

Bibliothèques et accès à l'information multilingue : pour en finir avec Babel

Libraries and Access to Multilingual Information: The End of Babel

Bibliotecas y acceso a la información multilingüe: para ponerle fin a Babel

Elaine Ménard

Volume 57, numéro 4, octobre–décembre 2011

URI : <https://id.erudit.org/iderudit/1028989ar>

DOI : <https://doi.org/10.7202/1028989ar>

[Aller au sommaire du numéro](#)

Résumé de l'article

La bibliothèque est de plus en plus confrontée à des demandes complexes sur les plans multimédia, multiculturel et multilingue. Cet article évoque les défis auxquels font face les bibliothèques à l'heure du numérique. La problématique de l'accès à l'information multilingue est décrite, de même que les solutions proposées par divers travaux de recherche.

Éditeur(s)

Association pour l'avancement des sciences et des techniques de la documentation (ASTED)

ISSN

0315-2340 (imprimé)

2291-8949 (numérique)

[Découvrir la revue](#)

Citer cet article

Ménard, E. (2011). Bibliothèques et accès à l'information multilingue : pour en finir avec Babel. *Documentation et bibliothèques*, 57(4), 205–211.
<https://doi.org/10.7202/1028989ar>

Tous droits réservés © Association pour l'avancement des sciences et des techniques de la documentation (ASTED), 2011

Ce document est protégé par la loi sur le droit d'auteur. L'utilisation des services d'Érudit (y compris la reproduction) est assujettie à sa politique d'utilisation que vous pouvez consulter en ligne.

<https://apropos.erudit.org/fr/usagers/politique-dutilisation/>

érudit

Cet article est diffusé et préservé par Érudit.

Érudit est un consortium interuniversitaire sans but lucratif composé de l'Université de Montréal, l'Université Laval et l'Université du Québec à Montréal. Il a pour mission la promotion et la valorisation de la recherche.

<https://www.erudit.org/fr/>

Bibliothèques et accès à l'information
multilingue : pour en finir avec Babel *

ELAINE MÉNARD

School of Information Studies
McGill University, Montréal, Canada
elaine.menard@mcgill.ca

RÉSUMÉ | ABSTRACTS | RESUMEN

La bibliothèque est de plus en plus confrontée à des demandes complexes sur les plans multimédia, multiculturel et multilingue. Cet article évoque les défis auxquels font face les bibliothèques à l'heure du numérique. La problématique de l'accès à l'information multilingue est décrite, de même que les solutions proposées par divers travaux de recherche.

***Libraries and Access to Multilingual
Information : The End of Babel*****

The library is being increasingly confronted by complex demands for multimedia, multicultural and multilingual materials. This article describes some of the challenges faced by libraries in a digital era. The problems associated with the access to multilingual information are described as well as solutions proposed by various research projects.

***Bibliotecas y acceso a la información
multilingüe : para ponerle fin a Babel******

La biblioteca enfrenta cada día pedidos más complejos en el ámbito multimedial, multicultural y multilingüe. Este artículo evoca los desafíos que enfrentan las bibliotecas en la era digital. Se describe la problemática del acceso a la información multilingüe, así como las soluciones propuestas a través de varios trabajos de investigación.

La bibliothèque du XXI^e siècle

TRADITIONNELLEMENT, la bibliothèque acquiert des documents sur supports imprimés tels que des livres, journaux, cartes, illustrations, etc. Cette image un peu poussiéreuse, un peu surannée de la bibliothèque a pourtant évolué depuis quelques années. Malgré cette conception quelque peu erronée mais largement répandue, la bibliothèque ne se limite plus seulement à l'acquisition de documents sur support papier. À l'heure du numérique, l'effort des bibliothèques se porte maintenant sur l'organisation et la mise en valeur des collections fort différentes. La bibliothèque, avec l'avènement des documents à contenu numérique, entre de plain-pied en compétition avec les intervenants aussi nouveaux que différents que sont les sites commerciaux, les blogs et autres sites Web issus de réseaux sociaux offrant une panoplie de contenus. La bibliothèque doit faire face à de nouveaux défis et s'adapter à de nouvelles réalités. Si elle veut survivre, la bibliothèque doit se doter des outils nécessaires pour mettre en valeur de nouveaux contenus et répondre aux besoins d'utilisateurs appartenant à des communautés linguistiques diverses.

La bibliothèque a pour mission de fournir les ressources et les services documentaires requis aux fins d'enseignement, de recherche et d'activités culturelles diverses. De plus en plus, les bibliothèques du monde entier se préoccupent de la cueillette, du traitement, de la conservation et de la diffusion des publications numériques diffusées sur Internet. La bibliothèque doit aussi organiser ses documents de manière à ce que ceux-ci soient accessibles localement, mais également à distance. L'organisation présuppose évidemment la description, l'indexation et la classification des ressources acquises. Le but premier de cette organisation documentaire est de permettre une exploitation et une diffusion efficaces et efficaces de cette documentation accessible par l'ensemble de la communauté.

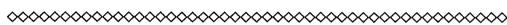
David Crystal (2006), un spécialiste des langues, mentionnait que le Web est un moyen de communication changeant qui accueille tous les styles linguistiques au sein d'une langue donnée, mais également toutes les langues. La connaissance des comportements linguistiques sur Internet et la mesure de leur diversité constituent des questions importantes. Il est toutefois difficile d'estimer avec précision la diversité

* Cette réflexion a été présentée lors de la table ronde « Bibliothèques et Sciences de l'information : les projets de recherche comme espaces de dialogue », dans le cadre des Entretiens Jacques-Cartier, Lyon, 24 novembre 2010.

** This paper was presented at the « Libraries and Information Sciences : Research Projects as a Places of Dialogue » round table, held at the Entretiens Jacques-Cartier, in Lyon, November 24, 2010.

*** Esta reflexión fue presentada durante la mesa redonda « Bibliotecas y ciencias de la información : proyectos de investigación como espacios de diálogo » en el marco de las Conversaciones en el centro Jacques-Cartier, que se desarrollaron en Lyon, el 24 de noviembre de 2010.

Le passage de la bibliothèque traditionnelle au numérique ne se caractérise pas strictement par une évolution technologique, mais nécessite un changement dans le paradigme par lequel les individus accèdent et interagissent avec l'information.



linguistique du Web. Certaines statistiques annuelles rapportent régulièrement le recul du nombre d'internautes de langue maternelle anglaise. Par exemple, en 2010, on rapportait que la proportion d'utilisateurs anglophones du Web n'était plus que de 27,3 %, comparativement à 35,9 % en 2004 (Internet World Stats 2010). Cette mutation implique nécessairement une réorganisation des services offerts sur Internet, y compris ceux des bibliothèques.

Les défis de la bibliothèque : de l'étagère à l'écran

Avec l'avènement du numérique, le document imprimé n'est plus au centre des préoccupations des bibliothèques ; l'accès à d'autres formes documentaires, jusqu'ici moins accessibles, s'avère maintenant possible. D'autres types de documents (tableaux, cartes, images, dessins, vidéos, fichiers sonores) font désormais partie des collections privilégiées par les utilisateurs des bibliothèques. Si de nombreux documents composés de textes et d'autres types d'informations multimédias peuvent souvent être repérés grâce à la présence de texte lié, l'accès à des documents strictement non-textuels continue de se faire plus difficilement. En outre, ce problème se complique lorsque les documents combinent des éléments linguistiques de diverses natures.

Le passage de la bibliothèque traditionnelle au numérique ne se caractérise pas strictement par une évolution technologique, mais nécessite un changement dans le paradigme par lequel les individus accèdent et interagissent avec l'information. La bibliothèque traditionnelle s'est longtemps concentrée sur l'entreposage et la conservation des éléments physiques, notamment des livres et des périodiques, sur le catalogage de haut niveau de détail et sur un système de navigation basé sur la proximité physique des sujets liés. Par exemple, les livres portant sur la religion étaient regroupés sur les rayons. L'organisation passive de l'information faisait en sorte que l'utilisateur devait venir à la bibliothèque pour en faire usage, puisque celle-ci était physiquement accessible en un seul endroit.

En revanche, le contenu numérique des bibliothèques diffère puisque l'accès aux documents numériques peut se faire de n'importe où et les exigences en termes d'entreposage du support physique sont élimi-

nées. Le catalogage peut exploiter le texte intégral et la navigation est basée sur des liens hypertextes, des mots clés, des systèmes de classifications ou de mise en grappes (*clustering*). Les documents traitant d'un même sujet n'ont plus besoin d'être physiquement situés les uns près des autres. Grâce à des moyens technologiques de plus en plus sophistiqués, les utilisateurs n'ont plus nécessairement besoin de visiter la bibliothèque, puisque l'accès et la diffusion de l'information se fait par voie électronique. Phénomène des temps nouveaux, la bibliothèque se trouve à tout endroit où l'utilisateur peut y avoir accès, de la maison à l'école, en passant par le bureau, l'aéroport, et à partir de presque tous les moyens de transports terrestres et autres. Il faut également constater que l'avènement des téléphones mobiles, ordinateurs portatifs et bloc-notes électroniques a considérablement modifié les habitudes des internautes, de même que le mode d'accès aux catalogues en ligne.

Le catalogue de bibliothèque à l'heure du numérique

Un document numérique se présente comme un ensemble de fichiers identifiés par un identifiant unique. Ce document, quel qu'il soit, doit être décrit de manière à ce qu'il soit repérable par une majorité d'individus. Cette description comporte un ensemble de métadonnées. Sans ses métadonnées, le document numérique n'a aucune existence réelle puisqu'il demeure inaccessible. Les métadonnées descriptives ont pour principale fonction de fournir une description bibliographique approfondie et détaillée dans un format normalisé, permettant l'accès au document de même que l'échange avec d'autres institutions. Les métadonnées structurelles ont pour objectif de reconstituer la structure du document, c'est-à-dire de faire connaître tous les fichiers qui composent un document (textes, images, vidéos, sons) et d'établir les relations qui existent entre eux. Finalement, les métadonnées administratives sont nécessaires pour la gestion des droits d'accès (droits d'auteur, confidentialité) et d'utilisation (droits d'impression, de reproduction, de modification). Elles ont également pour objectif de garantir l'intégrité des fichiers et le suivi de leurs éventuelles modifications et de préserver les informations techniques nécessaires à la lecture des fichiers (Haynes 2004).

Tel un outil clé, le catalogue est le point d'entrée par excellence des bibliothèques. Il permet de repérer les documents grâce à des points d'accès divers : auteurs (personnes physiques ou collectivités), sujets, indices de classification ou collections. Tous les types de documents sont décrits et inclus dans le catalogue : texte imprimé (livres, périodiques, etc.), musique imprimée et manuscrite, documents iconographiques, enregistrements sonores, images animées, documents électroniques, documents multisupports et multimédias,

Figure 1

Interface de recherche - Catalogue IRIS (BANQ)

The screenshot shows two search sections. The top section, 'Recherche simple', has radio buttons for 'Tout', 'Livres', 'Musique', 'Films', 'Revue et journaux', and 'Images'. Below these is a search input field, a 'Chercher' button, and an 'Effacer' button. The bottom section, 'Quick search', has radio buttons for 'All', 'Books', 'Music', 'Films', 'Magazines and newspapers', and 'Images'. It also features a search input field, a 'Search' button, and a 'Clear' button. Both sections include links for 'Aide' and 'Poser une question'.

documents cartographiques, documents d'archives et manuscrits, etc.

Même si la recherche dans le catalogue en ligne ne donne pas nécessairement accès à l'objet documentaire en tant que tel, il permet un premier contact entre l'utilisateur et le document. La notice bibliographique contient différents types d'information, les méta-données, relatives à un document. Plus précisément, la notice comporte les renseignements permettant à l'utilisateur d'identifier rapidement un document (titre ou nom du ou des auteurs), des informations destinées à le localiser avec précision (la cote d'un livre dans une bibliothèque, le numéro de volume, de fascicule et le numéro de la première page pour un article de périodique), des identifiants décrivant une entité bibliographique de façon unique (un numéro ISBN ou ISSN, un lien URL), de même que des informations permettant à l'utilisateur du catalogue de se faire une idée concrète du contenu (termes d'indexation, résumé).

L'utilisateur du catalogue en ligne accède à ces informations grâce à une interface offrant des options de recherche permettant la recherche simple (titre de l'ouvrage, nom de l'auteur) et la recherche dite « avancée » lui permettant de combiner plusieurs clés d'accès de manière à raffiner la requête. L'examen des interfaces des catalogues de bibliothèques révèle que celles-ci varient considérablement d'une bibliothèque à l'autre, en ce qui concerne les fonctionnalités offertes. Trop souvent, le manque flagrant de convivialité et des clés de recherche offrant un vocabulaire irrémédiablement hermétique engendrent beaucoup d'incertitude et de frustration chez l'utilisateur. De plus en plus habitué à chercher avec des moteurs de recherche de type Google, l'utilisateur qui tente sa chance dans un catalogue de bibliothèque abandonnera rapidement ses recherches et se tournera vers d'autres outils où il se sentira définitivement plus en confiance et pourra se dire en lançant sa requête « *J'ai de la chance* ».

Une autre difficulté à laquelle se heurte l'utilisateur du catalogue est d'ordre linguistique. Par exemple, le catalogue IRIS de Bibliothèque et Archives nationales du Québec (BANQ 2011) offre la possibilité de chercher des notices bibliographiques avec des fonctionnalités de recherche en langue française et anglaise (Figure 1).

En plus de modifier l'interface de recherche, le catalogue IRIS, à l'instar de nombreux catalogues, offre

Figure 2

Affichage des résultats - Catalogue IRIS (BANQ)

The screenshot shows a bibliographic record with the following fields:

- Titre**: Contes du lundi / Alphonse Daudet ; préface et commentaires de Catherine Eugène.
- Auteur**: Daudet, Alphonse, 1840-1897 [277]
- Éditeur**: Paris : Pocket, 1999. [4741]
- Description**: xvi, 364, xxxiii p. ; 18 cm.
- Sujets**: Guerre franco-allemande, 1870-1871--Romans, nouvelles, etc. [33]; Paris (France)--Histoire--1871 (Commune)--Romans, nouvelles, etc. [43]

 The same record is shown again below with slightly different field labels:

- Title**: Contes du lundi / Alphonse Daudet ; préface et commentaires de Catherine Eugène.
- Author**: Daudet, Alphonse, 1840-1897 [277]
- Publisher**: Paris : Pocket, 1999. [4741]
- Description**: xvi, 364, xxxiii p. ; 18 cm.
- Subjects**: Guerre franco-allemande, 1870-1871--Romans, nouvelles, etc. [33]; Paris (France)--Histoire--1871 (Commune)--Romans, nouvelles, etc. [43]

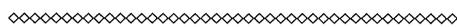
un affichage des résultats dans la langue choisie par l'utilisateur (Figure 2).

Une fois repérées, les notices bibliographiques présentent l'information de manière à peu près uniforme et consistante grâce à des règles de catalogue aussi complexes que strictes. Toutefois, dans la majorité des catalogues en ligne cette information est affichée en une seule langue, la langue de création de la notice. C'est le cas par exemple des vedettes-matière, mot ou groupe de mots indiquant le sujet sous lequel tous les écrits ou autres documents traitant du même thème sont entrés et regroupés (Office québécois de la langue française 2011). Les vedettes-matière sont tirés de répertoires ou de listes qui n'existent souvent qu'en une seule langue. Par conséquent, la notice bibliographique, une fois affichée, présentera l'information sujet en une seule langue dans les champs Sujets/*Subjects* (Figure 2). L'utilisateur unilingue du catalogue se heurtera ainsi à une difficulté importante. En effet, même s'il parvient à faire sa recherche grâce aux fonctionnalités de recherche multilingue offertes par le catalogue, l'examen des sujets demeurera un mystère pour lui puisque ces représentations de sujets sont extraites d'un répertoire élaboré dans une langue qui ne lui est pas nécessairement familière.

Accès à l'information multilingue numérique

L'accès à l'information multilingue est un problème de recherche pour lequel des solutions adaptées ne sont toujours pas mises en place. En 2001, l'UNESCO

*L'accès à l'information multilingue
numérique constitue sans aucun
doute le défi le plus sérieux auquel la
bibliothèque doit faire face, en particulier
dans les pays non anglophones.*



s'engageait, entre autres, à « *promouvoir la diversité linguistique dans l'espace numérique et encourager l'accès universel, à travers les réseaux mondiaux, à toutes les informations qui relèvent du domaine public* » (UNESCO 2003). Cet engagement ferme de l'UNESCO laisse déjà entrevoir un intérêt croissant envers l'information multilingue. Toutefois, dix ans plus tard, cet engagement ne demeure qu'un pieux souhait. Pourtant, selon Fluhr (2010), les applications nécessitant la prise en compte d'informations multilingues sont fort nombreuses : veille stratégique, commerciale, scientifique, brevet, surveillance d'opinion au niveau international, partage d'information entre sociétés multinationales, partage d'information dans des pays plurilingues, sécurité nationale (lutte contre le terrorisme, prolifération nucléaire), etc. L'accès à l'information multilingue numérique constitue sans aucun doute le défi le plus sérieux auquel la bibliothèque doit faire face, en particulier dans les pays non anglophones. La question linguistique soulève de nombreuses questions pouvant se résumer ainsi : comment les utilisateurs des catalogues en ligne peuvent-ils accéder à des documents numériques divers dans des langues étrangères de manière à en faire une utilisation efficace ?

La problématique de la recherche d'information multilingue n'est pas nouvelle puisque dans les années 1950 les documentalistes étaient déjà conscients de l'importance de se préoccuper de ce type d'information (Ménard 2008). Comme le suggérait déjà Suzanne Briet à cette période :

« L'importance de la connaissance des langues étrangères pour assurer l'une quelconque des formes documentaires n'a plus besoin d'être démontrée. Pour comprendre les documents, il faut pouvoir les lire [...]. Ce sera le rôle du documentaliste de mettre les documents en des langues diverses à la portée de ses usagers » (1951).

La valeur de l'information scientifique, culturelle et pédagogique des nombreuses collections numérisées des services patrimoniaux, des musées, des bibliothèques et des archives de différentes communautés linguistiques est incommensurable. Il faut mentionner que l'information multilingue est tout aussi capitale pour les entreprises qui désirent demeurer compétitives, surtout avec l'émergence du phénomène de mondialisa-

tion qui magnifie la concurrence entre celles-ci (Ménard 2008). De plus, les acteurs du commerce électronique reconnaissent progressivement la nécessité d'exploiter la langue et les traditions culturelles spécifiques des marchés visés (Hillier 2003). La communication multilingue permet la diffusion d'information au-delà des frontières des langues. Tous les secteurs de l'économie mondiale, de même que les populations de diverses cultures doivent maintenant composer et maîtriser la communication multilingue. D'une part, l'information numérique a été créée en plus d'une langue et, d'autre part, des populations d'utilisateurs linguistiquement différents et d'horizons culturels divers ont accès à cette information disponible au bout de leurs doigts. Les gouvernements, les entreprises et industries à rayonnement international recherchent constamment des informations dans des journaux et autres ressources Web rédigées en langues diverses (Chen et Bao 2009). L'étude des technologies et des ressources facilitant les échanges multilingues est donc devenue nécessaire afin d'aider les internautes, y compris les utilisateurs des catalogues de bibliothèques, à naviguer, chercher et utiliser ces ressources multilingues textuelles ou multimédias.

L'information est désormais disponible sur le Web en plusieurs langues et les utilisateurs d'Internet communiquent non seulement dans une *lingua franca* établie, l'anglais, mais aussi dans une multitude d'autres langues. Certains moteurs de recherche offrent à leurs utilisateurs les moyens d'accéder à cette information. Ces outils linguistiques fournissent aux utilisateurs une fonctionnalité dite « recherche traduite » qui permet, entre autres, de chercher dans d'autres langues et, dans la foulée, de traduire les documents textuels résultant de leurs requêtes.

Par exemple, Google a lancé son outil linguistique en mai 2007 et est maintenant identifié comme le moteur de recherche régulier offrant le meilleur support dédié à l'information multilingue (Zhang et Lin 2007). Ces outils linguistiques fournissent aux utilisateurs une fonctionnalité de « traduction des résultats » (Outils linguistiques Google 2011). Toutefois, ces fonctionnalités sont encore très peu répandues. Pourtant, Oard révélait en 1997 que les technologies de recherche d'information multilingue (RIML) pourraient considérablement améliorer les services commerciaux en ligne, tels que Dialog et Lexis/Nexis. Pour leur part, Petrelli, Beaulieu et Sanderson (2002) soulignaient que les traducteurs et les spécialistes de l'information, tels que les bibliothécaires de référence servant d'intermédiaires pour leurs usagers, pourraient tirer un grand avantage des mécanismes d'accès à des sources multilingues d'information.

Pour le moment, ces outils linguistiques fonctionnent en vase clos et ne sont pas systématiquement offerts à la recherche de tous les types de documents. L'accès aux images numériques se limite toujours à des résultats monolingues. Par exemple, les images affichées

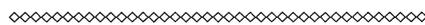
pour la requête « voiture » seront l'ensemble des images indexées avec le mot « voiture ». Si le chercheur d'images désire éventuellement augmenter ou varier les résultats avec le mot équivalent en anglais « cars », il lui faudra formuler une autre requête en anglais. Et s'il n'est pas familier avec le mot anglais, le chercheur d'image devra trouver un moyen de traduire sa requête « voiture » en anglais, puisque le moteur de recherche ne fournit pas nécessairement un lien direct entre les fonctionnalités de recherche d'images et les mécanismes de traduction. Cette gymnastique entre les options de recherche et de traduction n'est pas à la portée de tous les utilisateurs et provoque inévitablement de la frustration, voire même du découragement. La plupart des utilisateurs continuent donc de chercher des images numériques dans une seule langue qui les privera d'accès à de véritables trésors de ressources visuelles.

La recherche d'information multilingue (riml)

L'accès à l'information multilingue est un problème de recherche qui demande toujours qu'on s'y attarde. Cette information existe dans de nombreux contextes. On pensera d'abord aux bibliothèques traditionnelles ou numériques, sources privilégiées d'information depuis toujours. Comment celles-ci peuvent-elles mieux faire face aux nombreux défis engendrés par la prolifération de documents en langues diverses et par une diversification accrue de leurs clientèles en raison d'un accès désormais quasi universel à leurs ressources ? On pensera également aux trésors du Web. Ainsi, NETendances (2010) indique que la progression de l'utilisation d'Internet s'accompagne inévitablement d'une augmentation dans la recherche d'information (Centre francophone d'informatisation des organisations 2010). Par exemple, même si les Québécois préfèrent toujours connaître les actualités et les nouvelles par le truchement de la télévision, on remarque que celle-ci est en perte de vitesse au profit des journaux numériques et d'Internet. On peut se demander comment les internautes peuvent s'assurer un accès complet et approprié à un ensemble de documents numériques en langues diverses. La réponse à ces questions se trouve sans aucun doute dans la conception, le développement et l'implantation de solutions documentaires adaptées à la diversité linguistique des documents numériques accessibles en ligne.

Sur la lancée des campagnes d'évaluation du Text REtrieval Conference (TREC), la recherche d'information multilingue (RIML) a été abondamment étudiée au cours des 15 dernières années. La RIML est un type de recherche d'information permettant de repérer un document lorsque la langue de la requête est différente de la langue des documents repérés (Oard et Diekema 1998). Deux approches sont habituellement proposées en RIML. La première consiste à traduire les documents au

Certains travaux en RIML étudient les systèmes de questions réponses (QR) qui recherchent, dans un ensemble de documents, la réponse à une question posée en langage naturel (par exemple, quel est le premier homme à avoir marché sur la lune ?).

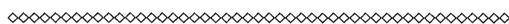


complet dans toutes les langues, alors que la deuxième suppose la traduction des requêtes dans la langue des documents à repérer (Kishida 2005). Ces deux approches font généralement appel à trois types de ressources linguistiques pour la traduction : les dictionnaires bilingues ou multilingues, les systèmes de traduction automatique et les corpus parallèles ou comparables (Ménard 2008). Ces ressources linguistiques ont chacune démontré leurs forces et leurs faiblesses. Or, il semble que la démarche la plus prometteuse consiste à combiner plusieurs ressources différentes pour la traduction (Braschler 2004).

De nombreux travaux reliés à la RIML textuelle ont été effectués depuis quelques années (Jones 2002 ; Nie 2003 ; Petrelli *et al.* 2004 ; Kishida 2007 ; Ménard 2008). Ces travaux mettent en relief plusieurs problèmes. Ainsi, certains chercheurs considèrent que peu importe la ressource linguistique utilisée, le principal problème associé à la RIML demeure l'ambiguïté sémantique et syntaxique découlant de la traduction (Gey, Kando et Peters 2005). Pendant longtemps, la RIML s'est surtout consacrée aux documents textuels. Cependant, on remarque depuis peu un intérêt grandissant envers les documents multilingues autres que textuels, l'image numérique par exemple. En outre, certains travaux en RIML étudient les systèmes de questions réponses (QR) qui recherchent, dans un ensemble de documents, la réponse à une question posée en langage naturel (par exemple, quel est le premier homme à avoir marché sur la lune ?). Le mécanisme d'extraction des réponses consiste, dans un premier temps, à obtenir un court passage de texte. Puis, dans un second temps, la réponse est extraite de ce passage. Pour le moment, les systèmes de QR sont surtout monolingues. Toutefois, à mesure que les moteurs de traduction s'amélioreront, les systèmes de QR multilingues offriront des performances de plus en plus intéressantes. Ce type de mécanismes pourrait bien révolutionner un jour le travail de référence des bibliothécaires.

D'autres initiatives ont également vu le jour. Parmi celles-ci, mentionnons « *The European Library* » (La Bibliothèque Européenne) qui offre dans 35 langues un accès aux collections des 48 bibliothèques nationales de l'Europe. Les ressources des bibliothèques nationales partenaires sont à la fois numériques (livres, affiches,

Les bibliothèques numériques désirant offrir des ressources en ligne en plusieurs langues ne se sont toujours pas dotées d'une ontologie ou encore d'un schéma de métadonnées, deux ressources primordiales qui faciliteront l'accès à l'information multilingue.



enregistrements sonores, vidéos) et bibliographiques (European Library 2011). Le projet Europeana, quant à lui, propose six millions d'objets numériques, dont des images (tableaux, estampes, cartes, photographies et dessins), des textes (livres, journaux, lettres, archives), des fichiers sonores (musique et collections sonores de phonogrammes, bandes, disques et émissions de radio) et des vidéos (films, actualités et émissions de télévision) provenant de sources diverses (musées, galeries, centres d'archives, bibliothèques, collections audiovisuelles) (Europeana 2011). En outre, des tentatives pour créer des outils pour l'accès sujet multilingue ont également été entreprises. Le projet MACS est l'un des plus prometteurs. Son but est d'intégrer les systèmes d'indexation les plus développés et les plus utilisés : *Library of Congress Subject Headings* (LCSH), Répertoire d'autorités-matière encyclopédique et alphabétique unifié (RAMEAU) et Schlagwortnormdatei (SWD) (Landry 2006). La possibilité d'interconnecter ces langages matières a été étudiée aux fins de fournir l'accès multilingue aux collections des bibliothèques nationales. Par ailleurs, l'avènement du Web sémantique pourrait bien signifier une avancée sans précédent vers l'accès à l'information multilingue, avec entre autres, le langage de représentation SKOS (*Simple Knowledge Organisation System*). Construit sur la base du langage RDF (*Resource Description Framework*), SKOS a pour principal objectif de permettre l'utilisation simplifiée de vocabulaires structurés. Les ressources décrites sont des concepts comme des classes, des propriétés, ou des concepts utilisés pour l'indexation. SKOS pourrait éventuellement permettre, par l'entremise de l'encodage RDF, la mise en correspondance de thésaurus multilingues (W3C 2011) et, par voie de conséquence, régler en tout ou en partie les problèmes d'incompatibilité entre vocabulaires d'indexation.

Il faut toutefois constater que les projets et initiatives décrits précédemment demeurent tout de même isolés et l'accès à l'information multilingue reste limité à cause des coûts technologiques importants et du manque de formation des catalogueurs en matière de langues. La plupart des utilisateurs n'ont encore accès à la recherche d'information que dans leur propre langue. En outre, les lois régissant l'utilisation de l'information dans différentes langues sont encore loin d'être complètes, particulièrement en ce qui concerne le droit

d'auteur. Langues et sociétés se sentent encore menacées par certains efforts numériques, tels que le projet Google Book Search. Les bibliothèques numériques désirant offrir des ressources en ligne en plusieurs langues ne se sont toujours pas dotées d'une ontologie ou encore d'un schéma de métadonnées, deux ressources primordiales qui faciliteront l'accès à l'information multilingue.

De nouveaux outils, tels que les systèmes d'accès à l'information multilingue, les systèmes de traduction automatique, les thésaurus multilingues, etc., ont fait l'objet de nombreuses recherches. Cependant, les développements technologiques en découlant n'ont toujours pas complètement résolu tous les problèmes liés à la diversité linguistique, sans parler des questions liées aux communications et à l'exclusion sociale. Par exemple, aucun véritable mécanisme d'accès à l'information multilingue (par ex., un système de traduction automatique permettant de traduire les requêtes ou les documents résultant de la recherche, un thésaurus multilingue facilitant l'accès sujet, etc.) n'est disponible dans la plupart des catalogues de bibliothèques.

En outre, il faut souligner que même si certaines fonctionnalités permettant la recherche multilingue sont offertes aux utilisateurs, ceux-ci n'ont pas nécessairement le réflexe de les utiliser. Dans une étude récente, nous avons demandé à 30 participants de rechercher des images à l'aide de moteurs de recherche offrant des fonctionnalités dites « translinguistiques » pour le repérage d'images représentant des objets muséaux (Ménard 2010). Les résultats de notre recherche ont révélé que la majorité des participants ayant effectué les tâches de repérage d'images n'ont pas vraiment utilisé ces fonctionnalités de recherche multilingue, par exemple la possibilité de rechercher par « langue » ou par « pays ». En outre, la possibilité de sélectionner une interface de recherche dans différentes langues n'a que rarement été mentionnée par les répondants. Curieusement, les participants à l'étude ayant accompli leurs tâches de repérage d'images et fait usage du filtre par « langue » ont tout de même constaté que celui-ci contribue au succès du repérage. Ces résultats, même s'ils semblent contradictoires, laissent présager que le besoin d'explorer et de repérer l'information, peu importe la langue dans laquelle celle-ci existe, est bien réel et que si un véritable système de repérage d'information donnant la possibilité de repérer rapidement l'information multilingue voit le jour, les utilisateurs seront sans doute disposés à en faire usage : *If we built it they will come !*

Conclusion

Nous vivons une période fascinante de l'histoire des services offerts par les bibliothèques. Pour la première fois, il est possible pour celles-ci d'offrir un accès vaste, diversifié et universel à des collections numériques à l'échelle mondiale. S'appuyant sur des ressources hété-

rogènes, la bibliothèque traditionnelle ou numérique destine son contenu à des populations diverses d'utilisateurs et se doit de pourvoir à des demandes complexes sur les plans multimédia, multiculturel et multilingue. La valeur de l'information scientifique, culturelle et pédagogique des nombreuses collections numériques des bibliothèques de différentes communautés linguistiques est un fait bien établi. Toutefois, les enjeux engendrés par la diversité linguistique du Web sont toujours énormes. Tous les secteurs de l'économie mondiale, de même que les populations de diverses cultures doivent maintenant composer avec la communication multilingue.

L'information numérique existe en plus d'une langue et des populations d'utilisateurs d'horizons culturels diversifiés et linguistiquement distincts ont accès à cette information disponible au bout de leurs doigts. L'étude des technologies et des ressources facilitant les échanges multilingues est donc devenue nécessaire afin d'aider les utilisateurs des bibliothèques à naviguer, chercher et utiliser ses ressources multilingues textuelles ou multimédias. Toutefois, les outils d'indexation et d'interrogation multilingue ne peuvent être créés que sous condition d'un travail de coopération entre chercheurs et bibliothèques. Beaucoup de travail reste donc à faire pour briser l'isolement des individus et leur donner le plein accès à l'information. ☉

Sources consultées

- Bibliothèque et Archives nationales du Québec. 2011. *Catalogue IRIS*. <<http://iris.banq.qc.ca/iris.aspx>> (consulté le 10 janvier 2011).
- Braschler, Martin. 2004. Combination approaches for multilingual text retrieval. *Information Retrieval* 7(1/2) : 183-204.
- Briet, Suzanne. 1951. *Qu'est-ce que la documentation ?* Paris : Éditions documentaires industrielles et techniques.
- Centre francophone d'informatisation des organisations. 2010. *Internet : source d'information et modes de communication*. <http://www.cefrico.qc.ca/fileadmin/documents/Publication/NETendances-Vol1_4_liens-basse_.pdf> (consulté le 12 mars 2011).
- Chen, Jiangping et Yu Bao. 2009. Cross-language search : The case of Google Language Tools. *First Monday* 14(3). <<http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/2335/>> (consulté le 10 janvier 2011).
- Chu, Heting. 2003. *Information representation and retrieval in the digital age*. Medford, N.J. : Information Today.
- Crystal, David. 2006. *Language and the Internet*. 2nd ed. Cambridge : Cambridge University Press.
- European Library. 2011. *Frequently Asked Questions*. <<http://search.theeuropeanlibrary.org/portal/en/index.html>> (consulté le 10 janvier 2011).
- Europeana. 2011. *Page d'accueil*. <<http://www.europeana.eu/portal/>> (consulté le 10 janvier 2011).
- Fluhr, Christian. 2010. *Mondialisation et information multilingue*. <<http://docnum.info/je/2010/2juillet/C.Fluhr.pdf>> (consulté le 12 mars 2011).
- Gey, Frederic C., Noriko Kando et Carol Peters. 2005. Cross-language information retrieval : the way ahead. *Information Processing & Management* 41(3) : 415-431.
- Google. 2011. *Outils linguistiques*. <http://www.google.ca/language_tools?hl=fr> (consulté le 10 janvier 2011).
- Haynes, David. 2004. *Metadata for information management and retrieval*. London : Facet Publishing.
- Hillier, Mathew. 2003. The role of cultural context in multilingual website usability. *Electronic Commerce Research Applications* 2(1) : 2-14.
- Internet World Stats. 2010. Internet users by language. <<http://www.internetworldstats.com/stats7.htm>> (consulté le 10 janvier 2011).
- Jones, Gareth J. F. 2002. *Cross-language information retrieval : consolidating and moving forwards*. <<http://ucdata.berkeley.edu/~7101/sigir-2002/sigir2002CLIR-15-jones.pdf>> (consulté le 10 janvier 2011).
- Kishida, Kazuaki. 2005. Technical issues of cross-language information retrieval : a review. *Information Processing & Management*, 41(3) : 433-455.
- Landry, Patrice. 2006. Multilinguisme et langages documentaires : le projet MACS en contexte européen. *Documentation et bibliothèques* 52(2) : 121-129.
- Ménard, Elaine. 2008. *Étude sur l'influence du vocabulaire utilisé pour l'indexation des images en contexte de repérage multilingue*. Thèse de doctorat, Montréal : Université de Montréal. <<https://papyrus.bib.umontreal.ca/jspui/bitstream/1866/2611/1/menard-e-these-indexation-reperage-images.pdf>> (consulté le 10 janvier 2011).
- Ménard, Elaine. 2010. Étude comparative de moteurs de recherche pour le repérage d'images illustrant des objets muséaux. In *Document numériques entre permanence et mutations, Actes du 13^e Colloque International sur le Document Électronique*, sous la direction de Madjid Ihadjadene, Manuel Zacklad et Khaldoun Zreik. Paris : Europa production. 159-172.
- Nie, Jian-Yun. 2003. Cross-language information retrieval. *IEEE Computational Intelligence Bulletin* 2(1) : 19-24.
- Office québécois de la langue française. 2011. *Le grand dictionnaire terminologique*. <<http://www.granddictionnaire.com>> (consulté le 10 janvier 2011).
- Oard, Douglas W. 1997. Serving users in many languages : Cross-language information retrieval for digital libraries. *D-Lib Magazine* 3(12). <<http://www.dlib.org/dlib/december97/oard/12oard.html>> (consulté le 10 janvier 2011).
- Oard, Douglas W. et Anne R. Diekema. 1998. Cross-language information retrieval. In *Annual Review of Information Science and Technology* 33 : 223-256. Medford, N.J. : Information Today.
- Petrelli, Daniela, Micheline Beaulieu et Mark Sanderson. 2002. User-participation in CLIR research. In *Proceedings of the Hawaii International Conference on System Science - HICSS-33*. <<http://ucdata.berkeley.edu/~7101/sigir-2002/sigir2002CLIR-12-petrelli.pdf>> (consulté le 10 janvier 2011).
- Petrelli, Daniela et al. 2004. Observing users, designing clarity : a case study on the user-centered design of a cross-language information retrieval system. *Journal of the American Society for Information Science and Technology* 55(10) : 923-934.
- UNESCO. 2003. *Diversité culturelle et linguistique dans la société de l'information*. <<http://tinyurl.com/y837ne>> (consulté le 10 janvier 2011).
- W3C. 2011. *SKOS Simple Knowledge Organization System - Home Page*. <<http://www.w3.org/2004/02/skos/>> (consulté le 12 mars 2011).
- Zhang, Jin et Suyu Lin. 2007. Multiple language supports in search engines. *Online Information Review* 31(4) : 516-532.