

NUMÉRISATION ET LECTURE SAVANTE. LE PROJET DE COLLABORATION AVEC LE CENTRE INTERNATIONAL DE SYNTHÈSE AUTOUR DE LA REVUE DE SYNTHÈSE

Éric DUSSERT et Yannick MAIGNIEN

Extrêmement attentifs aux conditions matérielles d'organisation et de diffusion du savoir, Henri Berr et la *Revue de synthèse historique* se sont attachés à décrypter les éléments institutionnels et structurants du savoir. Dès 1911, dans les premières pages de *La Synthèse en histoire*, Henri Berr fait état des enquêtes entreprises sur l'enseignement supérieur de l'histoire. Il s'interroge sur l'organisation des archives, bibliothèques et musées¹, sur le travail bibliographique et les congrès internationaux. Préoccupé de « l'efficacité du travail », de « l'amélioration de l'outillage », Henri Berr cite longuement Pierre Caron et Philippe Sagnac² sur « les conditions de solidarité », de « collaboration » et de « travail collectif » qui doivent présider aux travaux de spécialisation puis de synthèse dans le domaine de l'érudition.

Près d'un siècle plus tard, l'informatisation des bibliothèques peut-elle être située dans une réflexion méthodologique procédant, toute proportion gardée, d'un projet d'interdisciplinarité retrouvée ?

LA CONSTITUTION DU FONDS NUMÉRIQUE

La Bibliothèque nationale de France constitue un fonds électronique de 100 000 ouvrages³ qui sera accessible sur ordinateur dans les mois qui suivront l'ouverture de la bibliothèque en octobre 1996.

1. Voir à ce sujet la série d'articles de Victor CHAPOT, « L'organisation des bibliothèques et des archives en France », t. XIX, 56 sqq., *Revue de synthèse*, 1909.

2. Pierre CARON et Philippe SAGNAC, *L'État actuel des études historiques*, Paris, Revue d'histoire moderne et contemporaine, 1902, 91 p.

3. Soit environ 30 millions de pages.

Ce fonds à caractère encyclopédique rassemble les sources documentaires dans plusieurs disciplines. On pourra le consulter au cœur même de la bibliothèque à partir de chacun des quatre départements thématiques tant au niveau recherche qu'au niveau libre accès destiné à un plus large public.

Si cette offre documentaire vient compléter les services traditionnels de la bibliothèque, elle présente cependant une forte spécificité qui tient à la nature électronique de ces nouveaux supports.

Le projet de numérisation, mené dans une perspective encyclopédique et patrimoniale, nécessite la définition de critères de sélection tant des ouvrages que des corpus et doit aussi tenir compte des finalités et usages.

L'opération de sélection des ensembles documentaires est délicate puisque les ouvrages sélectionnés le sont en fonction de leur rareté et de la difficulté d'accès tout autant que de l'importance intellectuelle et de la complétude du corpus proposé⁴.

Pratiquement, la première étape du programme a vu la constitution d'un fonds de référence classique dans chacune des disciplines offertes. Soit l'histoire, la littérature, la philosophie, les sciences humaines et sociales et l'histoire des sciences.

Si l'objectif initial est de 100000 ouvrages, le but est d'atteindre le seuil de 300000 unités documentaires, volume qui permettra de garantir une diversité suffisante des collections pour prétendre à un horizon encyclopédique.

Un objectif quantitatif d'une telle ampleur ne peut faire l'économie d'une réflexion approfondie basée sur le développement des usages et surtout sur les autres offres numériques qui ne manqueront pas de se développer à partir de fonds de bibliothèques ou d'universités dans les années à venir. Qui pourrait d'ailleurs prédire ce que sera dans cinq ans le panorama de l'offre électronique de documents, la diffusion des articles de revues, la constitution et le réseau des bases de données, les types d'outils et les conditions d'organisation d'ensembles virtuels de communication pour le monde scientifique ?

Pour l'heure, la finalité de ce programme est d'offrir, à travers le poste de lecture assistée par ordinateur (PLAO), une plus large accessibilité aux documents en favorisant la simultanéité des consultations et l'interrogation à distance.

Le poste de lecture assistée par ordinateur est l'élément central et nouveau de ce dispositif mis en chantier dès 1989 par Jean Gattegno, Alain Giffard et Bernard Stiegler. Entourés de nombreux chercheurs constituant

4. Une commission présidée par MM. Roger Chartier et André Miquel a évalué ce projet à la fin de l'année 1992 et proposé des orientations qui allaient également dans ce sens.

le groupe des « Grand lecteurs », ils ont permis l'émergence de la notion de lecture assistée par ordinateur — ou pour être plus exact de lecture/écriture assistée par ordinateur — et la conception générale du système.

La constitution du fonds numérisé ne peut évidemment être dissociée des finalités qui lui sont assignées et qui conditionnent parfois les choix qui sont faits. En effet, la démultiplication des accès ne serait rien si elle ne reposait sur l'ergonomie d'un système susceptible de répondre à une recherche d'efficacité, de productivité intellectuelles telles celles obtenues par la lecture assistée par ordinateur. Analysant les fonctions génériques du travail de lecture savante, ils ont permis le choix et l'émergence des automatismes d'assistance dont le but est d'offrir à ces tâches souvent répétitives des possibilités sans égales de mémorisation, d'organisation et de vitesse de réactivation des données.

C'est bien parce que le projet de la Bibliothèque nationale de France intègre ces finalités et en fait son principe directeur et régulateur, que la sélection des ouvrages procède d'un nouvel encyclopédisme.

En amont, les ouvrages sélectionnés ne le sont pas en fonction de leur appartenance à une classification disciplinaire (celle déterminée par sa localisation physique dans l'un ou l'autre département thématique de la bibliothèque) mais au regard de nombreux critères intellectuels croisés : l'appartenance à un ensemble thématique et/ou chronologique, les contextes de production, l'histoire du livre, l'interdisciplinarité des auteurs, l'intérêt potentiel ou éprouvé pour la recherche, ou encore la continuité monographies/périodiques comme c'est le cas pour « L'Évolution de l'humanité » et la *Revue de synthèse*.

Bibliothèque physique et bibliothèque virtuelle

Plus que l'appartenance arborescente à une classification — comme celle bibliothéconomique de Melvil Dewey — c'est l'intertextualité multiforme dans laquelle peut s'inscrire un ouvrage qui justifie son élection au sein de la collection numérisée.

De fait, alors même que Michel Foucault n'avait vraisemblablement pas en tête les possibilités en devenir de l'informatique textuelle, *L'Ordre du discours* est l'outil programmatique qui rendrait le mieux compte de l'enjeu du projet de la Bibliothèque nationale de France. Car il s'agit bien de porter un double regard sur les textes indépendamment de leur matérialité et de la singularité de leur contenu dès lors qu'ils appartiennent à cet autre support qu'est l'espace électronique.

Ordonnateur, ordinateur de textes, l'informatique permet presque — voire autorise — le démembrement des volumes, reliures, séries, collections pour bâtir avec ces matériaux une recomposition à partir de nouveaux liens, de

nouvelles circulations hypertextuelles. Opérations qui toutes créent des éditions ou des bibliothèques virtuelles et font apparaître des généalogies, des archéologies, des volontés de savoir non directement explicites.

Les stratégies de recherche au sein de la collection

Pour autant, la question se pose de savoir en quoi la mise à disposition d'un fonds électronique favoriserait une approche interdisciplinaire ? La formulation de cette question est aussi évidente qu'en est malaisée la réponse.

Outre les remarques précédentes, la réflexion peut s'orienter ainsi :

Au-delà du caractère exponentiel de la production des connaissances, les volumes énormes des masses textuelles envisagées dans le cadre de notre projet ou de l'offre d'information sur des réseaux ouverts comme Internet posent la question de la recherche d'information pertinente.

Que veut dire « retrouver » un ouvrage ? Re-trouver ce que l'on connaît déjà ou, au contraire, inventer (au sens de découvrir), sauver de l'oubli un document dont on a besoin sans le connaître dans le cadre d'une recherche ? C'est bien sûr ce second cas de figure qui est intéressant : l'instruction de la stratégie de recherche heuristique est l'une des questions majeures du projet de numérisation de la BNF. « Savoir ce que l'on cherche, c'est comprendre ce que l'on trouve », disait Bachelard. On sait depuis que ce savoir préalable est un horizon de complexité, un spectre de diversité et de sources multidisciplinaires, et que le résultat de la requête peut être aléatoire, non déterministe et pourtant compréhensible.

Toute l'organisation des disciplines et le classement bibliothéconomique reposent aujourd'hui sur la notion de « document papier ». À savoir unité physique mesurable, fiable, sérialisable, identifiable et décrite dans des notices bibliographiques. Pour se repérer physiquement dans la bibliothèque, on doit savoir où se localisent les sources⁵. Leur appartenance à un champ disciplinaire est donc présumée, d'autant que leur classement repose sur ce rattachement « universel » et préétabli. Dans le cas du libre accès, la pluralité d'appartenances disciplinaires de certains documents nécessite de les présenter dans plusieurs salles à la fois, et par conséquent de les doubler. Pour les ouvrages dont l'accès est réduit, la cote — fruit d'une typologie disciplinaire et thématique — permet au magasinier de retrouver le livre pour le communiquer au lecteur.

Dans une bibliothèque électronique, l'ouvrage et son classement sont

5. C'est particulièrement le cas de la topologie de la bibliothèque de Tolbiac dont l'architecture figure nettement l'organisation thématique en départements disciplinaires.

deux choses indépendantes. L'ouvrage appartient à un fonds indifférencié dont les mobiles et les stratégies d'acquisition ont disparu. Il est clair que le lecteur doit pouvoir disposer, lorsqu'il ignore à quel ouvrage se référer, d'instruments pour instruire sa requête. Certes, les données locales du catalogue lui indiqueront si celle-ci porte sur un exemplaire papier et/ou micrographié et/ou numérisé dont le lecteur exploitera la description bibliographique. Mais on sait à quel point faire un choix en ne disposant que du titre ou de l'auteur peut être insuffisant. Aussi, sauf à charger en aveugle de nombreuses sources non déjà repérées, le lecteur est devant son ordinateur comme devant une boîte noire, un écran sans profondeur.

L'instruction de sa recherche — la terminologie contemporaine userait du terme « navigation » qu'utilisait déjà Leibniz — suppose une cartographie, même sommaire, de ces *terrae incognitae* qui en permet l'exploration. De fait, c'est toute une hiérarchie d'aides documentaires qui peuvent être convoquées à la condition que ces différentes couches d'information intermédiaire soient elles-mêmes organisées. Pour être appréhendables, ces niveaux de description doivent s'inscrire dans un espace, une visibilité, risquons le terme : une synthèse, quand bien même elle serait provisoire ou précaire — une classification mobile, dirait Leibniz — mais une synthèse qui sauve de la dispersion, de la perte de sens.

On sait que le marquage, non seulement des formes logiques des contenus⁶ mais aussi des balisages des étapes d'une recherche elles-mêmes⁷ est une des questions majeures des recherches documentaires dans le domaine de l'hypertexte et de la structuration.

Cette question qui associe à la visibilité synthétique image et notions liées dans des arbres ou des cercles de savoir, était dès avant la Renaissance un enjeu majeur des techniques et des arts de la mémoire⁸. La problématique de « la restitution universelle » se posait déjà à Leibniz dans *De l'horizon de la doctrine humaine* alors qu'il abordait sa réflexion sur la finitude de la langue et sur la marque comme signe mémoriel : « Les mots ne sont pas seulement des signes de ma pensée pour les autres, mais aussi des marques de ma pensée pour moi-même. »

Opérer une meilleure visibilité sur de grands corpus de connaissance, tel

6. Les outils les plus appropriés semblent être les définitions de type de document (DTD) SGML (*Standard Generalized Marked-up Language*) utilisés d'abord dans l'édition et maintenant dans la description de tout document (*Text Encoding Initiative*).

7. Les logiciels du type Hytime ou HTML sont actuellement les meilleurs éléments de réponse à ce genre de préoccupations, même dans la navigation au sein de réseau comme Internet.

8. Cf. France YATES, *L'Art de la mémoire*, Paris, Gallimard, 1975, 440 p. et Paolo Rossi, *Clavis universalis. Arts de la mémoire, logique combinatoire et langue universelle de Lulle à Leibniz*, Grenoble, Jérôme Millon, 1993, 272 p.

est ce que permet l'ordinateur : grande recension de textes via l'indexation, l'annotation, le classement de ces connaissances et les possibilités de réactiver de façon automatique le capital d'un travail entrepris de longue haleine.

Pour faire écho à Henri Berr, on pourrait dire que l'informatique permet de mettre à sa place la synthèse d'érudition et de laisser toute liberté à la synthèse véritable, à l'histoire-science de se déployer. Une maîtrise synthétique des savoirs n'est peut-être plus de mise actuellement — mais l'était-elle au début du siècle ? — devant l'explosion des sciences hyper-spécialisées, du moins des outils informatiques de navigation hauturière dans l'« hyperdocument » sont-ils maintenant disponibles...

Henri Berr avait encore l'idéal scientifique que la vérité doit être énoncée en loi. « Le chaos n'a pas d'histoire, pas plus qu'il n'a de loi », dit-il dans *La Synthèse en histoire*. Nous savons au contraire que nous devons compter avec lui, intégrer l'instabilité, l'incertitude, la complexité et que rien n'inscrit nécessairement le changement dans un « développement ». Aussi, les données encyclopédiques actuelles requièrent-elles des précautions, des constructions de méthodologie précaires et révocables.

LE PROJET DE COLLABORATION AVEC LE CENTRE INTERNATIONAL
DE SYNTHÈSE AUTOUR DE LA REVUE DE SYNTHÈSE

L'une des raisons qui nous a conduits à cette collaboration pouvait être de conserver la proximité que le Centre de synthèse entretient avec la Bibliothèque nationale, rue Colbert. C'est là l'occasion de réchauffer les ardeurs centenaires de ces deux nobles personnages, ardeurs dont Olivier Corpet disait au début de ce colloque qu'elles étaient quelque peu assagies. Éric Brian est l'initiateur de ce projet, et il lui reviendra de développer les raisons de fond de cette entreprise, et de soumettre ce point à vos débats. Pour notre part, nous voyons quelques bonnes raisons à entreprendre avec le Centre de synthèse un travail d'envergure.

Le but des deux partenaires est de mener à bien, dans le cadre du projet de numérisation de la BNF, la saisie en mode texte de la *Revue de synthèse*, d'abord, puis de ses émanations. En premier chef, seraient traités les éléments éditoriaux structurants⁹ des différentes séries publiées sous l'égide du Centre. En parallèle, il s'agirait de structurer de façon raisonnée les différents sous-ensembles pertinents, ensuite d'aller vers une saisie intégrale

9. Tables des matières, bibliographies principales, notes critiques, index d'auteurs et d'articles, etc.

en fonction des différents besoins de recherche d'occurrences, thématiques ou statistiques par exemple.

L'idée est bien de faire de la *Revue de synthèse* un chantier de numérisation exemplaire, où l'approche et les besoins des chercheurs piloteraient une saisie structurée des textes. Pour les chercheurs associés à ce projet — et cette décision leur revient —, la mise à disposition progressive sur informatique de la revue pourrait accélérer et permettre d'approfondir le traitement des questions interdisciplinaires que vous avez abordées dans ce colloque.

Disposer à terme d'un tel corpus structuré et en mode texte constitue pour la Bibliothèque de France une des couches documentaires majeures dont nous parlions plus haut. Dans son exploration du fonds numérisé, le lecteur y trouvera une *visibilité* sans égale et une aide d'orientation, de sélection et d'accès à un pan essentiel des sciences de l'homme en ce qu'elles comptent de courants de pensée, de débats : orientation bibliographique sur les auteurs, les collaborations, les institutions qui ont fait la vie culturelle et scientifique française pendant près d'un siècle.

En concourant largement à la vitalité des réseaux de chercheurs par l'évaluation critique et les recensions bibliographiques générales qu'elle a su faire, la *Revue de synthèse* est sûrement l'une des instances qui a le mieux mis en relation acteurs, auteurs et disciplines des différents champs du savoir. De ce fait, établir grâce à l'informatique une structuration des différents champs scientifiques abordés, des diverses typologies d'écriture peut aboutir à une cartographie documentaire extrêmement utile pour interroger très largement les productions éditoriales de cette période.

En effet, à partir d'un tel corpus numérisé, pourraient être plus explicites les relations, les liens avec de nombreuses institutions savantes, écoles de pensées, rivales, conjointes ou parallèles et les publications concernées : *L'Année sociologique*, la *Revue de métaphysique et de morale*, la *Revue d'histoire des sciences*, l'*Encyclopédie française*, les *Annales*, etc.

Le programme de ce chantier reste donc à écrire.

Nous tenons à remercier encore MM. Dominique Bourel et Éric Brian, Mme Jacqueline Pluet-Despatin, les éditions Albin Michel et l'IMEC pour nous avoir offert de présenter notre travail et pour les perspectives de collaboration qui devraient assurer à la *Revue de synthèse* un centenaire ouvert sur l'avenir*.

* Lors du colloque, l'allocation de Yannick Maignien a été suivie d'une démonstration des fonctions génériques du poste de lecture assisté par ordinateur effectuée par Jean-Didier Wagner et Éric Dussert : évolution des liaisons thématiques et chronologiques, surlignage, annotation, balisage, création de classe de remarques, de liens, des chemins de lecture et des unités de découpage, passage du mode Image en mode Texte, construction de dictionnaires et d'index, etc.