L'indexation amateur des œuvres littéraires au service des algorithmes de recommandation sur *Babelio*

Nathalie Deley
Université Jean-Monnet, Saint-Étienne – laboratoire Elico

Introduction

Dans une perspective de compréhension des pratiques de recommandation portant sur des objets culturels, il nous est apparu opportun de nous interroger sur les nouvelles possibilités offertes par les plateformes numériques en vogue sur le net. Désormais, l'internaute peut donner son avis à peu près sur tout, sur la toile. Les modalités de transmission des jugements sur les œuvres culturelles s'en trouvent par conséquent, modifiées ; le mélomane peut partager ses *playlist*, le cinéphile conseiller telle sortie en salle ou en VOD et le lecteur partager ses lectures et ses critiques, ce qui pose la question de la reconfiguration de la recommandation culturelle. Grâce au Web collaboratif, la place des amateurs devient déterminante.

Nous souhaitons dans ce contexte, étudier comment les acteurs du numérique (à l'origine du lancement de ces plateformes) mobilisent les internautes dans les pratiques de recommandation. Pour cela, nous centrerons notre propos sur le site *Babelio.com*. Il s'affiche comme « un réseau social dédié aux livres et aux lecteurs [qui] permet de créer et d'organiser sa bibliothèque en ligne, d'obtenir des informations sur des œuvres, de partager et d'échanger ses goûts et impressions littéraires avec d'autres lecteurs ». À ce jour [1], il ne compte pas moins de 450 000 inscrits et peut se prévaloir de trois millions de visiteurs uniques [2] par mois, ce qui est considérable. Le phénomène est donc massif.

Nous montrerons de quelle manière *Babelio* met en place un espace d'échange de données sur les livres qui va bien au-delà du simple partage des coups de cœur des

1. Ce travail porte sur une analyse de *Babelio* réalisée en mars 2018.
2. Le nombre de visiteurs uniques ne doit pas être confondu avec le nombre de visites. Lorsqu'un visiteur revient plusieurs fois, il n'est compté qu'une seule fois.

lecteurs. Les connaissances produites par chaque internaute autour des livres de sa bibliothèque permettent en effet non seulement d'enrichir un patrimoine collectif autour de l'œuvre (visée d'archivage) mais également de créer un processus de prescription actif dans le choix à venir d'autres œuvres littéraires, ce que *Babelio* appelle le service de recommandation.

Ce qui nous intéressera ici sera donc de montrer le rôle des internautes dans la production de métadonnées sur le livre, et plus particulièrement à partir d'une pratique spécifique qui est celle de l'étiquetage. Celle-ci, également nommée *tagging* consiste à renseigner les livres de sa bibliothèque virtuelle à partir de mots-clefs, mots-clefs qui seront ensuite repris par le dispositif de recommandation de la plateforme. Ainsi, les internautes peuvent indexer leurs livres en langage naturel, ce qui leur permet ensuite de se repérer dans leur bibliothèque mais surtout ce qui constitue des indices sur leurs goûts de lecture, indices repris, croisés et complétés par d'autres données produites sur le site, comme les notations par exemple, pour leur suggérer de manière automatisée des titres à lire.

Notre questionnement s'inscrit dans la lignée de travaux déjà menés autour des problématiques d'usages du numérique et des incidences sur le marché du livre (Wiart, 2015), de la circulation de la parole dans l'espace public et de sa légitimité par rapport à la parole instituée et professionnelle des acteurs du secteur (Pasquier *et al.*, 2015) ou encore du rôle du contributeur dans la patrimonialisation des œuvres et dans la constitution d'une mémoire collective culturelle (Bonaccorsi, Croissant, 2015).

Dans un premier temps, pour situer notre problématique, nous décrirons la plateforme en lien avec le modèle économique visé. Le projet repose sur un modèle qui associe les maisons d'édition d'une part et les bibliothèques d'autre part, d'où son originalité [3]. La projection à destination des bibliothèques nous intéresse tout particulièrement pour faire le lien avec la structuration du champ professionnel de l'information documentaire. Ainsi, dans un second temps, nous nous interrogerons sur la manière dont la plateforme sollicite et encadre le lecteur dans l'élaboration d'un index pertinent. Enfin, dans une troisième partie, nous questionnerons le lien entre les pratiques d'indexation amateur et certaines méthodes de travail des bibliothécaires pour montrer la manière dont cela peut transformer la construction de référents culturels (index, thésaurus..) les obligeant à revoir leurs propres pratiques professionnelles.

1 Le modèle de la plateforme *Babelio* : la place des contributions amateurs dans le dispositif prescriptif

Dans cette partie, nous allons donner quelques clefs pour situer le projet *Babelio* dans le modèle économique qui a motivé sa création et décrire les grands traits du modèle éditorial qui lui sont liés. Les fondateurs du site déclarent qu'ils utilisent les contributions des amateurs pour leur faire des suggestions de lecture personnalisées : « *Babelio* vous propose des livres en fonction de vos goûts sur la page de recommandations. » Cette recommandation intervient sur le marché du livre et peut, nous le supposons, l'influencer en participant, aux-côtés des prescripteurs traditionnels que sont les li-

3. Entretien avec Pierre Fremaux, co-fondateur de la plateforme, le 6 avril 2017.

braires, les éditeurs, les journalistes ou les auteurs eux-mêmes, à l'orientation des choix des lecteurs.

1.1 Le projet *Babelio* : paroles amateurs et double marché

Sur le Web, il existe de nombreux lieux d'expression de la parole amateur sur les livres et les pratiques de lecture. Blogs personnels, forums de discussion associés à un site, plateformes communautaires, sont des espaces variés peuplant le Web et sur lesquels nous pouvons trouver des traces de l'expression d'avis amateurs. Pour distinguer ces formes et spécifier celles des plateformes qui nous intéressent ici, nous allons revenir sur quelques caractéristiques générales de trois types d'espaces d'expression des goûts et appréciations.

Les blogs tout d'abord sont des espaces d'auto-publication de « billets », associés ou non à des liens hypertextes (Pledel, 2007, Bois, Saunier, Vanhée, 2016). Ils peuvent introduire un échange avec d'autres internautes qui à leur tour peuvent y laisser des commentaires. Ils sont présentés de manière chronologique, du plus récent au plus ancien (ante-chronodatés). Ils sont associés à des bloggeurs reconnus par la communauté. Parmi eux, nous pouvons citer : *Mademoiselle Cordélia*, *Le coin d'Aristed*, *les vicissitudes de Samuel d'Halescourt*… Une déclinaison de ces blogs existe sous la forme vidéo sur *Youtube* : les *booktubes* qui regroupent plusieurs milliers d'abonnés, *Les lectures de Nine Gorman*, *Margaud liseuse*, *Bulledop*, *Lili bouquine*…

Les forums de discussion quant à eux sont généralement associés aux sites Web des grands médias mais également aux *pureplayer* d'informations en ligne ou aux sites de grandes institutions ou associations. Ils sont le lieu de l'expression de la parole profane qualifiée d'ordinaire (Marcoccia, 2003) et apparaissent sous différentes rubriques « expression en ligne », « réactions », « commentaires », « avis »… Quel que soit le nom donné à l'espace de l'expression, l'internaute va pouvoir dans ces lieux dédiés, transmettre une opinion, annoter une information ou tout simplement participer au débat et laisser une trace de ce qu'il pense ou ressent (le contenu importe peu dans certains cas).

Enfin, les plateformes communautaires sont des formes plus récentes. Dans le secteur des industries littéraires, il existe des plateformes généralistes : *Babelio*, *Critiques Libres*, *Booknode*, *SensCritique*, *Livraddict* et des plateformes spécialisées : *Bedethèque* (Bandes dessinées), *Coin BD* (Bandes dessinées), *Lecture academy* (ados), *ZazieWeb* (e-lecteurs). Cette liste n'est pas exhaustive et évolue régulièrement. Ces plateformes interviennent de plus en plus pour guider les lecteurs dans leur choix de livres.

La deuxième enquête menée par l'association Le MOTIF (Observatoire du livre et de l'écrit en Île-de-France) en 2012 sur le référencement naturel sur *Google* des livres et des auteurs a montré que les sites collaboratifs d'échange d'avis parmi lesquels sont en tête *Babelio.com* et *booknode.com,* interviennent de plus en plus, certes derrière les sites de e-commerce (*amazon.fr* et *fnac.com*) qui ressortent en premier dans les requêtes et devant les sites d'information associés aux grands médias comme *L'express* (*lexpress.fr*) et *Le Figaro* (*evene.fr*)[4] (Daval, 2012). À titre d'exemple, un requête

4. Ces sites d'informations générales ressortent assez fréquemment dans les requêtes car ils sont forts de dizaines de milliers de critiques dans leur base de données.

sur Google (13 mars 2017) menée sur le titre de Philippe Besson, *Arrête avec tes mensonges* publié aux éditions Julliard en 2017 fait ressortir en première position, l'éditeur, en deuxième la plateforme *Babelio*, en troisième et quatrième place deux sites de e-commerce, la *Fnac* et *Amazon*. Viennent ensuite, sur ce titre, les sites de médias de critiques littéraires. Ceci montre l'importance grandissante des plateformes telles que *Babelio* dans l'accès aux informations sur les œuvres. Un livre qui fait parler de lui sur les réseaux sociaux lui garantit un bon référencement sur *Google* (Cardon, 2010).

Le site *Babelio* repose sur un modèle économique spécifique. Pierre Fremaux interrogé sur ce point [5] explique que *Babelio* repose sur une double stratégie. Le site va par exemple offrir des espaces de visibilité aux maisons d'édition de manière assez classique, au travers d'événements notamment mais va également commercialiser des services d'enrichissement des outils aux bibliothèques.

Prenant modèle sur la plateforme américaine *Librarything* [6], le service *Babelthèque* loue à l'année, à une soixantaine de bibliothèques partenaires [7], un accès à trois types d'enrichissements possibles : le service « Recommandation et nuage de mots », le service « Critiques des internautes » et le service « Contenu multimédia ». Les bibliothèques peuvent souscrire à l'une de ces trois briques de contenus pour venir compléter leurs propres bases de données. Ainsi par exemple, en s'abonnant au service « Critiques », les bibliothèques peuvent récupérer les critiques postées par les internautes sur *Babelio* et les lier à leur propre catalogue numérique. En s'abonnant au service « Recommandation et nuage de mots », les bibliothèques font le choix de confier à la machine une partie des recommandations de lecture. Les abonnés à la bibliothèque peuvent, grâce aux nuages de mots, trouver des ouvrages susceptibles de les intéresser. La bibliothèque utilise alors le système de recommandation par mots-clefs proposé par *Babelio*.

On comprend que le modèle économique sur lequel repose ce projet est lié à la pratique de lecture d'une part et à la consommation des livres d'autre part. L'objectif est de faire consommer des livres, peu importe que le lecteur les achète ou les emprunte en bibliothèque.

1.2 La plateforme offre différentes entrées à la communauté d'inscrits

Babelio annonce en introduction du site être « un réseau social dédié aux livres et aux lecteurs. Il permet de créer et d'organiser sa bibliothèque en ligne, d'obtenir des informations sur des œuvres, de partager et d'échanger ses goûts et impressions littéraires avec d'autres lecteurs ». La dimension communautaire est donc au cœur du dispositif. Nous allons présenter dans cette partie descriptive ce qu'on y trouve comme contenus et fonctionnalités.

5. Pierre Fremaux, co-fondateur de la plateforme, entretien avril 2017.

6. *Librarysign* est à ce jour le plus grand réseau social de catalogage pour les amateurs de livres puisque le site rassemble 2 279 289 membres au 27 février 2018, 122 603 119 livres catalogués, 140 894 046 mots-clefs ajoutés, 3 236 448 critiques.

7. Josée Cabanis à Toulouse a été la première bibliothèque partenaire de *Babelio*. Le site héberge leur base de données, loue les nuages de *tags* et le service de recommandations sur les œuvres qui sont dans leur base.

Il faut noter tout d'abord que le site présente une multitude de rubriques offrant plusieurs entrées à l'internaute. Ainsi, le site croise divers aspects de l'activité liée aux œuvres : les sorties des livres (actualité littéraire), mais également les actualités des auteurs ou de leurs éditeurs. Il rend compte de ce qui fait événement copiant en cela les pratiques journalistiques habituelles.

La presse ainsi imitée l'est également dans la forme que prennent les billets. Les articles sont rédigés sous des formes journalistiques traditionnelles (entretien, information, compte-rendu). Ils sont pris en charge par les rédacteurs du site eux-mêmes. Les nombreux entretiens d'auteurs, très souvent filmés afin de donner du corps à l'information, permettent une mise en contact du lecteur avec l'actualité présentée. On va rencontrer l'auteur, l'écouter parler de son livre.

Mais c'est la dimension communautaire du site qui nous intéresse plus particulièrement. *Babelio* n'est pas un site d'information littéraire classique. Il s'agit avant tout d'une plateforme collaborative qui met à contribution les internautes préalablement inscrits. Les membres les plus actifs de la semaine sont mis en avant [8] ainsi que leurs différentes contributions à travers divers dispositifs :

- *Les listes :* il s'agit de bibliographies thématiques et personnelles créées et postées par les internautes. Par exemple, la liste « 8 mars journée des droits des femmes et littérature jeunesse » proposée par *Letilleul* comporte 33 titres d'ouvrages.
- *Les quizz :* il s'agit de questions rédigées par les internautes pour faire jouer les autres avec les livres et la lecture, les amener à tester leurs connaissances et leur culture. Par exemple, *Ondin062* a proposé un quizz de 10 questions sur le 37e tome d'Astérix, *Astérix et le Transilatique*. À la date du 3 novembre 2017, 44 internautes avaient participé à ce quizz testant ainsi leurs connaissances.
- *Les citations :* il s'agit de copies des passages préférés des internautes. Toutefois, ces citations ne font pas apparaître les guillemets.
- *Les critiques de livres :* la plateforme mentionne les lecteurs les plus actifs de la semaine en fonction du nombre de critiques publiées. Il est possible de voir également le nombre d'internautes qui ont apprécié la critique ainsi que la note attribuée sur une échelle de 5 étoiles. Ces critiques sont mises en parallèle sur la page avec les critiques publiées sur le livre dans la presse quotidienne, magazine ou spécialisée.

La dimension communautaire du site est fortement mise en scène. Ainsi dans l'onglet « Découvrir » (bandeau haut), le membre peut savoir qui sont les internautes connectés (onglet Découvrir > internautes). Il peut repérer qui sont les meilleurs auteurs de critiques, les meilleurs graphomanes (ceux qui en nombre de signes écrivent le plus), les meilleurs sélectionneurs de citations, les meilleurs *Babeliothécaires* [9] et ceux qui

8. Il s'agit des lecteurs qui publient le plus de critiques par semaine.

9. Les Babeliothécaires sont des titres, appelés insignes sur le site, définis comme suit : « Les Babeliothécaires sont les membres qui enrichissent et classent le contenu bibliographique sur *Babelio*. L'insigne est attribué après le calcul suivant : 10 points par ajout de vidéo pertinente, 5 points par résumé de livre ou biographie, 5 points par ajout de photographie ou couverture de livre, 5 points par association de noms

ont ajouté le plus de livres. Ces classements servent à mettre en avant le palmarès des internautes. La plateforme fait apparaître également la popularité des dépôts : « Les quizz les plus populaires », « Les listes les plus populaires » et insiste sur leur actualité : « Les dernières critiques », « Les listes à l'affiche »...

Enfin, le membre doit se sentir connecté aux autres. La plateforme lui propose une mise en lien de sa propre activité avec les activités des autres internautes. Il voit par exemple les *posts* (une critique, une citation, une mention dans un quizz ou sur une liste) qui portent sur les livres de sa bibliothèque [10] en précisant s'il s'agit de ses amis. L'internaute peut alors réagir en commentant le *post* (ajout de contenu) ou en l'appréciant (symbole du pouce levé). Il peut également recevoir des messages privés ou des demandes d'amis comme sur n'importe quel réseau social numérique. Renseigner son profil est donc une étape importante. À son inscription, chaque babélien doit paramétrer son compte et choisir dans une liste de 50 titres de livres qu'il a lu. La rubrique « Mes livres » renvoie à la bibliothèque d'archivage de l'internaute. Cet espace permet d'avoir une vision des titres qui sont référencés dans sa base bibliographique et de les retrouver facilement. L'internaute peut également faire évoluer à tout moment le statut de ses livres et les notes associées à son appréciation. L'intitulé de cette rubrique rappelle que le lecteur en est la principale cible mais également le principal énonciateur. Dans La rubrique « Profil » l'internaute peut retrouver ses propres données : les auteurs de sa bibliothèque, les étiquettes qu'il a utilisées pour la classification de ses livres et ses actualités [11] mais également tous les insignes associés au compte, les contributions qu'il a appréciées (critiques et citations), les livres en cours de lecture, ses cinq titres phares et les lecteurs proches de lui. Ces lecteurs sont suggérés par le système à partir des notes portées sur les livres montrant à quel point le règne de la métrique domine.

D'une manière assez classique dans le fonctionnement des plateformes, l'interface et les fonctionnalités croisent de manière complexe toute une série de modalités d'entrées, de classement, de repérage pour tenter d'accrocher une grande diversité de profils d'utilisateurs, mais aussi avec pour objectif d'articuler la dimension personnelle des pratiques autour du livre à une dimension sociale et collective.

1.3 Les informations stockées comme base des recommandations à venir

La plateforme est pour chaque membre un espace de stockage de ses références et un espace de recommandations de livres à lire. Les titres sont proposés à chaque internaute en fonction des données qui sont inscrites sur son compte et qui sont agrégées, associées et comparées aux données d'autres internautes. On recherche la proximité dans les centres d'intérêt des uns et des autres.

Il est nécessaire de préciser qu'il y a une différence entre une bibliothèque dématérialisée et la plateforme *Babelio*. Le site parle de « bibliothèque en ligne » alors qu'on ne

d'auteurs et 1 point par ajout d'étiquette de classement. » Les suffixes en -aire, utilisés pour les mots libraire et bibliothécaire indiquent un agent d'action. Le babélien devient ainsi un acteur du site, reconnu.

10. Les listes quizz qui apparaissent dans la fenêtre de l'internaute sont donc différentes d'un internaute à l'autre.

11. Les actualités sont toutes les actions menées sur le compte, que ce soit l'ajout d'un livre, d'une critique ou une modification de paramétrage de son compte.

peut y déposer aucun livre. Cet espace permet simplement d'avoir une vision des titres que l'internaute a eu entre les mains ou qu'il souhaiterait avoir. Le livre n'est pas accessible via la plateforme. Seuls les extraits retranscrits par les internautes eux-mêmes sont une trace indicielle du livre (au sens de Pierce) et les jaquettes une représentation de sa matérialité. Par conséquent nous préférons parler de liste bibliographique plutôt que de bibliothèque numérique. Chaque liste bibliographique déposée par les internautes vient enrichir la base de données de la plateforme. Elle est suffisamment renseignée pour permettre ensuite aux algorithmes du site d'opérer des tris et des associations afin de proposer des titres à lire.

Pour optimiser sa base de références, *Babelio* encourage les internautes à l'enrichir en lui donnant des règles simples d'usage bibliographique. Dans l'onglet « Ajouter des livres » le site explique qu'il est facile d'ajouter rapidement des livres dans sa base d'archivage grâce au numéro ISBN [12]. Ce dernier facilite le travail de tous les acteurs du marché du livre. Les libraires l'utilisent pour passer leur commande, les bibliothécaires pour gérer leur catalogue, et les babéliens peuvent s'en servir pour constituer leur bibliothèque de référencement d'ouvrages. Le site s'appuie ainsi sur les normes du secteur pour faciliter le travail amateur d'archivage et éviter ainsi les erreurs de référencement dans sa bibliothèque. Cette évolution dans la méthode d'archivage n'a été mise en ligne que récemment [13]. L'équipe de *Babelio* cherche en effet, à faire en sorte que la bibliothèque virtuelle de l'internaute soit une représentation la plus fidèle possible des livres qu'il a lu ou qu'il a eu entre les mains. Des écarts d'édition avec des jaquettes de livres différentes pourraient perturber l'internaute dans l'image qu'il a des livres qu'il connaît. Grâce au numéro ISNB, les images correspondent exactement aux titres édités rendant plus fidèle sa représentation.

Enfin, un espace sur la page d'accueil en haut à droite est dédié aux étiquettes [14] présentées sous la forme d'un nuage de mots de tailles différentes (en fonction de la fréquence du *tag*, c'est à dire sa popularité) et de couleurs différentes pour assurer la lisibilité de l'ensemble. Les couleurs sont associées à quatre grandes familles de contenu : la nature du livre, le thème, les références géographiques, la période. Elles doivent aider à s'y retrouver dans la base de données collaborative et faciliter la recherche d'un titre en fonction de ses centres d'intérêts.

Le site se situe donc au croisement d'informations prises en charge par l'équipe rédactionnelle du site et d'informations proposées par les internautes eux-mêmes. Il n'y a pas de hiérarchie entre les deux niveaux d'information. En revanche, « plus le site est renseigné par les internautes, plus les recommandations sont facilitées », explique Pierre Fremaux (entretien, avril 2017). En effet, le site se sert des associations et de la proximité des informations pour proposer des titres et suggérer des lectures à l'internaute. Les étiquettes sont un des indicateurs pris en compte par le dispositif de recommandation mais ce n'est bien évidemment pas le seul. Nous avons choisi de travailler sur celui-ci car il fait référence à des pratiques professionnelles anciennes,

12. Le numéro ISBN (*International Standard Book Number*) est différent d'un titre à l'autre et d'une édition à l'autre. C'est un numéro à 13 chiffres normalisé qui figure sur chacun des livres (décret n° 81-1068, du 3 décembre 1981). Il est attribué par l'AFNIL (Agence francophone pour la numérotation international du livre) et imprimé sur chacun des livres. Il change à chacune des rééditions.

13. Elle n'existait pas au début de notre étude.

14. Les étiquettes, ou *tags*, sont des mots-clefs qui permettent de classer les livres.

en usage chez les bibliothécaires, et nous renseigne sur les modalités de recherche d'informations des internautes qui peuvent conduire les professionnels de l'indexation à revoir leurs propres méthodes de classement.

2 Les pratiques de folksonomie : entre classification pour soi et partage communautaire

Dans cette partie, nous allons montrer que les pratiques de *tagging* facilitent le repérage dans sa propre liste de livres et qu'elles conduisent le site à proposer des suggestions à l'ensemble de la communauté de lecteurs.

2.1 Les principes de l'étiquetage et la nature des *tags*

L'étiquetage ou *tagging* est une pratique relativement récente (début des années 2000) qui consiste à associer des mots-clefs à un objet documentaire. Apparentée aux pratiques de folksonomie [15], en usage pour référencer, organiser et classer des contenus Web comme les blogs ou les sites Web [16], l'étiquetage permet de construire un système de connaissances afin de retrouver facilement des informations sur un objet.

Dans le cadre des organisations productives, cette organisation du savoir et des connaissances augmente la rentabilité et l'efficacité du travail. D'une part, il s'agit d'archiver les connaissances et de préserver les données dans le but de garder une mémoire de l'organisation (visée patrimoniale). D'autre part, le savoir ainsi produit doit générer du développement, de l'invention, de la créativité (visée innovante). Il s'agit d'une création active de savoirs nouveaux sur la base de données archivées (Grundstein, 2003, cité par Desfriches Doria, 2013).

Dans le cadre de *Babelio*, le dispositif invite à étiqueter ses livres. L'internaute associe des mots-clefs à ses ouvrages, ce qui va lui permettre de se repérer dans sa propre bibliothèque [17]. Le *tag* est sans contrainte formelle. Le lexique choisi est illimité et propre à chacun. Ainsi dans l'exemple du livre de Céline Couriol, *Les Vieux ne pleurent jamais*, aux éditions Actes Sud, les internautes ont utilisé les mots-clefs suivants : « littérature française », « littérature francophone » et « littérature africaine » pour qualifier le même livre indiquant ainsi qu'il n'y a pas d'unanimité sur le classement. Or il s'agit bien d'un roman écrit en langue française (et non traduit) d'une auteure d'origine camerounaise, installée en France et qui a demandé la nationalité française il y a vingt ans. Faut-il dès lors classer son ouvrage d'abord parmi les œuvres françaises, francophones, africaines ? Les internautes semblent hésiter et le choix retenu est fonction de l'univers culturel du *tagger*. Chacun propose ce qu'il veut en matière d'indexation collaborative.

15. La folksonomie consiste en l'élaboration d'une liste de termes, appelés *tags* ou étiquettes. Le terme est né de la contraction de folk et taxonomie. Elle se distingue d'un thésaurus qui établit des relations entre les termes.

16. *Flickr*, *Delicious*, *Technorati* et bien d'autres sites permettent de retrouver l'information rapidement à partir de mots-clefs. Ces pratiques ont été étudiées dès les années 2005 et fait l'objet de nombreux articles (voir références en bibliographie).

17. Il faut noter qu'au départ, les folksonomies servaient à faciliter sa propre recherche. Ce n'est que progressivement que la pratique est devenue communautaire.



Figure 1 – Exemple de fiche signalétique de livre sur *Babelio*

Pour le guider dans l'élaboration de ses mots-clefs, les explications qui lui sont données par la plateforme sont relativement succinctes et les exemples de *tags* peu nombreux. Il doit donc imaginer son propre index.

Les mots-clefs apparaissent ensuite associés au livre dans la fiche signalétique de la bibliothèque de l'utilisateur, aux côtés de la jaquette, du titre, de l'éditeur, de l'auteur, de la note donnée par l'internaute, du statut du livre (lu, à lire, en cours, pense-bête, abandonné, à échanger, emprunté) et du nombre de lecteurs ayant ce livre dans leur bibliothèque (voir figure 1).

Les *tags* sont librement définis et donc forcément subjectifs. Ils reposent sur les compétences langagières et les connaissances de celui qui les définit. Ils supposent une prise de distance par rapport à l'œuvre et une élaboration cognitive complexe (créer des rubriques, des classements). Ils sont marqués par le cadre culturel du concepteur. En effet, le *tag* ne repose pas sur une ontologie de l'objet, c'est-à-dire sur ses attributs [18], ses caractères intrinsèques, incontestables et exhaustifs. Il ne prétend pas dire l'objet, il rend juste compte d'un point de vue sur l'objet et sur son contenu à partir d'un de ses usages possibles.

Il est à noter que les *tags* sont essentiellement descriptifs. Certains vont résumer le sujet de l'œuvre (ce qui se passe dans le récit, les protagonistes impliqués ou les objets importants). D'autres vont qualifier le genre du document. Pour certains, le genre correspond à la nature de l'œuvre (roman, BD, essai, correspondance, nouvelles etc.), pour d'autres à la nature du récit (roman comique, roman d'aventure, roman d'amour, roman de science fiction, roman policier etc.).

Nous trouvons également des critères chronologiques qui renvoient soit à la date d'impression de l'œuvre soit à la période à laquelle le récit a eu lieu, ou encore des étiquettes temporelles par année de lecture ou par année d'écriture. Enfin des *tags* géographiques peuvent confondre les origines géographiques des auteurs et les lieux du récit.

Pour illustrer ces principes, nous avons pris l'exemple du roman de Alessandro Barrico, *Soie*, publié en 2001 chez Gallimard. Il s'agit d'une œuvre de fiction, que l'éditeur résume ainsi :

> « Vers 1860, pour sauver les élevages de vers à soie contaminés par une épidémie, Hervé Joncour entreprend quatre expéditions au Japon pour

18. Les attributs d'un document correspondent à ce qui lui est propre.

> acheter des œufs sains. Entre les monts du Vivarais et le Japon, c'est le choc de deux mondes, une histoire d'amour et de guerre, une alchimie merveilleuse qui tisse le roman de fils impalpables. Des voyages longs et dangereux, des amours impossibles qui se poursuivent sans jamais avoir commencé, des personnages de désirs et de passions, le velours d'une voix, la sacralisation d'un tissu magnifique et sensuel, et la lenteur, la lenteur des saisons et du temps immuable. »

Les étiquettes postées par les internautes sont les suivantes : adapté au cinéma / littérature / aventure / roman / roman historique / poésie / roman d'amour / sensualité / soie / vers à soie / route de la soie / amour / voyages / France / Asie / Italie / japon / littérature italienne / XX^{e} siècle / XIX^{e} siècle. Nous observons donc que ces étiquettes ont été choisies librement par les internautes, sans aucune suggestion préalable, selon les principes évoqués précédemment.

Si nous analysons plus en détail leurs contenus sémantiques, nous trouvons des mots portant sur le genre du document : roman / roman historique / littérature / littérature italienne / roman d'amour, d'autres faisant référence à l'origine de l'œuvre : Italie / XX^{e} siècle / littérature italienne et d'autres encore au contenu du récit : roman d'amour / voyage / vers à soie / route de la soie / XIX^{e} siècle... Les étiquettes utilisées par les internautes mélangent donc différents niveaux de classification puisque rien ne nous permet de faire la distinction par exemple, entre le livre écrit au XX^{e} siècle par un auteur contemporain italien (étiquette : XX^{e} siècle) et la datation du récit, il s'agit de l'histoire d'un marchand de soie au XIX^{e} siècle (étiquette : XIX^{e} siècle). Enfin certaines peuvent être imprécises à un moment donné. Lors de notre première consultation, nous avions repéré les étiquettes Japon et Chine. Le contrôle fait à posteriori sur signalement des internautes eux-mêmes, repris ensuite par l'équipe technique de *Babelio* a finalement supprimé l'étiquette Chine qui ne correspondait à rien par rapport au livre.

À partir de cet exemple, nous remarquons une évolution par rapport aux catégories thématiques des *tags* observées par Scott Golder et Bernardo Huberman (2006, cités par Durieux, 2010) sur le site *del-icio-us*. Dans leurs observations les auteurs avaient repéré que les *tags* pouvaient être classés en quatre grandes catégories : (1) le sujet de l'œuvre (2) le type d'œuvre (3) l'auteur de l'œuvre et (4) l'opinion par rapport à l'œuvre. Sur *Babelio*, l'opinion n'est pas reprise dans les *tags*, elle passe par les notations ou les critiques. Le *tag* sert plus à qualifier l'œuvre elle-même que la relation que le lecteur entretient avec elle. Ainsi les *tags*, indices pertinents pour soi, deviennent quasi inexistants sur *Babelio*. Rares sont en effet les *tags* privés, comme le *tag* Babeliodocs, par exemple, utilisés par des documentalistes scolaires pour créer une communauté de partage [19]. Dans ce cas, le *tag* n'apparaît pas dans le nuage de *tags* et n'est pas repris dans l'indexation de l'ensemble des références sur la plateforme pour un usage collectif et collaboratif des collections. Il est privatif.

19. Ce *tag* est utilisé par les enseignants et les collégiens afin de se repérer dans une liste d'ouvrages.

2.2 Au-delà des données, une mise en relation des œuvres

Les *tags* produisent des données sur les livres qui s'articulent et s'enchevêtrent pour produire des relations signifiantes en vue des suggestions de lecture à venir. Dans un article sur les *tags* du site *Flickr* (base de données de photographies numériques), Dominique Bouiller et Maxime Crepel (2013) expliquent qu'une base de données est composée à la fois de données mais également de métadonnées. Ce sont les métadonnées qui permettent de mettre en relation les données entre elles.

Il ne s'agit plus seulement d'associer un contenu sémantique à une œuvre (en l'occurrence un livre dans le cas de *Babelio*) mais de mettre en place des dispositifs relationnels grâce aux algorithmes pensés par la plateforme. Le *tagging* est donc non seulement un outil de classement (ce n'est pas le seul, il en existe de nombreux autres comme les listes « Les livres les plus populaires », « Les prix littéraires », « Les livres dans des listes des internautes ») mais c'est surtout un dispositif de mise en réseau des œuvres à partir d'une mise en réseau des individus produisant des métadonnées.

Ainsi, la classification qui résulte de ces *tags* amateurs a une finalité précise : elle guide la plateforme dans la suggestion des œuvres qu'elle propose aux internautes :

> « Ces classifications sont exclusivement guidées par la subjectivité des utilisateurs. Elles sont multiples et sans contraintes. Et elles servent à chacun des utilisateurs à classer leur bibliothèque personnelle. La mise en commun de toutes ces données permet de créer un ensemble synthétique d'informations qui pourra être exploité par un utilisateur tiers, n'appartenant pas forcément au réseau. »

Les responsables de *Babelio* pensent ici aux usagers des bibliothèques. Par ailleurs, nous pouvons lire l'explication suivante sur le site :

> « Les mots-clefs sont classés en catégories majeures [...] elles servent à améliorer les services de suggestions automatisées de lectures : ainsi deux titres ayant le même genre, le même thème et la même origine géographique seront ainsi considérés comme fortement similaires. » (*Babelio*, septembre 2010.)

Cette pratique singulière de la lecture des internautes intègre ainsi une dimension subjective beaucoup plus proche des usagers-lecteurs de la plateforme que les dispositifs de classification proposées par les professionnels. Elle s'éloigne de la seule approche documentaire du classement des ouvrages et c'est pourquoi elle est de plus en plus plébiscitée par les documentalistes et les bibliothécaires qui souhaitent répondre aux attentes de leurs publics.

2.3 Le *tagging* au service de la recommandation

Les mots-clefs des uns et des autres doivent enrichir la base de données qui sera accessible à tous. Nous passons d'une pratique de *tagging* pour soi à une pratique de *tagging* pour les autres. On parle d'ailleurs de *collaborative tagging* ou d'indexa-

tion sociale, par opposition au système d'indexation traditionnelle car comme nous l'explique Pierre Fremaux :

> « Nous avons mis en place des outils et des technologies qui permettent de retraiter l'information communautaire afin d'en supprimer les étiquettes et mots-clés n'ayant pas de valeur d'indexation pour une bibliothèque publique ou pour un libraire. »

Dès lors, nous ne nous situons plus dans une pratique de *tagging* justifiant la relation intime d'un lecteur à des objets archivés mais dans une pratique qui va justifier de futures orientations prescriptives. En effet, derrière la fonction d'archivage se profile une volonté de production de connaissances sur le document indexé. Il s'agit de permettre à la communauté d'organiser les connaissances produites sur les livres en vue d'une prescription à venir. Dans la rubrique « Aide » de la plateforme *Babelio*, les internautes sont d'ailleurs encouragés à enrichir le contenu associé aux œuvres. On peut lire :

> « *Babelio* vous propose des livres en fonction de vos goûts sur la page de recommandations [20]. Ceci est le fruit de savants algorithmes en fonction de vos lectures, de vos notes, des thématiques abordées etc. Pour améliorer ces recommandations, ajoutez encore plus de livres et ajoutez des étiquettes à vos livres. » (*Babelio*, consulté le 20 mars 2017.)

Dans ce contexte, nous comprenons que les étiquettes proposées par l'internaute servent non seulement à classer ses livres et à les indexer dans la base de données mais sont également utilisées dans les algorithmes qui vont participer aux propositions de lecture qui lui seront faites ensuite. Ainsi l'internaute, en classant ses lectures dans des catégories qu'il définit lui-même, propose un cadre sémantique à la plateforme qui va servir à orienter les prescriptions de l'ensemble de la communauté (nous y reviendrons dans la partie suivante).

Ce système de recommandation construit par associations de mots est, du point de vue de Pierre Fremaux, bien plus intéressant que le système classique fondé sur les notes. La recommandation construite sur la base de strates de notes suppose une masse critique de lecteurs pour permettre des rapprochements statistiques pertinents. En fondant un modèle de recommandation sur du contenu rédactionnel, il suffit de deux occurrences similaires pour que le système les prenne en compte. Ce filtrage sur du contenu est utilisé en complément de la première méthode, plus ancienne, construite sur l'appréciation des œuvres par notation. Il s'agit d'un système complexe d'organisation des connaissances autour d'un livre et de son lecteur. Les données produites par les *tags* vont en effet être retraitées par les équipes techniques de la plateforme et reclassées en quatre catégories : genre de l'œuvre, thème, lieux et nationalités, période historique.

> « Cette première structuration permet notamment de faciliter la lisibilité des nuages de mots-clés en rapprochant les étiquettes avec une proximité

20. Cette page se trouve dans la sous-rubrique Accueil qui est un onglet personnalisé.

> de sens. Mais en outre, elle est exploitable pour améliorer les services de suggestions automatisées de lectures : deux titres ayant le même genre, le même thème et la même origine géographique seront ainsi considérés comme fortement similaires. » (Le livre blanc de *Babelio*, 2010.)

Toutefois, ce dispositif pose la question de la pertinence du mot choisi par l'internaute. Il représente pour certains documentalistes un frein à la recommandation car la confier à un algorithme sur la base d'indexation prise en charge par des amateurs ne va pas de soi, surtout pour des professionnels dont c'est la spécialité. Toutefois, il peut être considéré par d'autres comme une forme d'enrichissement. Ainsi certaines bibliothèques n'hésitent pas à adhérer sans crainte à ces briques de contenu proposées par la plateforme, considérant qu'elles pourront les accompagner dans leur travail de recommandation des œuvres à leurs usagers. Cité sur le *Google Drive* promotionnel de Babelthèque, Jérôme Pouchol, directeur des collections et du système d'information documentaire de la MIOP (Médiathèque intercommunale Ouest-Provence) explique :

> « Dans un contexte de surabondance de l'offre, amplifiée par la mise en réseau des catalogues, Babelthèque permet à l'usager de choisir le mieux dans le trop, en complément des conseils prodigués par les bibliothécaires de la MIOP. Ce service participe aussi grandement de la refonte des catalogues, qui ne se limitent plus à de simples notices bibliographiques mais deviennent de vrais ressources enrichies. »

3 Le *tagging*, concurrence ou complémentarité de l'indexation professionnelle

Dans cette partie nous allons tenter de répondre à la question suivante : l'indexation collaborative peut-elle être considérée comme une nouvelle norme de classement ? L'étiquetage proposé par les membres de la communauté participe à l'organisation et au classement des livres dans des catégories. Ce classement est en dehors du système de classification professionnelle normé et en usage parmi les professionnels des bibliothèques et des centres de documentation. Il est élaboré par les internautes eux-mêmes et repris par les équipes techniques de la plateforme qui va traiter les occurrences avant de les agréger de manière automatique. Par conséquent, en demandant aux internautes d'étiqueter leurs livres, *Babelio* réfère à la pratique de l'indexation professionnelle. Même si les pratiques diffèrent car elles n'interviennent pas dans les mêmes espaces et pour les mêmes usages, les lecteurs qui participent au dispositif de *tagging* sont souvent familiers des recherches documentaires et peuvent être tentés parfois de s'inspirer des catégories qu'ils rencontrent par ailleurs. Cohabitent alors des systèmes de classification au langage codifié par les professionnels et « les folksonomies qui relèvent de langage documentaire libre » (Desfriches Doria, 2013).

3.1 Trouver un document dans une base de données

Pour trouver un ouvrage dans une base de données, dans une bibliothèque ou sur une plateforme collaborative, l'usager peut tout d'abord se servir de descripteurs formels simples comme le nom de l'auteur, le titre, la date de la publication, l'éditeur,

parfois même le format, ce que permettent la plupart des outils de recherches bibliographiques et documentaires ou ensuite rechercher le livre à partir de son contenu. Les professionnels de l'indexation peuvent utiliser des thésaurus hiérarchisés [21] ou des index de mots codifiés.

Le thésaurus est un ensemble de termes normalisés et ordonnés, une taxonomie. Il s'agit d'« un cadre d'organisation (et pas seulement documentaire) pour les ressources destinées à en permettre une représentation ordonnée » (Menon, 2007). Ils permettent de classer les documents selon une arborescence thématique à partir de descripteurs allant du général au particulier (par exemple, équidé > cheval > cheval de trait > cheval ardennais). Les relations sémantiques entre le document et le descripteur sont de nature hypernymique / hyponyque [22]. L'élaboration des thésaurus suppose précision et structuration sémantique. Chacun des termes utilisés a été préalablement validé et est reconnu par la communauté professionnelle.

Les index de mots-clefs permettent de qualifier les œuvres selon une arborescence différente. Le sujet du document est classé dans des catégories appelées des « vedettes matières ». Un même document peut comporter plusieurs « vedettes matière ».

Une vedette matière est une suite de mots-clefs hiérarchisés selon le principe suivant : thème principal (tête de vedette, mot utilisé pour exprimer l'essentiel du sujet) – sujet (subdivision 1, un seul mot pour limiter, préciser la tête de vedette et rendre compte d'un des aspects du sujet) – géographie (subdivision 2, mot pour exprimer le lieu) – chronologie (subdivision 3, mot pour exprimer la période envisagée dans le document) – forme (subdivision 4, mot qui porte sur la nature du document et sa présentation formelle). Ainsi l'exemple de « vedette matière » suivant : Femmes – travail – France – XX^{e} siècle – bibliographie.

Elle permet ainsi au documentaliste / bibliothécaire de retrouver facilement une œuvre et de permettre une recherche par mots-clefs. Généralement, les professionnels veillent à ne pas reprendre les mots du titre dans les mots-clefs qu'ils utilisent car ils sont déjà automatiquement indexés, sauf quand le livre de l'œuvre est l'objet d'une étude. Il est à noter qu'il y a rarement de classement thématique pour les œuvres de fiction. Les thésaurus thématiques sont plutôt réservés aux œuvres documentaires.

Dans les index par mots-clefs, chaque mot est unique. Des catalogues de mots existent pour aider les professionnels, le plus connu et celui qui est usité dans les bibliothèques publiques, est l'index RAMEAU [23]. Il est actualisé régulièrement [24] au fil des éditions. En cas de synonymie, le catalogue suggère des renvois à des mots de référence. Ces

21. Il existe de nombreux thésaurus bibliographiques élaborés par des experts. Leur point commun : tous visent à la structuration globale de l'archivage documentaire.

22. L'hyperonymie signifie qu'un terme général en inclut un autre, plus spécifique. Par exemple, le mot siège englobe le mot chaise. On dira que siège est l'hyperonyme de chaise. L'hyponymie renvoie à la relation inverse. Par conséquent, les catégories ontologiques mettent en évidence les relations hiérarchiques entre les descripteurs et permettent de classer les documents selon une arborescence complexe. Dans les folksonomies cette relation de dépendance hyperonymique n'existe pas.

23. Répertoire d'autorité, matière-encyclopédique et alphabétique unifié. Ce langage documentaire est édité par la BnF (Bibliothèque Nationale de France) et est la référence pour le classement matière de nombreuses bibliothèques publiques.

24. 7^{e} édition, 2016.

outils utilisés par les professionnels de la documentation sont caractérisés par des règles et des référents communs qui font loi dans les pratiques.

3.2 Le *tagging* sur *Babelio*, un dispositif sans structure ?

Du côté des dispositifs de *tags* communautaires en fonction sur le Web, nous observons que les sites qui les utilisent n'imposent pas aux internautes une structuration préétablie par des professionnels de l'indexation. C'est l'usager lui-même qui pense ses *tags* en langage naturel. Il n'est pas contraint par un mode d'organisation élaboré ailleurs par des professionnels de l'indexation (thésaurus ou vedette matière). Il s'agit d'une liste de mots sans hiérarchisation qui ne nécessite aucun accord préalable (autour de la norme qui serait acceptée par tous). L'internaute ne discute pas au préalable la pertinence du mot qu'il propose, il n'y a pas de glossaire préétabli et aucune acceptation préalable à obtenir pour l'usage de tel ou tel mot (Le Deuff, 2006). Ils prennent appui sur le langage courant des utilisateurs et sont par conséquent forcément polysémiques. Un même terme peut avoir un sens différent en fonction de l'utilisateur et du contexte social dans lequel il est utilisé.

Du point de vue sémantique, certains de ces mots apparaissent bien comme des sous-catégories de catégories globales permettant la création d'un début de thésaurus : voyage initiatique appartenant à la catégorie « Voyage [25] ». Tout est fait comme si la communauté toute entière établissait un thésaurus géant qui associe les mots les uns avec les autres en fonction de leur fréquence d'utilisation par les internautes eux-mêmes.

Cette structuration est essentielle pour faciliter ensuite la recherche d'un livre ou pour suggérer des conseils de lecture. Elle est prise en charge par les professionnels de la plateforme. En effet, les *tags* produisent des métadonnées qui permettront ensuite d'associer les données entre elles et de les comparer. Deux livres qui auront la même carte d'identité seront rapprochés et suggérés aux internautes grâce au système de recommandation numérique. Il est donc essentiel qu'il y ait un traitement de ces mots-clefs pour en faciliter leur exploitation par la machine [26].

3.3 Le traitement des tags par les techniciens de la plateforme

Toutefois, les mots-clefs proposés par les internautes sont réinterprétés par les équipes techniques de *Babelio* en vue de leur exploitation sur la plateforme. Certains seront utilisés par le système de recommandation, d'autres resteront au niveau de la classification privative.

Les opérations de retraitement par les équipes de *Babelio* passent par quatre étapes :

- *Le nettoyage.* Cette opération consiste à retirer du nuage de mots les étiquettes à caractère privatif. Il faut s'assurer que l'étiquette porte une signification pour l'ensemble de la communauté. Ainsi, « on ne va retenir que les étiquettes à valeur publique dans les nuages de mots et dans les recommandations, les

25. Ainsi, l'étiquette « Voyages » est associée aux mots suivants : voyage initiatique / voyages autour du monde / voyageurs / voyage en Italie / voyage en Russie / voyage en Asie / voyage aux États-Unis / voyages en mer / voyageurs français / voyage en Chine.

26. Entretien avec Pierre Fremaux, 6 avril 2017.

autres resteront affichées au niveau de la bibliothèque privée de l'internaute mais ne seront pas utilisées dans une recommandation » (Pierre Fremaux, entretien, avril 2017). Pour être susceptible de rentrer dans la catégorie des mots-clefs à usage public et collectif le mot doit être proposé au moins par deux internautes. C'est également à cette étape que les occurrences vont être retraitées pour ne retenir que les mots-clefs « canoniques » dont on peut penser qu'ils paraîtront naturels et instinctifs au plus grand nombre d'internautes et que seront corrigées les fautes d'orthographe et les abrégés remplacés.

- *La distribution dans des* clusters. Dans cette deuxième étape, les mots-clefs sont associés à des noyaux ontologiques, appelés *clusters*. Il s'agit de catégories d'appartenance. Le choix d'appartenance à des noyaux prend appui sur les performances des internautes eux-mêmes. Les équipes techniques font confiance à l'expertise des internautes, expertise qui est révélée par des badges. Les badges sont des insignes de reconnaissance que la plateforme attribue aux membres pour récompenser leur activité et valoriser certaines de leurs compétences, elles-mêmes repérées à partir de leurs activités sur le site. Ces experts sont donc bien connus des équipes techniques de la plateforme. Ils sont repérables par l'intensité et la fréquence de leurs interventions sur le site. Ainsi explique Pierre Fremaux, « si un internaute reçoit le badge "Amateur de mangas", les administrateurs de la plateforme sauront être attentifs aux mots-clefs qu'il utilise et lieront les mots "shônen", "shôjo", "seinen" au noyau ontologique "manga" ». Ces noyaux ontologiques sont exclusifs, ceci signifie qu'il n'y a pas pour le moment de règles de double appartenance. Ainsi le mot-clef « littérature allemande du 18è s. » appartiendra seulement à la catégorie « Littérature du 18è s. ».

- *L'attribution de la valeur prescriptive d'un mot.* Dans un troisième temps, les équipes techniques décident des mots à retenir dans leur système de recommandation. Certains mots-clefs seront donc ôtés de la base car ils n'apparaissent pas pertinents. Ils ne seront dès lors pas utilisés pour suggérer des livres. Par exemple, le mot « prix Nobel » n'est pas considéré par les administrateurs de la plateforme comme pertinent dans le guidage des choix de lecture des internautes parce qu'il n'apporte pas de plus-value. Il ne sera pas utilisé dans le système de recommandation car le prix Nobel est en lui-même déjà un système de recommandation. Dans la politique de la plateforme, les mots-clefs doivent être liés à la relation d'un lecteur au contenu littéraire et non à une distinction *ad hoc* donnée par une instance critique extérieure au site.

- *La classification catégorielle en vue de la représentation du mot dans le nuage de mots.* Enfin et en dernière étape, chaque mot-clef est associé à une des quatre catégories suivantes : genre de l'œuvre (1), thème (2), période (3), lieux et nationalités (4). Cette classification formelle doit « faciliter la lisibilité des images de mots-clefs en rapprochant les étiquettes par une proximité de sens » (*Babelio*, septembre 2010). Cette classification se rapproche des « vedettes matières » utilisées par les bibliothécaires à l'exception de la notion de genre qui n'est pas utilisée en indexation professionnelle. En effet, les bibliothécaires lui préfèrent la notion de forme englobant ainsi la nature du document (bibliogra-



Figure 2 – Exemple d'étiquettes relevées le 30 mars 2017

phie, dictionnaire, ouvrage jeunesse) et sa présentation matérielle, c'est-à-dire sa forme physique (manuscrit, DVD, Microfiches). Les trois autres catégories en revanche se rapprochent des subdivisions des vedettes matières : le thème (2) renvoyant à la subdivision sujet (index RAMEAU), la période (3) à la subdivision du même nom et les lieux et nationalités (4) à la subdivision géographie. *Babelio* intervient donc sur l'indexation collaborative et se rapproche des pratiques des professionnels en réorganisant les interventions des internautes (voir figure 2).

3.4 Concurrence ou complémentarité ?

Margaret Kips (2005, cité par Durieux, 2010) et de Valérie Durieux (2009), ont comparé les *tags* utilisateurs et les index professionnels. Elles ont constaté tout d'abord que des écarts importants existaient. Un quart des *tags* des descripteurs professionnels n'est pas représenté parmi les *tags* des utilisateurs. Par ailleurs, dans une étude de 2005 révélée par le *New York Times* (mars 2007) et citée par Olivier Ertzscheid [27] sur son blog, le Metropolitan Museum of Art de New York indique que « sur trente œuvres d'art indexées par les usagers, plus de 80 % des *tags* ne figuraient pas dans le vocabulaire documentaire utilisé par le musée ». La conclusion commune à ces études est donc que les *tags* des utilisateurs se différencient des indexations professionnelles et que le fossé sémantique est important entre le langage documentaire et le langage naturel.

Les *tags* appartiendraient donc aux *taggers* en dehors de tout cadre de référence professionnel. Ils évolueraient par ailleurs plus vite que les outils documentaires traditionnels. Sur *Babelio*, nous avons pu remarquer que ces mots-clefs changeaient régulièrement. En comparant les nuages de mots à deux semaines d'intervalle, autour du livre de Leonora Minao, *Crépuscule du tourment*, aux éditions Grasset (sortie pour la rentrée littéraire de septembre 2016), nous avons pu observer une évolution dans les mots-clefs présentés dans le nuage. Certains *tags* avaient disparu : « fiction », « émancipation », « compagne », « Afrique subsaharienne », « adultère » et d'autres étaient apparus : « magie », « violence », « mère », « domination », « épanouissement ». Les

27. O'Connel, Pamela Licalzi, « One picture, 1 000 tags », *The New Yorks Times*, 28 mars 2007.



Figure 3 – Étiquettes relevées le 20 mars 2017

catégories de classification sont donc des espaces mouvants qui évoluent au fur et à mesure des inscriptions sur la plateforme (voir figure 3)

Par ailleurs, les mots utilisés ne font pas référence à leur contexte d'énonciation même si *Babelio* permet aux internautes de décrire leurs *tags* pour palier cette absence de contextualisation énonciative. Ainsi, ils peuvent poster de petits textes explicatifs que chacun pourra ensuite compléter à sa guise. En cliquant sur le *tag* « Voyages » par exemple, nous pouvons lire en introduction, le texte explicatif suivant : « Le voyage est un déplacement hors de son domicile habituel. On retrouve dans les livres suivants à la fois des guides de voyage et des livres sur le thème du voyage. » (voir figure 4)

À partir du bouton « Ajouter des informations », chacun peut apporter un complément d'information à cette définition. Dès lors, il s'agit de constituer un système de référencement produit par les internautes eux-mêmes pour leurs propres usages, en dehors du système classique de qualification des œuvres et de leur archivage. Permettre à chacun de décrire les étiquettes facilite leur appropriation. Ainsi la communauté construit son propre système d'indexation et de classement des œuvres présentes dans le fond commun.

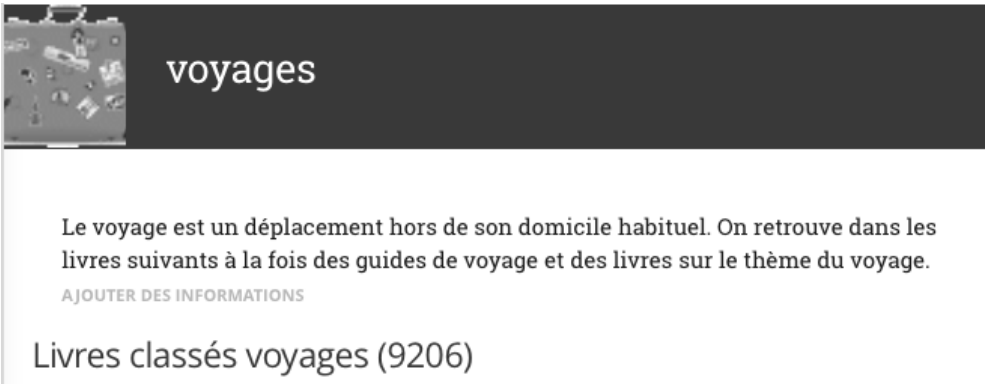


Figure 4 – Descripteur « Voyage », consulté le 31 mars 2017

Par conséquent, face à l'indexation des œuvres par les amateurs, les professionnels se positionnent. Pour certains et compte tenu de l'expansion d'Internet et de l'impossibilité pour les documentalistes, bibliothécaires et autres professionnels de l'information de traiter l'ensemble des ressources accessibles en ligne, le *collaborative tagging* est une bonne chose. Il permet un repérage plus aisé et un accès de tous à un index qui leur ressemble. En effet, il n'est pas utile de maîtriser les règles de l'indexation pour se repérer, pas d'apprentissage nécessaire. Les mots-clefs répondent aux besoins des utilisateurs et à la manière dont ils conduisent eux-mêmes leur recherche. Le *tag-cloud* (nuage de mots) facilite leur navigation. Guidé par les mots-clefs, l'internaute peut trouver des choses inattendues dans la base de données (Mathes, 2004). Certains bibliothécaires-documentalistes pensent même que l'indexation professionnelle est totalement inappropriée au Web et que par extension le *collaborative tagging* va modifier le rapport de l'homme à la connaissance (Shirky, 2005).

Toutefois pour d'autres, le langage utilisé étant libre et non contrôlé ne peut pas remplacer l'indexation traditionnelle. Différents problèmes sont posés autour de la polysémie notamment, de la synonymie, du nombre (singulier / pluriel) et des fautes d'orthographe (Durieux, 2010). Il n'est guère envisageable pour ces professionnels, de considérer que le *tagging* va remplacer l'indexation traditionnelle. Il serait pour eux, associé à une pratique de loisirs et non pas à un travail d'archivage et de repérage des livres dans un catalogue. Pour les professionnels des bibliothèques universitaires [28] qui utilisent le catalogue SUDOC notamment, les mots-clefs utilisés dans les index ne servent pas à faire de la recommandation. Ils sont uniquement utilisés pour pointer la ressource recherchée et faire la différence entre l'œuvre et l'étude documentaire sur l'œuvre. Actuellement, un travail en cours et des réflexions sont menées pour que les catalogues des bibliothèques soient visibles et apparaissent mieux sur le Web.

Enfin, il est important de noter que sur *Babelio* aucune obligation n'est faite aux internautes d'indexer les ouvrages de leur bibliothèque. En effet, nombreux sont les babéliens à ne pas attribuer de *tags* aux ouvrages de leur bibliothèque. Sur un échan-

28. Entretien mené auprès d'une bibliothécaire universitaire (mars 2018).

tillon de 223 internautes qui ont dans leur bibliothèque deux des ouvrages que nous avons déjà cités précédemment, Leonora Miano, *Crépuscule du Tourment*, Grasset et Céline Curiol, *Les Vieux ne pleurent jamais*, Actes Sud, nous avons repéré que 10 personnes sur 116 ont étiqueté le livre de Miano et 9 sur 107 (en date du 27 mars 2017) le livre de Céline Curiol soit un peu moins de 1 babélien sur 12. Ce pourcentage est relativement faible et indique que la pratique du *tagging* n'est pas encore centrale pour les babéliens. Le *tagger*-utilisateur de la plateforme est celui qui va éprouver le besoin de le faire en fonction de l'usage ultérieur qu'il fera de sa base de données ou en fonction de l'intérêt qu'il porte à ce genre de pratique.

Conclusion

Nous venons de voir avec l'étiquetage des livres de littérature générale sur *Babelio*, un exemple de mobilisation d'amateurs de livres pour classer dans des catégories des œuvres littéraires de façon à automatiser les suggestions que pourra leur faire en retour la plateforme. Les *tags* fonctionnent comme des métadonnées reliées entre elles aux côtés d'autres dispositifs comme les notes ou les avis. Ils sont réinterprétés grâce à des algorithmes complexes pour produire des associations de livres, associations qui se transformeront ensuite en suggestion de lecture par un jeu de mise en relation complexe des données les unes par rapport aux autres. Nous pouvons dire de cette pratique, érigée ici en dispositif (le *tagging*), qu'elle renvoie à une modalité d'appréciation du contenu des livres plus qualitative que la sempiternelle notation évaluative et moins complexe que la critique littéraire qui impose des standards d'écriture très normés. Elle est également un outil d'étiquetage social affichant la manière dont les individus lisent et interprètent les livres, outil qui s'avère utile aux acteurs du marché.

Dans la lignée des mouvements *pro-ams* (Miller, Leadbeater, 2004), l'internaute-babélien participe à la construction d'un réseau de conseil de lectures, une activité de production pour le modèle économique de la plateforme qui tente de le rendre rentable. Il travaille à alimenter gratuitement la base de données et enrichit les ressources en vue de la prescription automatisée. Cette mise au travail est économiquement désintéressée. Toutefois, pour encourager les plus actifs des internautes, *Babelio* a mis en place l'opération « Masse critique » qui permet d'offrir les livres mis à disposition par les éditeurs partenaires : « Pour participer, il [...] suffit de lire le livre et d'écrire une critique dans le mois qui suit sa réception. » (Babelio.com, 2018.)

Mais au-delà de cet aspect économique, cette activité de l'amateur en lien avec les évolutions numériques, pose la question de la notion même de métier et de compétences (Zarifian, 2004). L'exemple limité de cette étude nous a montré qu'il existait une forme de dissémination des pratiques documentaires à la sphère amateur. Avec l'apparition de nouveaux lieux d'expression et d'échanges, chacun peut se prévaloir d'une compétence. Le Web collaboratif pose la question de l'expertise, dans toutes les activités, y compris les activités culturelles que nous venons d'étudier. La sociologie des professions qui depuis une dizaine d'années travaille sur la manière dont se structure un marché professionnel et comment il évolue, a, de notre point de vue, peu pris en compte l'intervention des amateurs sur ce marché et les interactions entre les pratiques professionnelles et les pratiques amateurs.

Pour finir, s'il est aisé à l'issue de cette étude d'entrevoir les enjeux économiques que cela peut représenter et ainsi les motivations des acteurs du champ des industries culturelles à développer et soutenir de tels dispositifs, les motivations des individus qui passent du temps à construire leur bibliothèque, à partager des avis ou à classer des documents sont plus complexes à saisir. Tout d'abord, nous pouvons remarquer que le nombre d'internautes actifs est relativement limité. Alors que la plateforme compte des milliers d'inscrits, ils ne sont qu'une minorité à consacrer du temps à l'animation du réseau. Il y a un fort déséquilibre entre une minorité d'internautes qui produit du contenu (5 000 personnes environ) et la majorité de lecteurs qui consomme et qui occasionnellement va noter ses livres et encore plus exceptionnellement les *taggers*. Les motivations des individus qui passent du temps à construire leur bibliothèque, à partager des avis ou à classer des documents sont par conséquent complexes à saisir. Certains gros usagers (en nombre d'ouvrages dans leur bibliothèque) n'ont jamais apposé une seule étiquette. D'autres ont fait un usage très restrictif de ce *tagging* et seuls quelques-uns sont très prolifiques en la matière. Cet effort de catégorisation thématique n'est donc pas déployé par tous. En effet, construire un mode de classification suppose une démarche spécifique et une habitude qu'une étude complémentaire sur les usages pourrait éclairer. Par ailleurs, la masse importante des usagers non contributeurs est tout aussi importante pour la plateforme. Nous pouvons d'ailleurs supposer que c'est elle qui assure la motivation des forts contributeurs dans une logique d'audience. Pour que les contributeurs restent engagés et motivés, il est nécessaire que la masse des usagers soit importante et visible ; et inversement, afin que celle-ci reste présente et fidèle il faut que les contributeurs poursuivent leur activité sur la plateforme. Tel est certainement l'équilibre essentiel auquel travaillent en permanence les propriétaires de la plateforme.

Bibliographie

BOIS, Géraldine ; SAUNIER, Emilie ; VANHÉE, Olivier (2016), « La critique littéraire amateur sur les blogs de lecteurs. Une concurrence limitée de la critique littéraire professionnelle », *RESET*, disponible en ligne : http ://reset.revues.org/

BONACCORSI, Julia ; CROISSANT, Valérie (2015), « Votre mémoire culturelle » : entre logistique numérique de la recommandation et médiation patrimoniale. Le cas de *SensCritique* », *Études de communication*, p. 129-148.

BOULLIER, Dominique ; CREPEL, Maxime (2013), « Biographie d'une photo numérique et pouvoir des tags. Classer / circuler », *Revue d'anthropologie des connaissances*, vol. 7, n° 4, p. 785-813.

CARDON, Dominique (2010), *La démocratie Internet, promesses et limites*, Paris, Edition du Seuil, coll. République des idées.

DAVAL, Mathias (2012), *E-médiation. Le référencement des livres sur Internet*, Rapport d'études, Le Motif, 2è édition, disponible en ligne :

http ://www.lemotif.fr/fichier/motif_fichier/446/fichier_fichier_etude.ra.fa.rencement.na.2.pdf.

DESFRICHES DORIA, Orélie (2013), *La classification à facettes pour la gestion des connaissances métier : méthodologie d'élaboration de Folkclassifications à facettes*, thèse de doctorat sous la direction de Manuel Zacklad, CNAM.

DURIEUX, Valérie ; KERDELHUÉ, Gaétan (juin 2009), *Looking for health information on the Internet : can social bookmarking systems replace expert gateways ?,* Workshop of the European Association of Health Information and Libraries, Dublin.

DURIEUX, Valérie (2010), « Collaborative tagging et folksonomies. L'organisation du Web par les internautes », *Les Cahiers du numérique,* vol. 6, n°1, p. 69-80.

GOLDER, Scott A. ; HUBERMAN Bernardo A. (2006), « Usage patterns of collaborative tagging systems », *Journal of Information Science*, vol. 58, n° 8, p. 1175-1187.

GRUNDSTEIN, Michel (2003), *De la capitalisation des connaissances au management des connaissances dans l'entreprise. Les fondamentaux du* knowledge management. Disponible à l'adresse URL : http ://basepub.dauphine.fr/bitstream/handle/123456789/2620/INTKM0304_finalversion_modifiee.pdf ?sequence=2.

LE DEUFF, Olivier (2006), « Folksonomies : les usagers indexent le Web », *Bulletin des Bibliothèques de France*, vol. 51, n° 4, p. 66-70.

LEADBEATER, Charles ; MILLER, Paul, *The Pro-Am Revolution. How enthusiasts are changing our economy and society*, DEMOS, 74 p. Disponible en ligne : https ://www.demos.co.uk/files/proamrevolution final.pdf.

MARCOCCIA, Michel (2003), « Parler politique dans un forum de discussion », *Langage et société*, vol. 104, no. 2, p. 9-55.

MATHES, Adam (décembre 2004), « Folksonomies – cooperative classification and communication through Shared metadata », *Computer Mediated Communication*, disponible en ligne : http ://adammathes.com/academic/computer-mediated-communication/folksonomies.html.

MENON, Bruno (2007), « Les langages documentaires. Un panorama, quelques remarques critiques et un essai de bilan », *Documentaliste-Sciences de l'Information*, vol. 44, n° 1, p. 18-28.

PASQUIER, Dominique ; BEAUDOIN, Valérie ; LEGON, Thomas ; (2015), « *Moi, je lui donne 5/5 : Les paradoxes de la critique amateur en ligne* », Presse de l'École des mines, coll. I3.

PLEDEL, Iannis (2007), « Le blog face à la clôture informationnelle : la délibération au service de l'apprentissage », *Distances et* savoirs, volume 5, n° 2, p. 275-288

WIARD, Louis (2015), *La prescription littéraire sur les réseaux socionumériques de lecteurs,* Thèse de doctorat sous la direction de Bertrand Legendre et Christian Robin, Sorbonne Paris Nouvelle.

ZARIFIAN, Philippe (2004), *Le modèle de la compétence : trajectoire historique, enjeux actuels et propositions*, Paris, Editions Liaisons.

L'indexation communautaire en bibliothèque, livre blanc réalisé par Babelthèque (2010), Présentation *Babelio*, disponible en ligne : https ://fr.scribd.com/document/37162752/L-indexation-communautaire-en-bibliotheque.

Références des œuvres littéraires citées

Alessandro Barrico, *Soie*, Éditions Gallimard, 2001.

Philippe Besson, *Arrête avec tes mensonges*, Éditions Julliard, 2017.

Céline Couriol, *Les Vieux ne pleurent jamais*, Éditions Actes Sud, 2016.

Leonora Miano, *Crépuscule du Tourment*, Éditions Grasset, 2016.