

LE FICHER D'AUTORITÉS ANGLO-AMÉRICAIN

UN PROJET À MATURITÉ ?

Qu'est-ce que le fichier d'autorités anglo-américain ? Il s'agit en fait de l'intitulé d'un projet plus couramment désigné sous le sigle AAAF (Anglo-American Authority File). Y collaborent la British Library et la Bibliothèque du Congrès afin de mettre au point un fichier d'autorités commun à partir de l'actuel fichier d'autorités noms propres¹ utilisé aux États-Unis. La British Library en possédera à terme un double constamment mis à jour par transfert de fichiers.

Après avoir analysé les raisons qui sont à l'origine de cette décision, je décrirai l'état d'avancement du projet et examinerai les conséquences qu'il aura à terme sur la coopération internationale en matière de fichiers d'autorités.

Usage recommandé et usage toléré

Il existe, paraît-il, une chanson populaire qui déclare en substance : « *Vous dites tomaytoe nous disons tomato,*

vous dites potaytoe nous disons potato »².

Un thème qui n'est pas sans rappeler le bon mot de Churchill qui voyait dans les États-Unis et le Royaume-Uni deux pays séparés par une langue commune.

Pour minimiser l'effet de ces différences sur la recherche documentaire, les bibliothécaires ont trouvé des solutions. C'est notamment le cas de l'indexation matière. Il est possible d'établir des liens entre usage recommandé et usage « toléré » (en renvoyant par exemple *Railroad* à *Railway*), ou entre concepts proches (*Welfare*, voir aussi *Social Security*).

Les difficultés surgissent lorsque des mots identiques recouvrent des sens différents : il en est ainsi du mot *Football*, qui en Grande-Bretagne correspond au *Soccer*³, alors qu'aux États-Unis il désigne le *Gridiron Football*⁴.

Ces problèmes, particulièrement difficiles à résoudre de façon satisfaisante, expliquent en partie que, jusqu'à une époque relativement récente, les pays anglophones non américains aient rechigné à adopter les vedettes matières de la Bibliothèque du Congrès.

* Traduction d'Oristelle Bonis.

Cet article est le texte d'une communication présentée par l'auteur au séminaire international de l'IFLA, *Fichiers d'autorités – leur création et leur utilisation dans les catalogues*, 4-6 octobre 1995, Saint-Petersbourg, Russie.

Le compte rendu de ce séminaire est paru dans le *Bulletin des bibliothèques de France*, t. 41, n° 2, 1996, p. 94-95.

1. Par noms propres, on entend ici les noms de personnes, les noms géographiques et les noms de collectivités.

2. George GERSHWIN, « Let's call the whole thing off », from the musical *Shall we dance ?*, 1937.

3. L'équivalent du « foot » tel qu'on le pratique en France.

4. Ou « football américain », jeu plus proche du rugby, qui se joue à deux équipes de onze joueurs, mais avec un ballon ovale.

ALAN DANSKIN
The British Library

Un nom propre reste un nom propre

Mais, en première approche au moins, le contrôle des autorités des noms propres ne devrait pas buter sur les mêmes écueils et, étant donné que la Grande-Bretagne et les États-Unis ont un héritage culturel et linguistique commun, il serait somme toute logique que les vedettes noms propres autorisées par la British Library correspondent à celles de la Bibliothèque du Congrès. Après tout, pourrait-on penser, un nom propre reste un nom propre.

Or il n'en va pas ainsi, comme le montre l'étude⁵ consacrée par Ed A. Jones à la cohérence des entrées principales des notices bibliographiques créées par les agences bibliographiques nationales anglo-américaines en 1982 et 1989.

Pour Ed A. Jones, dans 55 % des cas environ, les entrées principales utilisées par la British Library et la Bibliothèque du Congrès ne sont pas harmonisées, du fait des variantes autorisées dans l'orthographe des noms propres de part et d'autre de l'Atlantique.

Aussi a-t-il préconisé que les deux pays s'associent pour élaborer conjointement un fichier d'autorités garantissant une meilleure cohérence entre les principales vedettes.

Politiques et usages

A quoi sont donc dues les distorsions incriminées ? Elles tiennent d'abord aux différentes politiques de catalogage. C'est ainsi que la British Library translittère systématiquement les noms propres écrits en caractères cyrilliques, alors que la Bibliothèque du Congrès préfère utiliser la forme anglaise traditionnelle chaque fois qu'elle existe. Typique est à cet égard l'exemple *Chaikovskii/Tchaïkovsky*.

5. Ed A. JONES, *Consistency in Choice and Form of Main Entry, 1982 and 1989: a Comparison of Library of Congress Monograph Cataloguing with that of the British Library and of the National Libraries of Australia and Canada* (Ph. D. dissertation, University of Illinois at Urbana-Champaign, 1994).

D'autres différences s'expliquent par l'usage : s'agissant de la Chambre basse du Congrès américain, la bibliothèque a adopté le terme *House* au lieu de la forme complète *House of Representatives* prônée par la British Library. Et si, en Grande-Bretagne, l'adjectif *Saint* souvent accolé aux noms de lieu est couramment abrégé en *St* (*St Andrews*, par exemple), aux États-Unis, on l'écrit en entier (comme dans *Saint Louis*).

Enfin, d'autres incohérences ont pour origine les différentes sources utilisées pour établir la forme d'un nom propre. Les formes figurant sur les pages de titre des œuvres d'un même auteur peuvent être différentes. Le contexte dans lequel un nom propre est introduit peut également en déterminer la forme. Dans un fichier d'autorités établi sur une base aussi large que celui de la Bibliothèque du Congrès, il est parfois nécessaire de donner des précisions sur un nom propre afin d'éviter toute ambiguïté. Dans le fichier moins important de la British Library, l'opération est superflue puisque ce même nom propre ne se rencontre qu'une fois.

La coopération

Pourquoi cela est-il si important ? Créer un fichier est une opération fort coûteuse. Une des méthodes les plus efficaces pour réduire le coût du catalogage tout en maintenant la qualité de l'accès aux collections consiste donc à travailler en coopération.

Dans un monde idéal, chaque document ne devrait être catalogué qu'une fois. La coopération se pratique désormais à l'échelle mondiale, mais l'incohérence des vedettes autorisées est un obstacle à l'échange des données bibliographiques. Alors que la British Library puise 50 % de son catalogage dans des sources déjà constituées, essentiellement d'ailleurs dans les notices de la Bibliothèque du Congrès, l'ampleur des interventions manuelles à accomplir pour harmoniser les vedettes avec celles de son propre fichier d'autorités l'empêche de réaliser les économies à la mesure

de ce gain de temps. Il serait bien sûr souhaitable de recourir plus largement au catalogage existant, mais, à l'heure actuelle, cette solution n'est pas financièrement avantageuse.

En mai 1993, Pat Oddy (Head of Cataloguing, British Library) et Sarah Thomas (Director for Cataloguing, Bibliothèque du Congrès) se sont entendues sur le principe d'un fichier d'autorités élaboré conjointement par leurs institutions respectives. Une décision qui revient à reconnaître non seulement une nécessité économique, mais aussi une opportunité pratique.

LA COOPÉRATION SE PRATIQUE DÉSORMAIS À L'ÉCHELLE MONDIALE, MAIS L'INCOHÉRENCE DES VEDETTES AUTORISÉES EST UN OBSTACLE À L'ÉCHANGE DES DONNÉES BIBLIOGRAPHIQUES

Isaac Newton, symbole de la nouvelle British Library qui va ouvrir à St Pancras, disait que, s'il voyait plus loin que d'autres, c'était pour s'être hissé sur les épaules de géants. De même on peut affirmer que, s'il est devenu possible d'envisager la création, certes plus modeste, d'un fichier d'autorités commun, c'est grâce à l'heureuse conjonction des progrès accomplis dans divers domaines. Les bibliothécaires ont ouvert la voie en définissant des normes de catalogage communes (Anglo-American

Cataloguing Rules-2nd ed., ou AACR2), et en développant d'importantes bases de données bibliographiques dans un format d'échange compatible (le format MARC). Quant aux technologies de l'information, elles ont fourni l'infrastructure avec Internet et les protocoles qui lui sont associés.

Vu d'un peu loin, tout paraît aller pour le mieux : des règles de catalogage communes, un format commun, des normes de communication internationales... Mais, à y regarder de plus près, on s'aperçoit que, si l'AAAF n'est toujours pas devenu réalité, c'est parce qu'il reste à régler bien des problèmes de fond.

Il a été mentionné précédemment que, malgré l'AACR2, il existe une pluralité de politiques et d'interprétations en matière de catalogage : il n'y a toujours pas d'accord sur l'orthographe du mot « catalogue » : les Anglais écrivent *catalogue*, les Américains *catalog*.

Fichiers d'autorités et formats d'échange

Je voudrais maintenant parler des différences qui existent entre les fichiers d'autorités eux-mêmes et entre les formats d'échange. Très dissemblables de par leur présentation, les fichiers d'autorités de la British Library et de la Bibliothèque du Congrès sont bien les produits de leurs institutions respectives.

Monolithique, utilisé sur tout le territoire des États-Unis (et au-delà) par des bibliothèques de toutes natures, le fichier d'autorités américain concerne la quasi-totalité des collections de la Bibliothèque du Congrès. A la fois complet et varié, il comprend 3,5 millions de notices couvrant la plupart des langues et des littératures du monde. S'il fait la part belle aux autorités auteurs noms de personne, on y trouve aussi des notices d'autorités pour les titres uniformes et les collections. Et, bien qu'il soit propriété de la Bibliothèque du Congrès, il est le résultat d'une collaboration entre plusieurs institutions travaillant sous les auspices du programme NACO (Name Authority Cooperative Project).

Le fichier d'autorités noms propres de la British Library fut lancé en 1981 sous forme de liste d'autorités noms propres (Name Authority List) pour la British National Bibliography (BNB), afin de répondre aux demandes des bibliothécaires du Royaume-Uni. Ceux-ci avaient en effet besoin de conseils pour rédiger les vedettes conformément aux directives de l'AACR2.

NI LA BRITISH LIBRARY NI LA BIBLIOTHÈQUE DU CONGRÈS N'ONT JAMAIS RELIÉ ENTRE EUX LEURS FICHIERS D'AUTORITÉS

Puis, en même temps que la British Library s'orientait progressivement vers la centralisation des opérations de catalogage, le fichier d'autorités s'est peu à peu élargi aux documents publiés à l'étranger. En 1993, on y intégra, pour la première fois, des vedettes établies par des institutions extérieures.

Au moment où ces lignes sont écrites, le fichier d'autorités britannique contient près de 600 000 vedettes autorisées, mais il ne couvre pas l'ensemble des collections de la British Library.

Choix des vedettes

Une des questions qui se posent d'emblée concerne le choix préalable des vedettes qui seront en définitive retenues. Pour reprendre les exemples donnés plus haut, l'AAAF doit-il accorder la préférence à *Saint Andrews* ou à *St Andrews*, à *Chaikovskii* ou à *Tchaikovsky*, à *House* ou à *House of Representatives* ?

Il semble de prime abord évident que la taille et l'étendue de la liste d'autorités de la Bibliothèque du Congrès imposent de la choisir comme base pour l'AAAF et d'opter en règle générale pour la graphie américaine. Je reviendrai un peu plus loin sur le sort des exemples cités.

Notons d'abord que ni la British Library ni la Bibliothèque du Congrès n'ont jamais relié entre eux leurs fichiers d'autorités. Ce qui fut un obstacle à l'AAAF. De même, la conséquence inévitable du fichier autorités commun sera, pour la British Library, l'alignement sur plusieurs milliers de notices de la Bibliothèque du Congrès. Toutefois, la British Library dispose maintenant d'un système d'édition en format PC (PACE) qui a considérablement accéléré le processus de correction. La tâche n'en reste pas moins immense. On peut envisager que, chaque fois qu'il y a concordance exacte, les vedettes reprises dans le fichier de la Bibliothèque du Congrès écrasent celles du fichier de la British Library. Les recherches de Sharon Agar⁶ montrent que, s'agissant des documents courants, les vedettes noms de personne se recoupent exactement dans 73 % des cas.

Pour les noms de collectivités, la proportion diminue tout en restant relativement élevée (57 %). Ces chiffres dissimulent toutefois un certain nombre de problèmes, notamment le fait que rien ne garantit qu'un recouplement à l'identique soit un recouplement correct. Sharon Agar précise ainsi que 10 % des vedettes de l'échantillon sur lequel elle a travaillé étaient des erreurs.

Qui plus est, toute vedette de la Bibliothèque du Congrès serait rejetée si elle recoupait exactement un renvoi du fichier d'autorités de la British Library. Cette restriction toucherait notamment les noms propres chinois, car la Bibliothèque du Congrès utilise de préférence le système de transcription Wade-Giles,

6. Sharon AGAR, *Survey of British Library and Library of Congress Name Authority Headings*, British Library, Unpublished Internal Report, 1994.

alors que la British Library a opté pour la transcription *pinyin*, avec des renvois pour les termes transcrits à partir du système Wade-Giles.

Procédure prudente

Il a été décidé de procéder avec plus de prudence et de préserver les données jusqu'à ce que leur fusion soit approuvée. Les vedettes de la Bibliothèque du Congrès serviront prioritairement à créer les nouvelles vedettes du catalogage courant. En parallèle, un programme de conversion rétrospective sera ciblé sur des entrées abondamment utilisées parce qu'elles concernent des individus ou des collectivités particulièrement prolifiques ou connus. Je suis chargé d'étudier les possibilités d'automatisation partielle de ce procédé.

J'ai dit plus haut qu'en règle générale, les vedettes de la Bibliothèque du Congrès prendraient le pas sur les formes correspondantes de la British Library, mais lorsque les différences d'usage sont identifiées, la Biblio-

thèque du Congrès accepte la forme britannique. On écrira donc *St Andrews* plutôt que *Saint Andrews*, alors que *Saint Louis* restera inchangé. De son côté, la British Library acceptera *House* au lieu de *House of Representatives*.

Les différences d'approche dans le traitement des données ont fait l'objet de discussions entre la British Library et la Bibliothèque du Congrès, ce qui a permis d'arriver à une interprétation commune des règles relatives à la rédaction des points d'accès pour les noms propres (la Bibliothèque nationale du Canada doit bientôt décider si elle y souscrit ou non).

Cet accord prévoit deux cas d'exception. D'abord les noms propres que la British Library translittère en caractères latins, et que la Bibliothèque du Congrès entre sous leur forme anglaise traditionnelle. Cela signifie qu'à court terme au moins, le désaccord sur les formes *Chaikovskii* et *Tchaikovsky* va subsister.

La deuxième exception concerne les noms propres qu'il convient de trans-

littérer. La British Library a entrepris le repérage de ces différences afin d'éliminer les incohérences. Les pre-

**LES VEDETTES
DE LA BIBLIOTHÈQUE
DU CONGRÈS
SERVIRONT
PRIORITAIREMENT
À CRÉER
LES NOUVELLES
VEDETTES
DU CATALOGAGE
COURANT**

miers résultats sont encourageants puisque, pour la grande majorité des langues les plus pratiquées (hormis le

chinois, bien sûr), on ne constate que des écarts minimes dans la graphie translittérée. Ces exclusions sont certes regrettables, mais il ne faut pas exagérer leur importance.

La British Library et la Bibliothèque du Congrès sont en train de préciser et de diffuser les changements de politique dont elles sont convenues. La British Library, qui les applique depuis janvier 1996, a élargi aux titres uniformes et au catalogage auteur/titre les discussions en cours avec la Bibliothèque du Congrès sur cette convergence de politiques.

Les formats

Si seulement les problèmes posés par les formats MARC pouvaient être aussi facilement résolus ! A la vérité, au cours des trente dernières années, les bibliothécaires britanniques et américains ont réussi à définir des procédures de présentation telles que l'échange des données bibliographiques a été rendu aussi simple que celui des roubles de part et d'autre du Dniestr !

Il existe en effet de profondes différences entre le format USMARC, conçu pour pouvoir être utilisé indépendamment de tout code de catalogage particulier, et le format UKMARC, très proche de l'AACR2. Dans UKMARC, la ponctuation est implicite dans la codification de certains sous-champs. Dans USMARC, la ponctuation doit être indiquée et, de manière générale, les sous-champs comprennent beaucoup moins d'éléments. USMARC couvre davantage de documents et l'utilisation des éléments codés y est plus logique.

Faute d'une procédure de conversion réversible ou de formats compatibles, la mise en œuvre efficace de l'AAAF s'avère par conséquent impossible.

La British Library, la Bibliothèque du Congrès et la Bibliothèque du Canada multiplient les contacts pour définir un format commun, mais même si ces discussions devaient aboutir, il faudra vraisemblablement attendre au moins trois ans avant qu'il devienne réalité. C'est pourquoi la British Library étudie la possibilité de convertir les données d'autorités

d'un format à l'autre. Afin de préserver l'intégralité des données entre la copie « maître » du fichier de la Bibliothèque du Congrès et le double que possède la British Library, il est essentiel que la procédure de conversion ne provoque aucune perte. On sait d'ores et déjà que la chose est irréalisable pour les titres uniformes, qui ne pourront donc pas être intégrés à l'AAAF avant que les formats soient devenus compatibles. Si les études actuellement poursuivies sur la conversion des vedettes noms de personne donnent les résultats escomptés, le chargement du fichier pourra se faire comme prévu en 1996. Si, en revanche, la tentative de définition d'une procédure de conversion

**LA BRITISH LIBRARY
ET LA BIBLIOTHÈQUE
DU CONGRÈS SONT
EN TRAIN
DE PRÉCISER
ET DE DIFFUSER
LES CHANGEMENTS
DE POLITIQUE
DONT ELLES SONT
CONVENUES**

valable échoue, la British Library devra se contenter de charger le fichier d'autorités américain pour compléter ses ressources documentaires, et force sera de repousser la réalisation de l'AAAF jusqu'à l'harmonisation définitive des formats.

Des gains potentiels

Les recherches entreprises par Ed A. Jones et Sharon Agar sur les données démographiques des fichiers d'autorités permettent de chiffrer approximativement les gains poten-

tiels qu'entraînerait l'AAAF. Comme on pouvait s'y attendre, les deux fichiers se recoupent assez largement.

Selon Sharon Agar⁷, 44 % des documents reçus par la British Library sont déjà répertoriés dans le fichier autorités des noms propres de la Bibliothèque du Congrès. Ce qui, en pratique, signifie que la duplication de ces notices d'autorités est en cours.

En outre, Ed A. Jones et Sharon Agar ont démontré que la duplication des notices ne s'accompagne pas nécessairement de la création de vedettes identiques. D'après Ed A. Jones, le défaut de correspondance de 55 % des vedettes principales s'explique par la transcription différente des noms de personnes.

Sharon Agar arrive à des conclusions du même ordre, bien qu'elle fasse état d'une plus grande homogénéité des noms de personnes. L'élimination de ce manque de concordances entre les vedettes sera le principal avantage de l'AAAF, mais ce nouveau fichier entraînera également :

- la multiplication des notices d'autorités. Chaque année, la British Library crée environ 60 000 notices d'autorités. En s'associant à l'AAAF, via le programme NACO, elle pourra pour la première fois les mettre à disposition d'un public plus large dans un format MARC commun ;
- la cohérence des points d'accès. La possibilité de consulter les catalogues à distance existe depuis un certain temps, mais, bien que cette méthode de recherche documentaire se développe rapidement, son efficacité est limitée par le manque de cohérence entre les points d'accès dans les différents catalogues interrogeables en ligne. L'harmonisation de la rédaction des vedettes - objet de la création de l'AAAF - facilitera la recherche documentaire ;
- une plus grande fiabilité des notices d'autorités. La British Library et les institutions homologues de la République d'Irlande, d'Écosse et du Pays

7. Sharon AGAR, *Survey of Hit Rate Using an Anglo-American Authority File*, British Library, Unpublished Internal Report, 1994.

de Galles sont les sources d'information les plus aptes à rédiger les notices d'autorités relatives aux individus et collectivités des îles Britanniques. Le fichier comprendra moins d'erreurs, et, partant, les notices seront moins sujettes à révision, ce qui constitue un argument de poids pour un fichier d'autorités ;

– et maints développements ultérieurs... L'AAAF n'est que la première étape d'un processus qui, à terme, permettra d'intégrer les données bibliographiques du Royaume-Uni à celles des pays d'Amérique du Nord et d'Asie australe.

La British Library a d'ores et déjà réintroduit les vedettes matières de la Bibliothèque du Congrès. Elle participe aux discussions sur l'homogénéisation des formats MARC et a adopté l'essentiel des critères de description bibliographique définis par le Program for Cooperative Cataloguing. L'AAAF est un des éléments clef de ce processus d'intégration et, plus tard, le catalogage dérivé par l'échange des notices de la British Library sera tout à fait rentable pour les responsables du catalogage travaillant hors du Royaume-Uni.

Coopération, compatibilité

Si le projet de l'AAAF est enfin mûr, il tarde aussi à se réaliser. Coopération, compatibilité, tels doivent être les maîtres mots des bibliothécaires. Nous avons beaucoup à apprendre des concepteurs des protocoles de

communication qui utilisent les normes, telles que la Z39.50, d'une compatibilité bien supérieure à celle de nos divers formats MARC.

COOPÉRATION, COMPATIBILITÉ, TELS DOIVENT ÊTRE LES MAÎTRES MOTS DES BIBLIOTHÉCAIRES

Est-il envisageable d'élargir au monde non anglophone la coopération entre institutions chargées des fichiers d'autorités ? J'en suis pour ma part convaincu, mais les difficultés venues compliquer l'harmonisation des fichiers d'autorités de deux pays culturellement aussi proches que la Grande-Bretagne et les États-Unis rendent quelque peu improbable un partage plus général des fichiers d'autorités au-delà du monde anglophone, tant que nous resterons attachés au concept d'usage autorisé.

Dans un contexte planétaire, il serait injuste que quelques-uns prescrivent à tous la forme sous laquelle un nom propre est autorisé à figurer dans l'ensemble des catalogues. Autrement dit, les données doivent être adaptables aux besoins et aux

attentes des utilisateurs. Barbara Tillet⁸ a proposé de remplacer la présentation linéaire des vedettes d'autorités par un regroupement souple qui donnerait la même importance à toutes les variantes d'un même nom. S'il est adopté, ce modèle aura des conséquences importantes sur les formats et la conception des bases de données du futur. Il pose aussi le problème des instances à qui confier la responsabilité du suivi et du développement du fichier d'autorités international qui en serait l'aboutissement. La British Library et la Bibliothèque du Congrès envisagent de confier la gestion de l'AAAF à un comité de pilotage composé de représentants des agences bibliographiques nationales concernées. Ce type d'organisation conviendrait, me semble-t-il, également à l'éventuel fichier d'autorités international. Seules, en effet, les agences bibliographiques nationales sont en mesure de maîtriser les normes de catalogage dont dépend en dernier ressort la qualité du fichier. Or, et j'espère l'avoir démontré, en l'absence de normes, autant renoncer tout de suite à l'aventure.

Octobre 1995

8. Barbara TILLET, « 21st century authority control : what is it and how do we get there ? », paper presented at the oclc pre-conference seminar *The Future is Now : Reconciling Change and Continuity in Authority Control*, ALA midsummer, Chicago, 1995.