



L'indexation matière en transition : de la réforme de Rameau à l'indexation automatique : introduction et conclusion

Etienne Cavalié

► To cite this version:

Etienne Cavalié. L'indexation matière en transition : de la réforme de Rameau à l'indexation automatique : introduction et conclusion. L'indexation matière en transition - De la réforme de Rameau à l'indexation automatique, Etienne Cavalié dir., Paris, Editions du Cercle de la Librairie, 2019, (Collection "Bibliothèques") - 978-2-7654-1623-4, 2019. hal-02435623

HAL Id: hal-02435623

<https://hal.science/hal-02435623>

Submitted on 11 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

L'indexation matière en transition

De la réforme de Rameau à l'automatisation

Conclusion

Étienne Cavalié. Bibliothèque nationale de France.

Le 19 octobre 2013, dans la regrettée *Place de la Toile*, émission animée par Xavier de La Porte sur France Culture de 2007 à 2014, se trouvait parmi les invités le journaliste Stephen Des Aulnois, alors rédacteur en chef du site *Le Tag parfait* consacré à la « culture pornographique ». Stephen Des Aulnois a pris le temps d'explicitier le nom du site : pour trouver des vidéos, l'internaute va (sans en avoir forcément conscience) réaliser des recherches par sujet, et pour cela exploiter les tags qui décrivent les ressources, en espérant trouver un jour « le tag parfait », celui qui correspondrait parfaitement à sa personnalité et lui permettrait d'obtenir des résultats toujours pertinents. Ce tag parfait représente en réalité « une utopie, quelque chose qui n'existe pas, une quête »¹.

On retrouve magnifiquement illustré, assez loin du monde des bibliothèques, l'espoir qu'un tag, ou un mot-clé, puisse un jour correspondre parfaitement à ce qu'il est censé désigner. Il est vrai que le bibliothécaire est moins ambitieux, puisqu'il espère l'adéquation entre un ensemble de termes et un document, et non entre un terme et la nature intime d'un être humain.

On pourrait toutefois inverser la démarche, et plutôt que de chercher à fournir une indexation sujet correspondant au document, en chercher une qui correspondrait à son possible utilisateur : si je suis un lecteur qu'un tel document intéresserait, quels seraient les mots par lesquels je le chercherais ?

Telle qu'enseignée généralement dans les formations aux métiers des bibliothèques, l'indexation est partie intégrante de la description documentaire : on décrit un document, et dans ce cadre on s'efforce de donner à comprendre son contenu par le choix de quelques termes². L'adéquation au lecteur relève d'un autre volet, celui de la médiation. Et l'idée de se préoccuper du lecteur plutôt que du document pose évidemment de nombreuses questions, et en particulier :

- En « sélectionnant » un type de lecteur cible, ne va-t-on pas passer à côté d'un autre ?
- Validerait-t-on ainsi le principe d'une subjectivité complète de la pratique de l'indexation, le même document étant susceptible d'être décrit de manières complètement différentes d'un établissement à l'autre, voire d'un catalogueur à l'autre ?

¹ Source : <https://www.franceculture.fr/emissions/place-de-la-toile/tubes-et-culture-porno> : enregistrement à partir de 3'40, et plus précisément 6'50-7'50.

² Bruno Bachimont nous rappelle utilement que l'indexation ne relève pas tant des métiers de la documentation, mais des métiers de l'information : ce que l'on cherche à rendre accessible, ce n'est pas un document précis (que l'on cherchera par son auteur ou son titre), mais l'information qui s'y trouve.

En réalité, il semble nécessaire pour sortir de cette impasse de rappeler une autre question posée en introduction à ce présent ouvrage : comment concevoir dans un établissement une politique d'indexation si le catalogage local tend à disparaître, et si les notices sont systématiquement dérivées ? Ou plus exactement : comment envisager de construire une offre de services qui s'appuie (entre autres choses) sur l'indexation des notices, alors qu'on ne la maîtrise absolument pas ?

Quelle politique d'indexation locale sur des notices importées ?

Il nous faut en réalité distinguer deux types d'indexation sujet :

- Celle qui est pratiquée par les fournisseurs de notices bibliographiques (la Bibliothèque nationale de France, etc.), et qui pour concevoir une indexation sujet ne peut s'appuyer que sur le contenu de l'ouvrage, ignorant tout des publics destinataires. Pour la BnF, c'est cohérent avec le fait que, dans le cadre du signalement du dépôt légal, elle ne met en œuvre aucune politique documentaire : elle est simplement dépositaire de ce qui lui est adressé.
- Celle qui est réalisée au sein d'une bibliothèque, *après* récupération de la notice bibliographique, afin de contextualiser le document, de faciliter sa découverte d'après ce que l'on sait des usages du public desservi.

Cette démarche ne pose aucune difficulté quand il s'agit d'attribuer une cote : chaque bibliothèque a son plan de classement (souvent conçu à base d'indices Dewey), et l'ouvrage peut se trouver dans un tout autre rayon selon l'établissement.

Cette constatation n'a rien de révolutionnaire : elle entre au contraire en harmonie avec les principes de gestion des collections et de politique documentaire, selon lesquels un ouvrage n'est pas bon ou mauvais en soi, mais son acquisition peut se révéler pertinente ou non selon l'établissement qui l'achète. Pour autant, elle entre en contradiction avec la manière dont l'indexation est associée à la description documentaire dans le cursus des formations initiales³, alors qu'elle devrait plutôt relever de la gestion des collections, ou de la médiation auprès des publics. Comme en témoigne l'exemple de la médiathèque de Roubaix⁴, les bibliothèques n'ont pas attendu pour réindexer les notices dérivées, afin qu'elles prennent sens *en contexte*.

Toujours est-il que ce qui semblait une difficulté sur la possibilité d'une politique d'établissement concernant l'indexation trouve au contraire sa parfaite solution : c'est précisément la réduction au strict minimum de l'activité de catalogage local qui va permettre de s'intéresser à l'indexation pour ce qu'elle est.

La définition d'une politique de médiation adossée aux données (d'indexation ou autre) nécessite des compétences en interne : car elle nécessite de savoir manipuler les données existantes, et celles qu'il s'agit d'importer. Ces compétences ne portent plus tant sur les

³ Cf. *supra* la contribution de David Guillemin.

⁴ Cf. *supra* le chapitre de François Pichenot.

subtilités du format et de ses zones à positions fixes, mais sur l'aptitude à extraire des informations et à les analyser, à savoir les agréger à d'autres sources sans perdre d'information.

Ce type de profil et de compétences doit entrer en complémentarité avec l'indexation à des fins de médiation : il convient de faire collaborer gestion des collections et gestion des données. Les données importées méritent elles aussi être gérées, et il serait très dommageable d'abandonner toute compétence sur elles. Cela contraint à s'intéresser encore aux formats Marc, y compris au niveau d'un établissement – mais avec un autre regard, d'autres enjeux.

L'entrée de ces compétences dans une bibliothèque peut relever d'une politique de recrutement. Elle peut aussi⁵ relever de la gestion de ressources humaines en interne, de la formation continue et de temps de travail dégagé pour pouvoir les acquérir : car on reste bien dans des compétences qui relèvent du traitement de la donnée, donc des métiers des bibliothèques, de la documentation, et des sciences de l'information.

Quelle articulation entre catalogage national et indexation locale ?

Cette capacité à gérer les données en flux⁶, ou en lots, est d'autant plus importante dans le contexte de la Transition bibliographique, et plus précisément ici de la réforme Rameau.

Le choix de mettre en place un enrichissement des notices après leur récupération, qu'il soit réalisé manuellement (par des personnels de la bibliothèque) ou automatiquement (en ayant identifié une source pertinente que l'on irait interroger sur la base des ISBN, par exemple), nécessite d'être à tout moment capable de distinguer les informations initiales des ajouts postérieurs : il faut préserver les informations récupérées, et il faut également pouvoir préserver les informations ajoutées *a posteriori*, et pouvoir toujours distinguer les premières des secondes.

Pour une indexation locale, une saine utilisation du format Unimarc signifie utiliser une zone dédiée, à savoir la zone 609⁷.

L'indexation locale doit donc impérativement se faire dans des zones structurées de telle manière qu'on ne puisse les confondre avec l'indexation récupérée en même temps que la notice : dans la zone 609, ou au pire dans une zone 606 qualifiée \$2{indexation locale}⁸.

Cela permettra ultérieurement de récupérer la notice de la BnF (ou d'une autre source), actualisée entre temps, en ayant la possibilité technique de protéger l'indexation locale de tout écrasement.

Pourquoi vouloir récupérer de nouveau les notices qu'on a eu besoin d'importer plusieurs années auparavant ? Dans le cadre de la réforme Rameau les chaînes d'indexation dans les notices bibliographiques vont évoluer. Or, comme le rappelle Florence Ménard, cette réforme se conduit en plusieurs étapes, et sur plusieurs années : certaines étapes sont susceptibles de

⁵ Comme le rappelle la conclusion de Géraldine Geoffroy.

⁶ Selon des mécanismes décrits *supra* par Bernard Strainchamps.

⁷ Dans les formats Marc, les zones en XX9 sont réservées à un usage local, et donc par définition non définies dans l'Unimarc standard, et donc absentes, logiquement, des notices que l'on peut être amené à importer.

⁸ Mettre ici un code propre à l'établissement sans risque de confusion.

rendre fautives les indexations d'avant le début de la réforme. Cela se vérifie pour les évolutions de 2019-2020, notamment pour l'interversion de l'ordre Lieu -- Sujet en Sujet -- Lieu⁹, et qui déplacent la qualification du genre ou de la forme de l'œuvre dans une zone dédiée.

Pour une bibliothèque, l'adoption de ces règles dans le catalogage courant implique de se pencher sur le cas des notices déjà existantes : qu'un même sujet puisse apparaître (dans la facette d'un catalogue en ligne) avec un ordre des termes différent d'une ligne à l'autre peut être source de confusion pour le lecteur.

La BnF entreprend pour son propre catalogue de réaliser un certain nombre d'opérations complexes pour appliquer aux notices existantes les nouvelles règles en vigueur. La reproduction de ces opérations n'est pas toujours à la portée de chaque bibliothèque.

C'est pourquoi le programme de la Transition bibliographique s'efforce d'accompagner dans toutes ses dimensions l'appropriation de cette réforme par les bibliothèques, qui concerne autant le travail courant que le traitement du rétrospectif.

- Pour les impacts de la réforme sur le catalogage courant : le site Rameau diffuse la documentation sur le processus global et sur chaque étape à venir (nature des évolutions, méthodologie, liste des notices concernées, etc).
- Pour le rétrospectif bibliographique :
 - Les traitements pour la reprise de données réalisés à la BnF font l'objet d'échanges avec l'Agence bibliographique de l'enseignement supérieur (Abes), ainsi qu'avec les éditeurs de SIGB et les bibliothèques réutilisatrices.
 - Le groupe Systèmes & Données fournit à travers le logiciel Bibliostratus¹⁰ un outil facilitateur pour la récupération des notices BnF/Sudoc hors traitement courant.

La réforme Rameau illustre bien une des principales problématiques dans la stratégie d'accompagnement du réseau des bibliothèques de France. La Transition bibliographique vise à transformer complètement les catalogues en adoptant un nouveau modèle de données. Les étapes de transformations successives des données existantes pour qu'elles correspondent peu à peu au nouveau modèle IFLA LRM sont prises en charge par la BnF et l'Abes, qui doivent ensuite diffuser leurs données converties, et faciliter leur récupération. Le dispositif pour faire « redescendre » ces futures entités dans les catalogues locaux est loin d'être précisé dans ses moindres détails. Mais déjà, les premières évolutions de 2019 (janvier puis mai) de la réforme Rameau permettent de faire le compte des acteurs à chaque maillon de la chaîne, et de leurs problèmes spécifiques. La réforme Rameau est en effet la première évolution à avoir un impact très fort sur les catalogues, puisqu'elle rend les données existantes incorrectes (parmi les étapes antérieures, l'ajout de nouvelles zones « Type de contenu », « Type de médiation » rend les notices existantes simplement incomplètes, mais non incorrectes).

Les mois et les années à venir, on peut l'espérer, permettront d'éprouver et d'améliorer ce mécanisme d'actualisation.

⁹ Aujourd'hui, selon les cas, l'ordre d'indexation sera tantôt Lieu -- Sujet, tantôt Sujet -- Lieu. Pour

¹⁰ À propos de Bibliostratus : <<https://www.transition-bibliographique.fr/systemes-et-donnees/bibliostratus-alignement-donnees-catalogues/>> sur le site de la Transition bibliographique.

Malgré tout, un tel dispositif ne sera possible que si les catalogues locaux sont en mesure de distinguer, au sein de chaque notice, ce qui relève de l'enrichissement local de ce qui peut être actualisé par un rechargement (distinction claire entre la zone 606 et la zone 609). On en revient donc bien à la nécessité de compétences locales pour piloter, et surtout *anticiper* un tel mécanisme.

La prédictibilité d'une indexation Rameau ?

Si les chapitres de Florence Ménard et Ewa Nieszkowska-Serlan permettent d'identifier différents objectifs et étapes dans la mise en œuvre de la réforme Rameau, une idée forte se dégage nettement : l'ambition de la prédictibilité d'une chaîne d'indexation à partir de concepts Rameau, et l'obsolescence d'une grande partie des 600 pages du *Guide d'indexation Rameau*. L'objectif est de limiter à un très petit nombre les règles de structuration d'une chaîne d'indexation.

La systématisation de l'organisation des concepts dans le même ordre Concept > Lieu > Temps, ainsi que le chantier du « lissage terminologique »¹¹ vont dans le sens de cette prédictibilité.

Celle-ci est identifiée comme une condition pour une ouverture (non plus technique, mais intellectuelle) de Rameau à de nouveaux réutilisateurs.

Cette ambition n'est pas sans générer des tensions logiques. Rameau sert à indexer les documents en vue de les trouver lors d'une recherche. Il propose aux bibliothécaires et documentalistes une liste de valeurs permettant de supprimer les risques d'ambiguïté. Or « l'impératif documentaire », qui énonce qu'un terme Rameau n'est créé que s'il y a au moins un document à décrire (ceci pour éviter la prolifération inutile de termes qui ne serviraient jamais à personne, et qu'il faudrait pourtant bien gérer), interdit à Rameau d'être aussi cohérent que la réalité même : on ne crée de notice Rameau au nom d'un fleuve sud-américain ou d'un concept à la mode que si arrive la demande d'un catalogueur du réseau des bibliothèques. Cela signifie que, par nature, il manquera toujours par exemple des chaînons intermédiaires qui ne font l'objet d'aucune publication¹².

Parallèlement, Rameau veut s'ouvrir à d'autres communautés d'utilisateurs, à commencer par le monde de la recherche. La promotion de Rameau comme langage documentaire doit donc se donner à voir comme un ensemble cohérent, indépendamment de son utilisation dans le cadre de l'indexation documentaire.

Le projet d'intégration de Rameau au Fichier national d'entités sera vraisemblablement aussi une raison de plus pour formaliser davantage la cohérence interne de Rameau et les règles d'évolution au quotidien.

¹¹ Cf. *supra* les explications de Florence Ménard.

¹² Ewa Nieszkowska-Serlan évoque l'absence d'une modalité de navigation dans Rameau en le parcourant par sous-ensembles successifs. Cette expérience n'est pas envisageable tant que ne sont pas systématisés ces éléments intermédiaires, qui sont interdits par cet « impératif documentaire ».

Le chantier du « lissage terminologique » (qui vise à faciliter la prédictibilité de Rameau), vise à restreindre ou supprimer la distinction entre notices recouvrant la même notion mais appliquée à des thématiques (dits « domaines d'application ») différents. La fusion des notions en un seul concept applicable en tout contexte rend moins long le travail du catalogueur dans la sélection des termes. Le Centre national Rameau cherche aussi à distinguer les sous-ensembles « Nom commun », « Nom géographique », « Période » (ou « Concept, Lieu, Temps ») au sein des vedettes construites. En effet s'il est interdit d'associer à l'indexation les deux concepts « Politique sociale » et « France » parce que la notice préconstruite « Politique sociale -- France » existe (tandis que « Politique sociale -- Italie » n'existe pas), alors Rameau n'est pas prédictible : pour indexer un document sur la politique sociale en France coexistent deux possibilités concurrentes, dont une seule est correcte (la vedette préconstruite l'emporte) mais uniquement pour des raisons historiques (la notice a été créée, alors qu'elle aurait pu ne pas l'être). Ce « lissage » doit donc avoir, parmi ses effets, de réduire le nombre des notices préconstruites.

Peut-on et faut-il donc supprimer toutes les vedettes préconstruites ? Il serait alors impossible d'enregistrer les « formes rejetées » ? Car un concept « Femmes -- Travail » doit exister pour que l'on puisse stocker la chaîne de caractères « Travail féminin » susceptible elle aussi d'être utilisée par des lecteurs.

Peut-on alors au moins supprimer les vedettes préconstruites qui n'ont pas de forme rejetée ? Il y en a plusieurs centaines. En voici une : "Acier -- Fatigue" (ark:/12148/cb11954523w)¹³. Or le concept "Fatigue" (ark:/12148/cb11967926z) dans Rameau est réservé au synonyme d'épuisement (et ses formes rejetées sont : Lassitude, Fatigabilité, Surmenage).

Si demain on supprime la vedette préconstruite "Acier -- Fatigue", et que la règle d'indexation consiste à utiliser l'association des deux concepts, "Acier" et "Fatigue", l'utilisateur pourra continuer à chercher "fatigue de l'acier" dans le catalogue et retrouver les mêmes documents. En revanche cela modifie le sens du concept "Fatigue" tel que Rameau le définissait jusque-là.

La question se pose selon deux points de vue différents, celui du lecteur et celui du gestionnaire du référentiel :

- Y a-t-il une dégradation de service pour l'utilisateur ? (ici : non, il semble qu'il trouve toujours ses documents avec les mêmes termes de recherche, puisque « Acier – Fatigue » ne comportait pas de forme rejetée)
- Comment gérer Rameau en tant que thésaurus de concepts, si la définition de ces concepts peut se trouver détournée lors de leur utilisation en situation ?

On peut aussi poser l'alternative suivante : Rameau est-il un référentiel de termes (liste des termes autorisés pour indexer des documents) ou un référentiel de concepts ? Est-il un dictionnaire (une entrée par terme) ou une encyclopédie (une entrée par notion) ?

À l'heure actuelle Rameau est un référentiel de concepts, mais le lissage terminologique repose la question et nécessitera de clarifier certains choix.

¹³ La « fatigue » appliquée aux matériaux désigne l'endommagement de ces matériaux pour différentes causes (forces exercées, etc.). Cf. l'article « Fatigue (matériau) », *Wikipédia*, [En ligne] <[https://fr.wikipedia.org/wiki/Fatigue_\(mat%C3%A9riau\)](https://fr.wikipedia.org/wiki/Fatigue_(mat%C3%A9riau))> (Consulté le 10/03/2019).

Un pas vers l'automatisation et l'intelligence artificielle ?

La simplification des règles de syntaxe de Rameau permet d'assurer plus facilement la prédictibilité des chaînes d'indexation. Par ailleurs, les consignes d'utilisation devraient progressivement être transférées intégralement dans des zones codées (c'est d'autant

plus facile que les règles sont plus simples). C'est une étape préalable pour envisager d'entraîner une intelligence artificielle à « fabriquer de l'indexation Rameau ». A l'heure actuelle il n'est pas envisageable de faire apprendre à une machine toutes les règles d'usage telles que décrites dans le *Guide d'indexation*. Mais dès lors que la liste de ces règles se réduit (dans quelques années, donc), la question d'une automatisation mérite d'être posée, et l'hypothèse de faisabilité d'être expérimentée.

Sans se lancer pour le moment dans une prédiction prophétique sur la question, soulignons tout de même que cet avenir est, dans d'autres pays, déjà un passé (et un présent). Il implique en France encore un certain nombre d'étapes intermédiaires, pour mettre en œuvre et évaluer les résultats d'un tel mécanisme. Nous n'en sommes donc pas là. Les questions d'évaluation seront certainement tout aussi complexes que la réalisation elle-même.

Si l'on en revient à la distinction entre les deux types d'indexation (description documentaire / médiation), évoquée plus haut, on peut s'interroger sur la place de l'IA : au niveau de l'indexation de description documentaire ou au niveau l'indexation de médiation ?

Dans les exemples donnés plus haut par Jean-Philippe Moreux sur l'indexation automatique, ces procédés s'appuient souvent sur du *machine learning* : des professionnels accompagnent l'apprentissage de l'algorithme en lui fournissant peu à peu une mémoire de ce qui est attendu, afin que l'algorithme puisse faire des choix peu à peu plus pertinents au fur à et mesure que sa base de connaissance s'accroît. Il est donc évident que les résultats d'une telle indexation dépendront directement du contenu de cette *mémoire*. Cette intelligence, même artificielle, fournira donc des résultats subjectifs, ou, pour mieux dire, contextuels à l'environnement dans lequel on la conditionne. Cette solution n'entrerait donc nullement en contradiction avec l'hypothèse d'une indexation locale, de terrain. Ce pourrait même être précisément dans sa capacité à fournir des résultats relatifs qu'il faudrait juger de la pertinence d'une telle indexation.

La mise en œuvre de tels processus, au niveau des agences bibliographiques comme des établissements, nécessitera bien évidemment des compétences en traitement de données, dans la continuité de celles déjà évoquées.