

# À quoi avons-nous renoncé ?

→ À PROPOS D'UN ARTICLE D'ÉRIC DE GROLIER PARU  
DANS LE BULLETIN DES BIBLIOTHÈQUES DE FRANCE<sup>1</sup>

## YVES DESRICHARD

École nationale supérieure des sciences  
de l'information et des bibliothèques  
(Enssib)  
yves.desrichard@enssib.fr

Titulaire d'un DESS option informatique  
documentaire, Yves Desrichard  
travaille au département des Études  
et de la Recherche à l'Enssib. Il a  
été auparavant conservateur à la  
bibliothèque interuniversitaire de  
Montpellier, a travaillé à Médiadix, à  
l'Abes, à la bibliothèque interuniversitaire  
scientifique Jussieu, au CNC et à la  
DLL. Il est l'auteur en 2001 de Julien  
Duvivier, cinquante ans de noirs  
destins et en 2003 de Henri Decoin,  
un artisan du cinéma populaire  
(Durante/Bibliothèque du film). Il a  
publié en 2006 une nouvelle édition  
d'Administration et bibliothèques  
(Éd. du Cercle de la librairie). Et il a  
écrit de nombreux articles dans le BBF.

Vingt ans après... L'exercice peut être cocasse, salutaire ou anodin. Relire le *Bulletin des bibliothèques de France* avec ce recul est une proposition qui n'a de sens que si, d'une part, on veut bien replacer le texte dans son contexte (évidence aujourd'hui disparue, et qui vaut aussi pour les images, les sons, etc.), d'autre part, on n'adopte pas (en tout cas, pas d'emblée) une attitude surplombante, vaguement méprisante, entre le « *Je vous l'avais bien dit* » et le « *Quelle naïveté de prévoir que...* ». Inutile de préciser que l'exercice vaudrait dans tant de domaines, et pas forcément avec une telle ancienneté, qu'on s'étonne qu'il soit si peu pratiqué. Internet y est pour beaucoup – mais c'est une autre histoire.

## Le regard diachronique

Vraiment, c'est une autre histoire ? Pas si sûr. Car, en relisant « Taxilogie et classification : un essai de mise au point et quelques notes de prospective », écrit à l'aube du tout-informatique, on s'essaie sans vergogne à ce regard diachronique qui fait cruellement défaut. Non dans la recension de tous ces projets mirifiques et disparus dans les limbes de l'histoire des techniques documentaires (UMLS, Nephis, Smart, Intercolta, Tacitus...) qui se proposaient d'utiliser les possibilités de recherche informatiques, alors perçues comme sans limites, pour améliorer la recherche et l'identification de documents qui ne pouvaient être que textuels. Ni même dans l'ironie facile qui consisterait à placer en exergue la citation du plus qu'estimable Éric de Grolier sur le fait que « *Le livre électronique* » a commencé à devenir une réalité ».

Non, ce qui oblige à la réflexion, c'est de constater qu'Éric de Grolier n'envisage le développement du document électronique *que* par le biais du livre alors que (vingt ans plus tard, tout de même) celui-ci reste le dernier *bastion* à conquérir et – surtout – que ce sont de nouvelles formes de documents qui ont assuré la domination de la documentation numérique, certaines « traditionnelles » (revues) d'autres spécifiques (sites web). Il ne s'agit pas de s'en gausser, juste d'inciter à la prudence, au risque d'un procès en archaïsme : aujourd'hui est au tout-numérique. Mais que sera demain ?

Plus moderne et plus troublant encore, le fait que de Grolier se prononce pour la « *persistance de l'imprimé et du papier* », annonçant qu'il ne croit « *guère à cette paperless society dont [...] tant de futurologues américains nous avaient annoncé le prochain avènement* ». Ce faisant, il confond, ou feint de confondre, écriture, livre et support imprimé – l'erreur est commune<sup>2</sup>. Ce qui est plus intéressant, c'est qu'il se place d'emblée dans une posture où les bibliothèques risquent d'être « *transformées en musées de l'imprimé* », autrement dit qu'il n'envisage absolument pas qu'elles puissent se saisir de ces nouveaux types de support, dont le *nec plus*

1. Voir : Éric de Grolier, « Taxilogie et classification : un essai de mise au point et quelques notes de prospective », 1988, n° 6, p. 468-489.

2. Dont on trouve d'amusants échos dans les confusions journalistiques entre le papier électronique et les *e-books* façon Kindle d'Amazon, qui relèvent pourtant de techniques très différentes.

## Réflexions

**LES SYSTEMES** traditionnels de classification du savoir, qui reposaient sur une division disciplinaire des connaissances, semblent de moins en moins adaptés aux nouvelles exigences de la recherche d'informations, notamment sur banques de données. Les sciences de l'information ne sont d'ailleurs pas seules concernées par la taxilogie, ou étude scientifique des classifications, à laquelle contribuent d'autres disciplines, telles que la philosophie, la sémiotique, l'intelligence artificielle...

Présenter brièvement les grands systèmes de classification adoptés dans le monde et replacer les nouvelles tendances, qui se font jour dans les bibliothèques, dans le cadre plus vaste d'une recherche interdisciplinaire, tel est le propos de cette mise à jour concernant un domaine en pleine mutation.

Pourquoi distinguer classification de taxilogie ? La classification, comme activité et pratique humaine — pas seulement humaine d'ailleurs : tout être vivant exerce une activité classificatoire, et l'« attitude catégorielle » lui est indispensable pour orienter son action dans l'environnement — peut être l'objet d'une étude scientifique. Celle-ci, dénommée « taxinomie » par Durand (de Gros) en 1899, ou « classologie » par d'autres, constitue un champ de recherches interdisciplinaires, pour lequel je préfère le terme de « taxilogie ».

### COMPONENTES DISCIPLINAIRES

On pourrait énumérer une dizaine ou une douzaine de disciplines intéressées par les recherches taxilogiques<sup>3</sup> depuis la philosophie jusqu'à la science de l'information. Je me limiterai ici à quelques notes concernant celles où la recherche semble avoir été la plus active au cours des deux dernières décennies.

### Philosophie et « théorie des systèmes »

Wählin<sup>2</sup> regrette que les philosophes actuels, contrairement à ceux du XIX<sup>e</sup> siècle, s'intéressent peu aux problèmes de classification, et note qu'après Vannérus (214), aucun professeur de philosophie suédois n'a repris ce sujet. On trouve bien, en allemand, un livre de Rochhausen sur « la classification des sciences comme problème philosophique » (179), paru en 1968, mais il est vrai qu'il n'y a rien de comparable, depuis la Deuxième Guerre mondiale, à la floraison de « systèmes des sciences » entre la fin du XVIII<sup>e</sup> siècle et les années 1920. J'avais noté ce fait en 1974<sup>3</sup>. Dolby<sup>4</sup> a cherché à l'expliquer, sans peut-être insister suffisamment sur la raison probablement essentielle de cette désaffection : l'éclatement des disciplines traditionnelles et la prolifération des disciplines « mixtes », depuis l'astrophysique jusqu'à la sociobiologie.

En octobre 1971, un colloque mémorable, organisé par la Faculté de philosophie de l'Université d'Ottawa, cherchait à définir « les fondements de la classification des savoirs » (en anglais : *The conceptual basis of the classification of knowledge*, dont une traduction française plus précise aurait été *Les fondements conceptuels de la classification des connaissances*). La circulaire qui l'annonçait constatait que, si l'on disposait de « puissants modèles techniques de classification et de classifications très développées de diverses branches du savoir », il n'existait pas de « classification générale satisfaisante des connaissances » et que « les problèmes philosophiques fondamentaux posés par une telle classification n'étaient pas étudiés de manière adéquate<sup>5</sup>. » Un des objectifs des organisateurs du collo-

aussi de l'ethnologie, de la linguistique, de la sémiotique, voire de la philosophie.

Le grand débat qui sous-tend le texte peut (très grossièrement) se résumer dans la question suivante : « la connaissance précède-t-elle le langage ? » ou, traduit en langage bibliothéconomique, « peut-on concevoir une notion sans mots ? » (et si oui, que fait-on ?). Vaste sujet, qui excède très largement les efforts plus ou moins fastidieux déployés à l'époque, dans les associations de bibliothécaires et de documentalistes, pour le traiter. Comment ne pas percevoir que, dans un monde où l'audio et le visuel ont désormais la *primauté* sur les moyens d'information et de divertissement du plus grand nombre, le questionnement gagne en intensité, se force en résonance ? La collecte, le stockage, la gestion, la recherche et la restitution d'ensembles d'informations aussi vastes et aussi peu compréhensibles obligent à poser à nouveau ces questions : taxonomies, ontologies, folksonomies, *tags* et autres web sémantiques (on y revient) trahissent ces détresses conceptuelles. Vont-ils apporter des solutions décisives là où, vingt ans plus tôt, rien de tel ne se dessinait ? Bonne question.

Eric de Grolier  
Consultant scientifique  
Conseil international des sciences sociales

## TAXILOGIE ET CLASSIFICATION

UN ESSAI DE MISE AU POINT ET QUELQUES NOTES DE PROSPECTIVE\*

\* Les chiffres entre parenthèses, dans le texte et dans les notes en bas de pages, renvoient à la bibliographie en fin d'article.  
1. Cf. Eric de GROLIER (94) p. 21-26.

2. (216), p. 2. 19.  
3. (94), p. 38.  
4. (83), p. 187.  
5. (222), p. 7.

L'article d'Eric de Grolier paru dans le *BBF*, 1988, n° 6.

*ultra* de l'époque est de les appeler « hypermédias ». Vingt ans plus tard, tout le chemin a pourtant été parcouru... alors pourquoi désespérer de demain ? La mission d'une bibliothèque n'est pas dans les supports qu'elle acquiert (ou non) et qu'elle communique. Les peurs non avérées d'hier doivent soulager de celles d'aujourd'hui même si tout n'est pas transposable.

Ces apartés ne sont pas, cela dit, le sujet de l'article. Il est essentiellement question de la classification du monde et, accessoirement, de sa représentation sous forme de mots. La phrase qui précède n'est pas une confusion. Loin de se cantonner à la sphère documentaire, de Grolier, suivant en cela ceux qu'il cite<sup>3</sup>, va puiser ses références du côté de la sociologie ou des sciences de l'éducation, mais

3. Plus de 200 références pour un texte d'une douzaine de pages ! Beaucoup de rédacteurs du *Bulletin des bibliothèques de France* pourraient s'en sentir confus...



## Articuler

Ce qui frappe dans le débat tel qu'il est traité, c'est que l'auteur y convoque sans la moindre hésitation toutes les sciences qui peuvent avoir à faire avec : « On pourrait énumérer une dizaine ou une douzaine de disciplines intéressées », écrit-il. Et, sans embrasser si large, le propos se nourrit de philosophie, s'alimente de linguistique et de psychologie. Ce faisant, et dans l'évidence de cette démarche, il prouve que, vingt ans plus tôt, on se souciait mieux d'articuler les sciences de l'information et les autres sciences.

Autre remarque qui reste profondément juste, et qui est un peu le corollaire du *souci* de faire le lien entre les sciences de l'information et d'autres sciences, le fait que la taxonomie a été abandonnée par le champ des autres sciences, en d'autres termes, qu'elle est aujourd'hui « entièrement dominée par les mathématiciens et les informaticiens ». Plus que la présence des premiers, c'est l'alliance entre les deux qui étonne aujourd'hui. On a parfois l'impression, à lire certains discours, que l'informatique est devenue une science sociale, voire une science des comportements. Il y a vingt ans, ce type d'ambiguïté n'existait pas, n'avait pas lieu d'être. On ne s'était pas éloigné des présupposés algorithmiques (et donc mathématiques) de la science informatique. Il n'était pas question de qualifier de « réseaux sociaux » la création de liens hypertextes entre des pages écrites en html.

Un tel article aurait-il sa place aujourd'hui dans le *Bulletin*, alors même que la presse professionnelle est parfois sommée d'adopter une perspective utilitariste, voire égo-techniciste ? La question mérite d'autant plus d'être posée qu'elle n'est pas sans conséquence sur la mise en œuvre de la formation professionnelle, du rapport de chacun à son exercice et à son métier, et donc du rapport de chacun à l'autre, décideur, usager ou non-usager. Retenons la leçon (qui ne s'érige pas comme telle) et pensons *en transversalité* – dans les limites humbles et égoïstes de sa propre « érudition », mais en convoquant justement ses « marges disciplinaires », au risque assumé de sa maladresse.

Mais qu'en est-il – donc – vingt ans après, de tous ces efforts taxologiques des années 90 ? À lire le compte rendu, parfois fastidieux, fait par Éric de Grolier des projets alors importants menés dans les différentes parties du monde, on se prend à se demander où en sont les tentatives de classification des connaissances qui ne seraient pas

« polluées » par une vision mercantile, ou par une approche uniquement utilitariste (comment l'usager veut chercher) du sujet. S'ils existent encore, ces efforts semblent singulièrement modestes. Grisé par la possibilité de chercher partout, dans n'importe quoi, n'importe comment, on n'a plus l'ennui de penser/classer les choses, activité considérée comme fondamentalement conservatrice, en ce qu'elle induit (pense-t-on) non seulement la mise en perspective des productions mais, plus grave, une nécessaire hiérarchisation entre elles.

De fait, il y a un abîme entre l'importance accordée alors aux thésaurus dans la pratique documentaire et la réalité d'aujourd'hui. Il y avait le besoin, à l'aube d'internet, d'organiser les domaines de connaissance et leurs interrelations. Si l'outil documentaire semble désormais largement obsolète, faute de s'être traduit de manière convaincante dans l'aire informatique, la démarche d'organisation des domaines de connaissance et de leurs interrelations a-t-elle cessé d'être pertinente ? Comment l'affirmer alors qu'on peut multiplier les relations graphiques de représentation des résultats d'une recherche, dans des formes bien frustrées si on les compare aux possibilités offertes par les versions... imprimées des thésaurus. Ce travail conceptuel, qui avait (a) le mérite d'une modélisation des connaissances dans une discipline donnée, gagnerait à être valorisé, et non écarté, par

les gestionnaires à la recherche de techniques de plus en plus automatisées de formalisation de l'information et de ses contenus.

## Traiter

Car, si l'informatique a (indéniablement) fait d'immenses progrès pour ce qui est de l'analyse morphologique, voire de l'analyse syntaxique, du vocabulaire, qu'en est-il exactement de l'analyse sémantique, la plus conceptuelle bien sûr ? Les moteurs de recherche les plus utilisés, et dont on suppose qu'ils sont les plus performants, devraient plutôt nous convaincre d'une régression dans ce domaine. On utilise souvent les « méthodes statistiques fondées sur la plus ou moins grande probabilité d'occurrence et/ou de co-occurrence utilisées pour la recherche d'informations en ligne », qu'Éric de Grolier présente en notant qu'on les préconisait déjà... en 1958. Certes, les capacités de traitement ont changé, et dans de gigantesques proportions. Mais, dans la mesure

Le grand débat  
qui sous-tend  
le texte peut se  
résumer dans la  
question suivante :  
« la connaissance  
précède-t-elle le  
langage ? » ou,  
traduit en langage  
bibliothéconomique,  
« peut-on concevoir  
une notion sans  
mots ? »

où le nombre de documents à traiter (souvent redondants, ce qui était moins vrai dans le « monde papier ») a explosé, tout comme la complexité des formes, des supports et des niveaux de ces documents, où est, exactement, le progrès ?

De plus, vingt ans après, a-t-on vraiment avancé quant à l'adéquation entre le vocabulaire d'interrogation et le contenu du document ? Paradoxalement, l'effacement de la notion d'index par la plupart des moteurs donne l'illusion de la recherche constante dans le texte intégral, alors que, le plus souvent, seule une partie du contenu est indexée, sans qu'on sache comment. Comme le dit bien Éric de Grolier : « *Comme les techniques d'indexation automatique sont presque exclusivement fondées sur l'extraction de vocabulaires, il n'est nullement évident que ces techniques, si complexes soient-elles, puissent être efficaces en vue de fournir des représentations adéquates du contenu des documents.* »

On ne saurait mieux dire, et l'article fournit d'autres exemples de ces pertinences anticipées qui sont le sel de l'exercice rétrospectif. Ainsi, on est étonné de voir évoquée la méthode des « réseaux sémantiques », sur laquelle on souhaiterait en savoir un peu plus, puisqu'elle ne peut qu'irrésistiblement amener à celle de « web sémantique », cœur du futur web 3.0.

Hélas, l'auteur l'écarte de son propos en écrivant « *qu'il est cependant douteux que la théorie des réseaux sémantiques suffise à fournir une représentation adéquate des connaissances* ». Le doute pouvait s'appliquer à la modélisation d'informations qui, ne l'oublions pas, étaient essentiellement secondaires, descriptives ou analytiques. Avec l'accès direct aux documents primaires, nous sommes passés à d'autres schémas de représentation et, par conséquent, de recherche. L'important n'est plus de savoir ce qu'on cherche, mais ce qu'on trouve, autrement dit non plus de *conceptualiser* sa recherche en en déterminant la correspondance avec un cadre préétabli (ontologie ou théaurus), mais d'extraire de ce qu'on a trouvé les indices et les traces de son appartenance conceptuelle.

Autre prescience qu'il faut saluer, celle d'un universitaire américain ayant élaboré un « *système de recherche d'information automatisé [...] fondé sur l'extraction de termes pris, d'une part, dans les demandes de recherche des usagers, et d'autre part, dans les documents susceptibles de répondre à ces demandes* ». Cette fois, nous y sommes.

“Paradoxalement,  
l'effacement de  
la notion d'index  
par la plupart des  
moteurs donne  
l'illusion de la  
recherche constante  
dans le texte  
intégral, alors que,  
le plus souvent,  
seule une partie  
du contenu est  
indexée, sans qu'on  
sache comment”

## Renoncer ?

Arrivé à la fin de l'article, on sent bien que ce n'est pas – pas forcément – ce questionnement sur l'efficacité de la recherche – ou non – automatisée, dans un contexte de tout-numérique, qui en fait le prix et la nostalgie. Si on peut s'interroger sur la place qu'occuperait, aujourd'hui, un tel article dans le *Bulletin*, sur ce que cet article traduit et que nous avons peut-être perdu, c'est au fond un rapport privilégié à l'analyse de ce que nous conservons. La possibilité de voir l'univers à travers nos collections, et la volonté de traduire « *de manière concise, contrôlée et logique [...] ce qui a été énoncé verbalement dans les textes compliqués [sic] des livres et des articles* », comme l'indique un peu naïvement une documentaliste citée, qui souhaite (*resic*) « *maîtriser la complexité de notre univers* ».

Fortuite ou pas, la méprise entre l'univers et ce qui le décrit possède une *vertu dynamique*. L'univers peut encore être maîtrisé par ce qui le décrit à l'aube des années 90. Vingt ans plus tard, il n'en est plus même question. Nous avons largement renoncé à la mission de médiation, qui semble pourtant plus que jamais à l'ordre du jour, et plus grand monde ne va chercher dans ce qu'il lit, dans ce qu'il voit, dans ce qu'il entend, une connaissance de l'univers. Il ne s'agit pas de réhabiliter la perspective utilitariste, encore moins pédagogique, des collections, il s'agit de souligner que l'on ne pose plus, tout simplement, la question en ces termes. Nous avons renoncé. À quoi, je n'en sais trop rien. Mais nous avons renoncé. ●

Novembre 2008