinaghe Ircam de base
	Tifinaghe Ircam étendu
	Autres lettres néotifinaghes
	Lettres touarègues modernes attestées
	Réservé pour un codage ultérieur

G = 00
P = 00

proposition d'ajout de l'écriture tifinaghe

Bibliographie commune aux quatre articles

- [Aghali-Zakara, 1994] Mohamed Aghali-Zakara, « Graphies berbères et dilemme de diffusion : interaction des alphabets latin, ajami et tifinagh », *Études et documents berbères*, n° 11, 1994, p. 107-121.
- [Ameur, 1993] Muftaha Ameur, « Diversité des transcriptions : pour une notation usuelle et normalisée de la langue berbère », *Études et documents berbères*, n° 11, 1994, p. 20-32.
- [André, 1996] Jacques André, « ISO Latin-1, norme de codage des caractères européens ? trois caractères français en sont absents ! », *Cahiers Gutenberg*, n° 25, novembre 1996, p. 65-77. <http://www.gutenberg.eu.org/pub/GUTenberg/publicationsPDF/25-andre.pdf>
- [André, 2002] Jacques André, « Caractères, codage et normalisation – de Chappé à Unicode », in [André-Hudrisier, 2002], p. 13-50. http://www.cairn.info/article.php?ID_REVUE=DN&ID_NUMPUBLIE=DN_063&ID_ARTICLE=DN_063_0013
- [André-Hudrisier, 2002] Jacques André et Henri Hudrisier (sous la direction de), Unicode, écriture du monde ?, numéro spécial de Document numérique, vol. 6, n° 3-4, Éditions Lavoisier+Hermès, ISBN 2-7462-0594-7, 2002. http://www.cairn.info/article.php?ID_REVUE=DN&ID_NUMPUBLIE=DN_063
- [Andries, 2000] Patrick Andries, *Unicode et ISO 10646 en français*, <http://hapax.qc.ca/> (mise à jour en continu).
- [Andries, 2002] Patrick Andries, « Entretien avec Ken Whistler, directeur technique du consortium Unicode », in [André-Hudrisier, 2002], 2002, p. 329-351. <http://www.cairn.info/revue-document-numerique-2002-3-page-329.htm>.
- [Andries, 2004] Patrick Andries, « Proposition d'ajouter l'écriture tifinaghe », ISO / IEC JTC 1 / SC 2 / WG 2 N2739, 21 juin 2004.
- [Andries, 2008] Patrick Andries, *Unicode 5.0 en pratique – codage des caractères et internationalisation des logiciels et des documents*, Dunod, 2008.
- [Anis, 1999] Jacques Anis, *Internet, communication et langue française*, éditions Hermès, 1999.
- [Baste, 2002] Sylvie Baste, « Enjeux de la normalisation internationale de la numérisation des écritures et langue française, appropriation d'Unicode », in [André-Hudrisier, 2002], p. 285-328.
- [Benlakhdar, 1994] M. Benlakhdar, « Écrire le berbère : une nécessité scientifique ou pratique », *Études et documents berbères*, n° 11, 1994, p. 19-23.
- [Bottin, 2000] Michel Bottin. *Écritures du Monde*. Semaine de la langue française et de la francophonie : les normes et le plurilinguisme. 29 mars 2000, AFNOR.
- [Calvet, 1996] Louis-Jean Calvet, *Histoire de l'écriture*, Hachette littératures, Paris, 1996.
- [S. Chaker, 1994] S. Chaker, « Pour une notation usuelle à base tifinagh », *Études et documents berbères*, n° 11, 1994, p. 31-42.

- [Cohen, 1981] Marcel Cohen, « Les langues chamito-sémitiques », *Les langues du monde* (Ed. Slatkine), 2 vol. p. 85-87.
- [O.Durand, 1994] O.Durand, « Promotion du berbère : problèmes de standardisation et d'orthographe : expériences européennes », *Études et documents berbères*, n° 11, 1994, p. 7-11.
- [Gaur, 1987] A. Gaur, *A History of writing*. Londres, British Library. 1987.
- [Haralambous, 1994] Yannis Haralambous, « Un système TeX berbère », *Études et documents berbères*, n° 11, 1994, p. 40-57.
- [Hudrisier, 1999] Henri Hudrisier, « Le document numérique normalisé XML, TEI, Unicode : nouvelles automatisations de l'intelligence et opportunités de prospérité : enjeux pédagogiques, nouvelles organisations de la recherche en ingénierie du document », *Colloque AUEPLF: Universités virtuelles, vers un enseignement égalitaire*, Edmundston, 26-30 août 1999.
- [Hudrisier, 2002] Henri Hudrisier, « La bibliothèque virtuelle berbère », *Études et documents berbères*, n° 21, 2002, p. 9-23.
- [ISO 10646] *International Standard ISO/CEI 10646-1: information technology-universal multiple-octet coded character set (UCS): part1 architecture and basic multilingual plane*, Genève: ISO, 1993, p. 3.
- [ISO-IEC, WG20] ISO/IEC, JTC1 SC22/WG20 Internationalization, Official home page, <http://anubis.dkuug.dk/JTC1/SC22/WG20/>.
- [Malherbe, 1983] M. Malherbe, *Les langages de l'humanité*, Éd. Seghers, Paris, 1983.
- [Marti, 1990] Bernard Marti et coauteurs, *Télématique - techniques, normes, services*, Dunod, 1990.
- [Nedgin and Bworn, 1984] P.J. Nedgin and T.L. Bworn, « CLOG: An Ada Package for Automatic Footnote Generation in UNIX », *Communications of the ACM*, 27 (4), April 1984, p. 351.
- [Simonsen, 1992] K. Simonsen, RFC 1345: *Character mnemonics et character sets*, juin 1992, 103 p. <http://www.ietf.org/rfc/rfc1345>
- [Unicode] Consortium UNICODE. *Specification for the Unicode Standard*, Version 5.1.0, <http://www.unicode.org/versions/Unicode5.1.0/>
- [Zghibi 2000] Rachid Zghibi, « Unicode et ISO/CEI 10646: introduction au codage informatique de l'écriture arabe et des alphabets berbères (latin et tifinagh) », *Études et documents berbères*, n° 18, 2000, pp. 87-100.
- [Zghibi 2002] Rachid Zghibi, « Le codage informatique de l'écriture arabe: d'ASMO 449 à Unicode et ISO/CEI 10646 », *Document Numérique*, in [André-Hudrisier, 2002], p. 155-182.
- [Zghibi 2005] Rachid Zghibi, *Normalisation des TIC dans le contexte arabophone*, Thèse de doctorat, Université de Paris 8, 6 avril 2005.

Articles « Normalisation », rassemblés et commentés par Jean-Michel Borde.

Nous avons fêté, dans une assez grande indifférence l'an passé, en 2007, les 20 ans du JTC1 de l'ISO et de la CEI, pourtant créé à l'époque avec le retentissement de la polémique. Ce comité technique commun a permis d'établir les fondamentaux de l'informatique ouverte, telle que nous la pratiquons banalement avec l'évidence du quotidien, aujourd'hui. À l'époque, la perspective semblait... révolutionnaire, pour certains, révoltante pour d'autres. En effet, l'horizon économique annonçait un avenir marqué par la fin de bien des monopoles, la remise en question du rôle de l'état, en particulier dans les télécommunications. Les normes et la demande de compatibilité informatique n'étaient pas beaucoup la préoccupation des industriels, à l'époque. Il n'en allait pas de même de leurs clients. Il devenait en effet de plus en plus difficile de trouver des gammes de fournitures totalement intégrées, couvrant toute la demande des besoins d'utilisateurs de plus en plus exigeants. Comment les satisfaire alors? C'est le dossier du « multimédia numérique » qui mit tout le monde d'accord. On peut être bon, voire excellent dans beaucoup de performances mais comment maîtriser des décennies de connaissances et de maîtrises professionnelles dans le domaine de la photo, de la musique et du film, surtout quand ce n'est pas votre cœur de métier? Le rôle le plus essentiel du JTC1 aura été tant, de pacifier tous les acteurs de ces différents marchés, que de leur permettre d'établir d'incontournables autant que précieuses passerelles entre leurs différentes cultures et spécialités. Si la performance

fut poly-technique elle fut aussi et surtout affaire de femmes et d'hommes, exemplairement déterminés. Nous avons compté parmi ceux-ci certains de nos membres. Que ce florilège soit un modeste hommage à des efforts méconnus, peu récompensés ou rarement honorés, à une représentation française qui n'a jamais démerité. Bien au contraire, il faudrait reconnaître à notre culture une maîtrise conceptuelle d'exception, qui aura servi de garde-fou à certaines perspectives socio-économiques auxquelles nous aurons ainsi pu échapper.... Parce que tous nos membres ont peu ou prou été acteurs en normalisation internationale, il était normal que nos colloques, leur donnent accès à leurs tribunes. C'est pourquoi cette thématique est l'un des trois piliers de ce florilège.

Nous avons très tôt, dès 1990, donné la parole à l'AFNOR, organisme qui fédère les positions françaises. Dans l'article partiellement reproduit ci-après, Jean Hyenne, alors chef du service technologies de l'information et applications, nous rappelle quelques fondamentaux et singularité de la discipline. En effet, l'une des résolutions fondatrices du JTC1 était l'abandon du sacro saint plurilinguisme natif des organisations internationales. Rappelons que les directives de ces organisations ne citent jamais le terme « traduction » puisque les travaux étaient contractuellement développés en plusieurs langues de travail, le russe le français et l'anglais, à l'ISO. Pourquoi n'avoir retenu que l'anglais pour l'informatique? On imagine l'intérêt de la question pour une association comme la nôtre.

LP1990 - Exposé pour le colloque Lexipraxis 1990

Jean Hyenne Chef du Service Technologies de l'Information et Applications, AFNOR.

Le rôle de l'afnor

L'AFNOR est l'association française de normalisation dont la mission est de piloter le système français de normalisation, de le représenter dans les instances de normalisation mondiales (ISO) et européennes (CEN). C'est elle qui décide de la publication des normes françaises, souvent élaborées dans des bureaux de normalisations sectoriels. Sa deuxième mission consiste à promouvoir et développer l'usage de la normalisation en particulier par les moyens de l'édition technique, de la formation aux normes, de la certification d'entreprise ou de produits. Les normes sont des documents techniques de référence élaborés collectivement par toutes les parties intéressées. À ce titre, elles capitalisent le savoir faire de l'ensemble du tissu industriel. Il n'est pas indifférent que les travaux de normalisation commencent souvent par déterminer un vocabulaire qui permet ensuite d'élaborer les spécifications sur lesquels les spécialistes se mettront d'accord.

D'une certaine manière, on peut dire que la normalisation développe une langue, un dialecte spécifique à des ensembles de professionnels, nécessaire pour assurer une compréhension sans ambiguïté. La capacité à générer un tel dialecte à partir de notre langue naturelle, le français, est une question par-

ticulièrement délicate dans les technologies de l'information (étendue à celles des télécommunications). C'est en effet le seul secteur technique dans lequel l'AFNOR pourrait se voir contrainte, à publier des normes en anglais. Mais il n'est pas exclu que la sensibilisation des milieux concernés puisse conduire à un sursaut et permette à l'AFNOR de continuer sans faillir sa tâche de mise à disposition du plus grand nombre des technologies contemporaines. C'est l'objet de mon exposé auprès de vous aujourd'hui.

État de la situation

Au 31 décembre 1989, la collection des normes françaises dans le domaine des Technologies de l'Information (TI) comprenait 293 normes, totalisant 10 600 pages environ, essentiellement en français. Seules 6 normes sont totalement ou partiellement en anglais. En 1990, environ 60 normes nouvelles ou révisées seront publiées, représentant près de 2 500 pages également intégralement en français. Dans les T.I., l'élaboration des normes se fait en quasi-totalité aux niveaux international et européen. Compte tenu de la stratégie mondiale des constructeurs et de l'internationalisation croissante des marchés, il n'y a pratiquement pas de normalisation purement nationale sauf dans quelques cas ponctuels où il apparaît intéressant de consacrer une position originale nationale par une norme française avant de la proposer

pour une validation européenne ou internationale. Les organisations internationales de normalisation l'ISO/CEI JTC1 et CCITT publient leurs normes en deux langues: l'anglais et le français (soit sous forme de 2 versions différentes, soit sous forme bilingue).

L'AFNOR assure notamment le secrétariat international des sous-comités ISO/CEI JTC 1/SC 1 «Vocabulaire», ISO/CEI JTC 1/SC 2 «Jeux de caractère et codage de l'information»

Ce dernier permettant notamment de contrôler que les jeux de caractères adoptés permettent d'assurer la transmissibilité des textes français compte tenu des particularités du français accentuations, signes diacritiques, Æ, æ, ... Le rythme actuel de production des normes en T.I. est d'environ 60 normes internationales par an (environ 4000 pages), 20 normes européennes par an (environ 1 200 pages). Dans le domaine des télécommunications, les normes internationales du CCITT sont publiées sous forme de livres réédités tous les 4 ans. Le volume est d'environ 20000 pages. Au niveau européen, en 1990, 19 normes ont été adoptées par l'ETSI soit environ 1 000 pages.

Les contraintes

Elles sont de plusieurs ordres

- Le volume et la vitesse d'évolution des documents font qu'il n'est généralement pas possible de faire évoluer une version française des documents internationaux parallèlement aux versions successives des projets élaborés en anglais. Pour ces raisons,

l'AFNOR n'entreprend la préparation des versions françaises que sur la base des versions définitives et stabilisées des projets de normes. Actuellement, le «retard» de traduction de l'AFNOR pour les versions françaises des documents de l'ISO/CEI JTC1 s'élève à 294 documents totalisant plus de 17 500 pages (au 21/08/90).

- Les programmes de publication des normes françaises étaient alors décidés par les commissions de normalisation, en fonction de l'intérêt direct des normes pour l'industrie française (constructeurs et utilisateurs) et des ressources humaines et financières disponibles pour élaborer les versions françaises. Depuis le démarrage des travaux de normalisation européens, les contraintes, liées aux règles de la normalisation européenne, se sont faites plus fortes. Les comités membres nationaux ont en effet, l'obligation de reprendre intégralement, même en cas de désapprobation minoritaire, les normes européennes adoptées et de supprimer toute autre norme en contradiction.
- L'expertise disponible pour assurer, outre la traduction, la relecture et la validation technique indispensable est parfois totalement absente, parfois insuffisante pour assurer la mise au point des versions françaises d'un sous-ensemble de normes. Tel est le cas particulier de l'ensemble des normes concernant les réseaux locaux d'entreprise. Ces normes (actuellement 17 documents recensés représentant plus de 1000 pages)

sont en effet préparés par l'Institut IEEE (Institute of Electrical and Electronics Engineers) sans participation française. Le volume de ces documents, leur instabilité et la non-maîtrise du processus d'élaboration, jointe à la faible expertise disponible en France, rendent impossible la traduction intégrale de l'ensemble de cette documentation.

- L'importance et l'urgence de la normalisation dans les T.I. est telle que les industriels et les utilisateurs ont fait pression sur l'ISO pour obtenir des procédures de publication les plus rapides possibles: préparation des normes par des éditeurs, relecture par un petit nombre de spécialiste, publication immédiate en anglais si le français n'est pas disponible. Il n'est généralement pas possible d'envisager de traduire avant de connaître le texte stabilisé définitif. La version anglaise est donc fatalement publiée (par le Secrétariat Central de l'ISO) en avance sur la version française. Ce délai, fonction du volume des documents et des difficultés de composition des textes, était de l'ordre de 18 mois. Devant cette situation pénalisante pour les intérêts des industriels et des utilisateurs français, l'AFNOR a pris en 1990 des mesures, notamment en assurant elle-même la composition des versions françaises, pour ramener ce retard à 6-8 mois, correspondant aux délais quasi incompressibles de traduction, validation et fabrication. Un tel délai ne peut cependant être tenu que lorsque l'expertise nécessaire est

disponible. Lorsque tel n'est pas le cas, le retard de la version française peut être notablement allongé, et devenir tel que l'intérêt de la publication d'une norme française est fortement diminué. À cet égard, la publication rapide d'une traduction « partielle » peut constituer un bon compromis entre le maintien du français et la diffusion rapide de l'information.

Se pose enfin un problème de moyens, à 3 niveaux :

- moyens humains en termes d'expertise extérieure pour la validation technique des traductions ;
- moyens humains en termes de personnel de l'AFNOR pour contrôler l'exacte correspondance entre les versions anglaise et française, dont l'AFNOR détient la responsabilité vis-à-vis de la communauté internationale. Les ressources l'AFNOR ne lui permettent pas d'avoir une équipe d'ingénieurs suffisante pour assurer à la fois la gestion des travaux de normalisation et le contrôle des traductions de l'ensemble des normes informatiques souvent fort volumineuses ;
- moyens financiers pour financer les dépenses de traduction. L'AFNOR dispose d'une équipe interne de cinq traducteurs mais doit faire appel à des traducteurs extérieurs, notamment pour faire face au flot important de traduction dans les T.I. qu'après s'être assuré de disposer des moyens humains interne et externe pour assurer les contrôles et validations indispensables à la publication de textes d'une telle technicité.

Conclusion

Dans une très grande majorité des cas et conformément à sa mission, l'AFNOR assure la mise au point des versions françaises des normes internationales et européennes dans le domaine des Technologies de l'Information. Dans quelques cas ponctuels et identifiés

- reprise par le CEN de normes internationales qui n'avaient pas fait l'objet de reprise en NF sur des sujets ne correspondant pas aux priorités du programme de normalisation du Comité d'Orientation Stratégique des Technologies de l'Information et Applications;
- normes de base adoptées au niveau international, citées en référence dans des normes européennes mais pour lesquelles il s'avère impossible de réunir une expertise suffisante pour la participation aux travaux;
- normes européennes de télécommunication sur des sujets ne correspondant pas aux priorités fixées par la Commission Française pour l'ETSI;

L'AFNOR envisage de publier des normes françaises reproduisant le texte en anglais des documents internationaux, précédés d'un avant-propos détaillé en français présentant la norme et en décrivant les principaux aspects. Cette possibilité permettrait, dans les cas spécifiques décrits ci-dessus, de concilier la nécessaire francisation des termes et concepts informatiques avec la diffusion rapide des normes dans le tissu économique national.

Merci à Jean Hyenne, devenu depuis le directeur adjoint de l'AFNOR, d'avoir su si bien expliquer et motiver la situation de l'époque. L'article suivant a été commis, en 1992 soit deux ans plus tard, par l'un de ses collaborateurs, qui s'est trouvé confronté à de multiples dilemmes linguistico terminologiques précisément posés par les documents relatifs à la normalisation de la représentation des médias images et son. À ce moment, les techniques de codage entropique sortent de la confidentialité des laboratoires, pour être accaparées par une nouvelle relation client-fournisseur, entre des fondateurs de silicium, d'un côté, et des opérateurs de télécommunications de l'autre. Nous assistons à cette époque à la mise en place des marchés d'équipement des infrastructures de réseaux qui, devenues autoroutes de l'information, s'ouvriront à ce que l'on appelle aujourd'hui globalement « Internet ». Le multimédia numérique était en train de naître. Pas forcément bien d'ailleurs, car il est parti sur un consensus bien établi autant qu'étrange: le mot multimédia n'a pas trouvé de définition à l'ISO. Nous pouvons même dater l'événement au 14 décembre 1990, lors d'une réunion à l'ANSI, à New York. Bernard Marti, nous le confirmera également. Il convenait d'être d'autant plus vigilant que l'objet du mandat n'était pas clairement établi. Les deux points de vues restitués ici ne doivent pas être vu comme antinomiques. D'abord parce que deux années riches en événements les séparent mais surtout parce que l'un fait le point d'une situation, alors que l'autre réagit à l'actualité. On y lira la preuve que la normalisation est un outil efficace, d'anticipation et totalement adaptable aux réalités de l'instant.

LP1992 - Le français, langue officielle de la normalisation internationale

Jean-Michel Borde AFNOR
décembre 1992

La langue française joue un rôle important dans la réglementation des échanges techniques internationaux. Plus particulièrement au sein de deux instances en charge de l'élaboration de textes déterminants pour l'établissement des normes sur le multimédia, l'ISO (Organisation Internationale pour la Normalisation) et l'UIT (Union Internationale des Télécommunications). Dans les statuts et directives de ces deux organismes, les réunions de travail doivent se tenir dans les langues officielles, françaises et anglaises à l'UIT et français anglais et russe à l'ISO. Ainsi, les travaux doivent être développés conjointement dans deux langues. En 1987, lors de la création du JTC1, une dérogation a été admise pour la soumission des projets de normes internationales relevant des technologies de l'information. Depuis cette date, il est admis que les votes puissent avoir lieu sur l'anglais seul, la version française pouvant arriver ultérieurement, du fait des délais de traduction. L'objet de cet exposé n'est certainement pas de contester cette mesure louable puisque, de bonne foi, elle reconnaissait la source de la plupart des termes employés, dans le but de ne pas retarder des travaux jugés pressants. Cependant, au fil des mois on a pu constater une inflation des travaux et du volume des normes produites. Celles-ci deviennent

de véritables projets requérant les compétences les plus sophistiquées pour leur réalisation et leur maintenance. Face à ce développement, le problème du maintien de la qualité de ces documents devient un corollaire préoccupant. En effet, sur la forme, les catalogues s'alourdissent de « Technical Corrigendum » et, sur le fond, d'amendements de plus en plus nombreux. En terme de communication, cet obscurcissement du statut réel d'un document, fondamental mais déjà peu simple à établir, devient d'une complexité incontrôlable pour l'utilisateur peu familiarisé avec la recherche de l'information. On commence même à voir poindre une ingénierie documentaire inverse pour expertiser l'état des bibliothèques techniques.

Que devrait-on en penser quelque vingt ans plus tard ? Qu'Internet aurait permis de tout régler ? Que SGML / XML s'est bien révélé être la panacée espérée de l'époque ? C'est pourquoi nous concluons cette partie du florilège par la communication de Laurent Romary qui établira un point d'état de l'art brillant à ce sujet. Il n'y a qu'en abordant ces niveaux de complexités extrêmes, que l'on peut comprendre les pertinences et portées des choix techniques effectués pour concevoir, et normaliser, nos systèmes d'information actuels, en assumant la prédiction des futurs.

Le constat, oblige à mener une réflexion à caractère économique pour évaluer les meilleures modalités de

maintien de la qualité de la production normative. Dans cette perspective, nous allons tenter de démontrer que la traduction, par la méthodologie qu'elle induit, est finalement un outil assez rationnel de contrôle de qualité. Il conviendra dans cette approche, résolument économique visant à l'optimisation des efforts des contributeurs, de garder une certaine souplesse pour ne pas dire un certain pragmatisme. Nous ne prôtons pas ici de recourir à une traduction systématique de tous les documents de travail qui aboutissent au vote international. Ni même d'ailleurs systématiquement de toutes les normes, ce qui implique un choix éclairé et par certains aspects, stratégique.

Le rôle de la terminologie :

Celui-ci est déterminant et a toujours été reconnu comme tel par la normalisation puisqu'il s'agit d'ailleurs d'un domaine à part entière couvert, au JTC1, par le Sous-comité 1. On notera que celui-ci s'est réuni, à Paris, en séance de travail du 2 au 6 novembre dernier et qu'à son ordre du jour, a été évoquée la possibilité d'établir une terminologie multimédia. L'ensemble des définitions des termes inclus dans les normes ISO/CEI 10918-1, 11544 et 11172-1 connues respectivement sous les acronymes JPEG, JBIG et MPEG seront sans doute proposés à la réflexion des experts de ce comité. Ces définitions, pourraient être accompagnées d'une extraction terminologique dont l'objet sera explicité plus loin. La soumission de ce sujet au SC1 interviendrait au

moment le plus favorable, alors que les normes sont en cours de vérification en vue de leur imminente publication. Dès lors toutes les suggestions, critiques et remarques issues de l'autorité de ce SC seraient susceptibles d'être examinées, traitées et adoptées. Dans le cas précis des technologies liées au multimédia, cette perspective ne peut que concourir à faciliter la diffusion fiable d'une connaissance qui risque de provoquer bien des évolutions dans le domaine du traitement de l'information. Cet effort systématique est aussi intéressant pour les organismes francophones, comme l'Afnor, en leur permettant de capitaliser une compétence à forte valeur ajoutée.

Propos rédigé il y a quinze ans, le recul du temps permettra d'apprécier ce que d'aucuns traitaient alors de « futurologie », ou parfois, et nous l'avons entendu, de « franc délire technocratique » !

Les travaux mixtes, un contexte particulier :

Si le JTC1 est né de la nécessité de coordonner les efforts normatifs de l'ISO et de la CEI sur les technologies de l'information, on a pu constater depuis quelque temps des recoupements de préoccupation de plus en plus importants avec l'UIT. Ce phénomène est lié au poids de plus prépondérant des contraintes de transmission de l'information numérique issue de traitements, ou en cours de calcul. Ainsi, les nouveaux médias posent des problèmes normatifs intrinsèquement liés à leur traitement,

leur transport et leur échange à travers la planète, à des coûts économiques raisonnables. C'est pourquoi les travaux sur l'image fixe ont été fusionnés entre ces organismes et ont donné naissance à deux groupes mixtes: le 'JPEG', pour Joint Photographic information coding Experts Group, et le 'JBIG', pour Joint Bi-level Image information coding Experts Group. La constitution officielle de ces groupes de travail s'est faite au sein du Sous-Comité 2 «Jeux de Caractères et Codage de l'Information». Depuis, celui-ci s'est divisé et les normes de codage des nouveaux médias sont dorénavant développées au sein du SC29 «Codage de l'Image, du Son et de l'Information Multi et Hypermédia».

À l'ISO/CEI, le groupe JPEG a proposé dès mars 1991, un 'projet de comité' le CD 10918, référencé au CCITT, Rec. T.81, dont le titre est «Technologies de l'Information — Codage de l'image de nature photographique». C'est en décembre 1991 qu'a été soumis à l'approbation internationale le CD 11544, référencé CCITT Rec. T.82, «Technologies de l'Information — Codage de l'image et du son — Compression progressive des images en deux tons». Notons que les deux organismes ne fonctionnant pas selon les mêmes règles, ne produisent pas des documents dont les effets soient assimilables. Les Avis et Recommandations du CCITT ont valeur de traités internationaux applicables entre des états, alors que les normes sont de mise en application volontaire par des entreprises, acteurs de leurs marchés.

La traduction des projets, contribution aux travaux à part entière :

Comme il a été évoqué en préambule de cet exposé, les normes sont co-élaborées dans deux ou plusieurs langues. Une traduction ne peut prétendre être assimilée à un tel concours que si elle est menée avant la publication du document final. Par contre la traduction de la norme, à l'un des stades quelconques de l'élaboration préalable à sa publication, permet des réflexions sur la qualité et des révisions à temps sur le texte source. À l'IUT, les documents de travail étant traduits en français, nous avons conclu, par échange de bons procédés d'assurer un relais dans la fourniture des projets dans cette langue assurant de cette manière une diminution sensible des charges pour les deux organismes. Les deux versions ISO et CCITT en français soumises aux votes des pays membres étaient rigoureusement semblables. De cette manière nous avons permis la diffusion des projets de normes JPEG et JBIG en français pendant une année entière avant leur parution définitive. Nous avons noté, car nous l'avons souvent sollicité, une aimable collaboration des lecteurs, qui nous ont retourné leurs commentaires et corrections. Notons que ceux-ci n'étaient pas tous français, ni même francophone, leurs encouragements nous ont confortés dans notre jugement sur l'intérêt et le bénéfice de cette démarche. Aujourd'hui, nous sommes en mesure de produire une version française en même temps que la version anglaise avec le plus haut niveau de qualité raisonnable.

Harmonisation terminologique :

La recherche d'une harmonisation de la terminologie a été envisagée dans le cadre de la traduction des documents sur le multimédia. Pour cela une méthodologie particulière a été établie. Un chef de projet unique, assumant la traduction et la validation par des experts identifiés, sous le contrôle de coordination de l'Afnor, a traité tous les documents en utilisant des régies uniques de saisie et d'élaboration. Ainsi, il a été relativement facile, car cela faisait partie du cahier des charges, de prévoir l'extraction de tous les termes spécifiques et d'en faire un glossaire. Celui-ci en cours d'achèvement sera, bien sûr, transmis au SC1 pour appréciation. Il sera aussi, sous une forme à déterminer, mis à la disposition des utilisateurs, au premier rang desquels, les traducteurs de textes relatifs aux technologies de l'information, normatifs ou non. Un glossaire de cette nature, suffisamment riche et utilisé, peut connaître une évolution séparée de ses textes sources.

Le futur proche :

Si les travaux sur le codage de l'image fixe numérique ont été menés de concert, l'image animée numérique a été normalisée à l'ISO uniquement. Cependant, si toutes les implications n'en étaient pas perceptibles a priori, la technologie de l'image animée ayant peu de portée sans le son l'accompagnant, l'approche normative de ces deux médias a été menée de front pour leur codage numérique. Ce travail a conduit la rédaction de la norme ISO/CEI 11172

« Technologies de l'information — Codage de l'image et du son Associé pour des débits allant jusqu'à 1,5 Mb/s ». Or, de longue date les technologies audiovisuelles ont évolué avec un état de l'art comparable. Il est apparu, au fil du temps, que la partie codage audionumérique devrait être « implémentable » séparément. Il n'est donc pas étonnant qu'aujourd'hui les algorithmes proposés par cette norme soient étudiés de près par le CCIR, pour la radiodiffusion sonore numérique. Dans cette perspective, la version ISO en français précède les documents de travail UIT sur cette question. Souhaitons, qu'il soit fait bon accueil à la version française ISO de ce document. La dernière filière normative impliquée dans les technologies multimédia est celle de l'hypermédia. Au SC 29, c'est le WG 12 qui est en charge du développement de la norme MHEG. Ce document spécifie une représentation codée des objets d'information hypermédia mais elle a des implications, qui sont déjà étudiées dans d'autres groupes de normalisation (ISO/CEI JTC1/SC18). Il n'est pas sûr que dans ce secteur particulier, l'on arrive à harmoniser le vocabulaire bilingue en une seule étape. L'actualité semble confirmer l'avenir de ces filières normatives avec des perspectives dans le domaine de la télédiffusion numérique. S'il le fallait, l'effort de mise en français pourrait être poursuivi pour ces travaux également. L'essentiel aura été fait pour que les notions de base de ces technologies soient pesées et éclaircies, lorsque cela était nécessaire.

Conclusion :

L'auteur remercie l'AILF et les organisateurs de ce colloque qui lui ont permis d'exprimer un aspect particulier de travaux de normalisation dont les bases sont aujourd'hui en train d'être consolidées. Dans le cas présent, il est facile de pronostiquer des développements impliquant le long terme et l'on sera rassuré de savoir que cette consolidation repose sur des critères de qualité. Il sera aussi flatteur pour la communauté francophone de connaître l'intérêt et l'apport de la langue française à cet effort industriel. Nous rappellerons le rôle majeur de tels efforts pour l'essor de la langue française dans le domaine de la technologie.

Après avoir exposé les problèmes vécus à l'époque par les organismes de normalisation comme l'AFNOR et par ses ingénieurs, voyons maintenant le contenu des travaux. C'est ce que nous détaille brillamment le père du Minitel Français, Bernard Marti, lorsque directeur technique délégué du CCETT de Rennes, il était président du sous-comité du JTC1 qui donnera naissance à JPEG et MPEG ! L'AILF aura eu la chance à travers de tels témoignages autant détaillés que sincères, de jalonner ainsi des étapes de chantiers, qui auront autant bouleversé nos pratiques professionnelles et personnelles. Nous espérons convaincre, grâce à cela, des motivations de vigilance qui animent l'AILF et tous ses membres, depuis ses fondateurs. Tout en rappelant à ce sujet que nous n'en sommes, encore aujourd'hui, qu'au milieu du gué des perspectives et promesses diverses contenues dans le savoir informatique ! Notre florilège se veut ainsi, plus qu'un satisfecit auto proclamé, un appel à la vigilance future.

LP1992 - Normalisation des différents média et problèmes de coordination

Bernard Marti, CCETT France Télécom

Mutation des métiers de l'image

Les métiers dont l'image est la matière première sont plus nombreux qu'on ne l'imagine à première vue. Mais, jusqu'à une date récente ces métiers constituaient des filières séparées sans grands intérêts communs, parlant des langages différents et manipulant des concepts éloignés les uns des autres. C'est ainsi que les photographes voyaient la notion de l'image au travers de la chimie des gélatinobromures d'argent, des oxydoréducteurs et des agents de blanchiment, les imprimeurs ne considéraient que les propriétés des encres et les gens de télécommunications ou télévision ne connaissaient que les décibels et les mégahertz. La numérisation de l'image (ceci vaut aussi pour le son avec de moindres conséquences), et son introduction progressive dans les différentes applications, oblige ces professions jadis séparées à se rapprocher. Elles doivent apprendre à utiliser des outils communs et des méthodes communes, donc à se poser la question de disposer de normes communes pour traiter de problèmes devenant brutalement communs. Cette situation nouvelle bouleverse des habitudes anciennes et, les rapports de force étant également modifiés, de vieilles citadelles renforcent leurs remparts, ce qui ne crée pas les conditions idéales pour une coopération ouverte permettant

un usage économique et efficace de la force de travail limitée disponible. La numérisation des média facilite leur manipulation conjointe et le concept de multimédia a pu ainsi apparaître et se développer, timidement encore en terme des applications arrivées à maturité, mais porteur de beaucoup d'espérances dont toutes ne seront pas déçues. Les pouvoirs nouveaux devraient supposer de devoirs nouveaux, et les concepteurs de méthodes ou de systèmes numériques doivent se sentir investis de la responsabilité de tenir compte des besoins d'autres utilisateurs qu'eux-mêmes.

La norme est une discipline du droit, il faut le rappeler. De la souveraineté des états, à la déontologie des professions et organisations, l'approche technique se complique singulièrement mais donne une dimension culturelle et humaine hors pair au travail de l'ingénieur. Les questions sociétales et morales sont ainsi intimement mêlées aux enjeux techniques les plus fondamentaux.

L'état de la normalisation des différents média

Media, multimédia, hypermédia.

Le mot média est lui-même ambigu en ce sens qu'on l'emploie pour désigner des concepts très différents. Ainsi on parlera de média d'échange (d'enregistrement ou de communication) pour désigner les supports d'enregistrement

ou les réseaux. On parlera de média de présentation pour désigner le support par lequel l'information sera soumise à l'utilisateur, écran ou papier par exemple. Mais c'est le média de représentation, significatif du type d'objet utilisé, qui sera ici utilisé pour désigner les différents types d'information.

Multimédia : Il n'existe pas actuellement de bonne définition qui permette de décerner dans tous les cas et sans hésitation le label Multimédia à un terminal, un système ou un service d'information ; il y a par contre un large consensus sur la liste des médias élémentaires : le texte, le graphique, l'image fixe, le son et la vidéo. Cette définition est pragmatique et traduit le côté nouveau du multimédia : concrètement, selon ce critère, la microinformatique domestique ou professionnelle devient multimédia lorsqu'elle commence à manipuler des images, du son ou une certaine forme de vidéo numérique alors qu'aujourd'hui elle se limite en général à offrir des fonctions d'éditeurs de textes, de tableurs, d'applicatifs de gestion et de jeux simples à dominante graphique ; la télématique deviendra multimédia lorsque l'information consultable (aujourd'hui des pages alphamosaïques sur nos Minitels) s'enrichira de commentaires sonores, de photographies voire de séquences vidéo ; l'évolution multimédia des services de messageries s'imagine aisément quant aux services de télévision, bien qu'ils s'appuient d'emblée sur l'image et le son, nous ne les qualifierons de multimédia ou plu-

tôt d'hypermédia (voir ci-après) que lorsqu'ils seront enrichis d'un minimum de fonctions d'interaction ou de manipulation autres que le zapping ou le contrôle du volume.

Hypermédia : Le livre est le premier système de consultation d'information textuelle ; la « navigation » du lecteur dans le livre repose quasi-totalement sur la table des matières et la numérotation des pages : choix de la rubrique dans la table des matières puis consultation séquentielle (fonction page suivante) avec de temps en temps un renvoi du type « cf. page n ». C'est Apple qui avec son logiciel HYPERCARD a mis à la mode une méthode de navigation très simple mais plus souple pour naviguer dans un livre électronique : l'HYPertexte. L'hypermédia désigne simplement l'extension au multimédia du concept de navigation Hypertexte initialement développé pour l'information textuelle ; on notera que les travaux actuels sur l'hypermédia ont pour objectif central la définition d'un modèle général d'outils de navigation, acceptable par l'ensemble des environnements applicatifs.

Les acteurs principaux

La normalisation des systèmes multimédia est conditionnée par le double contexte du développement de ces systèmes : le contexte physique (matériel) et le contexte communicant. Il est clair que cette coexistence induit une participation de deux familles d'organismes normatifs, les organismes de l'industrie informatique, représentés par le Comité

Mixte ISO-CEI (JTC1) et les organismes du monde des télécommunications, l'Union Internationale des Télécommunications et ses deux comités consultatifs, le CCIIT et le CCIR. L'Organisation Mondiale de la Normalisation (ISO) et le Comité Électrotechnique international (CEI) ont créé en 1986 un comité technique commun appelé JTC1 pour s'occuper de tous les sujets relatifs aux techniques de l'information. Dans ce comité, des sous-comités spécialisés s'occupaient du codage de l'information (SC2), des applications bureautiques (SCI8), infographiques (SC24) pour ne citer que les principaux impliqués dans les développements menant au Multimédia.

L'infléchissement lié à l'apparition du Multimédia

En 1988, le CCITT s'est proposé d'étudier les premières applications multimédia (dites applications audiovisuelles interactives ou AVIs). En France, un comité formé de nombreux acteurs de l'industrie, de l'édition, de l'éducation et des télécommunications avaient mis au point des spécifications communes pour une exploitation expérimentale de ces types d'applications pour des stations de travail ou des bornes communicantes. La RAVI, nom de cette spécification devenue depuis norme expérimentale française (éditée par l'AFNOR) a servi de point de départ à la création d'un groupe de travail (groupe d'experts) du SC2, le groupe MHEG (Multi and Hypertext Experts Group). Le SC2 travaillait depuis quelques temps sur le codage des images fixes (groupes JPEG et

JBIG) et animée (MPEG), groupes dont nous verrons plus bas les résultats.

Les objectifs contradictoires de la normalisation

La normalisation est, comme le sabre de M. Prudhomme, destiné à faciliter la libre circulation des marchandises et des services et, au besoin, à la combattre. Si l'objectif officiel est en effet de faciliter la portabilité, l'interchangeabilité des sous-systèmes pour faciliter l'ouverture des marchés, les participants au processus de normalisation n'hésitent ni à proposer un système propriétaire comme norme, ni à ralentir suffisamment les travaux pour permettre à un système propriétaire de conserver ses parts du marché en l'absence de norme, voire de devenir un standard (ou norme de fait).

Normalisation relative au media textuel

La normalisation du media textuel n'est pas aussi aisée qu'on le croit à cause de l'importance de l'environnement culturel et surtout linguistique. Je renverrai le lecteur à une conférence Lexipraxis précédente où fut faite une présentation des problèmes de normalisation des jeux de caractères dans le contexte du multilinguisme. Les problèmes de coordination posés par ces travaux sont gigantesques mais sortent du domaine de cet exposé.

On retrouvera ce domaine en conclusion de cette partie du florilège. Cette remarque, autorisée, illustre la complexité des

domaines adressés. Elle existait auparavant mais le débat à son sujet était confiné au sein de corporations fermées. Celles-ci s'étaient jalousement appropriées des parts de cette complexité sans oser imaginer le niveau des problèmes liés à la notion d'interopérabilité. L'histoire de l'imprimerie montre que Gutenberg lui-même, a induit des hypothèses de normalisation du document, dans les options de reproductions de l'écrit qu'il prit à son époque. Alors qu'aujourd'hui, Les individus eux-mêmes, doivent aujourd'hui s'impliquer dans la complexité pour vivre, échanger, exister. Ces degrés d'implications sociaux sociétaux, qui préoccupent l'AILE, transforment notre approche de la connaissance. À cause de ces transformations, le monde est devenu continuum ou global. Des frontières ont été repoussées, apparemment, mais pour mieux se reconstruire en nous ou ailleurs et nous renvoyer une image étrangement déstructurée des réalités. Il est urgent de reconstruire et de garantir une continuité à travers la banalisation de la globalisation. Nos philosophes seront conduits à réfléchir sur le rôle futur de la norme pour ce faire !

Codage des Images à deux niveaux

Dans le cas des systèmes fax, la numérisation de l'image fournit un tableau de points à un bit, 0 indiquant un point blanc et 1 un point noir, par exemple. Le message est composé généralement d'un ensemble de lignes d'un nombre fixe de points (1728 points et donc 1728 bits par ligne dans le cas de la télécopie numérique usuelle impli-

quant une analyse à 8 points par mm). L'analyse statistique de ce genre de messages ne montre aucune stationnarité permettant l'utilisation d'une méthode simple de compression à base statistique. C'est ici que le choix d'une bonne transformation joue un rôle essentiel. Celui choisi au milieu des années 70 pour les images à deux niveaux consiste à compter les points successifs d'une ligne ayant la même polarité et donc à transformer chaque ligne de 1728 bits en une double liste de nombre de points blancs (a priori, le papier) et de points noirs (l'écrit). La statistique de ces deux listes montre alors une stabilité remarquable sur un grand nombre de documents de nature différente, manuscrits, tapuscrits ou imprimés...[...]...

Codage des Images fixes à demi-teintes

Si les activités de recherche sur ce sujet sont anciennes, son introduction est récente en normalisation. Les progrès réalisés en algorithme, le progrès des outils de calcul et notamment des micro-ordinateurs ont permis, après des années de domination des méthodes prédictives, (DPCM) l'émergence et la maturation d'une famille d'algorithmes dont l'usage était jugé utopique il y a quinze ans : cette famille, basée sur l'utilisation de transformées orthogonales constitue la base reconnue de toute la génération actuelle des systèmes de codage d'images naturelles, fixes ou animées. Ce choix possède une conséquence immédiate : les transformées de type Fourier utilisent des fonctions

transcendantes (sinus ou cosinus) et leur application dans un domaine de données entières (la valeur associée à un pixel est un nombre entier) interdit toute réversibilité du codage en raison des inévitables erreurs d'arrondis ou de troncation. Les codages associés poussent plus loin la non-réversibilité. En effet, le but est ici de présenter in fine une image acceptable à un observateur humain et non de reproduire à l'identique un fichier numérique...[...]... La norme ISO-CEI 10918 (connue sous le surnom de JPEG) fonctionne bien jusqu'à des taux de compression jusqu'à 35-40, la non-réversibilité de l'algorithme pose des problèmes de fond dans certaines applications, par exemple celle liées à l'imagerie médicale...]....

Problème toujours non résolu, techniquement s'entend, à ce jour. Les résolutions sont peut-être à rechercher ailleurs.

Codage des séquences d'images animées

Comme pour les images fixes à deux niveaux, le codage des séquences animées relève de deux traditions et deux familles de textes normatifs ont été produites. La première famille est celle des télécommunicants qui, depuis des années, recherche des méthodes pour le codage du visiophone. L'autre, plus récemment entré dans ce secteur, est celle de l'informatique pour les applications multimédia. La technique utilisée pour la Recommandation H.261 du CCITT et celle utilisée pour le CD 11172 (MPEG 1) sont assez voisines

pour qu'il suffise de décrire le principe de la seconde pour donner une idée assez précise des méthodes actuellement utilisées. On peut toujours utiliser une méthode de compression d'images fixes pour coder des images animées. Ceci présente deux défauts; le premier est qu'on ne tire pas profit des ressemblances entre images successives, le second est que certains défauts, invisibles à l'œil sur une image fixe, se mettent à vivre et deviennent désagréablement visibles. Pour améliorer à la fois la compression et la qualité, il est donc nécessaire de tenir compte du fait que deux images qui se suivent dans une séquence se ressemblent. La méthode normalisée est basée sur un codage de type JPEG (donc une transformée en cosinus appliquée à une image découpée en blocs) sur quelques images sélectionnées (la méthode de sélection n'est pas normalisée et peut dépendre de l'application. Pour MPEG dont l'objectif est d'enregistrer des programmes de vidéo sur des disques compacts, le débit fixé par les caractéristiques du disque (1,15 Mbit/s pour l'image et le reste soit 256kb/s pour un son stéréophonique) donne une qualité d'image assez voisine de ce qu'un enregistrement VHS.

Normes concernant le média sonore

Le monde du son a donné naissance à nombre de normes, souvent liées au support. L'introduction du CD Audio et les travaux sur les réseaux téléphoniques modernes ont provoqué des travaux nombreux sur le son numérique.

La radio numérique de demain utilisera la partie audio de la norme MPEG conçue (en termes de normalisation) pour représenter le son d'accompagnement d'un signal vidéo numérique. MPEG-audio n'est pas le premier système de compression sonore. Les télécommunicants du CCITT ont mis au point des systèmes de codage numérique du son téléphonique destiné au RNIS comme G721 ou G722 utilisés aussi pour la sonorisation du vidéotex. C'est cependant le premier à se focaliser sur le son et non sur la voix et à prendre comme référence de qualité celle du disque compact. Avec une efficacité de 7, il ramène le débit actuel du CD de 1,4 Mbits à moins de 200 kb/s pour un son stéréo, sans perte apparente de qualité, ces mots magiques se retrouvant ici aussi. Avec une perte audible mais limitée, il est possible d'obtenir à 64 kbit/s une qualité encore très supportable. C'est le son MPEG qui a été choisi pour la représentation des objets sonores de la future norme de représentation des objets multi ou hypermédia (MHEG). On voit donc que le monde du multimédia est déjà partagé entre deux familles de systèmes incompatibles.

Presque vingt ans plus tard, les options continuent de fluctuer entre des performances antagonistes. La différence tient cependant à ce qu'aujourd'hui l'on sache mieux de quoi l'on parle, malgré les nombreuses zones à explorer et élucider dans ces domaines. On doit cependant à ces efforts soutenus d'avoir aidé à développer, des concepts que les langues usuelles ont

pu s'approprier. Le transfert de technologie aura été rapide mais il faudra des décennies pour que l'activité économique, après en avoir pris la mesure arrive à le dominer. Le défi est traité aujourd'hui l'enseignement, qui doit transmettre des connaissances aux générations à venir. Le point de vue du philosophe, historien de la pensée, nous semble cependant plus que jamais garant d'un certain recul d'interprétation. On parle peu du principe de précaution dans les technologies de l'information, hors du traitement des données personnelles. L'AILF, qui avait su donner la parole très souvent à Madame le conseiller honoraire Louise Cadoux de la CNIL, auteur de la Loi de 1978 ne perdra pas de vue ces contraintes, parmi les fondamentales de l'économie et du commerce.

Conclusion

Le partage par plusieurs professions des normes touchant à un média est déjà difficile. De nombreuses preuves peuvent illustrer à titre d'exemple cette affirmation a priori gratuite: dans le monde du texte entre informaticiens et bibliographes, les malentendus durent depuis 20 ans, pour ne pas parler, au sein de la profession informatique des luttes internes entre gens des langages, experts es caractères et architectes en documents...[...]...

Souvenons-nous que Jean-François Abramic fut un des cofondateurs de notre association, aux côtés de Jean Bécam. Il devint, par la suite, dans les années 90 co-président, fondateur du W3C. Il n'est donc pas anormal que nous ayons aussi entretenu de forts rapports d'intérêt avec cette célèbre organisation, qui

assure la promotion de multiples travaux d'avant-garde dans le déploiement de l'Internet et de ses applications globales. L'article ci-après posait le problème de la mise en relation des contenus internationaux différenciés du point de vue culturel et du codage. Depuis, l'i18n (pour internationalisation), est devenue une des composantes de l'ergonomie d'accès aux contenus. Là encore, notre vigilance n'aura pas été prise en défaut car nous entendons que notre culture croisse et embellisse mais parmi les autres. Selon nos résolutions, non pas à leur détriment mais en complète harmonie.

Nous pensons qu'une approche dogmatique serait que porteuse de dangereuses conséquences. Place donc à la diversité et à l'adaptation dont il serait dangereux de se priver. Aujourd'hui avec la montée en puissance des productions industrielles et commerciales des mondes asiatiques l'heure est plus que jamais au réalisme et la nécessaire négociation. Les engagements réciproques évitent les mauvais compromis. Technologie ou diplomatie? Les deux sans doute récompensées, on l'espère par les gains d'enrichissements scientifiques et culturels croisés, garants d'un futur possible.

LP2000 - Être strict pour partager: les standards du W3C pour garantir le multilinguisme sur Internet

Karl Dubost Conformance manager - W3C , Sophia Antipolis

L'internationalisation fait partie du domaine interface-utilisateur du W3C et est réalisée en collaboration avec un certain nombre de groupes de travail au sein du W3C.

Introduction

Au moment où le World Wide Web prend le sens de « World Wide » (mondial et étendu), de nombreux secteurs de l'industrie informatique sont attentifs à la nécessité de produire du matériel et des logiciels pour le marché international, et produisent des efforts pour être certains que le Web est utilisable de façon mondiale. Les besoins d'édition et

de traitement de l'information pour une large audience doivent être facilités: pour publier en arabe, chinois, français, japonais, coréen, hébreux, ou thaï. Les langues, les systèmes d'écritures, les codes de caractères, et les conventions locales ne devraient pas former de barrières aux technologies du W3C. Le but général du travail du W3C sur l'internationalisation est d'assurer que les formats du W3C et les protocoles sont utilisables dans le monde entier, dans toutes les langues et dans tous les systèmes d'écritures. W3C a souligné le rôle d'Unicode comme la base de l'internationalisation du web. Les plus récentes recommandations du W3C pour les formats de données et les protocoles utilisent ISO 10646/ Unicode pour identifier et décrire des caractères.

Dans les implémentations, Unicode est le noyau permettant la conversion entre des codages différents de caractère. Unicode attribue des nombres uniques à environ 50 000 des caractères du monde - fournissant un encodage unique pour tous les caractères. Une fois vos données encodées en Unicode, elles peuvent être traitées de façon uniforme, affichées, recherchées, triées, et manipulées sans craindre une détérioration des données. Unicode couvre pratiquement tous les répertoires usuels de caractères, y compris ASCII, Latin-1, JIS X 0208, etc. Historiquement, la RFC 2070, développée en 1994 par le groupe de travail HTML de l'IETF (maintenant clos) a donné les bases des possibilités d'internationalisation de HTML 4. De nombreuses autres idées de la RFC 2070 ont également été utilisées dans les spécifications XML 1.0 et CSS2.

Concepts expliqués simplement

Un modèle de caractère pour le web

Un domaine important du travail est le modèle de caractère pour le World Wide Web du W3C (Document en préparation). C'est une tentative de donner des directions et des règles pour les autres recommandations du W3C, pour être sûr que les possibilités d'internationalisation des différentes spécifications du W3C fonctionnent ensemble. Un point fondamental du modèle de caractère du W3C est de prendre conscience que l'intégration et les données transférées sur le web augmentant, le web doit être vu de plus en plus comme une application unique.

Le modèle de caractère du W3C définit alors que tout traitement de texte de toute sorte est fait en utilisant l'Universal Character Set (UCS), défini de manière conjointe par l'Unicode Standard et l'ISO/IEC 10646. Quand les données textuelles sont actuellement transmises à travers un réseau, différents encodages de caractères peuvent être utilisés, mais toujours en indiquant clairement quel encodage de caractère a été utilisé. Pour un comportement cohérent sur le web, dans certains cas, des spécifications supplémentaires associées avec ISO 10646/Unicode sont nécessaires. Le modèle de caractère détermine également comment compter les caractères dans de nombreuses circonstances et définit une façon consistante et compatible avec l'existant d'inclure des caractères non-ASCII dans les URIs.

Balilage pour le japonais, chinois et autres écritures asiatiques.

Certaines structures de texte de l'Asie orientale n'ont pas encore été prévues dans XHTML, et certains effets sont impossibles à réaliser sans utiliser une parade ou en insérant un graphique. Un de ces effets est le texte de type ruby, le nom donné communément à un texte qui apparaît à la suite immédiate d'un autre qui sert de «base». Les textes ruby sont souvent visibles dans les magazines japonais, et utilisés intensément dans les manuels de lecture des enfants. Une séquence de caractères idéographiques (kanji) est secondée par l'hiragana plus simple qui montre comment le mot doit être prononcé.

Contrôle des feuilles de style pour les langues asiatiques orientales.

Le document de travail sur la présentation internationale introduit un ensemble de propriétés pour couvrir la typographie de l'Asie orientale, en particulier les directions variées d'écriture, l'utilisation d'une grille dans la présentation, et la justification de texte. Dans le cas de la direction d'écriture, considérez que la plupart des documents latins utilisent une direction horizontale d'écriture de la gauche vers la droite dans laquelle la ligne suivante apparaît toujours en dessous de la précédente. Les langues asiatiques orientales utilisent une composition différente; par exemple les textes japonais sont fréquemment écrits du haut vers le bas, avec des lignes ordonnées de la droite vers la gauche.

Situation actuelle

Un certain nombre d'options de l'internationalisation ont été incorporées dans les recommandations du W3C y compris pour HTML 4, CSS2, XML, RDF, SMIL, DOM, XPath, et XSLT. Le protocole HTTP/1.1 inclut également dans différentes langues et différents encodages.

Documents de travail

Les documents de travail suivants ont été publiés:

- Document de travail Ruby qui définit le balisage Ruby comme un module XHTML (voir la déclaration d'activité sur HTML), pour l'utilisation en japonais, chinois et autres écritures asiatiques orientales (voir également la section précédente).
 - Présentation internationale qui propose un nombre de propriétés de style pour la présentation interne. Ce document de travail n'est plus développé, mais ses composantes sont intégrées dans CSS3, XSL, et SVG, pour être utilisable au sein du style HTML et XML.
 - Unicode en XML et dans d'autres langages de balisages qui contient des règles sur l'utilisation du standard Unicode de pair avec des langages de balisage comme XML.
 - Autres Documents
- Une note intitulée « Une notation pour les ensembles de caractères pour le WWW » a été publiée. Elle contient une proposition sur la manière d'exprimer des ensembles de caractères avec une notation XML pour leur utilisation dans des contextes variés. Le profil XML japonais a été soumis au W3C et publié comme une note.

Rôle du W3C

Relecture des recommandations du W3C

Une des plus importantes fonctions de l'activité internationalisation est de relire et revoir le travail du W3C dans le cadre de l'internationalisation. Ceci est réalisé pendant la période de Last Call. L'activité d'internationalisation a

le bénéfice d'être capable de travailler avec des spécialistes, permettant aisément les discussions et les avis sur les spécifications.

Meilleure connaissance des problèmes spécifiques de l'internationalisation

Un des buts est de collaborer avec les autres groupes en particulier le comité technique Unicode et autres comités ISO relatifs, et pour augmenter la connaissance sur les problèmes d'internationalisation par des conférences et des ateliers.

Open source

Charlint, un script Perl pour normaliser les données de caractères tel que proposé dans le Modèle de caractère pour le World Wide Web, a été amélioré et testé. D'autres logiciels Open Source du W3C, tels que Jigsaw et Amaya, ont également été améliorés pour mieux supporter l'internationalisation.

- Groupes de travail associés avec une activité

Il existe deux groupes associés à l'activité d'internationalisation. Le groupe de travail sur l'Internationalisation dont le but est de proposer et de coordonner toutes les techniques, les conventions, les règles et les activités au sein du W3C et avec d'autres parties de l'internet qui peuvent aider à rendre le web plus international. Le groupe d'intérêt sur l'internationalisation (membres et invités expert uniquement) fournit un forum aux membres pour discuter des problèmes relatifs à l'internationalisation sur

le web, et pour aider le groupe de travail de l'internationalisation.

Que nous réserve l'avenir ?

- L'activité, le groupe de travail et du groupe d'intérêt continueront à s'impliquer dans le travail de revue des autres groupes du Consortium, avec l'internationalisation à l'esprit. Cela signifie que nous nous assurons que les recommandations du W3C prennent en considération les problèmes d'internationalisation.
- Le modèle de caractère du W3C et l'annotation Ruby obtiendront le statut de recommandation.
- Un travail est en cours pour améliorer le support de l'internationalisation dans SVG, XSL, et CSS3.
- W3C est un sponsor de la 18^e conférence internationale Unicode du 24-27 avril 2001 à Hong Kong.
- À WWW10, 1-5 May 2001, à Hong Kong, une session internationalisation sera centrée sur le sujet des sites web multilingues, des traductions, des recherches, et des noms multilingues. La date limite pour proposer des présentations est le 22 décembre 2000.

De l'image 2D au document structure, une quête multimédia pourrait résumer la carrière de notre vice-président Henri Hudrisier qui développe les tendances des systèmes d'enseignements électroniques en réseau. Ou comment déjouer les insidieuses perversités de la machine informationnelle appliquée au déterminant de notre futur qu'est l'éducation ?

Enseigner les concepts et concevoir la transmission du savoir à l'ère de la mondialisation numérique.

LP2000 - Normalisation des systèmes de médiation pédagogiques : vers l'évolution de l'édition sous forme d'information structurée

Henri Hudrisier Maître de conférences
- Université de Paris 8 - GEMME

Pendant la plus grande partie de l'ère industrielle l'important était de vendre des produits et le service gratuit au consommateur représenté par la garantie qui était attachée à ces produits n'était qu'un argument de vente supplémentaire. Cette relation est aujourd'hui inversée: de plus en plus d'entreprises font littéralement cadeau de leurs produits aux consommateurs dans l'espoir de s'attacher leur fidélité à long terme en leur fournissant toutes sortes de services...[...]... Avoir, posséder et accumuler n'ont plus guère de sens dans une économie où la seule constante est le changement. Avec l'avènement de l'âge de l'accès, l'économie repose sur toute une série de prémisses qui diffèrent totalement de celles qui régissaient l'ère du marché. Dans ce monde nouveau, les marchés cèdent la place aux réseaux, vendeurs et acheteurs sont remplacés par des prestataires et des usagers et pratiquement tout se trouve soumis à la logique de l'accès. Le passage d'un régime de propriété fondé sur une notion de patrimoine amplement distribué au sein de la société à un régime qui re-

pose sur l'usage à court terme de ressources contrôlées par des réseaux de prestataires introduit un changement fondamental dans notre perception de l'exercice du pouvoir économique.»

Le JTC1, comité technique commun à l'ISO et à la CEI vient de démarrer la mise en chantier d'une normalisation des systèmes d'information destinés à l'enseignement et à la formation: le sous-comité SC36. L'ingénierie éducative devient une activité mondiale, multinationale, multiculturelle et multilingue étroitement dépendante des NTIC. De très nombreux éditeurs, réseaux câblés, chaînes hertziennes... se sont spécialisés dans ce secteur. Normaliser l'ingénierie éducative, c'est créer le cadre fonctionnel, formel, structurel, mais aussi qualitatif et sémantique, qui permettra d'organiser et de faire fonctionner toute tâche pédagogique, mais aussi de mettre en relation un collègue d'enseignants et/ou d'apprenants, d'établir la communication, d'encadrer les évaluations, d'organiser les routines d'apprentissage, d'établir des profils d'apprenants, de proposer des stratégies d'acquisition de savoir... Ce sera aussi, créer des architectures logiques et documentaires

qui permettront de gérer et d'utiliser d'immenses patrimoines de supports d'enseignement découpés en modules, voire en grains de pédagogie médiatisée, et ce, de façon multidisciplinaire, multimodale, multimédia, multilingue, organisée en niveaux et disponible sur réseau de façon planétaire. Le SC36 se voit confier un champ de normalisation limité à la médiation pédagogique proprement dite, qui exclu tout ce qui est de l'ordre des contenus d'enseignement. Ceux-ci dépendent en effet sur le plan normatif de la bureautique (SC18 du JTC1) et de la documentation (TC46 de l'ISO). Certes, la normalisation des systèmes de médiation de l'enseignement aura un impact certain sur l'ingénierie éducative en général. Là cette séparation formelle du champ du système de celui des ressources induira une accélération par ricochet, une obligation pour les ressources éditées qui seront candidates à l'usage de telles architectures normalisées, de s'organiser sous forme d'information structurée. Il s'agit ainsi d'aménager (comme c'est par exemple possible en XML) un « appareil de métadonnées » rigoureusement normalisées par rapport aux contraintes d'exploitation que requerront les systèmes d'enseignement selon la norme SC36.

Ce qui est en jeu: la normalisation de la médiatisation du savoir-faire pédagogique

Le JTC1, où se préparent de nombreuses normes qui organisent l'univers des NTIC, d'années en années plus convergent et inter compatible,

est maintenant chargé de coordonner la normalisation des systèmes d'information destinés à l'enseignement et à la formation. En termes concrets, il s'agit d'éclaircir pour pouvoir le normaliser, un contexte de médiation aujourd'hui complexe et disparate, c'est la mission du SC36. L'ingénierie éducative est un fait inéluctable qu'un certain nombre d'enseignants et de décideurs de l'éducation et de la formation (depuis l'université jusqu'à l'entreprise) ne remettent pas en cause. Il existe en France et dans le monde quantité d'universités virtuelles, de réseaux d'enseignement à distance, de professeurs impliqués dans la médiation pédagogique par les NTIC. Il existe aussi quantité d'entreprises, d'éditeurs multimédia qui produisent des systèmes technologiques dédiés à l'enseignement et de l'information pédagogique...[...]... L'activité d'organisation techno-médiatique de la pédagogie multimédia utilise malheureusement une proportion bien trop importante du temps de l'enseignant ou du formateur impliqué dans la pédagogie par les NTIC. C'est sur ces bases que s'organise (comme cela s'est déjà fait pour le texte et pour l'image) un cadre d'industrialisation cohérente et interopérable permettant de produire à terme de 5 à 10 ans, des stations (des logiciels ou des composants) de médiation pédagogique en très grand nombre et à très bas coût.

Le SC36 n'est certainement pas une panacée. C'est le produit d'un consensus raisonnable, mondial et inéluctable.

On peut imaginer le sentiment dubitatif de ceux, qui jusqu'à ce jour, ont consacré une part importante de leur énergie universitaire, à construire des centres de ressources et des réseaux de diffusion de l'enseignement médiatisé et à distance. C'est pourquoi les décisions que pourront prendre les professionnels français de la pédagogie tant enseignants, que formateurs, qu'éditeurs ou industriels à travers leur participation au « groupe miroir AFNOR du SC36 » sont fondamentales et urgentes. On aura remarqué que les enjeux linguistiques et culturels sont fondamentaux dans ce type de développement normatif. La place de l'Europe, et de la France en particulier, n'est pas négligeable à ce jour.

Distinguer savoir-faire et savoir.

La dynamique de normalisation des technologies de l'information pour l'éducation suscite assez souvent un certain désarroi chez nombre d'acteurs de l'enseignement. Quelquefois même, chez ceux qui sont directement impliqués dans l'EAD et la médiation NTIC de la pédagogie. Nous sommes confrontés à un changement des paradigmes technologiques de l'information: une transformation radicale de l'environnement technoculturel dans sa totalité, qui va prendre beaucoup de temps à s'imposer. Dans cet univers technoculturel en cours de transformation, la médiation de l'enseignement et de la formation constituerait un champ d'application stratégique.

Des systèmes d'information et de communication pour la pédagogie normalisée, selon les méthodes que préconise le SC36, ne serviraient à rien, s'ils ne peuvent pas activer un patrimoine de ressources de savoir numériquement, et structurellement, organisé. Sauf exception, les ressources de connaissances qui sont utilisées dans un réseau d'EAD ou toute autre forme de médiation d'enseignement ou de formation par les NTIC, sont spécifiquement aménagées. Elles ne sont réutilisables et interopérables qu'à l'intérieur du réseau d'enseignement à distance qui les a conçues. Ces ressources sont trop souvent conçues pour une seule cible d'apprenant et souvent encore pour un seul niveau d'apprenant. Grâce à des efforts de standardisation de la médiation et de la mise à distance, certains acteurs de la formation et de l'enseignement par les NTIC parviennent à activer un même univers de ressources pour l'utiliser sur des cibles différentes selon plusieurs niveaux de difficulté et selon plusieurs facettes d'approche.

Les mutations attendues du savoir-faire éditorial

Si l'on peut prévoir que les NTIC sont destinées à transformer à terme les savoir-faire pédagogiques, il est encore plus évident et déjà radicalement à l'œuvre qu'elles modifient en profondeur toutes les formes d'éditions. Il serait injuste en effet de ne pas souligner que ce sont les savoir-faire éditoriaux les plus sophistiqués qui ont été à l'origine de la mise en concordance de « l'en-

semble technologique» que constitue l'information structurée. C'est bien en effet GML, langage de balisage de textes et documents destinés à l'édition et notamment au formatage de bandes de photocomposition, qui a servi de base au développement de SGML dont on connaît la prolifique descendance. L'information structurée est au fondement même de la culture d'entreprise éditoriale. Ils sont très soucieux d'en diffuser les paradigmes à tous les niveaux de la chaîne éditoriale: de l'élaboration des documents de rédaction (ici encore entendu au sens large), jusqu'à la mise à disposition sur toutes catégories de support et de réseaux. À l'exception de ces «éditeurs high-tech», les éditeurs (si l'on tient malheureusement pour négligeable le dernier carré des artisans éditeurs-imprimeurs travaillant au plomb) se sont presque intégralement informatisés, mais leur culture informatique d'entreprise est trop souvent hétéroclite, constituée d'apports superposés bien plus destinés à répondre au coup par coup aux introductions de couches techniques successives de la modernisation de leur métier. Confrontés à ces progrès informatiques rapides, beaucoup d'éditeurs n'ont souvent pas su préserver le savoir-faire de leur cœur de métier.

Scénarios de développement optimum de l'information structurée normalisée et interopérable

Actuellement on voit bien que c'est l'ensemble des technologies de l'édition qui s'organise de façon cohérente sous

l'influence de la généralisation de l'information structurée, du multimédia, du multilinguisme. Il s'agit là d'un enjeu fondamental pour les éditeurs et les producteurs de toutes les catégories de documents dans les prochaines décennies. Le développement optimum de l'information structurée normalisée et interopérable constitue un scénario prospectif vraisemblable. Il faut savoir le rapporter au contexte spécifique d'une évolution de l'éducation et de la formation technologiquement médiée et explicitement identifiée. La normalisation des systèmes d'information d'un certain nombre de grands secteurs (la bureautique, l'audiovisuel, l'interlinguistique, les réseaux de télécommunication, la documentation et maintenant l'éducation et la formation) joue le rôle d'une sorte d'urbanisme des domaines. Les institutions de normalisation définissent au préalable un cadre normatif qui est ensuite proposé au monde de la recherche-développement (industriel ou universitaire). Celui-ci se comporte dès lors comme des architectes et des entrepreneurs construisant un environnement bâti plus concrets.

Dans le champ défini de la pédagogie, comment ces nouveaux paradigmes peuvent se mettre en place ?

Le développement rapide des NTIC rend aujourd'hui difficile la transformation effective des savoir-faire qui sont seuls à même d'induire des usages. On peut ainsi observer que se mettent en place conjointement :

- des besoins grandissants en pédagogie et formation ;
- une augmentation parallèle et complémentaire de ressources structurées se déployant selon divers modes d'usage: fabrication ou production de services, commerce, production éditoriale, techniques encyclopédiques, loisirs, production éditoriale ciblée sur l'apprentissage...

Des besoins grandissant en pédagogie et formation

Constatons d'abord, c'est une banalité, que la complexité de notre environnement de la technoculture nous impose, dans les années qui viennent à apprendre régulièrement tout au long de notre vie. Le progrès médical est un domaine sur lequel les usagers ne peuvent pas admettre que les professionnels de la médecine puissent être sous-formés. Il devient d'année en années plus patent que la formation d'un médecin généraliste est de plus en plus lourde et cependant il est obligatoire qu'elle soit réactualisée en permanence si on ne veut pas voir ce maillon fondamental de la profession médicale basculer dans l'incompétence ce qui pourrait entraîner des risques pour nombre de patients. De plus il apparaît que si on veut que continue de progresser la recherche médicale les patients eux-mêmes devront progressivement savoir prendre à leur compte une part croissante des routines de contrôle préventif et apprendre à gérer et entretenir eux-mêmes leur propre « capital santé » sous le regard plus distancé de leur mé-

decin de référence en association éventuelle avec leur pharmacien les professionnels paramédicaux voire leur caisse d'assurance.

Comment les ressources d'information qui augmentent parallèlement à cette complexité grandissante de la réalité sociale pourraient se redéployer.

Il paraît naturel que les professionnels de la création de la distribution d'information, c'est-à-dire les éditeurs au sens large et dans l'acception multimédia du terme, perçoivent ces nouveaux besoins et ces nouveaux marchés. Du fait de la convergence (qui concerne non seulement les médias et les modalités mais aussi les métiers et toute la réalité sociale) les informations ne sont plus sectorisées (destinées à être vues sous un seul angle), mais elles sont susceptibles d'être utilisées et comprises sous de très nombreuses facettes. Par exemple les données horaires d'un indicateur de chemin de fer doivent pouvoir être rééditées en réseaux, sur CD-ROM ou sur papier pour l'utilisateur final, servir de façon intégrée à participer au pilotage des trains et à la gestion du trafic, s'afficher dans la cabine du conducteur, dans les wagons voyageurs et sur les quais, se diffuser sur des réseaux de distribution d'information assurant la fluidité intermodale des transports, proposer le re-routage intelligent des autres moyens de transport collectifs ou des automobiles particulières, participer de la proposition d'offre globale de déplacement (agence de voyages

propositions directes aux individus etc.). Les professionnels qui peuvent accompagner ces nouveaux développements intelligents des informations ne peuvent que très rarement être les producteurs des données elles-mêmes. L'exemple des horaires de chemin de fer nous apparaît réaliste parce qu'il rentre précisément dans cette dernière catégorie particulière. On comprend bien comment un producteur quasi unique (ou pour le moins un nombre relativement restreint d'entreprises ferroviaires au niveau européen) sont susceptibles de redéployer l'édition de leurs informations horaires pour irriguer la globalité des usages potentiels et participer ainsi d'un développement harmonieux de l'offre globale de transport ce qui leur permet de ce fait de faire grandir leur marché et leurs activités en donnant une plus value et en commercialisant des informations dont la rentabilité est considérée aujourd'hui comme marginale. Vues sous ces nouveaux angles, ces informations horaires deviennent précisément un marché de rentabilité directe.

Qui serait à même de payer une telle offre globale ?

La voie de développement qui paraît pouvoir se dessiner, consiste en ce que le monde de l'édition sache proposer ces offres correspondant à des modes de rentabilisation économique très innovants. Prenant en compte à la fois les paradigmes de l'information structurée et des savoir-faire de plus en plus sophistiqués du traitement intelligent des

métadonnées (Topics map, RDF) et des ontologies. Il faut que ces éditeurs sachent prendre en compte la production future de systèmes normalisés consacrés à la pédagogie. Cette transformation de l'offre ne pourra aussi se faire que si on réussit à innover en intégrant ces nouvelles offres éditoriales avec les futures couches normalisées qui permettront des rétributions des micro-échanges. Ce sont ces couches qu'il est prévu à moyen terme de définir et de développer pour l'échange médiatisé par des langages d'informations structurées.

Choix, stratégies, niveaux d'usages et de temporalité

Les informations techniques et technologiques ne peuvent en aucune manière être prises au même niveau et ne doivent en aucun cas être perçues comme des séries de propositions indifféremment permutable. En matière de norme et de standard, il faut savoir rigoureusement distinguer l'un de l'autre. Le W3C introduit d'ailleurs un niveau intermédiaire, celui des « recommandations », qui au-delà d'un simple distinguo supplémentaire accompagne une nouvelle tendance du développement technique collaboratif que l'on retrouve avec l'émergence des « logiciels libres ». Il faut aussi savoir apprécier les temporalités, la maturité ou les étapes précoces de définition d'un champ comme c'est le cas pour SC36. Ainsi la définition et l'exploitation du champ SC36 est prévue pour se développer pendant 4 ou 5 ans, date à laquelle commencera à apparaître toute une famille de normes.

Dans le champ particulier qui nous occupe, on peut par ailleurs distinguer des métiers et des facettes d'usage :

- choix d'organisation et de recherche-développement appliqués à la numérisation des ressources.
- choix d'organisation et de recherche-développement utilisables par les décideurs de l'EAD aujourd'hui et demain.
- choix d'organisation et de recherche-développement utilisables pour les éditeurs agissants comme fournisseurs ou réalisateurs.
- choix d'orientation plus généraux concernant le savoir-faire de l'ingénierie du document et plus particulièrement de la documentation : XML, RDF, Topics map, Dublin Core, TEI...

Nous pouvons ainsi conclure en faisant l'hypothèse que la normalisation des systèmes d'information pédagogiques de type SC36 est une des conditions du développement sur une large échelle de l'édition structurée et normalisée. Non moins indispensable sera le développement d'une normalisation opératoire du cadre de l'échange pour permettre de justes rétributions des futurs concepts éditoriaux. Le secteur de l'éducation de la formation restera un métier plein de promesses. De ce point de vue, la convergence des NTIC et de l'échange (paiement électronique, réseaux de transmissions à intégration de services) participe directement de ce redéploiement éditorial.

Comme nous l'avons vu dans la restitution des interventions précédentes, l'époque était en plein bouleversement conceptuel, certains parleront de changement de paradigme. Nous voyons avec l'intervention suivante que celui-ci n'épargnera aucun des aspects de la représentation des médias. Parmi le plus sensible pour l'entendement, il y a certainement celui du référencement, qui conditionne l'existence par la possibilité de l'accès. Une information peu accessible subit un sort peu différent d'une information confidentielle : elle a pour vocation simple de rester méconnue. Pour ceux dont le métier est précisément d'assurer la logistique de l'accès à la connaissance, le problème n'est pas simple. Les solutions retenues doivent être internationalement compatibles. Les succès retentissants des normes ISO ISBN et ISSN sont là pour le démontrer avec les objets comme le livre ou le périodique. Quid cependant du média numérique en réseau ? Quel type de garantie doit donner à la dissémination de la connaissance à ce niveau ? Comment défend-on des positions et intérêts, voire philosophies nationales ? C'est à ces questions complexes que s'est consacré Patrick Le Boeuf, conservateur à la BNF, lors de notre édition de 2002.

LP2002 - Des modèles et des identifiants : les travaux de normalisation (ISO/TC46, AFNOR/CG46) potentiellement profitables au Web sémantique

Patrick Le Boeuf Conservateur – Bibliothèque nationale de France, Service de normalisation documentaire.

La vision du Web sémantique a trouvé son expression la plus énergique dans un article de Tim Berners-Lee, James Hendler et Ora Lassila, publié dans *Scientific American* en 2001 et intitulé simplement : « The Semantic Web ». Cet article ébouriffant s'ouvre sur le récit science-fictionnesque des démêlés d'un quidam dénommé Pete avec son ordinateur, à qui il demande de fixer automatiquement des rendez-vous chez le kinésithérapeute pour sa pauvre vieille mère, à des dates et à des heures auxquelles il pourrait la conduire en voiture sans que cela perturbe trop ses propres activités : « Le système intégré de loisirs multimédia faisait beugler la chanson des Beatles « We can work it out », lorsque retentit la sonnerie du téléphone. Au moment où Pete décrocha, son combiné fit baisser le son en envoyant un message à tous les autres appareils « locaux » pourvus d'un dispositif de contrôle du volume sonore. À l'autre bout du fil, c'était sa soeur Lucy, qui l'appelait du cabinet médical », etc. La suite de l'article se fait plus technique et décrit en détail tous les moyens à mettre en œuvre pour que ce récit introductif quitte le domaine de la science-fiction

pour devenir réalité. Parmi ces moyens, les auteurs mettent l'accent sur « l'accès à des agrégats structurés d'informations et à des ensembles de règles d'inférence que [les ordinateurs] doivent pouvoir utiliser pour mener des raisonnements automatisés », autrement dit l'accès à des « représentations de connaissances » passant par des ontologies « qui définissent de manière formalisée les relations qui existent entre des termes ». Ces ontologies vont servir de point de référence et permettre d'exprimer l'idée que l'information qui est saisie dans le champ « PV » de telle base de données est équivalente à l'information qui est saisie dans le champ « 717Z80 » de telle autre base de données, et que l'une comme l'autre renvoient au concept de « pension de veuvage », avec en outre des règles d'inférence qui doivent permettre à un agent informatique de déduire tout seul que, s'il a trouvé dans un document l'information selon laquelle Roméo a épousé Juliette à telle date et l'information selon laquelle Juliette touche une pension de veuvage à partir de telle autre date, alors Roméo est mort à un moment donné entre ces deux dates.

Un autre point qui est crucial pour la construction du Web sémantique mais que les auteurs de l'article cité ne met-

tent pas spécifiquement en avant, c'est l'importance des identifiants associés de manière unique et non ambiguë aux instances de certaines classes de ces ontologies. L'existence de ces identifiants peut aider à déduire automatiquement le sens de ce qu'ils identifient; en outre, si leur attribution s'accompagne obligatoirement, comme c'est de plus en plus souvent le cas, de la constitution d'un ensemble de métadonnées décrivant l'objet identifié, ces identifiants constituent naturellement un point d'accès privilégié à ces métadonnées, et à la richesse sémantique des ontologies qui les structurent. Ainsi, si Pete demande à son agent informatique de commander pour lui sur le Web un enregistrement de l'opéra Faust et que son agent trouve mention d'un document qui est bien intitulé « Faust: l'immortel opéra de Gounod » mais auquel est associé un identifiant de type ISBN, l'agent devrait être capable d'inférer par lui-même que ce n'est pas tout à fait l'objet désiré, même si la page Web sur laquelle il l'a trouvée ne précise pas qu'il s'agit en fait d'un ouvrage imprimé consistant en photographies d'une représentation de Faust. ...[...]

Présentation d'ISO/TC46 et AFNOR/CG46

En octobre 2001, le TC46 fut réorganisé pour prendre la configuration suivante:

Trois groupes de travail (WG2: Codage des noms de pays; WG3: Conversion des langues écrites; WG4:

Terminologie de l'information et de la documentation) sont directement rattachés au TC, tandis que tous les autres sont regroupés dans quatre sous-comités thématiques (SC4: Applications informatiques en information et documentation; SC8: Qualité, statistiques et évaluation de la performance; SC9: Identification et description; SC11: Gestion des archives courantes et intermédiaires). Pour mémoire, on notera au passage que le Dublin Core (dans sa version 1.1) a été proposé pour devenir norme ISO (sous la référence DIS 15836) par la procédure accélérée au sein du SC4 sans être rattaché à un groupe de travail en particulier. Si le vote, dont le résultat est attendu pour la fin janvier 2003, est positif mais assorti de commentaires, un nouveau groupe de travail verra le jour au sein du SC4.

Au sein du TC46, les travaux auxquels on s'intéressera plus en détail dans la présente intervention parce qu'ils sont potentiellement en relation avec la construction du Web sémantique sont ceux du WG9 du SC4 et de l'ensemble du SC9.

Dans la phraséologie de la normalisation, la Commission générale Information et Documentation de l'AFNOR, ou CG46, est réputée être le « miroir » français du TC46. Sa structure n'est cependant pas intégralement calquée sur celle du TC46, même si l'on retrouve bien sûr un grand nombre d'éléments communs. Elle a pour missions « de définir la politique d'orientation et assurer la coordination des Commissions de normalisation

(CN) qui sont sous sa responsabilité, et d'être une force de proposition pour le TC46 [...]. Elle doit en effet, par ses contributions et participations aux réunions ISO, créer une dynamique au niveau international car elle a une place pivot au sein du TC46 qui a pour président et secrétaire des représentants français.» Le «miroir» que constitue la CG46 ne doit donc pas se contenter de refléter, il peut aussi réfléchir...

...[...]

Il existe un certain nombre de différences qui illustrent la «force de proposition» censée caractériser la CG46. Ainsi, la CN10 «Conservation des documents» subsiste-t-elle à l'AFNOR, alors que son pendant international SC10 a été «mis en sommeil» lors de la réorganisation d'octobre 2001, mise en sommeil expliquée en ces termes par Jean-Marie Arnoult, président du TC46: «C'est le signe manifeste du changement qui s'est opéré progressivement dans les préoccupations contemporaines: les supports traditionnels — et notamment le papier — qui ont généré de nombreux sujets d'étude au cours des décennies passées, ne suscitent plus actuellement de curiosité renouvelée, et les nouveaux supports, par leur spécificité, sont du domaine de secteurs différents des champs d'action du TC46.» Il ne fait cependant pas de doute que, même virtuel, même sémantisé (et, peut-être, surtout sémantisé), le Web posera à relativement court terme des problèmes aigus de conservation physique. Mais c'est là une autre histoire.

...[...]

Travaux dans le domaine des ontologies et de la modélisation

Modèle conceptuel de référence pour l'échange d'informations relatives au patrimoine culturel (CRM)

Le Modèle conceptuel de référence ou CRM est élaboré depuis 1994 par le Groupe de normalisation documentaire du CIDOC (Comité international pour la documentation) qui est lui-même une émanation de l'ICOM (Conseil international des musées). Il n'a cependant été soumis à l'ISO qu'en 2000 et était soumis au vote jusqu'au 15 novembre dernier sous la référence ISO/CD 21127. Le CRM est traité au niveau de l'ISO/TC46 par le WG9 du SC4; le dossier est suivi au niveau de l'AFNOR/CG46 par la CN357. Comme son nom l'indique, il ne s'agit pas tant d'un modèle de données destiné à être implémenté concrètement dans telle ou telle base de données effective, même si cette utilisation ne saurait bien sûr être exclue a priori, que de définir au niveau le plus abstrait possible les réseaux de sens qui se tissent entre les différents éléments d'information présents dans des bases de données muséographiques et entre ces bases, même (et surtout) si elles ne sont pas normalisées et qu'elles articulent diversement toutes ces informations. Nous sommes donc ici très près de l'idéal du Web sémantique: il s'agit de rapporter le contenu de données d'origines diverses à une ontologie unique qui en indique le sens et les relations. Le CRM est destiné en priorité à la communauté muséographique (musées des beaux-arts et musées d'histoire

naturelle) mais pourrait sans grande difficulté être étendu à d'autres champs de l'information relative au patrimoine culturel.

Métadonnées d'autorité

La CN357 anime un groupe d'experts qui n'a pas sa contrepartie au niveau de l'ISO, le groupe GE4 pompeusement dénommé « Métadonnées d'autorité ». Les travaux de ce groupe visent un modèle abstrait définissant le sens et les relations des données contenues dans tous les types de « fichiers d'autorité » élaborés par les bibliothèques, les musées, les archives et les centres de documentation. Il est difficile de définir en peu de mots ce que l'on entend exactement par « fichiers d'autorité » ; disons pour simplifier que les deux types de fichiers d'autorité « par excellence » sont, dans l'immense majorité des organismes, les fichiers des noms de personnes physiques et des noms de collectivités qui, comme disent les Anglo-Saxons, « originent » (to originate) les documents conservés et décrits par ailleurs. Mais il est également possible de constituer des fichiers d'autorité pour un grand nombre d'autres notions — en réalité, c'est la typologie entière des éléments de description de tout document et des points d'accès à ces descriptions qui pourrait donner lieu à l'établissement de fichiers d'autorité, dans la mesure où chacun de ces éléments de description est susceptible d'être accessible sous plusieurs formes, d'être mis en relation avec d'autres éléments, et de faire l'objet d'un commen-

taire. Ces fichiers d'autorité permettent : de gérer les formes multiples que peut revêtir le nom d'une personne ou d'une collectivité, d'apporter des informations complémentaires indispensables pour lever l'ambiguïté sur des noms identiques dans leur forme et partagés par des personnes distinctes, de tisser un réseau de liens au sein de ces personnes ou collectivités ou bien entre elles et toute une variété d'autres concepts — toutes informations qu'il serait impossible de faire figurer dans la description d'un document émanant de cette personne ou collectivité, sans alourdir considérablement cette description, ou sans déboucher sur de pénibles redondances lorsque cette personne ou collectivité est à l'origine de plusieurs centaines de documents... Le Groupe s'est appelé « Métadonnées d'autorité », non pas simplement pour « paraître à la mode » (encore que cette considération n'ait pas été entièrement étrangère à ce choix), mais surtout parce qu'à l'origine il avait l'ambition d'inclure dans son effort de modélisation ces points d'accès particuliers que sont les métadonnées d'une ressource électronique, et notamment les éléments du Dublin Core — où l'on retrouve aussi, entre autres, ces notions de nom de personne et de nom de collectivité, avec toutes les ambiguïtés potentielles qu'elles impliquent. Cette problématique a malheureusement été quelque peu oubliée en cours de route, ce qui enlève un peu de la pertinence de ce projet dans le contexte du Web sémantique, sans l'anéantir tout à fait cependant. Pour être honnête, il faut

reconnaître que les travaux du Groupe ne progressent pas très vite — bien que le Groupe lui-même se réunisse souvent et que les membres en soient dans l'ensemble très assidus. Cela est dû au fait que la modélisation, surtout à un niveau aussi abstrait et sans avoir en vue d'implémentation concrète à court terme, est une activité qui nécessite une formation initiale d'ingénieur; or, le Groupe n'est composé que de bibliothécaires, d'archivistes et de documentalistes... Si ce groupe d'experts n'a pas de contrepartie au sein de l'ISO, il en a une en revanche au sein de l'IFLA (Fédération internationale des associations de bibliothécaires et de bibliothèques): il s'agit du groupe FRANAR (Functional Requirements And Numbering of Authority Records — Spécifications fonctionnelles et numérotation des notices d'autorité), qui travaille plus vite que le GE4 mais qui présente l'inconvénient d'être quasi exclusivement orienté vers les besoins des seules bibliothèques. Les deux groupes communiquent par le biais de Françoise Bourdon, qui préside le GE4 et est membre du groupe FRANAR.

Métadonnées des thèses

L'objectif du Groupe d'experts AFNOR/CG46/CN357/GE5 «Métadonnées des thèses», créé à la fin de l'année 2001, n'est pas (ou du moins: pas encore? ou pas ouvertement?) d'établir une ontologie, mais de «définir l'ensemble des Métadonnées Thèses, en s'appuyant sur les initiatives internationales, pour faciliter le transfert des mé-

tadonnées d'un système vers un autre (souci d'interopérabilité) et favoriser ainsi la diffusion des thèses à travers plusieurs réseaux documentaires [...] [et] étudier l'encodage des métadonnées en HTML et XML/RDF». Il semble toutefois inévitable que les travaux de ce groupe débouchent un jour ou l'autre sur la production d'une ontologie permettant d'extraire la sémantique des métadonnées de publications électroniques à caractère universitaire quelle que soit la structure spécifique de ces métadonnées, même si cet objectif n'est pas ouvertement exprimé. Pour l'heure, les travaux du groupe consistent à élaborer un mapping des éléments et des qualificatifs du Dublin Core vers le format Unimarc, afin de transformer automatiquement les métadonnées associées à une thèse sous forme de fichier informatique en notice bibliographique destinée au catalogue du SUDOC.

Le CDRM (Content Delivery and Rights Management — Mise à disposition de contenus et gestion de droits)

Le modèle CDRM n'est pas à proprement parler une norme. Il s'agit d'un rapport technique, référencé ISO/TR 21449, produit non par un groupe mais par un individu isolé, Tom Delsey, pour le compte de l'ensemble des Groupes de travail réunis sous la bannière commune du Sous-comité ISO/TC46/SC9. Le titre complet, très long, en définit clairement la portée: «Mise à disposition de contenus et gestion de droits: spécifications fonctionnelles des identifiants et des descripteurs à l'usage des indus-

tries musicales, cinématographiques, vidéographiques, phonographiques et de l'édition». Il s'agit d'établir un modèle de référence pour toutes les activités liées à la production, à la distribution et à la consommation (y compris frauduleuse...) de produits culturels et à la gestion et la protection des droits de la propriété intellectuelle, ainsi que d'une explicitation du rôle que doivent jouer les identifiants assignés à différents types de ces produits culturels à différentes phases de leur production / distribution. Ce modèle n'est très certainement pas appelé à jouer un rôle direct dans la construction du Web sémantique — il n'est du reste pas prévu d'en faire une norme et il doit garder son statut de simple « Rapport technique »; mais il est sans doute appelé à jouer un grand rôle dans la structuration du Web commercial; or il s'avérera sans doute de plus en plus difficile de faire la part des choses entre Web commercial et Web sémantique.

Travaux dans le domaine de la définition d'identifiants numériques

Tous ces travaux relèvent, au niveau d'ISO/TC46, du Sous-comité SC9 et sont suivis, au niveau d'AFNOR/CG46, par la Commission de normalisation CN6, qui n'a pas mis sur pied de groupe d'experts spécifique mais qui a au moins un représentant dans chacun des Groupes de travail du SC9. Au grand dam des professionnels de la documentation et des spécialistes des sciences de l'information, l'objectif avoué de la majeure partie de ces identifiants numéri-

ques est clairement d'ordre mercantile. Il s'agit prioritairement de gérer des stocks et des commandes et de répartir des royalties, et très accessoirement d'identifier des unités documentaires à des fins documentaires.

...[...]...

L'univers abscons des identifiants numériques normalisés internationaux connaît une véritable ébullition depuis quelques années. Cela est dû aux défis lancés par la e-information, les e-contenus et autres e-publications. Il a fallu établir des distinctions plus marquées entre contenu et support, et s'interroger en profondeur sur ce que les identifiants avaient pour mission d'identifier. De nouveaux besoins sont apparus — et notamment celui d'identifier non seulement des produits culturels sous leurs aspects concrets et tangibles, mais aussi les idées abstraites et les créations intellectuelles ou artistiques qu'ils véhiculent. Comme l'écrit Élisabeth Giuliani, présidente de la CN6, « on constate, d'une part, une multiplication des objets réclamant une identification, d'autre part, une évolution dans les objectifs assignés à la numérotation et partant une modification des structures de numéros. » Les identifiants internationaux normalisés les plus anciens et les plus familiers sont l'ISBN et l'ISSN, respectivement pour les livres et pour les publications en série. Mais l'apparition des e-books, et la possibilité de décliner un même « livre » sous de multiples formats (PDF, HTML...), de le diffuser non pas « en masse » comme dans l'édition classique mais à la demande, voire

de le « tronçonner » et de permettre à un client de n'en acheter qu'un ou deux chapitres, ont poussé les éditeurs à réclamer une révision en profondeur de l'ensemble du « système ISBN ».

Avec l'ISTC (projet de norme ISO/CD 21047), il va être possible d'attribuer un numéro unique à Notre-Dame de Paris, quel qu'en soit l'éditeur commercial, la date d'édition, le support, ou le type de fichier informatique; avec l'ISRC (norme ISO 3901), il est d'ores et déjà possible d'attribuer un numéro unique à un enregistrement d'un air de Carmen par Maria Callas, que cet enregistrement soit commercialisé sur CD ou sur cassette, au sein d'une anthologie ou dans une intégrale de l'opéra; avec l'ISAN (norme ISO 15706) il est possible d'attribuer un numéro unique au Grand Bleu et avec le futur V-ISAN (projet ISO/WD 20925) il sera possible de distinguer entre la version longue et la version courte... L'attribution de ces numéros s'accompagne souvent d'une déclaration de métadonnées relatives à ce qu'ils identifient, et qui en augmentent la valeur.

Conclusion

Pour terminer ce panorama forcément rapide, incomplet et susceptible d'évoluer très vite (il faudrait aussi évoquer les travaux du groupe MPEG, hors du champ du TC46 mais auxquels il est possible que la CN357 participe — la décision n'en est pas prise à l'heure où ces lignes sont écrites), il pourrait être instructif de reprendre la trame du petit scénario de science-fiction imaginé par

Tim Berners-Lee, James Hendler et Ora Lassila et évoqué au début de cette intervention, pour s'essayer à faire le même exercice en intégrant les contributions potentielles du TC46 et de la CG46 au Web sémantique. Cela pourrait donner quelque chose qui ressemblerait à ceci :

« Le système intégré de loisirs multimédia faisait beugler la chanson des Beatles *We can work it out* lorsque retentit la sonnerie du téléphone. Pete décrocha. C'était sa sœur Lucy, qui avait vu l'après-midi même dans un musée une statuette du Minoen moyen et s'était mise en tête d'offrir à leur mère une reproduction de statuette du Minoen moyen. Pete fut immédiatement d'accord pour que le cadeau fût commun. « Je vais demander à mon agent de trouver une boutique en ligne qui vende des moulages de statuettes du Minoen moyen », dit Lucy. Pete n'avait pas la moindre idée de ce à quoi cela pouvait ressembler ni de ce qu'était le Minoen moyen, et demanda à son agent de lui trouver des pages de sites muséographiques qui lui expliqueraient ce que c'était et qui contiendraient des photographies numérisées de ces fameuses statuettes. Grâce au modèle CRM, il obtint aussitôt des pages pertinentes d'une dizaine de musées à travers le monde, dont les bases de données étaient pourtant toutes structurées différemment. Lucy lui demanda : « Tant que tu y es, est-ce que tu peux me trouver un livre sur le sujet ? » Pete trouva une référence sans grande difficulté, mais lorsqu'il voulut commander l'ouvrage le site de son fournisseur l'avertit que l'éditeur était

en rupture de stock. Son agent, sur la base de l'ISTC, chercha alors aussitôt si le même contenu était disponible dans d'autres éditions, voire sous forme numérique. À cause d'un dysfonctionnement dans l'installation domotique de Pete, la sonnerie du téléphone n'avait pas entraîné une diminution du volume sonore; Lucy lui demanda donc: « C'est sympa, ce que tu écoutes; est-ce que tu pourrais regarder si je peux m'en procurer une adaptation pour mandoline, en partition ou en CD? » Lucy ne savait jouer en effet que de la mandoline. L'agent de Pete, après avoir recherché l'ISWC de « We can work it out » direc-

tement sur le CD, alla interroger les métadonnées ISWC et put retrouver tous les arrangements de cette chanson ainsi que les ISRC correspondant à leurs interprétations. Pendant ce temps, Lucy, qui avait l'habitude de passer rapidement d'une idée à une autre, demanda à son agent les heures d'ouverture des musées et galeries où étaient exposées les œuvres picturales éventuellement inspirées par cette chanson des Beatles... Là encore, c'est grâce à l'ontologie CRM et à l'ISWC que le Web sémantique put lui donner la réponse. »

LP2002 - Métadonnées et standards en ingénierie des langues

Laurent Romary Directeur de Recherche INRIA – Laboratoire LORIA

Introduction

La large diffusion du métalangage XML a eu pour conséquence heureuse d'offrir d'un seul coup un cadre syntaxique unifié pour la représentation de données qui étaient jusqu'alors définies et gérées par des communautés de spécialistes relativement différentes. De l'astronome au généticien, en passant par le spécialiste de littérature grecque classique, tous peuvent trouver en ligne des données exprimées sous la même forme et donc manipulables avec les

mêmes outils. Cependant, ce rapprochement s'accompagne d'une évolution, voire d'une remise en cause, de certains concepts fondamentaux attachés aux anciennes pratiques de représentation de l'information. Il en est ainsi dans le domaine documentaire au sens large du terme, où les préoccupations classiques de nature bibliographique apparaissent de plus en plus complémentaire d'un besoin croissant d'archiver intégralement les contenus sous une forme électronique. Pourtant on perçoit très vite, quand on souhaite par exemple répertorier un corpus d'ouvrages informatisés, que les structures de

catalogage classiques offrent un cadre un peu étriqué pour, par exemple, gérer tous les problèmes de versions ou de niveaux d'annotation multiples que pose le matériau électronique. Par ailleurs, le besoin d'identifier de façon précise les informations les plus pertinentes au milieu de la masse de documents disponibles tels quels sur la Toile a motivé plusieurs initiatives pour définir des cadres de description des contenus de ces documents. On cherche ainsi à fournir des informations sur le contenu d'un document, ou métadonnées, sans qu'il soit nécessaire d'accéder effectivement au contenu lui-même.

Ces initiatives – on pense typiquement au Dublin Core – cherchent à répondre à des besoins de simplicité et de diffusion large qui sont à l'opposé des objectifs de formats documentaires tels qu'UNIMARC.

Enfin, un courant particulier s'est développé autour de la notion de Web sémantique : il caresse l'espoir de définir un cadre unifié qui permette d'accéder aux documents de la Toile par le biais de descripteurs reflétant leur contenu informationnel. On a vu ainsi apparaître des syntaxes dédiées à ce genre de représentations – schémas RDF, OIL-DAML, Topic Maps –, mais aussi des initiatives de création de bases de concepts permettant de décrire des domaines d'activité particuliers.

Dans ce contexte encore instable, l'objectif de ce chapitre est d'essayer

d'élargir le cadre d'analyse pour tenter, d'une part, d'aborder le problème des descripteurs de structures de documents, et, d'autre part, de dégager des pistes de convergence en montrant que l'on peut s'appuyer sur une base conceptuelle unifiée pour décrire la variété des métadonnées que l'on souhaiterait associer aux bases documentaires.

Cette introduction nous paraît prémonitrice, tant elle correspond à ce qui se passe actuellement au niveau des différentes applications majeures, que nous pouvons utiliser aujourd'hui. Pourtant Laurent avait entouré ses appréciations de précautions oratoires lors de son exposé. Nous retiendrons de sa communication et de ses travaux, que la complexité peut et doit être décrite avant tout chantier visant à son éventuelle organisation. On observera surtout, à l'aune de ce propos que les acteurs fondamentaux de l'organisation de l'information de demain n'existent pas encore. On peut anticiper, à cause de cela, des bouleversements brutaux sur les marchés et dans le paysage de l'économie générale.

Quels moyens pour identifier des documents sur la Toile

Métadonnées spécialisées

Même en ne considérant que les documents produits dans le cadre de domaines spécialisés (littéraire, scientifique, économique, etc.), on constate que la quantité d'information disponible est telle qu'il est souvent devenu impossible d'accéder à l'ensemble des fonds correspondants par le seul biais de moteurs généralistes tels que Goo-

gle ou Altavista, qui reposent tous sur une simple recherche par mots clés. Il est très vite apparu que la seule possibilité d'offrir un accès fiable et peu bruité était de développer des métadonnées spécifiques, définies par les spécialistes du domaine considéré, et permettant d'identifier ces documents.

L'une des initiatives pionnières en la matière est bien évidemment la TEI (Text Encoding Initiative) qui, dès le tout début des années 1990, impose que tout document conforme aux directives qu'elle publie possède un en-tête contenant un ensemble structuré de descripteurs de son contenu. Cet en-tête peut s'exprimer sous la forme d'une structure hiérarchique reposant sur quatre grandes parties :

- une section dédiée à la description du fichier électronique (<fileDesc>, ou file description), qui contient toutes les informations relatives au document électronique proprement dit (celui qui a été balisé) et permet à un documentaliste de cataloguer ce document ; -
- une section permettant de décrire les choix de transcription et/ou d'annotation adoptés dans le document par rapport à sa source (<encodingDesc>, encoding description). Ces informations portent autant sur le contenu (par exemple, la correction

des erreurs dans la source ou la normalisation de l'orthographe pour des sources anciennes) que sur le balisage ;

- une section qui regroupe les descriptions liées au contenu informationnel du document (<profileDesc>, text profile), et qui peut aussi bien contenir des mots clés, une description du sujet du texte ou, dans le cas par exemple de transcription de données audio, la liste des personnes qui interviennent ;
- une section (<revisionDesc>, revision history) dédiée à la gestion des versions du document électronique en donnant la possibilité d'enregistrer l'histoire des révisions qu'il a subies.

La richesse des métadonnées contenues dans l'en-tête TEI a conduit au développement de nombreuses archives textuelles⁴⁰ exploitant les différents champs disponibles pour offrir des recherches complexes ou des navigations dans des bases de textes. Elle n'a cependant pas conduit à l'organisation de véritables réseaux internationaux d'archives de textes, car il manquait une infrastructure de recueil et de centralisation des métadonnées qu'un utilisateur aurait pu interroger par le biais d'un portail unique. C'est ainsi qu'à vu le jour le projet OLAC (Open Language Archive Community)⁴¹. Ce projet est

⁴⁰ On peut citer par exemples le Women Writers Project (<http://www.wwp.brown.edu/>), le projet Silfide (<http://www.loria.fr/projets/Silfide>) ou encore le Perseus Project (<http://www.perseus.tufts.edu/>), exemplaire par ailleurs pour sa navigation multilingue.

⁴¹ <http://www.language-archives.org/>.

principalement issu de la communauté des linguistes. La communauté OLAC a décidé de s'appuyer au maximum sur des propositions génériques de description de métadonnées et d'infrastructure d'accès à celles-ci. OLAC repose donc sur deux piliers fondamentaux :

- le Dublin Core⁴² comme cadre de description des métadonnées associées à tout document répertorié dans le réseau de sites affiliés à OLAC. Les 15 descripteurs peuvent être associés à des facettes affinant leur sémantique et/ou leur portée. Par exemple, OLAC introduit le champ <Subject.language>, variante du champ <Subject> du Dublin Core, pour indiquer la langue décrite par le document⁴³ ;
- l'OAI (Open Archives Initiative)⁴⁴, comme protocole de moissonnage des métadonnées. Le principe de base utilisé par l'OAI est que chaque site affilié à une initiative de métadonnées procure un accès à un fichier XML – disponible en statique, ou généré dynamiquement à la demande via un script CGI, par exemple.

Parmi les champs offerts par OLAC, on peut noter la possibilité de fournir des indications techniques relatives au format des ressources identifiées par les descripteurs (Format.markup). Les informations peuvent ainsi porter sur le type de la ressource, sur le support

des fichiers primaires audio ou vidéo, ou encore sur le type de codage ou de transcription utilisé. Il est même possible, quand l'information est pertinente (et surtout disponible), de donner une référence à un outil permettant de manipuler la ressource.

Des contenus aux structures

OLAC n'est pas destiné à fournir un accès réel aux ressources, tout au plus une indication que tel ou tel document existe quelque part, accompagnée des principales caractéristiques de celui-ci. Si l'on veut effectivement manipuler la ressource linguistique qui se trouve derrière une description particulière, pour la visualiser, ou exploiter les annotations qui lui sont associées, il faudrait pouvoir connaître avec précision l'organisation du document. Dans ce cadre, OLAC va le plus souvent pointer sur une référence générale indiquant, par exemple, que le document est conforme aux directives de la TEI. En fait, nous rencontrons là le problème de la multiplicité des structures pouvant exister au sein d'un même groupe de documents. Pour des objets de natures extrêmement similaires, des formats très différents peuvent être utilisés. Comment, les identifier avec précision ? Comment garantir qu'un outil logiciel saura ou non traiter un format particulier ? Nous proposons quelques éléments de réponses parmi celles qui

⁴² <http://dublincore.org/documents/dces/>.

⁴³ Langue qui peut, bien sûr, différer de celle utilisée dans le document lui-même. On peut très bien identifier une grammaire en anglais portant sur un créole parlé en Afrique, par exemple..

⁴⁴ <http://www.openarchives.org/>.

ont été proposées jusqu'à présent. De façon générale, connaître la structure d'un document correspond à partager un niveau de connaissance relative à celle-ci entre un émetteur (ou producteur) et un récepteur ou consommateur. Si cette connaissance est partagée a priori, on parle alors d'échange aveugle. Dans le cas contraire, au moment de la transaction, on parle d'échange négocié. Ce dernier type ne garantit pas que l'information pourra être comprise à tout coup et nécessite, comme on le verra dans la partie suivante, des moyens conceptuels et techniques plus élaborés.

L'échange en aveugle peut s'opérer suivant deux modes.

- Le mode implicite repose sur l'hypothèse qu'émetteur et récepteur ont la même connaissance du format partagé, sans souci de vérification au moment de la transmission de l'information. Ce mode, qui est celui de la majorité des échanges de documents HTML sur le Web, requiert un fort taux de tolérance au niveau du destinataire qui n'a aucun moyen de garantir la conformité de la donnée qu'il reçoit au format annoncé.
- Le mode explicite est celui adopté par les différentes initiatives visant à créer des répertoires de DTD ou de schémas XML. Il suppose l'existence d'un espace centralisé où émetteur et récepteur vont trouver une même description de référence des structures qu'ils vont respectivement produire et consommer. Ainsi, Microsoft a ouvert un espace où toute

personne ou projet peut déposer un schéma XML et y faire référence avec la garantie que ce schéma sera toujours accessible.

Le diagnostic de conformité est donc du type tout ou rien, alors même que l'on souhaiterait ignorer par exemple le balisage superflu, ou surtout accepter des documents correspondant à un sous-ensemble du format de référence reconnu par le destinataire. Ce référentiel privilégie le développement de schémas surgénérateurs contenant un maximum d'éléments, et ne répondant donc pas au besoin qu'ont des applications particulières de contrôler finement leurs données.

Vers une description plus fine des structures – retour sur la TEI

L'une des difficultés que l'on rencontre, dès que l'on cherche à décrire plus précisément la structure des documents auxquels on souhaite donner accès en ligne, est de déterminer si les descripteurs doivent s'adresser à un lecteur humain ou à un processus automatique. Ces deux types de destinataires sont vus comme incompatibles : une documentation simplifiée mais explicite va satisfaire le lecteur ayant besoin, par exemple, de connaître le niveau d'annotation d'une ressource linguistique, tandis qu'un outil informatique aura besoin d'une description exhaustive, mais condensée, du document pour pouvoir effectuer son travail. Pour illustrer cette contradiction, nous pouvons nous tourner de nouveau vers la TEI, pionnière

en la matière. L'une de ses caractéristiques les plus importantes est que, malgré l'étendue des possibilités de codage qu'elle autorise, elle s'organise autour d'une DTD modulaire. Celle-ci permet de ne sélectionner qu'un certain sous-ensemble des éléments disponibles. Sur la base de quelques structures communes (le jeu de balise « noyau », dont fait partie l'en-tête), il est possible de sélectionner un type de texte particulier ainsi que des modules additionnels de balisage (noms et dates, pointeurs, etc.).

La configuration la plus simple contient ainsi par défaut une petite centaine d'éléments. Du point de vue de la documentation des formats, cette modularité, même relativement simple, rend quasi impossible l'utilisation de répertoires de DTD ou de schémas. On souhaiterait pouvoir reconnaître les propriétés communes à tous ces formats pour, au moins, savoir que l'on peut appliquer les mêmes feuilles de style, ou encore générer une base de données bibliographiques à partir des en-têtes compris dans l'ensemble des textes d'une archive. Paradoxalement, la TEI n'offre pas de mécanisme plus élaboré de documentation. L'en-tête d'un document donné permet bien d'indiquer la liste des éléments effectivement utilisés dans le corps du document, par le biais de l'élément déclaration de balisage. Cette description ne répond qu'à moitié aux besoins d'interopérabilité exprimés jusqu'à présent.

En prenant un peu de distance, on observe que les mécanismes proposés par la TEI sont intimement liés au fait que celle-ci repose intégralement sur XML pour la description de ses documents. De ce fait, il existe un répertoire d'éléments parfaitement identifiés au sein de la communauté TEI. Quand bien même les informations portées par l'en-tête seraient plus précises, elles ne rendraient la structure du document accessible qu'à un processeur connaissant a priori ce répertoire d'éléments. Aucune garantie d'interopérabilité n'est offerte avec d'autres types de documents. Il serait donc intéressant d'étendre la notion de descripteur de structures pour parvenir à dépasser le cadre spécifique d'un métalangage comme XML.

Modéliser et documenter des structures

Perspective générale

Le survol des initiatives montre bien qu'il reste encore du chemin à parcourir avant de systématiser l'intégration d'informations permettant de reconstituer exactement l'organisation d'un document donné. Avant de généraliser trop rapidement, il est nécessaire de revenir sur la notion même de métadonnée, sur le plan sémantique, et d'envisager de quelle façon il est possible de l'étendre aux informations structurelles. Le principe d'une métadonnée est de fournir une information structurée quant au contenu d'un document destiné à être transmis d'une source A à un destinataire B (celui-ci n'étant pas

nécessairement spécifié). Il faut que le vocabulaire utilisé pour décrire ces métadonnées soit connu de la même façon par A et par B, pour qu'un document annoncé soit reconnu comme tel par le récepteur potentiel. On retrouve ici l'argumentaire développé pour la notion de modèle de document. Le partage de métadonnées est un peu plus complexe. Il faut garantir qu'un descripteur donné sera compris de la même façon par A et par B. Ceci n'est possible que si le descripteur en question est certifié par une autorité tierce, qui en donnera une définition précise garantissant la portée exacte du concept ainsi représenté. La notion même de vocabulaire est restrictive, car rien ne garantit que toutes les communautés, linguistiques en particulier, souhaiteront utiliser les mêmes termes pour désigner les mêmes concepts. L'autorité doit alors devenir un véritable répertoire de métadonnées, permettant à chacun de mettre en relation ses propres descripteurs avec un concept central. Au-delà de la notion de vocabulaire partagé, une métadonnée ne peut être comprise que si elle correspond au partage d'une véritable ontologie de concepts. Que ceux-ci soient génériques (cas du Dublin Core), ou propres à un domaine d'application particulier (par exemple, OLAC⁴⁵).

Cette vision implique d'un point de vue technique, que l'émetteur A communique, en même temps que les informations relatives à son document, l'appariement de son propre vocabulaire avec les catégories du registre de référence. Soit en transmettant, comme une sorte d'en-tête à ses métadonnées, la liste de ses appariements, de façon à donner au récepteur une clé d'interprétation de ses propres données. Soit en neutralisant ses données (en les exprimant dans un format pivot indépendant d'un vocabulaire particulier et donc reposant uniquement sur les catégories abstraites du répertoire central). Bien que la seconde solution semble la plus facile à mettre en œuvre, on doit garder la première à l'esprit car elle correspond à toute situation où l'on accède à des données statiques propriétaires qui n'auraient pas été neutralisées. Supposons maintenant qu'il ne s'agisse pas seulement d'identifier un document au regard d'un certain nombre de descripteurs mais, de comprendre l'organisation interne des données contenues dans ce document. De son côté, le récepteur souhaite intégrer ce document dans une base où les paragraphes sont indiqués à l'aide de l'élément <para>. En l'absence de tout autre mécanisme, une telle situation nécessiterait à l'émetteur et au récepteur de définir ensemble la correspondance entre ces deux éléments

⁴⁵ On peut aussi mentionner les travaux réalisés par le projet européen ISLE pour la définition d'un ensemble de métadonnées pour les ressources linguistiques, qui va bien au-delà, en finesse et en précision, du projet OLAC. C'est en particulier la nécessité d'unifier les descripteurs d'OLAC et de ISLE qui conduit l'ISO, dans le cadre de son nouveau comité TC37/SC4, à mettre en place un répertoire conceptuel unique de métadonnées pour les ressources linguistiques.

(et probablement quelques autres...) pour aboutir à un filtrage éventuel d'un format vers l'autre. Nous sommes donc dans une situation d'échange négocié qui alourdit toute transmission et surtout doit être réitéré à chaque nouvelle situation d'échange. Si l'on suppose alors qu'il existe un registre de descripteurs de structures textuelles, c'est-à-dire un lieu où des notions telles que /paragraphe/ sont définies de façon précise, au sein d'une communauté, ou mieux au niveau d'une organisation internationale telle que l'ISO, la situation devient différente. La notion de paragraphe ayant été définie préalablement à la transmission, l'émetteur peut communiquer en même temps que son document un appariement entre éléments XML qu'il utilise et les catégories correspondantes au sein du registre. Un tel registre, pour peu qu'il contienne une base suffisamment large de descripteurs, décrite une véritable ontologie de structures documentaires contribuant à l'élaboration de la sémantique d'un document particulier.

Là encore, une possibilité – que nous rencontrerons avec le langage GMT – est de neutraliser la structure du document en passant par un format pivot plus abstrait. Reste que l'on perçoit que le passage de la métadonnée simple à l'échange explicite d'informations relatives à la structure des documents semble être envisageable. Le paysage n'est pourtant pas si rose qu'il paraît :

- nous n'avons pas encore les moyens de décrire de tels registres. C'est l'objet de la partie suivante ;

- nous n'avons pas encore de garantie que l'intuition de faisabilité de l'échange d'information sur les structures est valide.

Nous nous appuyerons sur l'expérience acquise dans le cadre de la plateforme TMF (Terminological Markup Framework) pour avancer dans ce sens. Ce style permet de gérer les espaces entre les titres et le premier paragraphe de votre texte.

Métadescription des métadonnées

Comme nous l'avons vu, le partage d'une métadonnée correspond à l'existence d'un système conceptuel connu à la fois de l'émetteur et du récepteur de cette métadonnée. Ce système conceptuel, suivant le domaine d'origine de celui qui va l'aborder, sera vu alors de deux façons relativement différentes :

- il pourra être considéré comme la donnée d'une véritable ontologie de concepts, tels qu'on peut en rencontrer dans le domaine de la représentation des connaissances. On envisagera alors des langages élaborés pour la représenter, langages qui, dans le monde SGML/XML, prendront les noms de Topic Maps, ou OIL-DAML ;
- il pourra être simplement vu comme un ensemble d'unités techniques qu'il faut créer, maintenir et rendre accessible de façon cohérente pour que le système de communication fonctionne. C'est typiquement l'approche adoptée par la norme ISO/IEC-11179.

L'ISO/IEC-11179 est une norme internationale en six parties qui propose un cadre de description d'éléments de données (data element), de champs pouvant être utilisés pour exprimer des métadonnées. Il s'agit donc d'un format de « méta-métadonnées ». L'objectif est :

- de permettre une description consistante des éléments de données d'une application à une autre (en fournissant, par exemple, des directives précises concentrant la formulation des définitions),
- d'autre part, de fournir des méthodes d'enregistrement et de mise à jour de ces éléments de données.

Dans sa troisième partie, l'ISO/IEC-11179 introduit un certain nombre d'attributs pouvant être utilisés pour la description de l'élément de donnée considéré. Dublin Core a adopté le cadre de la norme ISO-11179 pour décrire son ensemble d'éléments de métadonnées. Sur cette base, un projet plus spécifique tel qu'OLAC, gardant le Dublin Core comme base, va pouvoir raffiner certains champs.

Principes généraux

Nous reprenons ici les principaux éléments formant la base de la plate-forme TMF (Terminological Markup Framework)⁸, qui devrait devenir le futur standard ISO 16642 de description de données terminologiques informatisées. La proposition de cette norme était double :

- décrire et de comparer les formats d'échange terminologiques existants tels que MARTIF ou Geneter, tant en terme de couverture descriptive que pour identifier les conditions d'interopérabilité entre ces formats.
- répondre à un besoin pressant de l'industrie de disposer de plus de flexibilité dans la définition de ses propres formats, tout en gardant un maximum de compatibilité avec les normes existantes.

La plate-forme TMF permet de décrire un format quelconque pour l'échange de terminologies multilingues en utilisant, par exemple, XML comme syntaxe d'implémentation. TMF ne décrit aucun format particulier, mais fait office de langage de spécification, sur la base des notions élémentaires suivantes :

- un métamodèle : une organisation générale des données partagée par tous les formats⁴⁶ pouvant être engendrés par la plate-forme TMF, et qui décompose la structure d'une base de données terminologiques en composants élémentaires. Ce modèle est conforme à la vision traditionnelle d'une entrée terminologique composante d'un réseau conceptuel, vision d'Eugen Wüster largement adoptée dans la communauté de la terminologie ;
- des unités élémentaires d'information (ou catégories de données), qui sont extraites d'un répertoire de ca-

⁴⁶ Ou TML, Terminological Markup Language.

tégories de données (DCR – Data Category Registry) en fonction des besoins de l'application courante ;

- des méthodes d'implémentation permettant de décrire l'ensemble des caractéristiques du format considéré (ou TML) en instanciant le métamodèle et les catégories de données à l'aide de vocabulaires idoines.

Interopérabilité entre deux formats de représentation de terminologies

Étant donné deux TML conformes aux spécifications de la plate-forme TME, l'expression des conditions d'interopérabilité se réduit à la comparaison de leurs spécifications respectives en termes de catégories de données, de par le fait qu'elles partagent exactement le même métamodèle⁴⁷. De fait, les méthodes de spécification de la plate-forme TMF permettent de faire un diagnostic précis de la quantité d'information qui sera préservée ou perdue lorsque l'on passe d'un TML à un autre, ce que l'on nomme la « bande passante d'interopérabilité ».

Styles et vocabulaires

L'évaluation de la bande passante d'interopérabilité entre deux TML repose uniquement sur les catégories

de données proprement dites. Elle se trouve donc entièrement indépendante de l'implémentation de celles-ci en tant qu'objets XML. Pour cette raison, la plate-forme TMF possède un mécanisme complémentaire, permettant de décrire la façon dont un TML sera concrètement implémenté comme type de document XML. Il s'agit en fait d'associer à chaque catégorie de données un style⁴⁸ qui sélectionne une réalisation possible en XML, d'une part, et un vocabulaire, correspondant aux chaînes de caractères nécessaires à l'expression de ce style, d'autre part. Cela permet d'utiliser un élément <générique> pour représenter l'unité d'information correspondante.

Représentation des catégories de données en RDF

La comparaison de deux TML n'est possible que s'il existe un registre central de catégories de données, les représentant de façon consistante. Nous avons vu comment la norme ISO/IEC-11179 pouvait servir de cadre à la formalisation d'un tel registre. Cependant, afin que le processus de spécification et de comparaison de TML puisse être automatisé, la plate-forme TMF s'appuie⁴⁹ sur une formalisation accrue de l'ISO/CEI 11179 qui exprime ses différents attributs à l'aide du format RDF.

⁴⁷ La proposition de norme ISO/CD 16642 identifie plus précisément les conditions liées aux valeurs (types) des catégories de données, à l'ancrage de celles-ci sur les nœuds du métamodèle, etc.

⁴⁸ TMF autorise cinq styles (Attribute, Element, Typed Element, Valued Element, Typed Valued Element), pour couvrir les différentes possibilités rencontrées dans les formats en terminologie.

⁴⁹ Ce modèle sert de base à la partie 1 de la proposition de révision de la norme ISO-12620.

Cette représentation présente l'avantage de permettre à un utilisateur qui souhaite spécifier un format conforme à TMF d'ajouter ses propres catégories de données à celles qu'il aura extraites d'un registre de référence.

Les catégories de données vues comme métadonnées

On peut anticiper que les mécanismes décrits précédemment ouvrent le chemin à l'automatisation de nombreux traitements. Tout d'abord, une spécification complète permet de générer le schéma XML d'un TML ainsi décrit. Cette même spécification permet de générer un filtre XSLT permettant de transformer des données d'un TML donné vers le format pivot GMT (et inversement), et, par transitivité, de générer les filtres de transformation entre deux TML (dans la limite de la bande passante d'interopérabilité).

Bilan et perspectives dans un contexte de normalisation des formats de données

Nous avons essayé de présenter dans ce chapitre différents éclairages destinés à intégrer la représentation de la structure de documents dans le champ plus global des métadonnées. Ce faisant, nous avons pu voir que cela, d'une part, conduisait à une réflexion de fond sur l'organisation et le partage de métadonnées au travers de registres internationaux de référence, et permettait, d'autre part, d'aborder le problème relativement complexe de la spécification de formats de documents, notamment dans le cadre des possibi-

lités offertes par le métalangage XML. Bien que limitée à un type de données très particulier, l'expérience acquise dans le cadre de la plate-forme TMF semble donner l'espoir qu'il soit possible, pour certaines classes de document (qui restent à identifier), de généraliser les mécanismes présentés ici. Ceci est particulièrement important dans un contexte où il existe des forces antagonistes au sein de nombreuses communautés qui souhaitent à la fois normaliser de façon de plus en plus précise les données qu'elles échangent, et conserver une certaine souplesse dans la description même de ces formats. Il n'est donc plus question de normaliser au niveau international un schéma XML particulier, mais de fournir les moyens aux différents acteurs d'une communauté de comparer leurs usages et de maintenir éventuellement les formats de données qu'ils manipulaient précédemment. Il reste probablement encore beaucoup de chemin à parcourir pour que toute application puisse être intégrée dans cette perspective. Il faut en effet identifier les méta modèles correspondants, ajouter éventuellement certains mécanismes que le domaine des terminologies ne nécessitait pas, mettre en œuvre des répertoires thématiques de catégories de données, et probablement définir des modes plus rigoureux de gestion de ces catégories de données sur la base de l'expérience acquise en représentation des connaissances. Il semble malgré tout que ce soit là des activités indispensables si l'on veut accroître les niveaux d'interopérabilité entre applications dans les années à venir.

notes

mise en page XFcréation — Paris

achevé d'imprimer à Mérignac par

COPY-MEDIA

23 rue Francisco Ferrer

33 700 Mérignac

France

www.copy-media.net

ISBN : 978-2-9527724-1-9



9 782952 772419

La Délégation générale à la langue française et aux langues de France

De tous les liens que nouent les hommes dans la cité, le lien de la langue est le plus fort, parce qu'il fonde le sentiment d'appartenance à une communauté. Parce que la mondialisation des échanges et les progrès de la construction européenne ne cessent de le faire évoluer, les pouvoirs publics sont appelés à réaffirmer une politique de la langue qui, tout en veillant à garantir la primauté du français sur le territoire national, participe à l'effort de cohésion sociale et contribue à la promotion de la diversité culturelle en Europe et dans le monde.

La délégation générale à la langue française et aux langues de France élabore la politique linguistique du Gouvernement en liaison avec les autres départements ministériels.

Organe de réflexion, d'évaluation et d'action, elle anime et coordonne l'action des pouvoirs publics pour la promotion et l'emploi du français et veille à favoriser son utilisation comme langue de communication internationale. Elle s'efforce de valoriser les langues de France et de développer le plurilinguisme.

L'AILF (Association des Informaticiens de Langue Française) organise, depuis presque 20 ans, un colloque biennal, Lexipraxi. Celui-ci traite des TIC, notamment, de leur appropriation socioprofessionnelle, dans un contexte francophone et, plus largement, dans un environnement multilingue.

L'ensemble des textes produits tout au long de ces 20 années représente un patrimoine qu'il nous a semblé intéressant de fouiller, pour en mettre les visions à l'épreuve et en restituer les tendances.

ISBN : 978-2-9527724-1-9



collection **LEXIPRAXI**