

Introduction :

comprendre et maîtriser la redocumentarisation

1. PENSER LE DOCUMENT NUMÉRIQUE

Pratiquement plus aucun document contemporain n'échappe à un passage par un support électronique. Le contraste est très fort entre la stabilité relative qui a prévalu jusqu'ici et la bascule soudaine dans une forme radicalement nouvelle d'organisation et de conception des documents. Il serait pourtant erroné de considérer, dans une sorte de déterminisme technologique, que le numérique est seul à l'origine des transformations documentaires actuelles. Force est de constater que les mouvements repérables dans les multiples dimensions du document, comme dans les déplacements des contextes de médiations sont d'une telle ampleur que l'on doit les relier à des mouvements sociaux plus profonds.

Pour éclairer la relation entre le numérique et le social, on peut faire un parallèle entre le rôle du document-papier imprimé dans l'émergence des sociétés modernes et celui du document numérique dans les transformations sociales auxquelles nous assistons. C'est ce que nous appelons la *documentarisation*. La première documentarisation a accompagné l'industrialisation, la mise en place de l'État au sens moderne ou encore, parmi bien d'autres illustrations possibles, la montée des sciences positivistes. Une figure comme le belge Paul Otlet et sa tentative de fonder une « documentologie » au début du XX^e siècle est emblématique du besoin de rationaliser la prolifération documentaire qui s'est installée. Aujourd'hui certains pensent que nous sommes entrés dans un nouveau processus de modernisation. De la même manière, le document, en basculant dans le numérique, se transforme, accompagne et amplifie ces changements de plus grande ampleur. Tim Berners-Lee et sa

16 LA REDOCUMENTARISATION DU MONDE

volonté de développer un « Web sémantique » joue un rôle comparable à celui de Paul Otlet un siècle plus tôt.

Nous assistons, sans doute, bien à une *redocumentarisation*, c'est-à-dire une nouvelle forme de documentarisation qui reflète ou tente de refléter une organisation post-moderne de notre rapport au monde, repérable aussi bien dans les sphères privée, collective et publique. Comme dans la précédente modernisation, le document participe au processus et y joue même un rôle clé, mais il s'est transformé au point que l'on peut se demander s'il s'agit encore de la même entité.

Le développement des nouveaux standards, l'invention d'outils, sont réservés à une dialectique entre des experts très pointus et des industriels à l'affût de nouveaux créneaux de rentabilité. Et les succès résultent d'une prise en main très rapide par les usagers, facilitée par le caractère ouvert du réseau et par une diffusion massive et instantanée des informations.

Malheureusement, les chercheurs universitaires se trouvent souvent débordés par des développements issus d'experts autonomes et d'entreprises « technoréactives ». Et les institutions publiques sont souvent prises au dépourvu entre des mouvements puissants et complémentaires qui les ignorent tout aussi souvent. Un décalage s'est ainsi installé entre une dynamique technico-industrielle, d'autant plus sûre d'elle-même qu'elle peut se prévaloir de succès spectaculaires, un mouvement social spontané et multiforme et la compréhension et la maîtrise sociale des phénomènes. Les manifestations de ce hiatus sont très nombreuses. Pour s'en tenir à quelques illustrations prises dans l'actualité, citons la polémique entre les bibliothèques, les éditeurs et la firme Google, les positions contradictoires des défenseurs de la propriété intellectuelle et des promoteurs des échanges *peer-to-peer*, les tenants des systèmes d'archives ouvertes et les éditeurs commerciaux dans la publication scientifique, la « blogosphère » et les journalistes ou les atteintes à la vie privée, etc.

L'objectif du Réseau thématique pluridisciplinaire « Documents et contenu : création, indexation, navigation » (RTP-DOC) du CNRS a été et est toujours de réduire ce décalage en tentant de mieux comprendre et de maîtriser la redocumentarisation explosive et désordonnée actuelle et en permettant aux institutions publiques d'y prendre toute leur place. C'est ainsi que, en parallèle au travail de réflexion et de rédaction collective ayant abouti à la publication des trois articles de Roger T. Pédaque, les travaux du réseau se sont focalisés sur des points précis dans le cadre d'Actions spécifiques puis d'ateliers.

2. PROBLÉMATIQUE

Une façon simple de repérer les travaux sur le document numérique est de les regrouper en *trois catégories* par rapport à leur angle principal d'approche : la forme ou le signe (F), le texte ou le contenu (T), le médium ou la relation (M). Il est apparu au cours de notre travail collectif de réflexion que ce balisage des recherches permettait aussi de préciser la notion de document et, sans doute, de bâtir une méthodologie pour l'analyse et le développement, pouvant, par exemple, se concrétiser dans l'élaboration de guides de bonnes pratiques.

Dans cette perspective, un document peut être défini comme la représentation d'une vérité partagée au-delà du chaos (le silence et le bruit), de la cacophonie (la confusion et le sensible) et de l'oubli (l'intime et l'éphémère). Ainsi, les modalités anthropologiques (lisibilité-perception, forme/signe), cognitives (intelligibilité-assimilation, texte/contenu) et sociales (sociabilité-intégration, médium/relation) doivent non seulement être pertinentes prises chacune séparément, mais encore être cohérentes entre elles. S'il ne peut être « vu » ou repéré, « lu » ou compris, « su » ou retenu, un document n'est d'aucune utilité. Dans la mesure où il s'agit d'un artefact, on peut même dire qu'il n'existe pas comme réalité textuelle, mais seulement et éventuellement comme trace iconique.

2.1. Forme ou signe

Dans la première entrée sur le signe ou la forme, trois communautés de chercheurs en informatique se croisent sans toujours bien se connaître : les traiteurs d'image, les éditeurs numériques et la communauté du Web. Des travaux du PARC à Palo-Alto¹ au consortium W3C², le chemin parcouru en quelques dizaines d'années est impressionnant. Même si les chercheurs sont partis de problématiques différentes, ils convergent nettement aujourd'hui dans la compréhension des structures logiques du document et dans les interrogations sur les formes perceptibles. Le travail principal concerne les formats, au sens informatique.

Le numérique a déplacé la question du support du document, qui en assurait la stabilité grâce à la fixité de l'inscription, vers la problématique de sa structure. La popularité croissante du langage XML marque une étape mais elle laisse encore largement ouverte la problématique de la

¹ <http://www.parc.xerox.com/about/history/default.html>

² <http://www.w3c.org>

perception, de la visualisation et de la lecture. En effet, en séparant de façon radicale la structure logique d'un texte de sa représentation visuelle, elle autorise des traitements formels différents pour un même contenu à une échelle inédite. Et, sans toujours s'en rendre compte, ces chercheurs retrouvent des problématiques anciennes soulevées par des anthropologues comme Goody, des historiens du livre ou encore des sémioticiens.

En nous fondant sur des appréhensions diverses de la notion de texte, nous avons montré qu'un processus de « grammatisation », au sens de S. Auroux, était en cours³.

2.2. Texte ou contenu

La seconde entrée sur le texte (il faut comprendre ce terme dans un sens générique sans référence à un mode d'expression particulier, comme synonyme de contenu) fait se rencontrer les chercheurs en linguistique, en sciences de l'information, en recherche d'information et en ingénierie des connaissances. Le problème principal est de construire des modèles permettant de traiter le contenu des documents, afin de les retrouver, et éventuellement les réagencer pour en produire de nouveaux, adaptés à la demande du lecteur. Il s'agit de produire du sens en dépassant la confusion de l'accumulation des informations, à partir notamment d'une modélisation de données sur les documents, les métadonnées. Depuis les bases de données bibliographiques jusqu'aux « moteurs » d'aujourd'hui, ici aussi les développements et perfectionnements ont été très rapides.

Au-delà de la performance des outils, la question posée est celle de la construction des savoirs dans la relation entre les métalangages et les documents eux-mêmes. Celle-ci est posée aussi bien dans la relation entre un lecteur et un texte que dans celle de sa « navigation » entre plusieurs textes. Manipuler des ontologies, c'est-à-dire, entre autres, articuler des concepts, n'est pas sans effet sur l'organisation des connaissances, particulières à une communauté ou génériques pour une société. Là encore les chercheurs en informatique et ingénierie des connaissances croisent des préoccupations anciennes, mêmes très anciennes des chercheurs en sciences de l'homme depuis la philosophie jusqu'aux sciences de l'information, en passant par les différents courants des sciences du langage.

³ Voir le 2^e article de R. T. Pédaque, http://archivesic.ccsd.cnrs.fr/sic_00001401.

2.3. Médium ou relation

La troisième entrée sur le médium comprend trois espaces emboîtés, ceux *a*) de la construction identitaire (la relation à soi et aux autres), *b*) de la communication de groupe (communautés d'intérêt, organisations), *c*) de la publication (médias). Dans chacun de ces espaces, le document joue un rôle essentiel de mémoire, d'organisation, de créativité ou de transmission⁴.

Les thématiques de recherche autour de ce sujet sont majoritairement les télécommunications mais aussi les sciences sociales et les sciences de la communication. Le document y fonde son statut dans une diffusion dépassant l'intime et l'éphémère, en s'articulant avec d'autres dans des collections. Deux pratiques sociales très « normées » se modifient au travers du Web. D'une part, la frontière entre la communication interpersonnelle et la communication « flottante » ou publique, entre correspondance privée et mise en publicité, se déplace et tous les codes sociaux et modalités organisationnelles qui s'y rapportent sont ébranlés. D'autre part, illustrées par le préfixe « hyper » (-texte, -média), les relations entre documents se distribuent différemment entre partage et liens, entre mise en ligne et mise en réseau, entre mobilité et ubiquité.

La notion de publication, qui s'était stabilisée au cours des siècles, se transforme brutalement. Celle de bibliothèque, dont les contours étaient limités par la taille des collections et l'implantation physique et institutionnelle, explose en bibliothèque numérique, centralisée ou distribuée et largement accessible. Celle d'archive, autrefois réservée à un enregistrement pour une préservation à long terme, a subi elle aussi des changements tout aussi radicaux avec d'un côté le « record management » et, de l'autre, les « archives ouvertes » ou « dépôts institutionnels ».

Ce qui est en cause ici, c'est le média comme « prothèse » et toutes les questions de notre relation aux terminaux, celles des médiations et des métiers afférents et des régulations économiques et sociales. Le document se trouve souvent d'emblée inclus dans un espace régi par les lois des grands nombres : parce qu'il est mis en relation avec un nombre quasi-infini de ses semblables et parce qu'il est potentiellement visible par un nombre non-fini de lecteurs.

L'ensemble peut être résumé dans un tableau reprenant les trois entrées citées plus haut (*cf.* Tab. 1).

⁴ Voir le 3^e article de R. T. Pédaque, http://archivesic.ccsd.cnrs.fr/sic_00001741.

Tab. 1 – Domaines couverts par les recherches sur le document numérique.

Entrée Problématique	Chercheurs et Disciplines	Objet Résultats	Aboutissement Interrogation
Forme/Signe Grammaire (Vu)	Traiteurs d'image, édition numérique, W3C, anthropologie, histoire du livre	Support/Structure Formats	XML Perception ?
Texte/Contenu Référence (Lu)	Sciences du langage, Sciences de l'information, ingénierie des connaissances	Sens/Meta Moteurs, ontologies	Web sémantique Connaissances ?
Médium/Relation Prothèse (Su)	Télécoms, Sciences de la communication, gestion, Sciences sociales	Collection/hyper Portails, collaboratifs, sites	Bibliothèque et archives numériques Publication/communication ?

3. DES RECHERCHES SUR LES DOCUMENTS NUMÉRIQUES

A l'aune de la catégorisation proposée, nous allons résumer le contenu de chaque chapitre en respectant l'ordre du tableau précédent. Cette classification est évidemment subjective, d'autant plus que les recherches abordées dans la plupart des chapitres ressortissent souvent de plusieurs des catégories de disciplines décrites précédemment.

La forme ou le signe : chapitres 2, 3, 4 et 5

La numérisation en question (chapitre 2) aborde les problèmes que provoque la numérisation massive de fonds documentaires, eux-mêmes liés à la nécessaire indexation de ses nouvelles ressources numériques. Il faut, pour les résoudre, inventer une indexation multimodale, construire des représentations de haut niveau (sémantique) à partir d'une image scannée. Il y a un enjeu stratégique de mémoire collective autour de la numérisation historique et culturelle qui nécessite des actions de mutualisation, d'évaluation des performances et de gestion des connaissances sur les chaînes de traitement. Le texte synthétise les résultats de réflexions menées dans le cadre de l'Action concertée incitative Masses de données « Madonne » et de l'atelier Numérisation.

La lecture sur supports numériques : des repères pour une activité complexe qui se diversifie (chapitre 3) propose un état de l'art synthétique des recherches en France sur la lecture numérique. Il présente les avancées actuelles dans la description et la compréhension de ce que les lecteurs développent comme « pratiques et activités de lecture » lorsqu'ils interagissent avec des documents sur supports numériques. Quatre études y sont présentées : *a*) l'existence d'un mode prototypique de lecture, *b*) l'observation scientifique de l'acte de lecture, en particulier l'étude du mouvement des yeux, *c*) l'acte de lecture-écriture et *d*) la lecture comme activité intentionnelle en cours de transformation. Ces études suggèrent de nouvelles pistes de recherche.

Au-delà du Web : les interfaces de visualisation et d'annotation pour les bibliothèques numériques (chapitre 4) aborde les bibliothèques numériques via leurs limitations, en particulier au niveau des interfaces du Web. Il est difficile d'offrir une navigation fluide au sein d'un grand corpus de documents et la lecture à l'écran de fac-similés numériques d'ouvrage reste pénible. De nouvelles métaphores d'interaction peuvent conduire à une intégration plus fluide des tâches de navigation et de

22 LA REDOCUMENTARISATION DU MONDE

lecture, plus proche de l'expérience réelle des lecteurs en bibliothèque. Des prototypes d'applications sont présentés, ainsi qu'une discussion sur leur implémentation et une comparaison avec d'autres travaux. La montée en puissance des systèmes d'affichage et des connexions internet haut débit permettent d'espérer que l'accès aux bibliothèques numériques devienne dans un futur proche *as easy as a game*. Il reste auparavant à comparer, dans un cadre expérimental commun, les procédés proposés par les différentes équipes qui contribuent à ce thème de recherche et les pratiques de lecture sur les différents systèmes. Cette réflexion devra être étendue aux thématiques proches qui relèvent du document numérique en général, notamment celles concernant le son et la vidéo.

Les temps du document numérique (chapitre 5) présente les résultats d'une Action spécifique. Celle-ci est née de la volonté de chercheurs venus d'horizons divers, de fédérer leurs interrogations sur « les temps » du document numérique. Alors qu'isolément, chacun se représente le temps comme une prise de repère sur un axe unique, la mise en perspective des différents points de vue des participants fait ressortir les nombreuses dimensions temporelles du document numérique. Entre autres recherches à conduire, il reste à trouver un formalisme apte à rendre compte de l'articulation entre ces différents axes.

Le texte ou le contenu : chapitres 6, 7, 8 et 9

Sur des aspects primordiaux du Web sémantique (chapitre 6) aborde quelques thèmes majeurs de l'approche Web sémantique. Ce projet entend dépasser les limites actuelles du Web en faisant le constat que même si de nombreux outils logiciels, fort utiles, existent, ils sont limités dans l'aide qu'ils peuvent fournir aux utilisateurs faute d'une description formelle suffisante des ressources Web. Cette évolution suppose encore des recherches dans des thèmes décrits dans le chapitre : méta-données / annotations, ontologies, langages de représentation de connaissances, intégration de sources d'information hétérogènes, personnalisation et adaptation à un utilisateur et enfin des services Web sémantiques. Des analyses sur les usages du Web, sémantique ou non, doivent impérativement compléter les recherches pour obtenir un meilleur niveau de confiance et de sécurité dans l'utilisation du Web.

Corpus scientifiques numérisés : savoirs de référence et points de vue des experts (chapitre 7) aborde l'instrumentation de l'usage du corpus documentaire par la communauté scientifique. En complément aux approches ontologique, il propose de travailler sur le fait qu'il y a des points de vue différents sous-entendant des ontologies différentes et de

voir comment on peut proposer des outils qui tiennent compte de ces points de vue mais aussi d'un savoir de référence. Une approche « éminemment » documentaire pourrait être de réinstrumenter l'activité d'éditeurs de documents au sens large – bibliothécaires, documentalistes, président de comité de programme, etc. – sur un espace éditorial cohérent. Les recherches dans ce domaine doivent permettre de voir comment se construisent les savoirs de référence à partir des multiples points de vue et, ensuite, d'instrumenter cette activité.

Corpus et terminologie (chapitre 8) étudie les liens entre ressources termino-ontologiques et corpus. Ils intéressent la linguistique de corpus, la terminologie, l'informatique et les sciences de l'information. A partir de l'identification des difficultés théoriques et techniques communes, se dégage la nécessité d'approfondir la notion de genre, de mieux prendre en compte les questions d'usage et de maintenance ainsi que les modalités d'évaluation des ressources et des outils de construction.

Processus d'annotation dans les documents pour l'action (chapitre 9) explicite l'évolution des documents numériques dans une perspective communicationnelle et pragmatique qui vise à rendre compte des activités coopératives. Le document est l'objet d'une transaction. C'est *une production sémiotique transcrite ou enregistrée sur un support pérenne qui est équipée d'attributs spécifiques visant à faciliter les pratiques liées à son exploitation présente et ultérieure*. Ces documents pour l'action se retrouvent dans de nombreux domaines avec des caractéristiques, entre autres, d'inachèvement, de pérennité, de rapport argumentatif souvent complexe aux autres documents et forment avec ceux-ci des dossiers. Les annotations explicitent la liaison d'un document (fragment) avec le reste de la production (le dossier). Elles sont contributives comme dans le Web sémantique ou attentionnelles. Dans ce chapitre, les auteurs approfondissent les domaines médicaux, de conception, des bibliothèques numériques et les pratiques scientifiques.

Le médium ou la relation : chapitres 10, 11 et 12

Interactions document organisation : document enaction (chapitre 10) essaie de répondre à la question « que pouvons nous attendre d'une approche du document par le biais organisationnel ? » à partir de deux études de terrain – travailleurs sociaux au sein d'un Centre d'information et d'accueil départemental et ingénieurs brevet dans un cabinet conseil en brevets d'invention – en proposant les bases d'un enrichissement inédit du document numérique. Le document est en interaction avec l'action et donc, *in fine*, avec l'organisation qui le préconise. Son importance est

24 LA REDOCUMENTARISATION DU MONDE

signalée par une montée du travail d'éditorialisation concomitant avec la montée des écrits d'organisation. De ces études découlent des perspectives de recherche présentés dans le chapitre.

Auctorialité : production, réception et publication de documents numériques (chapitre 11) étudie l'auctorialité selon trois processus : la production de documents par un ou plusieurs auteurs, la réception par les lecteurs de l'auteur dans l'œuvre ou le document et enfin la liaison permettant de retrouver auteur et document. L'autorité est précisément attachée à l'auctorialité par le lien susceptible d'unir un auteur à un discours ou à un document. Les conclusions temporaires de cette recherche peuvent être décrites selon trois points de vues : *a*) l'auctorialité numérique qui montre un déplacement des conditions de reproductibilité du papier vers les conditions de lisibilité et de structure temporelle des documents ainsi qu'une extension de l'activité de l'auteur vers la labellisation et l'auto-archivage ; *b*) l'interaction entre auteur et lecteurs qui affaiblit la position de l'auteur au profit d'une forte interaction avec les lecteurs (critique, lien, annotation, etc.) ; enfin, *c*) les outils qui créent et amplifient une activité d'auto-observation des auteurs en même temps qu'ils bouleversent les étapes de production et de diffusion des textes.

La publication sur le web, entre filiations et innovations éditoriales (chapitre 12) vise à repérer les modalités du processus de publication – entendu comme « la mise en forme d'un contenu préalablement sélectionné, en vue de sa diffusion collective » –, depuis son développement sur le support Web. Dans un double cadre de filiation avec les modèles de publication précédents et d'intégration des nouvelles logiques socio-techniques de l'Internet, l'analyse interroge, à un niveau à la fois diachronique et synchronique, le degré de pérennité et de généralisation trans-sectorielle des différentes expériences de publication sur le Web. Cette analyse est d'autant plus difficile que le dispositif socio-technique est récent : les concepts sont souvent empruntés aux configurations antérieures (l'édition imprimée ou la diffusion audiovisuelle), et pas forcément adaptés pour appréhender ces nouvelles réalités. En même temps que l'observation empirique, c'est donc à un travail de redéfinition des notions que sont confrontés les chercheurs se penchant sur le numérique.

4. EN GUISE DE CONCLUSION

Ce livre termine un cycle de travail commencé en 2002 avec la création du RTP-DOC. Il est le second et, pour l'instant, dernier ouvrage de Roger T. Pédaque. Ces deux livres retracent le cheminement du réseau des chercheurs sur le *Document numérique*, que ce soit en proposant une mise en perspective des trois articles de réflexion de Roger⁵ ou, ici, en présentant les travaux des Action spécifiques et ateliers du réseau. Durant cette période, le réseau a eu bien d'autres activités comme l'organisation d'une école d'été à Tromsø en Norvège ou la mise en route des *Semaines du numérique* dont la seconde édition vient de se dérouler à Fribourg en Suisse. Ces activités et bien d'autres sont répertoriés sur le site du réseau, <http://rtp-doc.enssib.fr/>. Enfin, la liste de diffusion, rtp-doc@enssib.fr, est le medium d'échange du réseau.

Comme le dit Michel Melot en préface du précédent livre, le réseau a encore du travail, beaucoup de travail, s'il veut explorer toutes les facettes des documents numériques et traquer l'opacité des processus technologiques en œuvre autour d'eux. Nous espérons pouvoir, avec l'aide de nos tutelles et en particulier du CNRS, réactiver les travaux de Roger et trouver les moyens de poursuivre cette aventure de la compréhension de la redocumentarisation du monde.

⁵ Roger T. Pédaque, *Le document à la lumière du numérique*, C&F éditions, 2006.

