

La bibliographie d'histoire de l'art

Indexation et thesaurus

par Maryse Bideault

Responsable scientifique de BHA, CNRS

L'exposition de quelques uns des aspects et problèmes présentés par le vocabulaire de la *Bibliographie d'histoire de l'art* (BHA), bibliographie courante produite conjointement par l'INIST et le J. Paul Getty Trust nécessite un bref rappel historique de cette base de données, afin d'en définir ses spécificités. Problèmes semblables et différents de ceux posés par l'indexation du fonds d'une bibliothèque d'art et par les documents d'une base de données factuelles.

La fusion du *Répertoire d'art et d'archéologie* (RAA) et du *Répertoire international de la littérature d'art* (RILA) a non seule-

ment bénéficié de la confrontation de deux expériences contemporaines en matière de bibliographie courante informatisée mais aussi a été à l'origine d'un travail poussé sur les problèmes d'une indexation bilingue des documents bibliographiques.

• Le **RAA**, bibliographie d'histoire de l'art, publiée en français de 1910 à 1989, informatisée en 1973, a connu des politiques diverses dans la conception de l'indexation, celle des noms propres (artistes et lieux) ayant toujours été privilégiée par rapport à celle des concepts. Les index manuels ont précédé l'informatisation du

système bibliographique et le passage de l'un à l'autre des systèmes s'est fait sans une remise en cause fondamentale du vocabulaire ni de sa structuration. Il est vrai que si, dès le début, l'informatisation s'accompagnait d'une nouvelle méthode de dissémination de l'information (interrogation de bases de données sur Questel) elle concernait principalement la gestion des données et l'édition de volumes imprimés. Le vocabulaire a évolué et s'est enrichi avec le temps, malgré l'injonction, à une certaine époque, de limiter les concepts au nombre de 1 000, ce qui entraîna une pratique abusive du

renvoi et du terme générique, politique défendable dans le cas du seul produit imprimé mais indéfendable quand on pense interrogation informatisée.

• Le **RILA**, bibliographie de langue anglaise publiée aux États-Unis de 1975 à 1989, offrait non seulement des analyses longues mais aussi des indexations sophistiquées aux multiples chaînages, faisant souvent appel à des descripteurs composés (par exemple « Baroque Italian painting » là où BHA, aujourd'hui, utilise « Painting ; Italy ; Baroque ») ; certains des systèmes d'indexation étaient empruntés aux bibliothèques américaines. Par ailleurs, la finesse de l'indexation du RILA la rapprochait de celles de données factuelles.

Dès que la fusion fut décidée, un travail de rapprochement des deux bibliographies fut entrepris : travail sur le domaine à couvrir, sur la fonction de l'analyse par rapport au document primaire, sur la relation entre le document primaire, l'analyse et l'indexation, réflexion enrichie par la connaissance que l'on avait, de chaque côté, de l'attente des utilisateurs de ces deux bases. La conclusion fut de tendre vers une indexation extrêmement fine et d'abandonner tout niveau de grande généralité. Dans l'optique de la construction d'une base bilingue, on procéda à la constitution d'un noyau né de la comparaison des vocabulaires respectifs, comparaison qui mit en évidence des spécificités propres aux deux bibliographies (davantage de termes propres à l'archéologie du Moyen Âge et des temps modernes ou à l'art des pays de l'Est pour le RAA, davantage de termes propres à l'art de l'Amérique du Nord ou bien encore à l'art contemporain pour le RILA).

Plus des deux tiers de la totalité des références que les bibliographies ont le devoir de traiter sont des articles de périodiques, des contributions dans les mélanges, dans des actes de congrès, dans des ouvrages thématiques ou dans des catalogues d'exposition. Ces documents sont potentiellement riches d'une infinité de thèmes ; ils traitent de sujets les plus variés et les plus pointus et engendrent une indexation complexe. Il n'est pas rare qu'un document requiert 20 à 25 descripteurs : nom de l'artiste, ou du personnage historique, la nationalité, le siècle ou la période d'activité, le titre des

œuvres d'art ou des édifices, leur localisation, des concepts historiques, religieux, sociologiques, etc. Ainsi un article publiant six tableaux inédits, signés de quatre artistes différents, conservés dans trois églises différentes, oblige à d'autant de termes indexés. Une lecture du thesaurus ou encore des fichiers d'autorité de BHA et une comparaison même rapide avec tout thesaurus utilisé par une bibliothèque d'art convainc de l'infinité richesse de ceux-là. Jamais aucune monographie n'aura pour seul sujet le motif de la feuille d'ortie dans l'enluminure du XV^e siècle ou l'emploi du carton-pâte dans la sculpture de Jacopo Sansovino. Cette mine inépuisable explique la taille des fichiers d'autorité de BHA : 7 476 mots-matière ; 75 047 noms de personnages ; 20 088 noms géographiques ; 29 757 noms d'édifices et d'institutions ; 10 515 titres d'œuvres d'art ; 284 titres de congrès.

La nécessité pour BHA d'être bilingue a entraîné un travail de mise en commun des fichiers du RAA et du RILA, produisant un vocabulaire d'environ 5 000 termes qui fut déchargé dans le système et servit d'outil de référence pour la constitution des références bibliographiques. Ces références étant fournies tantôt en français, tantôt en anglais selon l'origine du travail de documentation, chaque rédacteur dispose, en ligne, des dictionnaires issus des fichiers d'autorité qui, pour le moment, ne sont consultables qu'à Williamstown, là où se trouve l'équipe de vérification des autorités ; un nouveau système informatique qui utilise les technologies les plus récentes doit bientôt remédier à ce problème. Ainsi, alors que le dictionnaire ne donne que la forme établie et quelques éventuelles variantes, le fichier d'autorité ajoute la nationalité, le rôle, toutes les variantes connues du nom, les sources diverses qui ont servi à l'établissement de la forme. Pour les besoins de l'indexation du document et en fonction de la méthodologie définie, le terme consulté dans le dictionnaire est soit validé et sélectionné, soit soumis à des modifications. Le terme est nouveau, il devient alors terme-candidat.

Toutes ces modifications ou créations sont soumises à des vérificateurs qui travaillent sur les fichiers d'autorité à Williamstown. Un travail minutieux et rigoureux de vérification est fait à travers le plus grand nombre de sources souvent

contradictoires ; le document analysé se révèle parfois être la seule source récente existante.

L'établissement des mots-matière, noms communs, pose deux autres problèmes : celui du bilinguisme et celui de la hiérarchisation dans un thesaurus. Ces termes qui servent à appréhender le contenu du contenu primaire servent aussi à l'établissement des chaînes d'index ou chaînes syntagmatiques, ce qui oblige à une parité dans le vocabulaire. Cette contrainte ne signifie nullement une traduction mais, du moins, une équivalence. Si certains termes gardent leur forme vernaculaire (par exemple *bonheur du jour*), la plupart des 7 500 termes ont un équivalent même si certaines traductions ne sont qu'approximatives et sont accompagnées de notes ; il peut se faire aussi qu'une langue n'utilise qu'un terme là où l'autre en use de trois (à *candlestick* correspondent en français, *flambeau*, *chandelier*, *bougeoir*).

Le thesaurus, lui, est constitué sur un seul site par l'équipe qui a la charge des autorités et en une seule langue, l'anglais ; le thesaurus français est obtenu automatiquement par le jeu de l'équivalence linguistique. Si le choix du terme appartient au rédacteur, si sa proposition se fait tantôt en anglais, tantôt en français, sa hiérarchisation se construit en anglais, ce qui est source de problèmes évidents, tant du point de vue linguistique que du point de vue du sens. Le thesaurus BHA n'a aucune prétention à l'universalité et ne fonctionne qu'en circuit fermé, chaque nouveau terme devant s'inscrire dans une hiérarchie existante qui se révèle valable dans la plupart des cas, mais qui génère parfois quelques distorsions sémantiques. On ne peut contester le choix d'*arts graphiques* comme terme générique de *cartes de vœux*, mais cette relation hiérarchique devient étrange, lorsque ces cartes de vœux sont réalisées en fonte de fer à Berlin au cours du XIX^e siècle !

À une époque où la technologie et les moyens nouveaux de collecte et de dissémination de l'information pourraient laisser envisager la constitution d'un vocabulaire non plus seulement bilingue mais plurilingue, ces quelques réflexions sur la complexité des problèmes laisse deviner l'énorme défi que représenterait un tel travail.