
Danièle Duclos-Faure

RÉFLEXIONS ET RECOMMANDATIONS POUR LE CATALOGAGE DES FONDS EN CARACTÈRES NON LATINS EN FRANCE DANS LE CATALOGUE COLLECTIF UNIVERSITAIRE, SUDOC

Selon une enquête menée en 2001 par le Ministère de l'éducation nationale auprès de l'ensemble des bibliothèques universitaires françaises relevant de sa tutelle, leurs collections d'ouvrages en caractères non latins s'élèvent à environ 1 million de titres répartis dans 44 bibliothèques.

20 bibliothèques sur les 103 interrogées possèdent des fonds en caractères arabes, ce qui représente 73 011 titres, en langue arabe pour la majorité, mais également en persan (16 550 titres) et quelques milliers en turc ottoman ou en pachto.

Le traitement bibliographique de ces collections n'est pas sans poser problème comme en attestent les chiffres du catalogue: 30% des titres ne sont pas catalogués, le plus souvent faute de personnel compétent dans ces langues.

La majorité des notices existantes sont sous forme de catalogue sur fiches et en caractères originaux, mais 44% des notices existantes sont translittérées et informatisées, celles-ci sont, pour la plupart aujourd'hui, visibles dans le catalogue collectif de l'enseignement supérieur: SUDOC.

SUDOC est un catalogue collectif et un outil de catalogue partagé installé dans les bibliothèques universitaires françaises de décembre 2000 à mars 2002. A cette date, 113 établissements collaborent à l'enrichissement de SUDOC qui contient plus de 5 millions de notices bibliographiques et 13 millions de localisations. Il est consultable sur le web à l'adresse suivante: www.sudoc.abes.fr.

Aujourd'hui dans SUDOC aucun caractère non latin n'est présent. Les notices des ouvrages en caractères non latins doivent y être translittérées selon les normes ISO. Les données sont codées selon un jeu de caractères propriétaire du constructeur PICA et compatible ISO 646 et ISO 5426. Le format de saisie des notices bibliographiques et d'autorité est le format UNIMARC officiel IFLA.

Le projet du "Pôle Langues et civilisations du monde" vise à constituer à Paris, un pôle interdisciplinaire d'envergure internationale sur les langues, les cultures, les civilisations d'Europe centrale et orientale, d'Asie, d'Océanie, d'Afrique et des Amériques. Il s'agit de regrouper des unités de formation et de recherche aujourd'hui dispersées et de développer leurs collections.

La bibliothèque du Pôle, nommée BULAC: bibliothèque universitaire des langues et des civilisations, rassemblera les collections aujourd'hui éparses dans les neuf établissements partenaires, soit 1,6 million d'ouvrages, et

réunira dans un catalogue unique, les catalogues actuels de ces bibliothèques qu'ils soient en caractères originaux ou en caractères translittérés.

Un groupe de travail a été constitué en 2001, à l'initiative du Ministère de l'éducation nationale, pour étudier la faisabilité d'un catalogue unique pour la BULAC et émettre des recommandations pour sa réalisation. Les dix membres du groupe, tous bibliothécaires, se sont appuyés sur les réflexions de 6 groupes d'experts en écritures et en langues, également professionnels de la documentation.

Les principales conclusions peuvent se résumer ainsi: le développement d'un catalogue multi-écritures participant au réseau de catalogue universitaire est une nécessité. Le rapport est disponible en ligne à l'adresse suivante: www.sup.adc.education.fr/bib/.

En conséquence, le développement de SUDOC pour permettre la saisie, l'indexation, la recherche et l'affichage de données en caractères non latins devient une priorité.

La difficulté principale, outre la variété des systèmes d'écritures et de langues présents dans la bibliothèque, réside dans le fait que le catalogue doit être accessible à la fois à un lectorat expert et à un plus grand public: comment organiser les données bibliographiques et les données d'autorité (accès auteurs et titres) pour que, tout en respectant les normes bibliothéconomiques, elles permettent à tout lecteur d'accéder à l'information, quelle que soit la langue et l'écriture de l'ouvrage recherché et la forme utilisée par le lecteur dans sa requête, écriture non latine ou écriture romanisée.

Plusieurs conditions nécessaires à la mise en œuvre de ce catalogue ont été identifiées:

— en matière de codage de caractères: la norme UNICODE (ISO/IEC 10646), dans sa variante UTF 16, pour le stockage, l'indexation et l'affichage des différents systèmes d'écritures utilisés assurera le codage univoque de chaque caractère [www.unicode.org];

— en matière de romanisation: une norme de translittération réversible (un caractère pour un caractère), permettra le passage d'une écriture à l'autre et facilitera la saisie; pour l'écriture arabe, on sait que seul le passage de la romanisation à l'écriture originale est automatiquement possible. La norme ISO 233-2: 1993 pour la translittération de