

Rendre lisible l'illisible

Esquisse d'un état de l'art en matière de translittération, transcription, romanisation, et autres conversions d'écritures

Depuis le début de la décennie, et à la faveur de la généralisation d'Unicode, on a de plus en plus fréquemment l'occasion de rencontrer des données bibliographiques en caractères hébreux, cyrilliques, arabes, chinois, etc. dans les catalogues des bibliothèques. Si la chose finit par être banale dans les bibliothèques universitaires, ou à la Bibliothèque nationale de France, certaines bibliothèques municipales ont fait aussi le choix d'élargir leurs fonds à des littératures extra-européennes.

**Philippe
Chevrant-Breton**

Bibliothèque nationale de France
philippe.chevrant@bnf.fr

Selon les établissements, les moyens techniques et humains disponibles et le public visé, se pose donc la question du traitement documentaire le plus adapté et, le cas échéant, de l'éventuelle translittération et/ou transcription à mettre en œuvre.

C'est une question ancienne – elle a d'ailleurs été traitée dans ces colonnes par Maxime Rodinson en 1964¹ et, en un sens, la voie à suivre dans les bibliothèques françaises est largement balisée à l'heure actuelle²; elle se pose cependant dans des termes renouvelés ces dernières années³.

1. Maxime Rodinson, « Les principes de la translittération, la translittération de l'arabe et la nouvelle norme de l'ISO », *Bulletin des bibliothèques de France*, 1964, p. 1-24.

2. La liste des normes et usages en vue de la restitution en caractères latins se trouve sur le site de la BnF: www.bnf.fr/pages/infopro/produits/pb-aidprod_normes.htm ainsi que sur celui de la Sous-direction des bibliothèques et de l'information scientifique du ministère de l'Enseignement supérieur: www.sup.adc.education.fr/bib/Info/Norm/2.4-Translitteration.htm

3. Voir en particulier les deux références suivantes: *Catalogage des documents en*

La simplicité au prix d'un compromis

En préambule, notons que la simplification qu'on recherche par l'usage des caractères latins n'est jamais totale. Le degré de simplicité atteint par tout système de romanisation résulte d'un compromis avec la double exigence de rigueur et de réversibilité.

Pratiquer une romanisation rigoureuse suppose le plus souvent d'avoir recours à des combinaisons typographiques complexes. Privilégier des graphies latines courantes, comme « ch » en français, plutôt que « š » est par exemple un choix dicté par la simplicité plutôt que la sophistication, mais ceci limite la diffusion potentielle de ce système au sein des

caractères non latins: rapport du groupe de travail, rédigé par Danièle Duclos-Faure, janvier 2002 et Vincent Hachard, « Enjeux et défis d'un système de catalogage multilingue et multiécriture », *Arabesques*, n° 35, juillet-août-septembre 2004, p. 12-16.

Archiviste paléographe de formation, **Philippe Chevrant-Breton** est chargé de collections en arabe au département Littérature et Art de la BnF après avoir été chargé de mission pour le livre à l'ambassade de France au Caire. Il est l'auteur d'un précédent article dans le BBF (n° 6, 1998), « Mutations récentes des bibliothèques en Égypte ».

utilisateurs des caractères latins : des italophones ou germanophones seront certainement déroutés par l'emploi du « ch » et auront recours à un autre système (par exemple « sch » pour un Allemand, « sc(i) » pour un Italien). Inversement, le choix de « š » limite la diffusion du système de romanisation parmi les utilisateurs francophones, parce que ce choix typographique sera jugé rebutant par un certain nombre de lecteurs.

En 1868, l'orientaliste Mac Guckin de Slane observe, dans la préface du *Vocabulaire destiné à fixer la transcription en français des noms de personnes et de lieux usités chez les indigènes d'Algérie*, que le système de transcription savant utilisé par les auteurs de l'*Exploration scientifique de l'Algérie* (publiée entre 1840 et 1845) est trop complexe pour être généralisé : « La complication qui résulte des points ou signes distinctifs ajoutés à tant de lettres lui donne une sorte de caractère scientifique qui l'empêchera toujours d'être accepté par l'usage ordinaire⁴. »

Terminologie

Les pratiques actuelles des bibliothèques françaises en matière de romanisation tendent à suivre l'usage savant. Elles sont notamment déterminées par les caractéristiques de l'écriture convertie.

Vincent Hachard propose un classement des écritures en quatre types : écriture alphabétique, écriture

consonantique, écriture syllabique et écriture idéo(phono)graphique⁵. Le comité technique 46 de l'ISO, « Information et documentation », coordonne en son groupe de travail 3, « Conversion des langues écrites » (TC46/WG3), les efforts de normalisation en la matière émanant notamment du monde des bibliothèques. Les mêmes précisions terminologiques se retrouvent dans l'introduction de la plupart des normes de conversion, en substance les suivantes :

« La *translittération* est l'opération qui consiste à représenter les caractères d'une écriture alphabétique ou syllabique par les caractères d'un alphabet de conversion. En principe, cette conversion doit se faire caractère par caractère. [...] »

La *retranslittération* est l'opération qui consiste à convertir les caractères d'un alphabet de conversion dans le système d'écriture converti. C'est l'opération exactement inverse de la translittération [...] de façon à reconstituer le texte translittéré sous sa forme originale.

La *transcription* est l'opération visant à noter la prononciation d'une langue donnée au moyen du système de signes d'une langue de conversion. Un système de transcription repose nécessairement sur les conventions orthographiques de la langue de conversion. La transcription n'est pas strictement réversible. [...] La transcription est la seule méthode utilisable pour les systèmes non entièrement alphabétiques ou syllabiques, et pour toutes les écritures idéophographiques, comme le chinois. »

Enfin, dans ce contexte ISO, on entend par *romanisation* la conversion d'écritures non latines dans l'alphabet latin, soit en transcription, soit en translittération, soit selon un système mixte.

Réversibilité

Si les définitions des concepts restent pertinentes, les assertions sur la réversibilité dans les normes ISO, qui remontent à des conclusions des années 60, mériteraient d'être très largement nuancées au vu des développements technologiques actuels. Il s'agissait en effet de spéculations sur la réversibilité *théorique*, alors que les outils à notre disposition nous permettent à présent d'évaluer la réversibilité *pratique*, ce qui conduit à des conclusions largement divergentes.

Pour caractériser ce contexte, très schématiquement, il s'agit d'un cadre dominé par l'ISBD et les Principes de Paris (1961). Selon l'ISBD, titre et adresse bibliographique sont décrits dans la langue et l'écriture du document. S'agissant des vedettes, les Principes de Paris stipulent : « *Quand des éditions ont paru dans plusieurs langues, la préférence doit en général être donnée à la vedette fondée sur les éditions dans la langue originale; mais si cette langue n'est pas employée couramment dans le catalogue, la vedette pourra être empruntée à des éditions et références dans une des langues couramment utilisée dans le catalogue*⁶. »

4. Mac G. de Slane, *Vocabulaire destiné à fixer la transcription en français des noms de personnes et de lieux usités chez les indigènes d'Algérie*, Paris, Imprimerie impériale, 1868, p. III-IV.

5. Pour plus de détails, voir « Enjeux et défis d'un système de catalogage multilingue et multiécriture », *op. cit.*, p. 15.

6. Voir *Bulletin des bibliothèques de France*, n° 12, décembre 1961, p. 567. La formulation actuelle de ces *Principes de Paris*, en cours de révision est la suivante : « *Quand un nom a été exprimé dans plusieurs langues, la préférence est donnée à une vedette établie à partir des informations trouvées sur les manifestations de l'expression dans la langue et l'écriture originales; mais si celles-ci ne font pas partie de celles qui sont habituellement utilisées dans le catalogue, la vedette peut être établie à partir des formes trouvées sur les manifestations ou dans les ouvrages de référence dans l'une des langues et écritures qui conviennent le mieux aux utilisateurs du catalogue. Il faut permettre à chaque fois que c'est possible un accès dans la langue et l'écriture d'origine, soit par la vedette autorisée soit par le biais d'un renvoi. Si des translittérations sont souhaitables, elles doivent suivre une norme internationale de conversion d'écriture.* »

L'œuvre de Volney

C'est au cours du XVIII^e siècle, à la faveur de l'intérêt de l'Europe des Lumières pour les philosophies et littératures orientales, que se pose pour la première fois dans des termes modernes la question de l'emploi systématique de l'alphabet latin, y compris pour des langues auxquelles cette écriture n'est pas naturelle.

Si les informations réunies par Barthélemy d'Herbelot dans la *Bibliothèque orientale* (1697), premier vaste dictionnaire bibliographique en français consacré aux civilisations orientales, sont d'un niveau remarquable pour leur époque, la restitution en caractères latins des noms de personnes ou de lieux est encore dépourvue de systématique et impropre aux recoupements.

Ce n'est qu'un siècle plus tard que Volney, dans son ouvrage *Simplification des langues orientales* (1795), pense la conversion vers l'alphabet latin de façon méthodique : un recensement des sons aussi exhaustif que possible lui sert à composer ce qu'il nomme « l'alphabet européen », c'est-à-dire un alphabet latin augmenté par le recours à des combinaisons diacritiques, lointain prédécesseur du latin étendu des catalogues informatisés de nos bibliothèques (ISO 5426). Cet « alphabet européen » est ensuite utilisé par Volney pour proposer des tables de conversion de diverses écritures orientales.

Même si le système que propose Volney est initialement conçu pour la cartographie qui accompagne la publication de la *Description de l'Égypte*, en réalité, ce système de transcription n'est pas seulement dans l'esprit de son auteur, idéologue de l'expédition d'Égypte de 1799, une commodité typographique à usage restreint, mais une technique dont la généralisation doit être l'instrument d'une « révolution morale ». Il souhaite en effet que « *les meilleurs livres ou fragments de livres asiatiques soient transcrits et multipliés par l'impression; et, ce qui est bien plus impor-*

tant, il faut que nos meilleurs livres d'Europe, traduits par d'habiles interprètes, soient également transcrits et imprimés dans cette forme [...] : le jour où les hommes d'Europe traduiront facilement leurs idées dans les langues d'Asie, ils acquerront partout en cette contrée une supériorité décidée sur les indigènes en tout genre d'affaires ».

Révolutionner les esprits de l'Orient en bouleversant la manière d'écrire, et en proposant une sorte d'équivalent du système métrique appliqué à l'écriture des langues, tel est donc l'ambitieux projet de Volney, qui finit par trouver un écho plus d'un siècle plus tard, lorsque Mustafa Kémal impose, en 1928, l'usage de l'alphabet latin à toutes les imprimeries de Turquie.

Nécessité de la standardisation

Plus modestement, dans les ouvrages imprimés du XIX^e siècle, le recours à l'alphabet latin en vue d'écrire les langues orientales s'improvise en réponse à diverses nécessités : nécessité didactique - faciliter l'apprentissage de la langue en éludant l'étape de la familiarisation avec un nouvel alphabet, mais aussi nécessité économique : il est beaucoup moins cher, en particulier en Europe, d'imprimer un ouvrage en caractères latins qu'en caractères arabes ; ainsi les caractères non latins sont à l'époque un obstacle au tirage et à une diffusion de masse. Inversement, l'usage de l'alphabet latin permet de vulgariser la connaissance des langues orientales, ce qui est indispensable dans le contexte de l'expansion coloniale européenne.

Or l'hétérogénéité des modes de translittération ne fait pas sentir ses inconvénients dans les seuls catalogues des bibliothèques. Où s'exprime le besoin de standardisation en la matière ? Se trouve assez naturellement concernée toute administration utilisant une écriture différente de

celle couramment pratiquée dans la société avec laquelle elle est en contact. Dans une lettre au gouverneur général d'Algérie, Napoléon III observait : « *Les noms arabes sont d'une transcription difficile en caractères français; en les écrivant comme on croit les entendre prononcer; tout le monde ne les orthographe pas de la même manière. [...] Ainsi le même nom figure tantôt d'une façon, tantôt d'une autre, sur les registres de l'état civil, sur les matricules de l'impôt ou sur les sommiers du domaine, ou dans les dossiers du tribunal*⁷. » Le choix de la francisation de l'Algérie, après 1870, et de façon plus nette encore, l'émancipation de cette nation en 1962 ont rendu caduc le souci de Napoléon III.

Mais notre environnement présent comporte lui aussi des exemples similaires. Ainsi, depuis l'adhésion de la Grèce et celle plus récente de la Bulgarie, et du fait de ses échanges avec le reste du monde, l'Union européenne n'échappe pas à cet ordre de préoccupations. En septembre 2004, dans un rapport au Parlement européen sur le modèle type de visa dans le cadre européen, le député Carlos Coelho relève : « *Le contrôle aux frontières peut prendre la forme d'une vérification sur le système d'information Schengen (SIS) au moment de la demande de visa [...] Cependant, il y a de nombreux noms fréquents, ainsi que des orthographes différentes pour le même nom selon les règles de translittération utilisées. En conséquence, non seulement il est de plus en plus difficile d'identifier une personne correctement, mais il est de plus en plus fréquent, également, de confondre une personne avec une autre, ce qui pose parfois de graves problèmes pour les intéressés*⁸. »

7. Mac G. de Slane, *op. cit.*, p. 1.

8. Rapport sur la proposition de règlement du Conseil modifiant le règlement (CE) n° 1683/95 établissant un modèle type de visa. - Rapporteur: Carlos Coelho, 24 octobre 2004.

www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONS/GML+REPORT+A6-2004-0029+0+DOC+PDF+V0//FR

En 2001, dans un rapport à Javier Solana sur l'application de la Convention de Dublin, le secrétaire général de la Commission européenne note que, dans leurs échanges bilatéraux d'informations sur les demandeurs d'asile, les pays européens recourent aux empreintes digitales, notamment pour dépasser les difficultés d'identification liées à la diversité des translittérations utilisées⁹.

Le projet Strabon

Deux expériences européennes d'échelle très diverse en matière de translittération peuvent être citées à titre d'exemple. La première est le projet Strabon : en février 1999, la Commission européenne a approuvé une initiative régionale globale dans le cadre de MEDA. Celle-ci avait pour but le développement d'une société euro-méditerranéenne de l'information (EUro-MEDiterranean Information Society, Eumedis¹⁰), conçue spécifiquement pour combler le fossé informationnel et technique qui sépare la région des pays voisins.

Dans ce cadre, le projet Strabon, doté de 3 millions d'euros et piloté par la Maison des sciences de l'homme de Paris, se proposait d'offrir un ensemble cohérent de systèmes d'informations en ligne, multimédias, multilingues, utilisables en arabe et accessibles à tous, sur le patrimoine culturel et l'activité tou-

ristique de chaque pays et d'offrir, à partir d'un « portail » commun, des informations publiques, des services spécifiques et des outils de travail coopératif innovants destinés aux professionnels de la culture et du tourisme. Le descriptif du projet précise : « *La cohérence et la pérennité de la plate-forme informatique reposent sur l'utilisation généralisée de normes documentaires multilingues, sur la structuration des documents grâce à l'emploi extensif du langage XML, sur l'utilisation d'une architecture ouverte et, dans la mesure du possible, de logiciels dont les sources sont libres*¹¹. »

Les techniques de romanisation de l'arabe ont intéressé les opérateurs du projet, car les noms personnels comme les noms d'auteurs sont évidemment des métadonnées précieuses en vue de regrouper les ressources. Mais en définitive, la programmation nécessaire pour parvenir à une translittération automatique de l'arabe vers le latin s'est révélée hors de portée des développeurs mobilisés autour de Strabon.

Le manuel Sirene

Autre cas, où cette fois les moyens investis sont d'un tout autre ordre : le manuel Sirene, qui régit les échanges de données informatiques entre les services de police nationaux et le Système d'information Schengen.

Le manuel Sirene entre dans le détail des procédures et spécifie notamment des règles de translittération, vraisemblablement utilisées pour l'identification de détenteurs de passeports ou documents d'identité en caractères non latins : « *Les règles de translittération, qui figurent à l'annexe 2, doivent être respectées*¹² »

11. www.eumedis.net/en/project/11

12. 2006/758/CE : Décision de la Commission du 22 septembre 2006 portant modification du manuel Sirene, paragraphe 1.4.6. dans *Journal officiel de l'Union européenne*, L317-10, 16 novembre 2006
http://eur-lex.europa.eu/LexUriServ/site/fr/oj/2006/l_317/l_31720061116fr00010040.pdf

précise la récente décision de la Commission, à l'occasion de la publication partielle du manuel Sirene.

Mais ces règles de translittération sont considérées comme confidentielles et maintenues secrètes, et il n'a donc pas été possible de les consulter dans le cadre de cet article. On devine que ce secret a trait au complexe dossier des données biométriques, de l'admission des étrangers dans l'espace Schengen, et des procédures de surveillance du transport aérien des personnes, etc., mais il est difficile d'en admettre la pertinence.

La même décision de la Commission prévoit, en son paragraphe 2.6, un droit d'accès et de rectification des personnes aux données la concernant : comment une personne peut-elle vérifier si la translittération de son nom est correcte si le système de translittération utilisé est confidentiel ? Accessoirement, par ailleurs, le secret qui entoure ce système de translittération interdit tout partage technologique susceptible de retomber dans nos catalogues de bibliothèques, même si là n'est pas l'essentiel.

Le serbo-croate

Voyons à présent deux autres exemples, où des états ont mené en la matière des politiques volontaires. L'État yougoslave, en premier lieu, a autant que possible encouragé après la Seconde Guerre mondiale l'émergence d'une langue commune aux Serbes, Croates et Bosniaques : le serbo-croate.

Historiquement, le latin prévalait en Croatie, le cyrillique en Serbie et, après 1945, les pouvoirs publics se sont efforcés de répandre l'usage des deux alphabets sur le territoire de la république de Yougoslavie, sur une base paritaire, et à l'aide de tables d'équivalence qui ont connu une grande diffusion.

Ainsi, le quotidien bosniaque *Oslobodenje* paraissait alternativement chaque semaine dans l'une ou

9. « *Bilateral exchanges of fingerprints are used successfully to overcome the difficulties that may result from changes in the spelling of names (a problem frequently associated with the transliteration of patronymics written in another alphabet)* » dans *Commission staff working paper: Evaluation of the Dublin convention: cover note from Bernhard Zepter, Deputy Secretary General of the European Commission to Mr Javier Solana, Secretary General.*
<http://register.consilium.europa.eu/pdf/en/01/st09/09269en1.pdf>

10. Programme doté de 63 millions d'euros. Le descriptif du programme peut être consulté en ligne : www.eumedis.net; ce programme, achevé, a donné lieu à la mise en ligne d'un « portail de capitalisation des connaissances » (*sic!*) consultable en ligne à l'adresse : <http://eumedis-meta.net/index.php/index.php/collections/default/index>

l'autre écriture entre 1945 et 1992¹³. En pareil cas, l'utilisation de ces règles de conversion s'impose au reste du monde, y compris au monde savant et aux catalogues de bibliothèque, et quand bien même il existe une norme ISO 9 de translittération du cyrillique, appliquée pour les autres langues concernées par cette écriture (russe, ukrainien, etc.).

Le pinyin

Deuxième cas notable: les conséquences de l'adoption de la norme hanyu pinyin. Ce système de transcription a été adopté dans les années 50 en République populaire de Chine, dans le cadre de l'alphabétisation de masse¹⁴. On a en effet jugé que la façon la plus simple d'enseigner la prononciation du mandarin aux écoliers chinois consistait à leur transcrire phonétiquement la valeur des idéogrammes selon le hanyu pinyin qui, dans ce cadre, comporte bien sûr la valeur tonale des syllabes.

Répandue massivement en Chine, utilisée à présent dans des logiciels comme TwinBridge® pour la recherche et la saisie des idéogrammes, la norme pinyin a progressivement supplanté les autres usages, et finit même par s'imposer à la presse: la plupart des noms de personnes provenant de Chine continentale, y compris les responsables politiques, sont à présent écrits dans les journaux européens d'une façon uniforme, ce qui est l'indice d'une diffusion très forte du standard.

Dès 1977, le groupe d'experts pour la romanisation des noms géographiques, sous les auspices de l'Organisation des Nations unies, approuve le pinyin¹⁵. La Bibliothèque

nationale à Paris l'adopte aussi pour le catalogage de ses collections imprimées chinoises à partir de 1979. La Library of Congress, de son côté, pratiquait de longue date le système de romanisation du chinois appelé Wade-Giles. En 1980, lorsqu'y débuta le catalogage informatisé des imprimés chinois, il fut envisagé d'adopter la transcription pinyin. Probablement parce que le pinyin n'était pas encore entré dans les habitudes des chercheurs américains, les bibliothécaires chargés des collections spécialisées sur l'Extrême-Orient aux États-Unis n'approuvèrent pas ce changement. Par la suite, et malgré la faveur croissante du pinyin aux États-Unis, le coût engendré par la conversion des données romanisées Wade-Giles a longtemps conduit à différer l'adoption de cette norme dans les catalogues américains, jusqu'en 1997. À cette date, un groupe s'est formé pour planifier l'opération, qui a effectivement eu lieu en 2000¹⁶.

À une époque où il est si facile d'échanger des données bibliographiques en réseau, et où la Chine elle-même développe des outils informatiques de conversion pinyin très accessibles, on pouvait sans doute estimer qu'il ne cesserait de devenir plus coûteux d'assurer localement la saisie selon le système Wade-Giles au regard du coût de l'opération de conversion en pinyin des données déjà présentes dans les catalogues américains.

Il ressort assez nettement de ces deux exemples que l'adoption officielle par le principal État concerné d'un standard de romanisation, accompagnée d'efforts en vue de sa diffusion effective, ne laisse presque aucune place à des usages alternatifs. En un sens, dans les catalogues de bibliothèque, ceci n'a que des avantages: le système de romanisation est familier aux usagers, l'interrogation

selon un protocole Z 39-30 est beaucoup moins aléatoire, etc.

Standards et divergences avec l'ISO

Tout autre est la situation de la romanisation de langues et d'écritures pour lesquelles il n'existe pas de standardisation d'application quasi universelle émanant des pays concernés.

Il existe, certes, des normes ISO de translittération pour un certain nombre de langues et d'écritures, dont la mise à jour relève du groupe de travail 3 de l'ISO au sein du TC46 Information et documentation. Relevons d'emblée qu'au sein de la Division statistique des Nations unies, le Groupe d'experts pour les noms géographiques¹⁷ comporte un groupe de travail aux visées très similaires, le UNGEGN Working Group on Romanization Systems. Ce groupe de travail dispose de son propre site web et a rendu librement accessible en ligne la plupart des standards préconisés pour la romanisation des langues en écritures non latines¹⁸.

Ces translittérations diffèrent des normes ISO de romanisation, sauf lorsqu'un standard national s'est imposé: sans surprise, le hanyu pinyin prévaut pour le chinois, le standard de conversion yougoslave pour la romanisation du serbe. En revanche, dans le cas de l'arabe, le standard UNGEGN de romanisation des toponymes arabes comporte de nombreuses divergences avec l'ISO 233:2 (1993), ce qui place sans doute les bibliothèques abritant des collections spécialisées en cartes et plans face à une délicate alternative. L'existence de standards concurrents, y compris au niveau national, pour des finalités très proches n'est évidemment pas de na-

13. Information communiquée par Nermina Loncarević, responsable à la BnF du secteur bulgare, serbe et croate.

14. La République de Chine à Taiwan a élaboré à la même époque son propre standard de romanisation.

15. www.eki.ee/wgrs/rom1_zh.htm

16. Philip Melzer, « Library of congress converting to pinyin for Chinese romanization », *Chinese Librarianship: an International Electronic Journal*, n° 3 (1997). <http://white-clouds.com/iclc/cliej/>

17. United Nations Group of Experts on Geographical Names, UNGEGN ou Genung selon son sigle francophone, <http://unstats.un.org/unsd/geoinfo/ungegn.htm>

18. www.eki.ee/wgrs/

ture à conforter le rayonnement des normes de romanisation ISO.

De fait, si on considère le titre de l'œuvre en arabe d'Adonis: أول الجسد آخر البحر, il apparaîtra dans les catalogues en ligne sous des formes romanisées variées :

Auwal al-ğasad, āħir al-baħr
en Allemagne,

Awwal al-ğasad, āħir al-baħr
aux États-Unis, en Grande-Bretagne,

Awwal al-ğasad, āħir al-baħr
en Espagne,

Awwal al-ğasad, āħir al-baħr
en France (ISO 233:2).

Ce sont des différences d'apparence minimale, et tout arabisant reconnaîtra le titre qui l'intéresse sous ces diverses formes. Mais cette situation est très handicapante dans la perspective de mise en commun des ressources bibliographiques à l'échelle européenne : dans ces conditions, pour un utilisateur qui aurait recours à un portail d'interrogation simultanée Z 39-50 afin de connaître la diffusion d'un ouvrage arabe dans les bibliothèques d'Europe, le critère titre est impraticable, le seul critère praticable est l'ISBN. « *Il faut se féliciter de voir un système cohérent et soigneusement étudié adopté à l'échelle internationale par un organisme qui a les moyens de diffusion de l'ISO. Il faut souhaiter que ce système soit de plus en plus adopté par les personnes et les institutions auxquelles se pose le problème* », écrivait Maxime Rodinson en 1964 à propos de la nouvelle norme de translittération de l'arabe ISO 233. Comme l'a montré l'exemple du titre du recueil d'Adonis cité plus haut, il faut reconnaître que le souhait de voir le système ISO de plus en plus adopté ne s'est pas réalisé.

À travers ce constat d'échec, se pose la question des moyens de diffusion de l'ISO, notamment pour les normes ayant trait à l'information. Depuis plusieurs années déjà, le comité technique 46 (TC46) de l'ISO déplore l'approche commerciale générale qui prévaut au sein de l'ISO : le mode de mise en vente des normes, à un coût

fort élevé, est un frein à leur diffusion et à leur pertinence. Actuellement, il en coûte environ 500 euros pour se procurer 11 normes de romanisation de l'ISO¹⁹.

Dans la mesure où l'ISO exerce ses droits de propriété intellectuelle de la façon la plus stricte, la BnF n'a pas la possibilité juridique d'afficher librement sur son site web les procédures de romanisation qu'elle emploie dans son catalogue, lorsque ce

L'existence
de standards concurrents,
y compris
au niveau national,
pour des finalités
très proches
n'est évidemment pas
de nature à conforter
le rayonnement
des normes
de romanisation ISO

sont des normes ISO ; tout au plus pourrait-elle réorienter vers le site de la librairie en ligne de l'ISO, ce qui ne serait pas satisfaisant ni nécessairement bien compris par le public. Ce serait pourtant une aide précieuse pour les lecteurs, puisqu'ils sont censés connaître les translittérations en vigueur pour interroger le catalogue en ligne.

Inversement, la Library of Congress publie sur son site les standards de romanisation qu'elle pratique, et cette rubrique de son site web, abondam-

ment citée dans la littérature spécialisée, s'impose comme une référence de fait en la matière. En théorie, la norme ISO 233:2 *Translittération simplifiée de l'écriture arabe en caractères latins: partie langue arabe* permet aisément, à partir de la forme romanisée, de reconstituer la forme originale : c'est ce qu'on appelle la réversibilité. Néanmoins, c'est le standard américain, moins rigoureux, qui permet en pratique d'effectuer cette opération, dans le cadre du module OCLC Arabic utilisé, sans doute pour cet avantage notamment, à la BIULO (Bibliothèque inter-universitaire des langues orientales)²⁰. La confidentialité des normes de romanisation ISO est de nature à dissuader tout développeur de bonne volonté de proposer un outil assurant la réversibilité réelle.

Outre les inconvénients pratiques qui découlent de cette faible diffusion des normes ISO, une autre raison pourrait justifier un assouplissement en la matière : face à l'accélération de la standardisation informelle, le Comité technique conjoint (JTC1) au sein duquel collaborent l'ISO et la Commission électrotechnique internationale ont rendu librement consultables en ligne un grand nombre de normes liées aux technologies de l'information²¹. Parmi ces normes librement consultables, la norme ISO 10646, jeu universel de caractères compatible avec Unicode. Étant donné qu'on pense nécessairement une norme de translittération comme une opération de correspondance d'un sous-ensemble de l'Unicode vers un autre, par exemple du sous-ensemble « grec » vers le sous-ensemble « latin », il serait logique que les normes ISO de romanisation soient, dans le même esprit, rendues librement accessibles.

20. Vincent Hachard, *op. cit.*, p. 16.

21. Sous la rubrique Freely Available Standards, on peut consulter des standards correspondant à des formats de fichiers bien connus, comme JPEG, MPEG, etc. http://isotc.iso.org/livelink/livelink/fetch/2000/2489/Ittf_Home/PubliclyAvailableStandards.htm

19. ISO 9, cyrillique; ISO 233-2, arabe simplifié; ISO 233-3, persan; ISO 259-2, hébreu simplifié; ISO 843, grec; ISO 3602, japonais; ISO 9984, géorgien; ISO 9985, arménien; ISO 11940, thaï; ISO 11941, coréen; ISO 15919, devanagari.

« La ville de l'homme d'acier »

Il peut sembler paradoxal avec la généralisation de l'Unicode et à l'heure où les écritures non latines sont de plus en plus faciles à implémenter dans les bases de données et accessoirement dans les SIGB, de se soucier de translittération en bibliothèques. Pourtant, la question conserve une certaine acuité; il s'agit peut-être même d'un enjeu industriel et commercial qui pourrait se préciser dans un avenir proche.

Les moteurs de recherche les plus en vue (Google, Altavista) se soucient, par la mise à disposition d'outils de traduction en ligne, de donner accès à des contenus multilingues, en substance, de permettre à un anglophone de rechercher des informations dans le web russe ou arabe, et vice versa. À cet effet, ces moteurs ou les consortiums auxquels ils appartiennent passent des accords avec des sociétés spécialisées dans la traduction automatique (Systran, Misbar, Sakhr). Ainsi, même si vous ne lisez pas du tout l'arabe, vous pouvez prendre une vague idée du contenu du journal égyptien *Al-Abram* avec les outils linguistiques de Google, ou du contenu du *Quotidien du peuple* = *Renmin ribao* avec l'utilitaire Babelfish d'Altavista²².

Pour être efficaces, ces utilitaires de traduction doivent impérativement comporter un outil de détection des noms propres (noms de personnes et noms de lieux), et se doivent de resti-

tuer ces noms propres russes, arabes, etc. selon une forme accessible au lecteur. Si le terme Сталинград apparaît dans un texte en ligne automatiquement traduit du russe, ce serait un bogue d'afficher « la ville de l'homme d'acier » là où tout un chacun attend « Stalingrad ».

Ainsi, pour les noms propres, les noms de lieux, et accessoirement les

Il peut sembler paradoxal avec la généralisation de l'Unicode de se soucier de translittération en bibliothèques

noms de collectivités, c'est à la transcription ou translittération selon les cas de prendre le relais de la traduction, et cela fait au moins une bonne raison pour considérer que la question de la conversion des écritures reste d'actualité. C'est sans doute pourquoi l'un des responsables de l'Unicode, Michael Everson, a pris en charge depuis 2003 la coordination du TC46/WG3 « Conversion des langues écrites au sein de l'ISO ».

Des francs suisses

Ajoutons que ces moteurs de recherche se soucient même de contribuer à créer des contenus multilingues et multiécritures: la Bayerische Staatsbibliothek, par exemple, dispose d'un fonds oriental

d'une grande réputation, et significativement, le communiqué de presse émis le 6 mars 2007 à l'occasion de son accord avec Google comporte la remarque suivante: « *Font également partie de ce projet de numérisation des textes exceptionnels et très rares appartenant aux fonds spéciaux de la bibliothèque en langues orientales d'Europe et d'Asie*²³. »

Parallèlement à l'édification d'une bibliothèque numérique francophone, à laquelle la BnF prend une part si active comme chacun sait, on voit donc se profiler des bibliothèques de contenus numériques multilingues de mieux en mieux capables d'adapter leurs outils de recherche aux diverses écritures utilisées par les langues les plus répandues sur la planète.

Actuellement, il est vrai, la traduction automatique, lorsqu'elle recourt à la romanisation, privilégie des formes courantes plutôt que des translittérations dites « rigoureuses », comme celles que s'efforcent de pratiquer les bibliothèques, mais il n'est pas exclu que la confusion qui risque fort d'en découler finisse à terme par susciter des solutions à la hauteur des enjeux. Si tel devait être le cas, on imagine que ce serait plutôt sous la forme de logiciels intégrés que sous la forme de normes ISO de romanisation théoriquement réversibles et facturées 8 francs suisses pour chaque page A4.

Mars 2007

22. *Al-Ahram*: www.ahram.org.eg
Google arabe anglais: www.google.fr/language_tools?q=bsb+m%C3%B4nchen&hl=fr&as_qdr=all
et *Le Quotidien du peuple* = *Renmin ribao*:
www.people.com.cn/ Babelfish chinois-anglais:
<http://babelfish.altavista.com/>

23. « *Einzigartige und sehr seltene Drucke der umfangreichen Sonderbestände der Bibliothek in osteuropäischen und asiatischen Sprachen sind ebenfalls Teil des Digitalisierungsprojekts* », www.bsb-muenchen.de/Einzeldarstellung.402+M5ce863f745c.0.html