

Que faire de Rameau ? Ou comment faire avec Rameau ?

par Brigitte Turquet*

Quelques remarques à propos de l'article de Dominique LAHARY (*Bulletin d'informations de l'ABF*, 1997, n°174 pp. 60-62)

Faut-il conserver RAMEAU, et si oui, comment « faire avec » ? L'article de Dominique Lahary éclaire tout à fait des « questions inextricablement mêlées », tenant soit au répertoire RAMEAU lui-même, soit à sa gestion dans les logiciels de bibliothèques. Mais puisque l'éventail des positions concernant RAMEAU est « incomplet » (cf. son avant-propos, pp. 4-5), un indexeur de base dans une BU en sciences sociales se permet de prendre la parole, non pas tant pour une défense mesurée de ce répertoire imposé, que pour relancer quelques questions, dont la solution ne nous paraît pas toujours évidente. Nous suivrons donc le plan de l'article de Dominique Lahary¹.

1. Sur les « caractéristiques Rameau » (Rameau en tant que thésaurus)

1.1 Sur la Syntaxe RAMEAU (c'est-à-dire sur la construction au moment de l'indexation, de la vedette-matière), il y a presque

unanimité : cette syntaxe (survivance de la construction des vedettes dans les catalogues sur fiches et héritée du RVM laval)¹ est trop complexe. En conséquence, les chaînes construites sont très longues et très éloignées du langage naturel (c'est le cas par exemple de toutes les vedettes-matière construites concernant l'histoire économique, sociale, politique d'un pays...).

On l'a donc déjà dit : l'ordre respectif des subdivisions géographiques et chronologiques, l'utilisation de subdivisions de sujet « affranchies » ou non obéissent à des règles souvent compliquées pour l'indexeur de base. Dans le domaine juridique, par exemple, la construction est différente selon que l'on se trouve en présence d'une TV juridique ou non juridique...

Peut-on alors simplifier cette syntaxe (question elle aussi trop difficile pour y répondre ici, malgré quelques éléments de réponse ébauchés dans les points ci-après)?

Ceux qui contestent RAMEAU feront d'ailleurs remarquer qu'on ne voit pas pourquoi s'astreindre à appliquer une syntaxe si compliquée, puisque ces vedettes-matières si bien structurées seront dé-

construites au niveau de la recherche par mots unitermes dans l'OPAC (voir point 2.2) ?

Pourtant, on verra qu'il n'est pas toujours dénué d'intérêt de respecter une syntaxe ou une norme (définie dans le guide d'indexation RAMEAU), et cela même au niveau de la recherche sujet dans un OPAC (voir pt. 2.3), car les VMC¹ ont en elles-mêmes un sens, qui est, logiquement, d'abord perdu dans la recherche par mots unitermes.

1.2 Sur la structure et sur l'arborescence des liens hiérarchiques et associatifs dans RAMEAU.

Nous ferons deux remarques :

- Il est possible parfois que « l'arrière petit fils soit aussi le fils de son arrière grand-père » (cf. Dominique Lahary p. 60) : mais cela ne gêne pas vraiment ni le lecteur, ni l'indexeur, il nous semble. Pour l'indexeur, l'expérience montre que la navigation à travers les liens hiérarchiques et associatifs donne le plus souvent des idées correctes d'indexation, permettant d'aboutir à la vedette adéquate (la complexité de la chaîne construite et sa gestion dans l'OPAC est un autre problème).

- Dans certains domaines (dont le domaine juridique), il faut signaler en outre la présence de très nombreuses « notes d'application » très utiles pour l'indexeur et qui constituent parfois un véritable dictionnaire terminologique.

* Bibliothèque interuniversitaire Cujas de Droit et Sciences économiques.

1. Nous utiliserons dans cet article les abréviations suivantes (parfois utilisées dans les stages) : VMA : Vedette matière d'autorité ; VMC : Vedette matière construite ; TV : Tête de vedette ; OCLC : On Line Computer Library Centre ; RVM Laval : Répertoire de vedettes matières de l'université Laval (Canada).

1.3 Sur la rigidité du langage RAMEAU

Tout vocabulaire normalisé et pré-coordonné engendre cette rigidité, « *d'où l'intérêt des renvois, pour peu qu'on puisse les gérer* » (cf. Dominique Lahary p. 60).

Trois remarques sur ce point :

- Il faudrait en effet entrer le maximum d'expressions synonymes ou quasi-synonymes (pourquoi ne pas entrer en formes rejetées les nombreuses expressions utilisées par les lecteurs eux-mêmes, quand on ne peut modifier toutes les chaînes construites déjà existantes : ce serait déjà une manière de traiter en partie et en amont le langage des requêtes, lequel devrait l'être aussi en aval, mais c'est un autre problème...)

- Il faut noter une évolution heureuse de RAMEAU vers un langage plus naturel : la BnF privilégie de plus en plus les expressions, plutôt que les chaînes construites. On s'en aperçoit nettement en lisant la liste des « modifications » et des « créations » dans le journal RAMEAU n° 1 et 2. Exemple : CO Communautés européennes
** Aide économique

devient :

NC Aide économique européenne

Exemple : NC Notaires ** Archives

devient :

NC Archives notariales

- La liste RAMEAU est-elle une liste fermée (art. cité, p. 60) ? RAMEAU ne nous semble pas être un langage documentaire fermé : il s'enrichit constamment des propositions de mises à jour des bibliothèques participantes, à commencer par celles de la BnF (voir le nombre de descripteurs créés depuis 10 ans et en augmentation constante). Et nous pouvons témoigner que pratiquement toutes les propositions de la Bibliothèque Cujas faites depuis 1988 ont été acceptées (et cela malgré une difficulté particulière en droit tenant à l'origine anglo-saxonne de la liste, et à la polysémie interne au domaine juridique). Si un descripteur « spécialisé » manque, toute bibliothèque participante au réseau Rameau peut donc faire une proposition de création (Exemples récents en sciences sociales : « contrats-cadre », « garanties de

passif », « staries », « surestaries », « portage d'actions », « vente fob », etc.).

Mais il est vrai que la procédure de mise à jour est longue, et que faire une « bonne » proposition est aussi un travail difficile qui demande des connaissances dans le domaine. Toutes les bibliothèques n'ont sans doute pas les moyens de le faire, et le nombre d'experts par grands domaines est peut-être également insuffisant ? Cela explique aussi un certain pessimisme et découragement de l'indexeur quotidien, dont on exige rapidité et rentabilité...

Il est donc possible que RAMEAU, référentiel universel, ne réponde pas aux besoins de tous les publics. Mais il nous semble bien répondre aux besoins d'une bibliothèque moyennement encyclopédique dans un domaine donné, et donc aux besoins des chercheurs relativement spécialisés .

La « contestation sourde » qui accompagne RAMEAU depuis le début et les reproches qui lui sont faits ne nous paraissent donc pas toujours dirigés vers la bonne cible... qui pourrait bien être plus souvent les logiciels de bibliothèque !

2. RAMEAU et les logiciels

Rappelons ici la petite phrase parue dans l'ouvrage collectif *Offrir aux publics un catalogue en ligne*, IFB, 1995, p. 75 : « *Mettre à la disposition des usagers un OPAC sans les renvois est leur donner une fourchette pour manger de la soupe.* »

Rappelons aussi que les catalogues sur fiches savaient déjà gérer un système de relations tel que RAMEAU — les « voir aussi » étant l'équivalent des termes associés (TA) et des termes spécifiques (TS) — et que, dans un premier temps, l'informatisation a abouti à une relative régression, du point de vue de la recherche sujet dans les OPAC, simplement parce que de nombreux logiciels ne savaient pas ou ne savent pas encore gérer correctement ces liens (et cela malgré l'intérêt et la richesse des accès sur les mots unitermes de la zone titre, ou autres zones).

Il est donc permis de poser la question : est-ce RAMEAU (ou tout autre thésaurus ou vocabulaire normalisé) qui doit s'adap-

ter aux logiciels, ou bien les logiciels qui doivent savoir prendre en compte un vocabulaire normalisé et un système de relations, tels que RAMEAU ? Sans doute chaque outil concerné (et le thésaurus et l'OPAC) peut et doit être amélioré, et comme le fait remarquer Jean-Paul Gaschi-gnard, dans ce n°174 du *Bulletin* de l'ABF, l'amélioration viendra aussi d'une pression et d'une demande collective des bibliothèques sur les fournisseurs de logiciels.

Ceci dit, nous voudrions revenir sur quelques difficultés soulevées dans l'article de Dominique Lahary, et dont la solution ne nous semble pas évidente.

2.1 La gestion des relations de synonymie (ou renvois d'équivalence) dans les OPAC

Deux problèmes concernant cette gestion peuvent être soulevés :

- Un « forme exclue » telle que « révolution française », doit-elle être liée aux notices bibliographiques :

— des documents indexés avec la seule « vedette-matière d'autorité » (VMA)¹.

— ou bien également aux notices des documents indexés avec toutes les « vedettes-matière construites » (VMC) à partir de cette VMA ?

Pour notre part ce dernier choix ne nous semble pas « aberrant », contrairement à ce que semble penser Dominique Lahary (art. cité, p. 61), à partir de l'exemple « Révolution française » (chaîne construite particulièrement complexe !).

Prenons un autre exemple pris dans l'OPAC de la Bibliothèque Cujas : la « fraude informatique » (forme exclue rameau) renvoie à la forme retenue « délits par ordinateur ». On obtient l'affichage suivant :

Requête : < fraude informatique >

- | | |
|---|--|
| 6 | Délits par ordinateur |
| 3 | Délits par ordinateur ** Etudes comparatives |
| 4 | Délits par ordinateur ** France |
| 1 | Délits par ordinateur ** Grande-Bretagne |

1 Délits par ordinateur ** Italie

concernés au total par sa recherche, faute de quoi il aurait un silence de 10 documents qui peuvent l'intéresser.

1 Délits par ordinateur ** Pays de l'Union européenne

16 Notices bibliographiques

Dans l'ensemble, il semble donc pertinent de lier les termes exclus à toutes les VMC, et pas seulement à la vedette matière d'autorité.

Il nous semble intéressant qu'un lecteur, qui formule sa requête de façon large, ait accès, à partir du terme exclu, non seulement aux 6 documents indexés avec la seule VMA « délits par ordinateur », mais aussi aux 5 VMC (vedettes matières construites), et donc aux 16 documents

• Un autre problème est celui de la **récupération ou de l'importation**, sur le plan **informatique**, des « termes exclus » RAMEAU dans un OPAC particulier : en effet chaque bibliothèque a besoin d'une petite partie de RAMEAU seulement.

Comment importer les seules formes rejetées dont telle bibliothèque a besoin et les lier aux notices bibliographiques, étant donné les problèmes de format que l'on sait (notamment pour les bibliothèques qui récupèrent des notices dans l'OCLC) ? On trouve des indications sur ce point dans l'ouvrage collectif *Contrôler la qualité et la cohérence du catalogue*, IFB, 1996, pp. 65-84. Mais des stages de formation permanente sur ce problème d'importation et de gestion des formes rejetées dans les OPAC seraient très souhaitables. Car c'est bien le minimum que l'on puisse attendre d'un OPAC : qu'il sache gérer les renvois d'équivalence. Une gestion performante de ces renvois ferait plus qu'« atténuer », à notre avis, les problèmes difficiles de la recherche sujet dans les OPAC.

2.2 Cela conduit à la deuxième condition de bonne gestion de RAMEAU signalée par Dominique Lahary (p. 61) : La recherche sujet doit se faire par mots UNITERMES, et non par le début (1^{er} mot) de la vedette-matière d'autorité, que le lecteur ne peut connaître. Mais faut-il proposer en première approche un accès indistinct par unitermes : « mots du titre, mots des vedettes RAMEAU, mots du résumé ou du sommaire » ? Il semble bien que ce genre d'accès « indistinct » soit relativement satisfaisant, lorsque la recherche utilise des mots significatifs et non polysémiques. Mais dans le cas contraire (et malgré l'utilisation implicite d'opérateurs booléens), le « bruit » est souvent considérable, (sauf à proposer aussi aux lecteurs l'utilisation d'autres opérateurs, comme l'adjacence et la proximité ?).

Par ailleurs, les temps de réponse, dans cet accès, sont généralement très longs. Exemple (pris dans l'OPAC de la bibliothèque Cujas, qui offre un accès indistinct par mots unitermes du titre et mots des descripteurs RAMEAU) : pour une recherche sur le « Conseil d'Etat », l'OPAC rappelle 280 documents pour 45 seulement qui sont pertinents par rapport à la requête : il y a donc un « bruit » de 235 documents ! Et ce cas n'est pas un cas limite rare...

2.3 Une dernière condition de bonne gestion de RAMEAU dans les OPAC signalée dans l'article cité (p. 61) est l'accès direct, à partir d'une requête sujet, à la liste des notices bibliographiques, sans passer d'abord par la liste des VMC (vedettes matières construites), car « un bruit de

vedettes [serait] pire qu'un bruit de notices» (art. cité p. 61). C'est vrai. Mais faut-il dans tous les cas préconiser cet accès direct aux notices ? D'une part, il n'y a pas forcément un bruit de vedettes (ici de nombreux paramètres entrent en jeu). Par ailleurs le passage par l'intermédiaire de cette liste (index des VMC) est souvent très intéressant. C'est notamment le cas lorsque la requête porte sur des domaines assez larges. Exemple : Pour une recherche sur le « droit commercial » en « France » (termes saisis dans la requête), on peut avoir l'affichage suivant :

Requête : < droit commercial France >

- 61 Droit commercial ** France
- 16 Droit commercial ** France ** Codes
- 1 Droit commercial ** France ** Dictionnaires
- 16 Droit commercial ** France ** Examens, questions, etc.
- 3 Droit commercial ** France ** Histoire
- 1 Droit commercial ** France ** Législation
- 45 Droit commercial ** France ** Manuels d'enseignement supérieur
- 143 Notices bibliographiques

L'OPAC affiche 7 VMC (vedettes matières construites) : le lecteur (notamment l'étudiant) pourra sélectionner d'un coup d'œil dans cette liste les 45 manuels dont il a besoin, plutôt que de lire l'une après l'autre les 143 notices bibliographiques proposées. Les requêtes étant souvent

posées de façon large, l'affichage des VMC est donc parfois très utile. Peut-être faut-il dire que chaque bibliothèque doit faire un choix en fonction de son public ?

2.4 Autre problème soulevé dans l'article cité (p. 62) : doit-on afficher l'arborescence des vedettes correspondant à une requête, c'est-à-dire les relations hiérarchiques et d'association ? La question mérite en effet d'être posée, mais elle reste théorique dans l'état actuel des OPAC, dit Dominique Lahary.

Nous nous permettrons seulement de redire ceci : les catalogues sur fiches comportaient l'accès à ces liens (sous la forme des « Voir aussi »), et les rétablir dans les OPAC ne sera que retrouver ce que nous avons (provisoirement) perdu... avec l'informatisation des catalogues !

Quoi qu'il en soit, la lecture des TA (termes associés) et TS (termes spécifiques) est d'une grande aide pour orienter et l'indexeur, et le lecteur. La BnF nous annonce d'ailleurs un « suspense » sur ce plan (art. de Françoise Bourdon, *Bulletin ABF*, 1997, n°174, p. 52).

Du bon usage de Rameau et/ou autres outils linguistiques

Pour une bibliothèque, indexer en RAMEAU est un « investissement lourd », qui prend du temps (art. cité p. 62). Mais ce qui prend du temps (on l'oublie), c'est d'abord la compréhension des ouvrages eux-mêmes, surtout en langues étrangères, d'autant plus que les indexeurs ne sont pas toujours assez spécialisés, dans les disciplines qu'ils traitent. Cela pose à nou-

veau le problème de la coopération insuffisante entre bibliothèques : il manque peut-être des spécialistes par grands domaines et par langues, qui puissent indexer assez rapidement pour tous, et traiter aussi les propositions de création.

Quant à la récupération de l'indexation RAMEAU dans les « réservoirs » bibliographiques, l'expérience montre qu'elle n'est pas exempte de problèmes. On nous dit qu'il ne faudrait pas « ré-indexer systématiquement les notices » (art. cité p. 63). C'est exact. Mais d'une part on constate parfois des « dérives » dans l'indexation (utilisation des formes rejetées, ou trop générales, ou indexation erronée parce qu'établie d'après le titre, etc.). D'autre part, la récupération des VMC anglaises (pour les bibliothèques qui cataloguent dans l'OCLC), et leur traduction en RAMEAU doit se faire avec beaucoup de prudence (si on veut éviter les contresens, dus à la polysémie de la langue anglaise et à d'autres facteurs). Récupérer l'indexation avec un esprit critique fait donc partie du « contrôle de la qualité », et c'est le lecteur qui en bénéficie. Comme il bénéficierait aussi d'une indexation additionnelle, non conforme à RAMEAU, et faite en fonction d'un fonds ou d'un public (idée très intéressante, relevée par Dominique Lahary, mais à condition de gérer le problème du « bruit » dans l'OPAC).

De même l'indexation pourrait être enrichie par la numérisation des sommaires, résumés, etc. : peut-on alors faire porter ses espoirs, se demande Dominique Lahary, sur d'autres traitements linguistiques, qui seraient capables de faire de l'analyse sémantique et de traiter le problème de la « jungle du sens ». On ne saurait rien ajouter (pour raison d'incompétence) à ce qui est déjà excellemment dit dans l'article cité, sinon cette remarque : lors d'une journée d'études de l'ADBS en décembre 1996, sur les « outils linguistiques et les nouvelles technologies », il était frappant de constater que, pour de nombreux professionnels de la documentation (documentalistes des journaux *Les Echos*, *Le Monde*, etc.), les deux techniques possibles de l'indexation par sujet (traitement linguistique du texte intégral et thésaurus) sont maintenant considérées comme des approches complémentaires : le texte intégral (du document primaire ou du résumé..., traité en amont, est sur-indexé à l'aide d'un thésaurus, lequel semble donc garder tout son intérêt pour