

Benoît Epron
Marcello Vitali-Rosati

L'édition à l'ère numérique
Fichier preprint bon à composer

LA DÉCOUVERTE
9 bis, rue Abel-Hovelacque
75013 Paris

Si vous désirez être tenu régulièrement informé des parutions de la collection « Repères », il vous suffit de vous abonner gratuitement à notre lettre d'information mensuelle par courriel, à partir de notre site **www.collectionreperes.com**, où vous retrouverez l'ensemble de notre catalogue.

ISBN 978-2-7071-9935-5

Ce logo a pour objet d'alerter le lecteur sur la menace que représente pour l'avenir du livre, tout particulièrement dans le domaine des sciences humaines et sociales, le développement massif du photocopillage. Nous rappelons donc qu'en application des articles L. 122-10 à L. 122-12 du code de la propriété intellectuelle, toute photocopie à usage collectif, intégrale ou partielle, du présent ouvrage est interdite sans autorisation du Centre français d'exploitation du droit de copie (CFC, 20, rue des Grands-Augustins, 75006 Paris). Toute autre forme de reproduction, intégrale ou partielle, est également interdite sans autorisation de l'éditeur.

© Éditions La Découverte, Paris, 2018.

Introduction

Textes, images, vidéos, données sont produits et circulent désormais en environnement numérique. Même les contenus destinés à l'imprimé sont dans leur totalité rédigés, structurés et mis en forme avec des outils numériques et sont ensuite commercialisés, rendus visibles et accessibles via des plates-formes en ligne.

En ce sens, il n'y a plus aucun contenu qui ne soit pas touché par les technologies informatiques.

C'est pour cette raison que l'objet de ce livre n'est pas principalement l'édition numérique, mais plutôt l'édition à l'ère numérique : son ambition est de donner un aperçu de l'impact non seulement des outils, mais plus largement de la culture numérique sur l'édition.

Dans cet ouvrage, le lecteur trouvera un survol des principaux aspects de l'édition et une analyse de la façon dont ces aspects sont en train d'être modifiés par les changements culturels, sociaux et économiques qui caractérisent notre époque.

Au cours des dernières années, la notion de « désintermédiation » a souvent été convoquée : le Web et les technologies numériques réduiraient les médiations entre la production et la publication des contenus. L'édition aurait donc perdu son importance, car n'importe quel usager pourrait, sans médiation, rendre disponibles les contenus qu'il souhaite. Au contraire, ce livre montre que la fonction éditoriale n'a jamais été aussi présente et aussi centrale qu'aujourd'hui. Il y a de l'édition partout : dans les différents médias en ligne, mais aussi dans la structuration des contenus sur les réseaux sociaux, dans les plates-formes de distribution, dans les blogs, dans les moteurs de recherche...

Pour démontrer cela, ce livre s'articule en quatre chapitres. Le chapitre i propose une définition de l'édition et une analyse des enjeux généraux liés à la culture numérique : le changement du métier d'éditeur, la question des droits d'auteur et l'émergence du concept d'éditorialisation. Ce chapitre montre que l'édition a principalement trois fonctions : celle de produire des contenus, celle de les faire circuler et celle de les légitimer. Chacun des autres trois chapitres analyse une de ces trois fonctions. Le chapitre ii analyse donc la production des contenus et montre comment l'environnement numérique détermine de nouveaux modes d'écriture et de structuration des contenus.

Le chapitre iii prend en compte le rôle de légitimation des contenus typique de l'édition : à la reconnaissance symbolique liée à la publication imprimée s'ajoutent de nouveaux

modèles de légitimation, en particulier celui fondé sur les recommandations de la communauté et celui fondé sur les celles des algorithmes. Le chapitre iv explique comment les nouveaux modes de diffusion des contenus — et en particulier le Web — ont changé les circuits et les dispositifs de circulation prénumérique en affectant le rôle des libraires et des bibliothèques.

L'édition change donc, mais elle reste fondamentale : cet ouvrage se veut une réflexion destinée à tous les lecteurs qui souhaitent mieux comprendre comment l'ensemble des connaissances, du savoir et des contenus en général sont produits, validés et diffusés à notre époque.

I / La fonction éditoriale

Qu'est-ce que l'édition aujourd'hui ? Les technologies numériques ont pris une place fondamentale dans notre vie et affectent profondément notre rapport au savoir. Dans cet ouvrage, nous voulons montrer comment les changements déterminés par ces technologies affectent le monde de l'édition. Il ne s'agit donc pas de donner une définition de l'édition numérique comme un champ distinct – et encore moins opposé – de l'édition « classique » ou « non numérique ». Il s'agit plutôt de comprendre ce que devient l'édition en général à l'ère des technologies numériques. Nous parlerons dans ce livre du « numérique », en substantivant l'adjectif pour signifier l'ensemble des changements déterminés par les technologies numériques, changements qui ne sont pas seulement techniques, car ils affectent en général toute la sphère culturelle, indépendamment des outils. « Le numérique » n'a donc pas seulement changé les techniques éditoriales, mais, plus généralement, le sens même de l'édition. Nous n'écrivons, ne lisons, ni n'accordons confiance aux auteurs et aux éditeurs de la même manière qu'auparavant. Pour comprendre ces changements et saisir le sens de l'édition aujourd'hui, il faut d'abord définir l'édition, ses objectifs et ses fonctions.

L'édition peut être comprise comme un processus de médiation qui permet à un contenu d'exister et d'être accessible. On peut distinguer trois étapes de ce processus qui correspondent à trois fonctions différentes de l'édition : une fonction de choix et de production, une fonction de légitimation et une fonction de diffusion.

Analyser ces trois fonctions nous permettra de comprendre à quoi sert l'édition pour s'interroger ensuite sur la façon dont les technologies numériques réagent le processus en le transformant.

Fonction de choix et de production

En premier lieu, l'édition a une fonction de choix et de production des contenus. C'est la différence la plus évidente entre un contenu édité et un contenu non édité. Entre le manuscrit dans le tiroir d'un auteur et le livre sur l'étagère d'une librairie ou d'une bibliothèque, il y a un processus de choix et de mise en forme qui distingue radicalement les deux objets. En quoi consiste exactement ce processus ? Il s'agit d'une sélection parmi plusieurs contenus existants, fondée sur la qualité ou sur des exigences de marché ou encore – le plus souvent – sur un mélange des deux. Éditer signifie décider quels

contenus sont dignes d'être rendu accessible à un public. Depuis plusieurs siècles – ou, plus précisément, depuis le XV^e siècle –, cette sélection a été le plus souvent garantie par des maisons d'édition et plus généralement des éditeurs : les auteurs leur proposent des manuscrits parmi lesquels elles choisissent ce qui leur semble valoir la peine d'être publié. Mais ce mécanisme de sélection peut être différent : par exemple, une maison d'édition peut demander à des auteurs de produire un texte nouveau, sur un sujet qu'elle considère comme important ou potentiellement lucratif. La sélection peut par ailleurs être faite par une instance différente d'une maison d'édition : dans le cas d'un article de journal ou de revue, il s'agira plutôt d'une rédaction, ou d'un directeur de publication. Dans le cas d'une anthologie, il s'agira de l'anthologiste. Et avant l'apparition des maisons d'édition, ce choix était réservé à des instances complètement différentes, par exemple un mécène qui décidait de financer un auteur pour qu'il produise des textes. Cela signifie que la fonction éditoriale, prise en charge en grande partie par des maisons d'édition dans les derniers siècles, ne leur est pas réservée par définition.

Cette phase de choix est toujours indissociable d'une instance de production. Considérons les cas de la commande et de la sélection. Dans le premier cas, il est évident que le contenu est produit par l'instance éditoriale : celle-ci (maison d'édition, directeur de publication, mécène...) pense la forme et la structure en amont de la réalisation. L'écriture est déterminée par l'instance qui commande. Par exemple, si une maison d'édition commande un manuel à un spécialiste de la grammaire française, elle décide du format du livre, établit, en dialogue avec l'auteur, les sujets à traiter, le nombre de pages, etc. Autre exemple, un mécène peut commander à un poète un poème à la gloire de sa famille – ce fut le cas de Virgile, dont l'*Énéide* devait légitimer l'autorité d'Auguste. Dans le second cas, celui de la sélection, un auteur propose son manuscrit à une maison d'édition. Cette dernière décide de publier le texte – fonction de choix –, mais le soumet en même temps à un travail de révision, de relecture, de mise en forme et de mise en page. Même dans le cas de la sélection d'un contenu déjà existant, l'instance éditoriale procède donc à une « production » du contenu, en le retravaillant en vue de sa publication.

Le choix et la production sont donc toujours les fonctions premières de l'édition. Éditer signifie d'abord choisir et produire.

Les changements techniques impliquent des mutations des instances chargées de ces fonctions : après l'invention de l'imprimerie moderne, le modèle des maisons

d'édition s'est peu à peu imposé, remplaçant les modèles anciens. Les technologies numériques ont, à leur tour, un impact sur ces instances. Dans le prochain chapitre, nous verrons comment les outils numériques conditionnent cette phase – choix et production – du processus éditorial.

Fonction de légitimation

Le deuxième aspect qui distingue le manuscrit dans un tiroir du livre publié est la valeur symbolique que nous attribuons à ce dernier. Cette différence s'explique par la reconnaissance de la médiation effectuée par l'instance éditoriale. C'est ce que nous appelons la fonction de légitimation. Cette légitimation est rendue possible par le travail de sélection et de production du contenu, formaté et retravaillé par une instance ayant le pouvoir symbolique d'en garantir la qualité. Selon la nature de l'instance éditoriale, cette légitimation aura un poids et une forme différents, mais dans tous les cas cette instance confère au contenu une légitimité reconnue par son lectorat. Par exemple, dans le cas d'un roman, la signature de la maison d'édition nous donne des indices sur le type de littérature et sur la qualité que l'on peut en attendre. Dans le cas d'un ouvrage universitaire, par exemple de médecine, l'instance de légitimation garantit une certaine objectivité scientifique du contenu, notamment grâce au processus d'évaluation par les pairs.

Cette fonction a une importance sociale et politique fondamentale : c'est le dispositif grâce auquel nous distinguons des formes d'autorité qui nous permettent de nous repérer dans les contenus, choisir les textes que nous voulons lire, et savoir quelle confiance leur accorder. La fonction de légitimation établit donc une différence entre les contenus et donne des indices sur leur valeur et finalement sur leur sens.

Fonction de diffusion

Reprenons l'exemple du manuscrit dans le tiroir de l'auteur pour identifier une troisième fonction de l'édition : la diffusion des contenus. Le manuscrit dans le tiroir n'est pas diffusé, il est invisible. Mais l'invisibilité n'est pas la seule caractéristique qui distingue ce manuscrit du livre publié : justement parce qu'il est dans le tiroir, le manuscrit ne s'adresse à personne, il n'est *pour* personne. La fonction de diffusion confère donc d'une part une visibilité et d'autre part une adresse. L'instance éditoriale a pour mission de créer une relation entre le producteur du contenu et son destinataire, le

lecteur. Il est important de bien analyser cette fonction pour ne pas en avoir une idée réductrice. Car, en effet, on pourrait identifier la fonction de diffusion au fait de rendre matériellement disponible un contenu – par exemple en le distribuant dans des librairies, des kiosques ou des bibliothèques. Mais la diffusion ne se limite pas à cette action matérielle. Il s’agit plus précisément d’identifier un lectorat, d’analyser ses besoins, ses compétences, ses désirs et ses pratiques, et de faire en sorte que le contenu lui soit adressé. Un contenu est édité quand il est *pour* quelqu’un. En ce sens, la fonction de diffusion empiète en partie sur la fonction de production : par exemple, les choix du titre, du format, de la mise en forme, mais aussi du langage d’un livre relèvent de la fonction de diffusion. Prenons l’exemple de ce livre : il fait partie d’une collection particulière, « Repères », qui s’adresse à un public précis : un lectorat qui porte un intérêt à une question précise sans forcément disposer de compétences savantes dans le domaine en question. La taille du livre, sa mise en page, son vocabulaire, sa structure sont faits pour convenir à ce lectorat. Si l’on s’était exclusivement adressé à un lectorat de spécialistes et d’universitaires, ce livre aurait été pensé et structuré autrement : on aurait fait les choix d’un grand format, d’un appareil critique important, d’un vocabulaire académique et d’un discours moins introductif.

La diffusion matérielle du contenu, sa distribution, dépend de ce premier aspect qu’est l’adresse. C’est à partir de l’identification d’un lectorat spécifique que l’on peut déterminer un mode concret de distribution qui réponde à la question : comment toucher le lectorat visé ? Quels sont ses usages et ses pratiques ? Dans le monde du livre papier, cette diffusion matérielle est prise en charge par deux instances différentes : le diffuseur, qui met les livres sur les catalogues, les présente aux détaillants et enregistre leurs commandes, etc., et le distributeur qui est en charge du stockage et du transport des livres vers les points de vente.

Pour résumer, la fonction de diffusion comporte l’adresse, la distribution et tous les dispositifs qui tendent à rendre un contenu matériellement accessible, mais aussi visible.

Instances éditoriales et maisons d’édition

Nous avons parlé jusqu’à présent d’instances éditoriales pour ne pas confondre les trois fonctions de l’édition avec les maisons d’édition. Il est en effet important de souligner que, si ces fonctions ont été prises en charge par des maisons d’édition au cours des derniers siècles, l’instance éditoriale ne doit pas intrinsèquement être identifiée à celles-ci. À partir de l’invention et de la diffusion en Occident de la presse à caractères

métalliques mobiles au xv^e siècle, s'est opéré un processus d'institutionnalisation des maisons d'édition qui sont devenues l'instance éditoriale par excellence. Or, avec le rapide développement du Web à partir des années 1990, on assiste à une progressive émergence d'autres instances éditoriales. En d'autres termes, d'autres entités, différentes des maisons d'édition, prennent en charge d'une manière ou d'une autre ces trois fonctions de production, de diffusion et de légitimation, en bouleversant le panorama qui s'était stabilisé au cours des derniers siècles.

Le Web amène de nouveaux moyens de production, de diffusion et de distribution qui modifient profondément les dispositifs d'autorité permettant la légitimation des contenus. Pour comprendre ces changements, il suffit d'observer l'une des pratiques les plus courantes vis-à-vis des contenus : la recherche d'information. Jusqu'aux années 1990 – et probablement même après, jusqu'au début des années 2000 – la recherche d'informations dépendait presque exclusivement du monde de l'imprimé. À partir des années 1990 et surtout à partir du rapide développement du Web et de services comme Google Search ou de plates-formes comme Wikipédia, le recours au livre s'est raréfié, la plupart des informations étant trouvable en ligne. Dans ce contexte, plusieurs questions se posent : peut-on considérer le contenu en ligne comme le fruit d'un processus d'édition ? D'après quels critères ? Quels sont les nouveaux acteurs de ce processus et quel rôle y jouent les acteurs du monde du papier (les maisons d'édition, les librairies, les bibliothèques) ?

Les éditeurs face au numérique

Des enjeux très hétérogènes

La filière de l'édition présente, au-delà de l'apparente homogénéité d'un secteur d'activité, une situation hétérogène dans son rapport au numérique. La structuration des secteurs éditoriaux et des marchés n'est pas la même et induit donc des approches différentes vis-à-vis du développement de l'écosystème numérique. Ces approches diffèrent notamment en raison des différences de statut des auteurs, utilisateurs, clients ou acheteurs.

Le rapport au numérique de chaque secteur est fonction de différents paramètres, comme l'existence d'un marché professionnel (édition juridique), le degré de concentration des acteurs (édition scientifique) ou la nature des acheteurs (édition

scolaire). Ces enjeux sont mouvants, l'édition numérique s'inscrivant dans l'évolution constante et rapide du contexte technologique et des usages.

Cette évolution présente des disparités importantes. En France, le Syndicat national de l'édition (SNE) estime la part du numérique dans le chiffre d'affaires global de l'édition à 8,65 % en 2016. Dans le même temps, la part du numérique dans le marché étatsunien de l'édition est estimée à plus de 33 %.

À grands traits, les enjeux des différents secteurs de la filière de l'édition sont les suivants :

1. La définition de modèles économiques hybrides qui permettent de gérer la coexistence à plus ou moins long terme d'une activité éditoriale sur supports numériques et sur supports papier. C'est notamment le cas pour l'édition de littérature grand public qui voit la part de son activité relative au numérique croître progressivement.
2. La définition de formes éditoriales numériques pertinentes et adaptées. C'est notamment le cas pour l'édition scolaire ou l'édition jeunesse dans lesquelles se posent des questions relatives à l'interactivité ou aux fonctionnalités embarquées. Ces questionnements portent sur l'efficience ou la valeur ajoutée réelle de ces fonctionnalités ou enrichissements. Depuis les années 2000, le concept d'*édition augmentée* est étudié et discuté. Ses applications sont encore limitées, mais constituent une voie intéressante pour certains secteurs.
3. L'adaptation des modèles, des filières et des acteurs. Ces enjeux politiques ne sont pas spécifiques à certains secteurs même si l'édition scientifique en est un bon exemple. En effet, il s'agit d'un secteur déjà quasiment entièrement numérique, dans lequel des débats politiques posent la question de la place des différents acteurs, publics ou privés, dans cette activité.

Ces enjeux couvrent également des aspects transversaux de l'édition numérique : le statut des auteurs de livres enrichis, la place des bibliothèques dans ce nouvel écosystème ou encore l'environnement législatif, concernant le prix unique du livre ou les durées d'embargo par exemple. Ces questionnements sont complexes car ils mêlent des aspects techniques, politiques et économiques.

La littérature : nouvelles formes de diffusion

La littérature est, en termes de visibilité et de poids symbolique, un secteur central dans l'analyse du passage au numérique de l'édition. La littérature représente en effet, d'après les chiffres du SNE, environ 25 % du chiffre d'affaire de l'édition en France, devant les secteurs de la jeunesse et du livre pratique. Ce secteur a longtemps concentré les attentions et analyses lorsqu'était évoqué le *livre numérique*. Cette focalisation s'explique notamment par la visibilité dont bénéficie la littérature auprès du grand public et par le rôle matriciel qu'elle pourrait jouer pour les autres secteurs. En effet, depuis le Village eBook de l'édition 2000 du Salon du livre de Paris, la question du livre numérique a longtemps été abordée sous le double angle du dispositif de lecture et du modèle économique.

Dans les années 2000, les expériences de développement de dispositifs techniques de lecture numérique se multiplient avec notamment le lancement du CyBook par la société Cytale en 2001. Ce dispositif, un écran d'ordinateur portable tactile intégré à une coque recouverte d'un rabat en cuir, ouvre plusieurs questionnements. Tout d'abord, il pose la question du matériel nécessaire à une activité de lecture numérique. Le rétroéclairage de l'écran, ses performances d'affichage, la taille ou le poids du dispositif deviennent des problématiques auxquelles il apparaît indispensable de répondre avant d'imaginer le développement d'un marché du livre numérique. Les acteurs du livre, en l'occurrence les éditeurs, se retrouvent donc confrontés à de nouveaux types d'interlocuteurs avec d'un côté les fabricants de dispositifs techniques de lecture qui conditionnent, de façon logicielle ou matérielle, l'expérience de lecture et, de l'autre, les plates-formes dédiées au livre numérique, qu'elles soient de commercialisation et/ou d'autoédition. Ce secteur de l'édition est ainsi confronté à de multiples formes émergentes de diffusion. Comme pour l'ensemble des modèles numériques de document, cette transformation de la diffusion se structure autour d'une tension entre désintermédiation et réintermédiation.

D'une part, il s'agit d'une dynamique de diffusion *directe* des productions éditoriales, de l'auteur au lecteur, dans une logique d'autoédition. Ce modèle n'est pas propre à la littérature numérique – il existait déjà des éditions à compte d'auteur dans le monde de l'imprimé – mais elle prend sur Internet une dimension particulièrement importante. D'autre part, la réintermédiation, est le mécanisme d'émergence de nouveaux acteurs intermédiaires dans la diffusion des produits éditoriaux numériques. Le point à souligner dans ce mouvement de réintermédiation est la nature des nouveaux acteurs,

parfois nouveaux entrants dans le domaine du livre, qui déploient des modèles économiques dans lesquels la part de leurs revenus issue de la vente des ouvrages numériques reste très limitée. Ainsi, en réintermédiant leur diffusion numérique, les éditeurs traditionnels se trouvent confrontés à des acteurs ayant des objectifs stratégiques potentiellement très éloignés. Dans le cas d'Apple par exemple, la diffusion des produits éditoriaux numériques ne représente qu'une part très réduite de son activité, au contraire de la vente de matériel. Le cas d'Amazon est différent : en soutenant un tarif relativement bas pour ses liseuses, c'est davantage la croissance de sa base de clients qui est recherchée. Cette divergence en termes d'enjeux stratégiques complique d'autant plus la relation de coopération entre les éditeurs et ces diffuseurs numériques qu'elle s'inscrit dans un rapport de forces nettement à l'avantage des plates-formes de diffusion. Des tentatives de reprise en main de la distribution numérique par les éditeurs existent toutefois. Certaines, comme Immatériel.fr ou Dilicom, se positionnent comme intermédiaire de distribution vers les plates-formes de commercialisation, tandis que d'autres se focalisent sur des secteurs spécifiques, comme Cairn pour l'universitaire.

L'édition jeunesse : des nouvelles productions multimédia

L'édition jeunesse est à considérer précisément dans sa confrontation au numérique. Si l'édition d'ouvrages traditionnels destinés à la jeunesse (Harry Potter par exemple) s'inscrit dans des logiques proches de celles de la littérature, le cas de l'édition illustrée, augmentée, interactive... est bien différent. Nous appréhenderons ici uniquement ce second cas pour mieux identifier les enjeux spécifiques auxquels ce secteur est aujourd'hui confronté.

La particularité de l'édition jeunesse dans son passage au numérique réside dans la spécificité des formes éditoriales qu'elle prend. Par formes éditoriales, nous entendons l'ensemble des fonctionnalités exploitant les potentialités interactives et multimédia propres au numérique. Dans ce domaine, les enjeux pour l'édition jeunesse se situent à plusieurs niveaux.

Dans un premier temps, la question de la pertinence de ces fonctionnalités nouvelles dans la création de productions éditoriales pour la jeunesse se pose. Il ne s'agit pas d'une problématique récente : les acteurs impliqués dans les productions *éditoriales* sur supports numériques (comme le CD par exemple) ont du traiter ces questions depuis plusieurs décennies déjà. Ainsi, l'apport d'éléments multimédias dans les trames narratives ou dans l'illustration de récits est une question qui s'est déjà largement posée à

la fois aux éditeurs et aux auteurs d'œuvres destinées à la jeunesse. Elle se pose aujourd'hui de manière différente en raison notamment du renouvellement permanent des possibilités offertes par le numérique (comme la réalité virtuelle ou augmentée, pour ne parler que des pistes les plus récentes).

Dans un second temps, l'évolution des formes éditoriales induit une complexification des processus éditoriaux. En effet, en intégrant des éléments de plus en plus variés dans les productions éditoriales numériques pour la jeunesse, les éditeurs et les auteurs se trouvent confrontés à des problématiques économiques et juridiques de plus en plus diversifiées. Ainsi, quels modèles de rémunération et de gestion des droits utiliser dans le cas de l'intégration d'une interactivité tactile dans un récit illustré ? Quels seront les statuts des différents contributeurs dès lors qu'un projet peut rassembler un auteur, un illustrateur, un animateur ou un développeur ?

Enfin, le dernier aspect à considérer pour l'édition jeunesse est celui des modèles économiques mis en œuvre lors du passage au numérique. Ainsi, en investissant les potentialités du numérique, les éditeurs jeunesse se trouvent confrontés à des choix stratégiques à la convergence des formats techniques, des circuits de distribution et de diffusion et des pratiques de lecture. Ces choix, complexes, conditionnent l'ensemble de la filière. Le choix du format, ePub ou application, est nécessairement associé à un modèle économique spécifique, et à une diffusion par les acteurs du livre ou par des plates-formes d'applications (iTunes Store ou Google Play). Les modèles tarifaires et fiscaux suivent ces logiques et poussent l'édition jeunesse soit vers le champ de l'édition de livres numériques soit vers celui des applications et du jeu vidéo.

Les manuels universitaires : nouvelles formes de consultation

Les manuels représentent un versant important de l'édition universitaire. À la différence des revues savantes destinées principalement à un public de chercheurs ou d'étudiants en fin de cursus pour accompagner un travail de recherche, les manuels universitaires visent un public d'étudiants dans un contexte d'apprentissage. Ce marché est donc tourné vers deux types de clientèles potentielles : étudiants faisant l'acquisition à titre personnel d'ouvrages qui vont les suivre tout au long d'un enseignement, et bibliothèques universitaires ou académiques. À cette première clé de segmentation du marché, il faut en ajouter une seconde qui prend en compte les différences entre les disciplines. On distingue principalement les sciences humaines et sociales (SHS) des sciences, techniques et médecine (STM). La confrontation au numérique du secteur des

manuels universitaires conduit à interroger à la fois les pratiques documentaires des étudiants, leurs comportements d'achat ainsi que leurs approches pédagogiques.

L'ensemble de ces critères dessine une situation de l'édition de manuels universitaires numériques très contrastée. Par exemple, le modèle de l'édition de manuels de gestion en premier cycle n'a que peu à voir avec celui des codes juridiques destinés aux étudiants en droit. Il est toutefois possible d'identifier certains invariants dans l'évolution des différents modèles. D'une part, les pratiques documentaires des étudiants sont fortement liées aux pratiques pédagogiques qu'ils rencontrent, notamment la prescription de lecture. Ces pratiques s'appuient aujourd'hui largement sur des plateformes pédagogiques dont les fonctionnalités techniques constituent elles aussi un ensemble de contraintes et permettent la mise à disposition des ressources directement par les enseignants. D'autre part, les pratiques d'acquisition par les étudiants restent, en France, largement conditionnées par le marché de l'occasion dans une logique d'achat-revente et par une forte attente, envers les Bibliothèques universitaires (BU), d'exhaustivité des ressources documentaires nécessaires au suivi des cursus. Les manuels numériques doivent donc trouver leur place selon des modalités qui correspondent à ces pratiques. Ainsi, la définition d'un modèle pour l'édition numérique de manuels universitaires doit prendre en compte à la fois les spécificités disciplinaires des pratiques documentaires des étudiants, les contraintes des environnements pédagogiques universitaires et les attentes des enseignants et des bibliothécaires dans la mise à disposition des ressources documentaires.

Les encyclopédies : de nouvelles formes de légitimation

Le secteur des encyclopédies a connu ces dernières décennies deux phases bien distinctes dans son passage au numérique, séparées par le lancement de Wikipédia qui marqua un tournant majeur pour ce secteur. La première phase a vu, à partir des années 1990, un « simple » changement de support avec l'arrivée des encyclopédies sur CD puis DVD. Cette évolution permit aux éditeurs d'exploiter les potentialités du format numérique comme la puissance de recherche ou l'intégration de contenus multimédias dans les notices. Cette phase fut accompagnée d'investissements importants des éditeurs dans des versions enrichies de leur produits, tout en conservant le modèle traditionnel combinant des auteurs identifiés, parfois connus, rédigeant des notices, et la vente de ces produits selon une logique classique de vente directe au consommateur. Cette situation a

conduit nombre d'éditeurs d'encyclopédie à renoncer, à partir des années 2000, à leurs versions imprimées voire dans certains cas à disparaître.

Ce changement de support a été suivi d'une seconde phase, modifiant bien plus en profondeur ce secteur éditorial, avec l'arrivée de Wikipédia. Cette encyclopédie collaborative, lancée en 2001, connaît un succès important, qui l'a élevée depuis plusieurs années au sixième rang des sites les plus visités dans le monde.

Plusieurs facteurs doivent être pris en compte pour analyser ce bouleversement. Le premier est la croissance d'une pratique de recherche en ligne qui passe de plus en plus par l'interrogation d'un moteur de recherche, au lieu de la saisie d'une adresse URL dans un navigateur. De plus, l'intégration de Wikipédia comme une des bases du Knowledge Graph (base sémantique de connaissances utilisée par Google) donne une visibilité extrêmement importante à ses notices. En 2016, Wikipédia proposait plus de 1 900 000 notices en français et plus de 46 millions au niveau mondial, dans plus de 280 langues.

Le principal facteur de changement introduit par Wikipédia du point de vue éditorial réside dans le mécanisme de création et de validation des contenus qu'elle propose. En effet, la logique contributive à la base de Wikipédia constitue un renversement du processus de légitimation des contenus encyclopédiques. Celui-ci passe ainsi d'une validation verticale descendante, apportée par l'éditeur et par la légitimité des contributeurs, à une logique de validation collective par l'ensemble des contributeurs. Il découle de cette approche une instabilité systémique des notices proposées, à l'opposé de la pérennité des encyclopédies traditionnelles inscrites dans une temporalité longue.

Ce n'est donc pas uniquement le changement de support au cours de la première phase qui explique ce bouleversement du secteur des encyclopédies, mais la conjonction d'une logique contributive et du développement de parcours de recherche d'information en ligne dans lesquels les contenus proposés par Wikipédia trouvent très efficacement leur place. Dans ce nouvel environnement, et en occupant la place prépondérante qu'elle occupe aujourd'hui, Wikipédia doit déployer des mécanismes de modération et de validation de plus en plus efficaces pour faire face aux tentatives multiples d'utilisation à des fins commerciales ou politiques.

Le secteur des encyclopédies présente donc, parfois de manière archétypale, plusieurs éléments caractéristiques des problématiques éditoriales à l'ère numérique. D'une part, le changement de support est à considérer à deux niveaux radicalement différents. Le premier niveau, formel, n'induit pas de modifications structurelles des modèles en place, les éditeurs étant souvent dans une logique de maintien d'un modèle

économique *classique*. Le second niveau est plus délicat : en passant d'un support imprimé à un format numérique, les encyclopédies ont dû appréhender la problématique nouvelle de la dispersion de leurs contenus et des parcours des utilisateurs en ligne. La capacité à être visible et repéré dans les pratiques de recherche d'informations est devenue un point-clé des stratégies numériques des éditeurs d'encyclopédies. D'autre part, la question de la légitimité éditoriale a constitué un point de bascule incontournable pour ce secteur. En passant d'une logique de légitimation verticale et descendante à une approche horizontale et collaborative avec Wikipédia, c'est une des bases du régime traditionnel de transmission du savoir qui est questionnée.

L'édition scientifique et savante : légitimation, modèles économiques, usages et nécessité de structuration

L'édition scientifique et savante – à savoir l'ensemble des contenus produits dans le cadre de la recherche et destinés à un public de spécialistes – compte parmi les champs de l'édition qui ont été le plus touchés par le développement du numérique. En particulier, le Web a profondément modifié les pratiques des chercheurs et, par conséquent, le panorama de l'édition scientifique. On peut identifier deux formes éditoriales principales dans le domaine scientifique et savant : la revue et la monographie. La revue savante – dont l'existence remonte au XVIII^e siècle – est le moyen le plus courant pour publier et faire circuler les résultats de la recherche, spécialement dans le champ des sciences exactes. L'article scientifique est la forme privilégiée pour partager avec la communauté scientifique ses recherches et les revues savantes ont été, depuis le XVIII^e siècle, le vecteur par excellence de ce type de contenu. Les monographies savantes caractérisent plutôt la production en sciences humaines et sociales – soit des disciplines qui requièrent le développement d'argumentations plus longues et discursives, ne pouvant pas toujours être contenues dans la forme courte de l'article. Dans les deux cas, celui de la revue comme de la monographie, nous pouvons identifier trois défis majeurs posés par les technologies numériques : en premier lieu, le changement des dispositifs de légitimation, en deuxième lieu, une transformation des modèles économiques et, en troisième lieu, l'apparition d'une nouvelle exigence de structuration des contenus qui n'était pas présente dans le cadre de l'édition imprimée.

L'enjeu de légitimation est crucial pour l'édition scientifique : il faut que le contenu soit validé scientifiquement. Cette validation se fait par le biais d'un système d'évaluation par les pairs : chaque contenu – monographie ou article – est donné à lire à

des spécialistes qui donnent leur avis sur la qualité et la rigueur du propos. Ce processus est normalement pris en charge par la communauté scientifique elle-même (ce sont des chercheurs qui font les évaluations), mais dans la plupart des cas ce sont des maisons d'édition qui garantissent la réussite du processus : elles se chargent d'identifier les évaluateurs, parfois (même si rarement) les payent, suivent le processus pour s'assurer de sa transparence et de son avancement. C'est grâce à ce travail qu'une maison d'édition acquiert une plus ou moins bonne réputation et parvient finalement à augmenter la valeur symbolique des textes qu'elle publie. Un ouvrage paru chez un grand éditeur a un fort prestige, car son niveau de légitimation sera supérieur. Or le Web met en question ce modèle, car il produit de nouveaux processus de légitimation : le blog d'un chercheur peut bénéficier d'une immense visibilité, par exemple, et peut peser dans la légitimation de contenus. Sans maison d'édition ni même processus d'évaluation, le contenu peut être cité dans des travaux scientifiques. Un autre aspect qui met en crise les modèles de légitimation provient des outils de recherche : les moteurs de recherche généralistes sont de plus en plus utilisés par les chercheurs pour leurs travaux scientifiques [Takševa, 2012] à la place des catalogues des bibliothèques, par exemple. Or cela implique que les moteurs de recherche – et en particulier Google Search – commencent à obtenir une fonction de légitimation : ce que l'on trouve sur Google est pertinent et, finalement, fiable. Les dispositifs de légitimation de l'édition papier sont donc remis en question et nous sommes appelés à repenser le processus de validation scientifique.

Un deuxième enjeu est lié aux modèles économiques. Le Web a rendu possibles des formes de publication à très bas coût. Un contenu peut en effet être publié en épargnant les frais d'impression – qui représentent une part importante des frais éditoriaux. De façon plus radicale encore, un chercheur peut déposer un texte sur un dépôt ouvert (comme HAL, ou les dépôts des universités, ou encore des dépôts privés comme Academia.edu ou ResearchGate) sans même avoir à payer un hébergement, s'il a par exemple recours aux services d'une plate-forme comme WordPress (la plus utilisée dans les années 2010 pour ce type d'activité). Dans ce cas, les frais éditoriaux sont réduits à zéro. Bien évidemment, une publication de ce type ne bénéficie pas (ou pas complètement) de la médiation de la fonction éditoriale : il n'y a pas de révision par les pairs, de mise en forme, de diffusion, etc. Mais cette possibilité pose une question importante à l'édition savante : quel est son rôle et quel est le bon prix que le lecteur peut être prêt à payer pour profiter du travail éditorial ? La possibilité de la diffusion gratuite impose une réflexion pour redéfinir à la fois le rôle de l'édition dans le domaine

scientifique et ses coûts. Cette réflexion devient encore plus urgente dans la mesure où les grands diffuseurs numériques (Elsevier, Springer, Wiley-Blackwell, Taylor & Francis), plutôt que de corrélérer la baisse des prix de production à une baisse des prix d'achat, ont au contraire augmenté ces derniers [cf. par exemple MacColl, 2014]. Les profits de ces grands diffuseurs sont passés de 665 millions de dollars en 1991 à environ deux milliards en 2012 [Haustein *et al.*, 2015].

Le troisième enjeu est lié aux possibilités qu'offre le numérique par rapport au papier. Les contenus peuvent notamment être structurés grâce à un balisage fin des informations qui permet ensuite une exploitation automatique – notamment à des fins de recherche. C'est ce que l'on fait, par exemple, avec des métadonnées – soit des informations structurées à propos du document : titre, auteur, date de création, etc. C'est aussi ce que l'on peut faire avec des langages de balisage comme le XML qui permet d'explicitier le sens et la valeur de certaines parties des documents : par exemple, spécifier que le mot « Athènes » est le nom français d'une ville qui est la capitale de la Grèce et qui a certaines coordonnées géographiques. Ajouter ce type d'informations dans un texte peut être très utile d'un point de vue scientifique, car les données à l'intérieur des documents sont exploitables : ainsi, un géographe pourrait vouloir repérer tous les textes qui parlent d'Athènes et de ses alentours et faire une recherche dans les textes à partir des coordonnées géographiques. Une autre fonctionnalité importante du numérique est de proposer des versions « augmentées » ou « enrichies » des textes, à savoir des versions qui contiennent plusieurs types de médias (vidéo, son, image) incorporés dans le texte. Exploiter ces possibilités demande des compétences – techniques et théoriques – qui ne faisaient pas partie des compétences éditoriales pour l'édition imprimée. Les acteurs de l'édition scientifique sont donc appelés à repenser leur fonction et à acquérir de nouvelles compétences.

Édition et droits d'auteur

Naissance des droits d'auteur et naissance de l'édition

Les analyses de l'impact du numérique dans les différents secteurs de l'édition nous ont fait remarquer l'importance du concept d'auteur et de son statut dans l'édition. Or ce statut ne s'est stabilisé qu'au terme d'une longue histoire, au cours de laquelle la question des droits a joué un rôle fondamental.

« Le statut d'Anne », également connu comme « Copyright Act de 1710 », est un texte adopté par le parlement britannique et qui est habituellement considéré comme l'acte de naissance du droit d'auteur : il s'agit là, du moins, de la première loi qui réglemente explicitement la publication et la republication de livres en limitant la possibilité de copier et de diffuser sans l'autorisation de l'auteur. Il est intéressant de constater que cette loi apparaît à une période clé de l'institutionnalisation de l'édition, processus auquel elle contribua fortement. Les droits d'auteur sont en effet un pilier de l'édition imprimée, car ils structurent la condition même d'existence des maisons d'édition : sans droit d'auteur, le modèle économique ayant permis à ces maisons de se développer et d'avoir une place si centrale dans la production et la circulation des contenus n'existerait pas.

L'invention de l'imprimerie a créé une certaine séparation entre le contenu et le support sur lequel il est publié – une séparation auparavant moins évidente. La mécanisation de la copie rendue possible par la presse à caractères mobiles permettait en effet d'imprimer un même contenu pratiquement sans effort et à un coût réduit. Au contraire, le manuscrit mettait sur le même plan la valeur de l'objet et la valeur de son contenu : les deux étaient difficilement dissociables. Dans le cas d'un livre imprimé, il est devenu plus aisé de distinguer le contenu – reproductible – de l'objet – unique. De là l'effort théorique pour inventer la notion de droit d'auteur : l'auteur n'est pas propriétaire de l'objet livre – qui peut donc être vendu –, mais de son contenu – ou pour être plus précis, d'une certaine manière de formuler des idées qui s'inscrit ensuite dans le texte imprimé [Rose, 1993]. L'idée contemporaine de droit d'auteur est donc née en réponse à un problème économique et social bien précis, lié surtout à une technologie singulière de production et diffusion des textes : l'imprimerie.

Souligner la contemporanéité de l'émergence des lois sur les droits d'auteur et l'institutionnalisation de l'édition papier implique de reconsidérer ce modèle juridique à une époque où émergent de nouvelles technologies de production et de circulation des contenus. Le numérique fait naître des problèmes et des enjeux économiques et sociaux très différents, c'est pourquoi il est nécessaire d'identifier ses caractéristiques spécifiques pour comprendre comment notre idée des droits d'auteurs est appelée à changer.

Formes de droit d'auteur – États-Unis/Europe

Avant d'identifier les défis posés par le numérique en termes de droit d'auteur, il est nécessaire de spécifier les deux différentes interprétations auxquelles cette notion a pu donner lieu, en Europe d'abord, puis aux États-Unis.

En général, s'agissant de l'édition de textes, le droit d'auteur se fonde sur trois principes [Rose, 1993] :

1. la propriété : l'auteur est propriétaire du contenu de l'œuvre qu'il a produite ;
2. la responsabilité : l'auteur est responsable, moralement et légalement, du contenu de son œuvre ;
3. la singularité : le contenu de l'œuvre représente la singularité et l'originalité de l'auteur.

On comprend que la propriété n'est possible que si on identifie une singularité du contenu qui dépend justement de l'auteur. N'importe qui pourrait écrire une histoire du droit d'auteur, alors pourquoi le texte que vous êtes en train de lire appartient-il aux auteurs de ce livre ? Parce qu'elle représente la singularité de ces auteurs – ce qui en fait quelque chose d'unique. Il est aussi évident que, pour en revendiquer la propriété, il est nécessaire d'accepter la responsabilité des contenus : l'auteur répond de ce qu'il écrit. Cette responsabilité – comme on le verra dans le chapitre IV – implique aussi la légitimation du contenu : l'auteur garantit la fiabilité de ce qu'il écrit.

Or la différence fondamentale entre l'interprétation européenne et l'interprétation américaine du droit d'auteur est que la première associe au droit patrimonial (le fait que l'auteur est propriétaire du contenu) un droit moral (le respect de la singularité de l'auteur), tandis que la seconde se concentre sur le droit patrimonial, sans mentionner le droit moral. Dans la législation européenne, le droit patrimonial est limité dans le temps – dans la plupart des cas 70 ans après la mort de l'auteur –, tandis que le droit moral est pérenne. Ce dernier impose le respect de l'auteur et de son œuvre au-delà des enjeux économiques. La législation européenne souligne, avec cette distinction entre droit patrimonial et droit moral, la différence entre les trois principes du droit d'auteur : la responsabilité et la singularité de l'auteur pouvant être pensées séparément par rapport à la propriété, elles peuvent perdurer après la fin des droits matériels.

Cette différence n'est pas sans importance : elle rend notamment problématique la gestion des droits d'auteurs au niveau international. Le principe du droit moral est en effet repris par la Convention de Berne pour la protection des œuvres littéraires et artistiques (1886), accord dont la fonction est de régler le droit d'auteur au niveau

international. La convention stipule que dans chaque État signataire les ouvrages étrangers bénéficient de la même protection que les œuvres locales. Par exemple : un ouvrage d'un auteur français sera dans le domaine public au Canada cinquante ans après la mort de l'auteur, selon les règles canadiennes, même si en France le droit d'auteur s'étend à soixante-dix ans après la mort de l'auteur. Inversement, un ouvrage canadien sera dans le domaine public en France soixante-dix ans après la mort de l'auteur, même si au Québec la loi vaut seulement pour cinquante ans.

Problèmes actuels

Les considérations sur la Convention de Berne permettent de comprendre comment les technologies numériques peuvent mettre en crise les lois sur le droit d'auteur. La convention s'appuyait sur l'idée que les livres circulent physiquement dans un État ou dans l'autre. L'accessibilité d'un livre dans un État n'impliquait pas son accessibilité dans l'autre. Le fait qu'un livre soit tombé dans le domaine public au Canada, mais pas en France ne posait donc que peu de problèmes : les Français n'auraient pu accéder à l'ouvrage qu'en allant physiquement au Canada. Or, dans le cadre de la circulation numérique, un problème se pose : si un livre est mis à disposition gratuitement sur un serveur au Québec, qu'empêche un Français de le télécharger, même s'il est encore sous droit d'auteur en France ? Cet exemple montre le premier défi posé au droit d'auteur à l'époque du numérique : la déterritorialisation des contenus. Un contenu sur le Web est en effet potentiellement accessible partout dans le monde. S'il est vrai que l'on peut limiter cette accessibilité – notamment en bloquant les adresses IP étrangères – il est aussi vrai que la structure même d'Internet – distribuée, non hiérarchisée – rend très difficile ce type de contrôle et très facile son contournement. Le premier défi posé au droit d'auteur par le numérique est donc la reconfiguration des frontières de circulation des contenus.

Le deuxième défi est lié au coût de reproduction – pratiquement inexistant – des objets numériques : un fichier peut être multiplié sans aucun coût réel. L'édition imprimée avait déjà posé en son temps le problème de la reproductibilité : c'est justement de cette possibilité de reproduction à un coût assez bas qu'était née l'idée moderne de droit d'auteur. Mais que se passe-t-il quand le support ne nécessite plus aucun coût de production ou de reproduction ? Si un lecteur est prêt à payer pour posséder un livre, c'est notamment parce qu'il reconnaît sa valeur objective en tant qu'objet. Lorsque l'on parle d'un fichier, le lecteur sait que l'objet en tant que tel n'a presque aucune valeur.

Le troisième défi est déterminé par la surabondance de contenus sur le Web. L'édition papier est basée sur une économie de la rareté : l'information est rare et donc précieuse. Le web met en place une économie de l'abondance : les contenus, les informations qu'on y trouve sont innombrables. Si dans le cadre d'une économie de la rareté on peut être prêt à payer un auteur pour produire de nouveaux contenus, cela a-t-il encore du sens dans notre culture numérique, où le lecteur, plutôt qu'on lui en propose sans cesse de nouveaux, a davantage besoin qu'on l'aide à se repérer parmi les contenus existants ?

Plusieurs modèles : du DRM aux creative commons

Le développement de l'édition numérique a été accompagné d'une réflexion importante sur les outils juridiques appropriés. Il s'agit de répondre à deux types de besoins *a priori* contradictoires : d'une part, le souhait de contrôler les usages des fichiers numériques contenant les œuvres et notamment de lutter contre le piratage, et d'autre part la volonté d'étendre la diffusion des œuvres en redonnant du pouvoir à leurs auteurs.

Pour les éditeurs la problématique du piratage était d'autant plus importante que leur passage au numérique eut lieu à une époque où le secteur de la musique était confronté à une explosion des téléchargements illégaux. Cette période a donné lieu à la création, entre autres, de solutions techniques de contrôle des usages, communément regroupées sous le terme de DRM (Digital rights management) ou MTP (Mesures techniques de protection). Les DRM sont des logiciels qui, associés aux fichiers des œuvres, permettent aux éditeurs de définir précisément les usages qu'ils souhaitent interdire ou autoriser. Cela peut concerner la possibilité d'imprimer, de copier ou d'accéder aux fichiers. Un autre outil est le *watermarking* (filigrane) des fichiers qui, s'il ne vise pas une limitation des usages, permet une traçabilité des fichiers (un usager sera moins enclin à diffuser illégalement un livre numérique sur lequel figure son nom).

Dans le même temps, de nouvelles propositions juridiques ont émergé pour, au contraire, faciliter la diffusion, le réemploi et le partage de contenus. Parmi ces propositions, la plus visible est celle des licences *creative commons*, lancées en 2001 à l'initiative du juriste américain Lawrence Lessig, qui permettent à un auteur de définir *a priori* les conditions d'usages et éventuellement de réutilisation de ses œuvres. Ces licences précisent les modalités de mention de l'auteur et les possibilités d'utilisation commerciale et de modification d'un contenu. Une licence CC0 permet même de renoncer à ses droits, dans la limite du cadre législatif de chaque pays.

Tableau 1. Les différents types de Digital rights management (DRM)

Schéma de DRM	Propriétaire	Plate-forme de distribution	Particularités
DRM d'Amazon	Amazon	Amazon (Kindle)	Interdiction de copier ou de transférer un e-book dans une autre bibliothèque. Certains e-books peuvent être prêtés à un autre utilisateur pour une durée de 14 jours.
Adobe Digital Experience Protection Technology (ADEPT)	Adobe	Adobe Digital Editions	Visionnement des contenus sur un maximum de six appareils.
FairPlay	Apple	iBooks Store	Les contenus achetés ne peuvent être lus que sur les liseuses d'Apple.
Marlin	Marlin Developer Community (MDC)	Kno	Marlin Developer Community est composé de cinq sociétés : Intertrust, Panasonic, Philips, Samsung et Sony. La vente de la licence est assurée par la Marlin trust management organization (MTMO)

L'éditorialisation

Une définition de l'éditorialisation : éditer dans l'espace numérique

Les changements et les problématiques soulevés par le numérique ont fait émerger une nouvelle terminologie et de nouvelles approches théoriques du fait éditorial. Une d'entre elles semble avoir acquis une importance fondamentale, au moins dans le domaine francophone : la théorie de l'éditorialisation.

Le terme « éditorialisation » a été de plus en plus employé par la communauté scientifique francophone, à partir du début des années 2000, pour faire référence à la production et à la circulation des contenus dans les environnements numériques, mais il est parfois difficile de saisir la signification exacte que les chercheurs lui attribuent.

Bruno Bachimont met pour la première fois en avant le concept en 2007 [Bachimont, 2007] : l'éditorialisation désigne alors une activité éditoriale réalisée à partir des fragments indexés d'un document. Bruno Bachimont se sert de l'éditorialisation pour décrire un passage important : celui du document non numérique au document numérique. Il s'agit d'un transfert d'informations restructurées et réagencées pour être adaptées à l'environnement numérique. Dans le processus d'adaptation s'opère une

transformation créative des ressources qui sont réutilisées à de fins nouvelles. Le terme est donc lié à cet environnement, et souligne la nécessité de l'adaptation dans le passage de contenus non numériques aux supports numérique.

En raison de l'augmentation progressive de l'activité éditoriale numérique, le concept d'éditorialisation a été repris par de nombreux chercheurs et acteurs de la publication numérique. Le travail d'un groupe de recherche dirigé par Gerard Wormser, lié en particulier à la revue numérique *Sens public*, a permis à partir de 2008 d'élargir la notion d'éditorialisation aux contenus nativement numériques [Vitali-Rosati, 2016].

Selon une première acception – plutôt restreinte – l'éditorialisation désigne l'ensemble des appareils techniques (réseau, serveurs, plates-formes, *Content management systems* (CMS), algorithmes de moteurs de recherche...), des structures (hypertexte, multimédia, métadonnées...) et des pratiques (annotation, commentaires, recommandations *via* les réseaux sociaux...) qui permettent de produire et d'organiser un contenu sur le Web. En d'autres termes, l'éditorialisation est un processus de mise en forme et de structuration d'un contenu dans un environnement numérique. En ce sens, on dira que l'éditorialisation qualifie ce que devient l'édition sous l'influence des technologies numériques. Cette définition rapproche l'éditorialisation de ce qu'on appelle aussi « curation de contenu » qui consiste à choisir, mettre en relation et mettre en forme une série de contenus dans un environnement numérique afin d'en faire une unité cohérente et compréhensible pour les lecteurs. C'est la définition de l'éditorialisation que donnaient en 2010 Dacos et Mounier : « La valorisation d'un corpus par la sélection des textes, par la mise en œuvre des collections, par l'établissement d'index thématiques, par la mise en place régulière de focus éditoriaux en fonction du type de public » [Dacos et Mounier, 2010].

Cette première définition comprend cependant un inconvénient majeur, en négligeant le fait que, dans notre culture numérique, presque tout devient édition : un restaurant est « éditorialisé » sur TripAdvisor ou sur Google Maps, notre identité est « éditorialisée » sur Facebook, sur le site de notre employeur, sur AirBnB et sur des milliers d'autres plates-formes. En d'autres termes, tout objet n'existe que parce qu'il est présenté et structuré dans l'espace numérique.

Par conséquent, nous pouvons formuler une définition plus large de l'éditorialisation, qui vient alors désigner l'ensemble des dynamiques – soit les interactions des actions individuelles et collectives avec un environnement numérique particulier – qui produisent et structurent l'espace numérique.

Cette définition s'appuie sur une hypothèse préalable : le fait qu'il existe une hybridation entre l'espace numérique et l'espace non numérique. Structurer l'espace numérique signifie donc structurer l'espace en général.

Différences par rapport à l'édition

À partir de cette définition, nous pouvons souligner une série de différences entre l'éditorialisation et l'édition, qui peuvent être résumées en trois points : premièrement le changement technique, deuxièmement le changement de l'objet éditorialisé par rapport à l'objet édité et troisièmement le caractère ouvert du processus d'éditorialisation par rapport au processus éditorial. Soulignons que ces deux dernières caractéristiques constituent aussi les deux différences fondamentales entre l'éditorialisation et la curation de contenu.

Le changement technique qui caractérise l'éditorialisation est évident : le développement et la large diffusion d'outils numériques impliquent un bouleversement des pratiques. L'édition imprimée est basée sur un ensemble déterminé de techniques et de technologies qui s'est développé à partir du XV^e siècle. C'est justement l'ensemble de ces technologies qui a permis l'émergence de l'édition telle que nous la connaissons, avec ses modèles conceptuels, ses règles – dont le droit d'auteur –, sa fonction sociale et politique. Mais le changement technique n'est que l'une des caractéristiques de l'éditorialisation. Si l'on se limitait à ce changement, on pourrait se passer du néologisme et tout simplement appeler « édition numérique » le processus éditorial refaçonné par les outils numériques.

Or les modèles de production et de gestion de l'information qui émergent avec le Web impliquent un autre changement majeur : l'action de médiation représentée par la fonction éditoriale s'est en effet élargie à toute sorte d'objets. On ne traite plus seulement des contenus, on traite plutôt des objets. Il ne s'agit pas tellement de structurer les informations que nous avons sur quelque chose, par exemple sur une personne, ou sur un objet ; il s'agit plutôt de structurer la place que cette chose occupe dans le monde : on éditorialise les choses et non les informations sur les choses. Ainsi une plate-forme comme AirBnB met en forme, diffuse et légitime non pas des informations sur des appartements, mais les appartements eux-mêmes, ce qui donne une valeur opérationnelle au résultat de l'éditorialisation : on peut ensuite louer un appartement depuis la plate-forme. Enfin, l'édition est un processus bien délimité et bien défini, tandis que, en raison

de la structure même des environnements numériques, l'éditorialisation est un processus qui reste ouvert, toujours inachevé et toujours collectif.

Un processus ouvert

Concentrons-nous sur cet aspect ouvert de l'éditorialisation qui la distingue plus clairement de l'édition classique. L'éditorialisation consiste en une série d'actions en mouvement qui n'ont ni commencement ni fin bien définis. Un processus d'éditorialisation est toujours en cours ; il est nécessairement dans une dynamique de mouvement. La nature processuelle de l'éditorialisation rend complexes l'identification et l'isolement d'un acte d'éditorialisation unique et particulier : chaque processus d'éditorialisation est lié d'une certaine façon à d'autres, et il est impossible de délimiter exactement une chaîne précise d'actions. Considérons par exemple la publication d'un article académique, qui montre que, même lorsque l'exigence éditoriale est élevée et le contrôle du contenu renforcé, le processus d'éditorialisation reste ouvert. Ceci est vrai, à plus forte raison, pour toutes les autres formes d'éditorialisation.

Publier un article scientifique dans une revue imprimée relève d'un processus qui peut être délimité et isolé : il implique successivement un auteur qui écrit l'article, le comité éditorial de la revue à qui est proposé cet article, probablement deux ou trois évaluateurs, ainsi qu'un groupe d'individus qui travaillent pour la maison d'édition de la revue. On peut identifier et compter ces individus : ils peuvent être plus ou moins nombreux selon la taille de la revue, mais leur nombre restera toujours bien défini. Le processus éditorial commence quand l'article est proposé à la publication par l'auteur, et s'achève quand la revue est imprimée. Après cela, l'article devient un objet stable et statique. On pourrait objecter que la distribution de la revue, les réactions de ses lecteurs, le nombre de citations, etc., sont aussi importants et mériteraient d'être considérés comme faisant partie du processus éditorial. Même si cela est sans doute vrai, il n'en demeure pas moins indéniable que l'étape de l'impression produit une rupture importante dans le processus : l'article ne changera plus d'état après sa publication.

Le groupe de personnes travaillant à la publication suit le processus du début à la fin, depuis le moment où l'auteur propose l'article jusqu'à l'impression.

Dans le cas d'un article en ligne, identifier sa publication comme un processus isolé et délimité est difficile sinon arbitraire, pour ne pas dire impossible. Considérons le processus éditorial à partir du moment où l'auteur a achevé d'écrire son texte et le propose à la publication. Mettons pour l'instant entre parenthèses la phase d'écriture elle-

même. Le début du processus est assez similaire à celui de la publication papier : il y a un comité éditorial, bien défini et composé d'un nombre précis de personnes, probablement deux ou trois évaluateurs, ainsi qu'un groupe d'assistants éditoriaux qui met en forme l'article et le publie sur la plate-forme en ligne. Mais le processus ne s'achève pas là. En premier lieu, la place que l'article occupe n'est pas encore déterminée. Dans le cas d'une revue papier, l'éditeur décide où l'article doit apparaître : il le place par rapport aux autres articles, choisit s'il doit figurer en couverture ou non de sa revue. Pour un article en ligne, sa position par rapport à d'autres contenus et, finalement, sa signification, dépendent d'un ensemble de facteurs qui échappent au contrôle de l'équipe éditoriale. Par exemple, la place qu'il occupera dans les listes de résultats des moteurs de recherche, d'un réseau social ou d'un site institutionnel. Si, dans le cas de l'édition papier, les relations entre les différents articles sont proposées et organisées par les éditeurs – qui rassemblent sciemment certains matériels dans le même objet imprimé –, dans le cas d'un article en ligne, nombre de ces relations sont au contraire produites par d'autres plates-formes qui agrègent les contenus. Par exemple, la plate-forme française Rechercheisidore indexera l'article et le placera dans des listes basées sur des métadonnées. L'article sera cité et repris sur différents réseaux, par exemple sur Twitter ou sur Academia, ou sur un dépôt institutionnel comme HAL. La conversation et les discussions sur le texte pourront alors se développer à l'intérieur des communautés en ligne.

Bien évidemment, la même structure de diffusion et de commentaires existe pour les articles imprimés, mais dans l'environnement numérique, ces structures font partie intégrante de l'article, puisqu'elles existent dans le même espace que celui-ci. Selon l'analyse de Genette [2002], dans une revue papier, on peut établir une distinction entre le texte et l'épitéxte, puisque ce dernier appartient à un autre espace – un autre livre ou une autre revue. Dans le cas d'un article numérique, en revanche, les commentaires, les listes des résultats et les recommandations sur les réseaux sociaux sont tous dans le même espace : le Web. Ils font partie intégrante de l'article. De plus, l'article n'est guère cristallisé, il peut être modifié très facilement. L'article imprimé ne peut, quant à lui, être modifié, à moins de produire une seconde édition. De plus, si l'article numérique peut être transformé facilement, il peut aussi être copié et réutilisé dans d'autres contextes. Même si l'éditeur essaie de limiter ce type de pratiques, elles sont inhérentes à l'environnement numérique : copier ne demande aucun effort et devient de fait une pratique commune. L'article peut alors être présenté sous différentes formes et dans différents espaces. Par exemple, la publication d'un article sur un dépôt institutionnel est

de plus en plus courante. La vie de l'article continue ainsi après la fin du travail des éditeurs, qui n'exercent plus aucun contrôle sur lui. Le processus éditorial est ouvert dans l'espace et dans le temps : dans l'espace, parce qu'il n'est pas limité à une plate-forme ou à un groupe spécifique de personnes comme c'était le cas pour la version papier ; dans le temps, parce que l'article n'est plus figé à un moment donné, comme c'était le cas dans le modèle imprimé.

En vérité, cette ouverture se manifeste en amont du processus éditorial, dès la phase d'écriture. En effet, l'auteur peut par exemple posséder un blog sur lequel il livre une première intuition à l'origine de l'article. Il peut ensuite recevoir des commentaires de la part de ses collègues ou de ses lecteurs, recevoir des suggestions de lecture, des informations complémentaires et tenir des discussions à ce sujet. Le travail d'écriture de l'article est ainsi partie prenante du processus d'éditorialisation.

Cette ouverture propre au processus d'éditorialisation implique une instabilité à plusieurs niveaux : une revue n'est plus une forme de circulation stable, puisque l'accès à l'article ne dépend plus seulement d'elle, mais d'autres plates-formes et outils, tels que les moteurs de recherche ou les réseaux sociaux. L'article devient un objet instable qui peut être fragmenté, car on peut notamment en extraire des données et considérer ces dernières comme des unités originaires pouvant être agrégées de façon différente. La nature processuelle de l'éditorialisation est donc profondément liée à l'idée de fragmentation déjà soulignée par Bruno Bachimont.

À coup sûr, ces qualités caractérisent aussi dans une certaine mesure l'édition papier. Mais encore une fois, ce principe ne déroge pas à l'idée selon laquelle il n'y a pas vraiment eu de « révolution numérique », car la culture numérique s'inscrit dans la continuité de la culture prénumérique. Les différences sont davantage une question de degré que de qualité. Les tendances culturelles auxquelles nous assistons aujourd'hui ont des origines qui remontent loin dans le passé, et qui ne sont donc jamais complètement nouvelles. L'aspect processuel de la culture numérique en est un exemple. Ce phénomène n'est pas complètement inédit, mais il était moins prononcé dans la culture papier, culture qui insistait plutôt sur la possibilité de contrôler un processus bien défini et délimité.

Notre exemple montre un processus ouvert, relevant d'une certaine fragmentarité et qui repose sur l'interaction d'un groupe très inclusif d'acteurs. L'éditorialisation met l'accent sur cette ouverture et donc sur une certaine perte de contrôle de l'éditeur sur le contenu éditorialisé.

Ces caractéristiques sont encore plus évidentes dans le cas de l'éditorialisation sur un blog, ou sur un wiki (*cf.* p. XXX). Le concept d'éditorialisation se présente donc comme un outil théorique très utile pour comprendre des pratiques qui s'apparentent à l'édition en changeant le sens de la fonction éditoriale et de ses acteurs. L'éditeur perd sa place centrale, les contenus deviennent un type d'informations à mettre en forme – parmi d'autres – et l'accent est finalement mis sur une généralisation de la présence de la fonction éditoriale dans les environnements numériques.

II / La production des contenus

La plate-forme et l'écriture

Les techniques et la pensée

Les outils façonnent la pensée. Ce que nous pouvons penser et ce que nous pouvons dire résulte d'une dynamique dans laquelle les outils et les techniques jouent un rôle fondamental. L'idée spiritualiste selon laquelle la pensée se construit au-delà de toute matérialité et que les outils et les techniques ne servent qu'à la faire apparaître a été profondément critiquée, en particulier au XX^e siècle : pensons par exemple à la célèbre phrase de McLuhan « le médium est le message » ou aux analyses de Jack Goody [1979] sur le rapport étroit entre l'écriture et la naissance de la pensée scientifique [Deseilligny, 2013].

Ce qui nous intéresse plus particulièrement ici est le fait que les outils et les techniques d'écriture ne sont pas neutres par rapport aux contenus que l'on peut produire. Pour être plus précis : les techniques et les outils d'écriture, mais aussi les modes de diffusion des textes, conditionnent et façonnent ce que l'on écrit.

Du côté des modes de diffusion, on peut ainsi souligner la différence fondamentale entre la forme orale et la forme écrite. Comme le remarque Christian Vandendorpe, les contenus transmis dans les traditions orales sont caractérisés par un aspect formulaire – c'est-à-dire ponctué de formules récurrentes telles que l'on peut en observer dans les poèmes homériques, par exemple – qui limite de fait l'éventail des potentialités des récits, lesquels ont tendance à se standardiser. « Les cultures orales ne s'expriment pas seulement en formules, elles pensent en formules » [Ong, 1977 cité par Vandendorpe, 1999, p. 26]. L'écriture permet l'émergence de contenus à l'organisation plus complexe, ainsi que leur conservation et leur consultation. Pour autant, le papyrus – support linéaire qui ne peut être consulté que dans un ordre prédéfini – n'offre pas encore les possibilités du codex, support divisé en pages, plus maniable, qui permet de passer facilement d'un point du texte à un autre. Cette distinction a des conséquences majeures : aurait-on pu penser une encyclopédie sur papyrus ? Le concept d'encyclopédie serait-il apparu sans l'invention du codex ?

Quant aux techniques et aux outils à proprement parler, on connaît déjà l'influence des instruments sur le statut de l'écriture : de la plume – réservée à une élite – à la

démocratisation du stylo à bille, un changement radical de la valeur symbolique de l'écriture s'est opéré. Entre l'imagerie de l'auteur à sa table de travail, plume à la main, et l'omniprésence industrielle du stylo à bille qui, comme l'affirme Umberto Eco, est probablement le « seul exemple de socialisme réalisé », on s'aperçoit qu'écrire n'a plus le même sens. Mais outre la dimension symbolique, il est évident que la vitesse de l'écriture, tout comme la valeur économique du papier et son format, ou encore les possibilités d'effacer ce qui a été écrit, changent non seulement les pratiques d'écriture, mais aussi le contenu même de ce que l'on peut écrire.

Les outils numériques

Les outils d'écriture et de diffusion n'ont cessé d'évoluer dans l'histoire de nos sociétés. Les changements survenus à la fin du XX^e siècle et au début du XXI^e sont cependant particulièrement impressionnants. On peut résumer ces changements en cinq grandes catégories : 1) la maniabilité du texte ; 2) le rapprochement entre texte manuscrit et texte diffusé ; 3) la multiplication des formats ; 4) les possibilités de structuration ; 5) les possibilités de recherche de ressources pour l'écriture.

Tout d'abord (point 1), l'introduction des logiciels de traitement de texte, dont la diffusion a été généralisée à partir des années 1980 [Kirschenbaum, 2016], a profondément changé les possibilités de manipulation du texte. À la différence de l'écriture au stylo ou à la machine à écrire, le texte numérique peut être facilement effacé, copié, collé, restructuré. Cela implique notamment (point 2) le brouillage de la frontière entre le brouillon et la version définitive du texte, ainsi que la disparition progressive de figures professionnelles comme le dactylographe. Entre l'écriture et la diffusion des contenus, le processus éditorial compte ainsi une médiation en moins. La forme même du texte – à partir des logiciels WYSIWYG (acronyme de l'anglais « *what you see is what you get* », « ce que vous voyez est ce que vous obtenez » parfois traduit par « tel-tel » en français) – se rapproche sensiblement de la forme finale du texte imprimé. Ces deux premiers changements œuvrent en faveur d'une plus forte accessibilité à la publication pour tous. Ils nourrissent par ailleurs un sentiment de marginalisation des processus de médiation entre le manuscrit et le texte publié.

Mais il ne faut pas faire l'erreur de considérer les outils numériques comme un tout homogène (point 3) : s'il est vrai que certains logiciels de traitement de texte – et en particulier Microsoft Word – se sont imposés dans les pratiques, il est aussi vrai que les logiciels et les formats n'ont cessé de se diversifier. Il est important de souligner que

chaque format et chaque logiciel porte une idée particulière de ce que signifie « écrire » et « publier ». Les différents types de visualisation du texte influent sur la perception de l'acte d'écrire, mais surtout, la structuration même du matériel écrit change selon les formats. En particulier, il y a une différence fondamentale entre les formats orientés vers la mise en forme graphique du texte et ceux orientés vers la structuration sémantique (point 4). Les premiers requièrent un balisage des attributs graphiques du texte – choix de la police et de la taille du caractère, usage de l'italique ou du gras, disposition du texte sur la page, etc. Typiquement, Word – et tous les autres logiciels WYSIWYG – est fondé sur ce principe. Les formats RTF (Rich text format) et doc (ou docx) sont représentatifs de cette approche. Le sens du texte est exprimé à travers sa mise en forme graphique. Ce type d'approche tend à considérer le texte numérique lors de sa production comme une étape avant l'impression.

Un autre approche consiste à baliser le texte selon sa valeur sémantique : c'est le cas des outils basés sur les technologies XML (Extendible markup language), où l'on indique par exemple qu'une partie du texte est un titre de niveau 1 ou 2, une note, ou, de manière plus précise, qu'un mot est le nom d'une ville ou d'une personne, etc. Les formats sémantiques ont la caractéristique de permettre une forte structuration logique du texte, ce qui était impossible sur papier. En ce sens, ces formats représentent un changement important par rapport au modèle de l'imprimé. Lors de l'écriture, il devient important de baliser ce que l'on écrit par rapport à son sens, afin d'être ensuite capable de manipuler automatiquement le texte.

Pour finir, écrire en environnement numérique signifie – au moins depuis les années 1990 – avoir accès, sur le même support qu'on utilise pour écrire (ordinateur, tablette...), à un grand nombre de sources potentiellement mobilisables (point 5) : le Web propose en effet une quantité impressionnante de contenus qui peuvent servir de documentation. Tous les types d'écriture sont concernés : de la littérature – pensons aux accusations de plagiat de Wikipédia adressées à Michel Houellebecq pour son roman *La Carte et le Territoire* – aux textes scientifiques – lors de l'écriture desquels le chercheur dispose d'un accès numérique à un grand nombre de sources.

Une réflexion s'impose quant à la façon dont le numérique modifie la phase de production des contenus éditoriaux, et aux enjeux intellectuels et politiques qui en découlent. Nous n'avons pas fait qu'inventer des outils permettant de réaliser plus rapidement et plus efficacement des tâches déjà connues : nous faisons désormais face à un environnement d'écriture qui bouleverse profondément ce que signifie écrire.

Parallèlement, il faut insister sur le fait que les outils numériques ne peuvent pas être considérés comme un tout. Il ne s'agit donc pas de chercher comment « le numérique » en général change notre façon de produire des contenus, mais plutôt de réfléchir aux différentes formes d'écriture que les outils numériques proposent, pour comprendre comment les choix technologiques reflètent nos conceptions de l'écriture. En d'autres termes, si l'on ne saurait dire ce que fait exactement le numérique à l'écriture – et en général à la pensée – il faut du moins analyser les visions du monde et de l'écriture au fondement de chaque outil.

Tableau 2. Principaux formats de textes associés aux logiciels qui permettent de les traiter

formats	Extension(s)	logiciels	propriétaire ou non	approche
Text	.txt	Éditeur de texte	Libre	Non balisé
Markdown	.md, .markdown	Atom, MarkPad, Haroopad, MarkdownPad, Sublime, Netbeans, etc.	Libre	Balisé
Rich Text Format	.rtf	Word LibreOffice OpenOffice	Propriétaire	Graphique
ODT	.odt	LibreOffice OpenOffice	Libre	Graphique
DOC	.doc, .docx	Microsoft Word	Propriétaire	Graphique
Hyper text markup language	.html, .htm	Éditeur de texte	Standard W3C, libre	Graphique et sémantique
XML	.xml	Éditeur de texte/Oxygen	Standard, libre	Graphique et sémantique
LaTeX	.tex	Winefish, LaTeXila, TeXShop, TeXmaker, Winshell, etc.	Libre	Graphique

Enjeux stratégiques de la structuration des documents

Les formats de fichiers

Avec le passage au numérique de l'activité éditoriale, l'ensemble des acteurs sont confrontés à un ensemble de problématiques nouvelles. Ces problématiques sont pour certaines techniques, mais elles renvoient également à des enjeux stratégiques,

économiques et éditoriaux. Ce rapport des éditeurs avec les aspects techniques de leur métier n'est pas lié au numérique. La taille des ouvrages, leur pagination, les choix de matériaux, de typographie... l'ensemble de ces choix a toujours fait partie de l'activité éditoriale. Ils restaient toutefois très spécifiques et surtout rattachés de façon exclusive à l'édition. Avec l'édition numérique, le rapport à ces éléments techniques est profondément transformé puisqu'ils relèvent tous de domaines d'activité qui dépassent largement le seul périmètre de l'édition. Les technologies déployées aujourd'hui dans l'édition numérique ont en effet avant tout été développées pour répondre aux besoins de l'informatique d'abord et du Web ensuite. Il s'agit d'un renversement important, puisqu'une part incontournable de l'activité éditoriale devient de fait conditionnée par des choix techniques issus de secteurs d'activité parfois très éloignés de l'édition.

Les conséquences de cette situation sont doubles : d'une part, la mise au point et l'adoption des standards techniques du métier ne relèvent plus uniquement d'acteurs du monde de l'édition ; d'autre part, les éditeurs sont contraints de composer avec des formats ou des technologies qui ne correspondent pas nécessairement à leurs enjeux. Parmi ceux-ci, la question des formats revêt un caractère particulier. En effet, au-delà du choix d'un standard, les formats utilisés pour les productions éditoriales numériques conditionnent à la fois la forme éditoriale proposée par l'éditeur, l'expérience utilisateur du lecteur ou encore les circuits de distribution envisageables et donc les modèles économiques afférents. Il est donc nécessaire d'appréhender cette question des formats en dépassant les seules potentialités techniques pour analyser les enjeux stratégiques qu'ils impliquent pour la production éditoriale et la structuration du secteur.

Lorsque l'on considère les formats de fichiers, il faut le faire sous le triple angle de leurs limites techniques – donc des contraintes qu'ils font peser sur les productions intellectuelles –, de leur degré d'ouverture et de leur statut de standard. Plutôt que de décrire dans le détail l'ensemble des formats envisageables pour une activité d'édition numérique, nous préférons éclairer cette question sous l'angle de la réflexion à déployer du point de vue des éditeurs. Cette approche nous semble la plus pertinente, l'évolution permanente et accélérée du paysage technique pouvant autrement rendre notre propos rapidement obsolète. De plus, d'autres ouvrages traitent de ces questions avec un niveau d'expertise élevé [cf. Boulétreau et Habert, 2014]. Nous opterons donc pour une approche stratégique en prenant comme exemples les formats tels qu'ils existent aujourd'hui. Historiquement, la question du format des fichiers pour l'édition numérique s'est posée sous la double problématique du choix d'un standard et de la protection des fichiers.

Ainsi, les premières productions éditoriales numériques commerciales se firent dans un paysage extrêmement foisonnant de formats utilisables pour le livre numérique. Des plus répandus comme le pdf, doc, odt ou txt à des formats plus spécialisés comme l'OEB (Open eBook), le mobi (format Mobipocket, racheté par Amazon), le azw (format d'Amazon pour ses Kindle) ou le BBoB (format utilisé par Canon ou Sony pour leurs propres liseuses), la question du format était avant tout pour l'éditeur numérique des années 2000 un choix du parc de dispositifs sur lesquels ses ouvrages pourraient être lus. Ce choix était d'autant plus délicat que la conversion d'un format à un autre n'était pas toujours une simple formalité. Ainsi, au travers des formats de fichiers, c'est la structuration des différents écosystèmes du livre numérique qui se jouait au début des années 2000. Ce phénomène n'était pas nouveau pour les acteurs de l'informatique. Cette question des formats et des phénomènes de verrouillage avait déjà été largement étudiée au début de la micro-informatique, quand les formats et le jeu des comptabilités jouaient déjà un rôle majeur. Mais pour les éditeurs, cette question était tout à fait nouvelle. Dans l'univers papier, la question des formats techniques était un enjeu majoritairement délégué aux acteurs de la fabrication, notamment les imprimeurs. Il s'agissait d'un enjeu qui conditionnait avant tout l'efficacité d'une chaîne de production et en aucune façon les marchés ou lectorats potentiels. Avec le livre numérique, l'ensemble des acteurs appréhenda assez tôt le poids de la question des formats en termes de structuration du marché.

La volonté initiale de maîtriser un écosystème qui deviendrait majoritaire se heurta à l'effet dissuasif de cet émiettement des formats sur les lecteurs. En effet, confrontés à ce nouveau marché, la question de la pratique de lecture numérique se combina, pour les « adopteurs précoces » (*early adopters*), à la question du choix d'un matériel en fonction des formats qu'il pouvait traiter et donc de l'offre d'ouvrages publiés dans ces formats – en résumé la question de l'interopérabilité. Cette difficulté dans le choix du matériel de lecture numérique freina le développement du marché et convainquit l'ensemble des acteurs de la filière de la nécessité d'un format standard, partagé par tous. C'est l'ambition du format ePub, dérivé du format OEB en 2005, et sur lequel nous reviendrons plus loin.

Cette dynamique s'enclencha en partie en marge de deux acteurs majeurs du livre numérique, Amazon et Apple. Ces deux entreprises avaient fait des choix de formats différents pour leurs activités de distribution de livres numériques. Amazon avait dès le départ opté pour un format propriétaire (aujourd'hui Amazon Kindle, ou azw). Ce choix

de format s'applique autant aux livres numériques vendus par Amazon qu'à ses liseuses Kindle. L'objectif était clairement de proposer un écosystème maîtrisé et clos fondé sur un format adapté exclusivement à ses dispositifs et intégrant son propre système de gestion des droits numériques (Topaz).

Apple choisit une approche un peu différente. L'application iBooks proposée sur ses téléphones et tablettes peut lire des fichiers aux formats ePub et pdf. Les ouvrages vendus par Apple sont en fait des fichiers ePub mais intégrant le système de gestion des droits FairPlay, spécifique à Apple. Enfin, Apple propose un outil de fabrication de livres numériques dédiés à son application, iBooks Author. Cet outil produit des fichiers dans un format spécifique (iba, pour « iBooks Author »), pour une diffusion et une commercialisation *via* la boutique en ligne d'Apple. Ces deux exemples montrent bien que la question des formats de livres numériques doit se lire à trois niveaux : celui des formats exploitables selon les dispositifs de lecture, celui des formats ouverts ou propriétaires pour la production des fichiers et enfin celui du système de gestion des droits numériques ajouté éventuellement aux fichiers.

Homothétique, fixed layout, reflowable

Du point de vue éditorial, la question des formats informatiques des livres numériques conduit à une réflexion qui dépasse les enjeux d'écosystèmes ou de gestion des droits. En effet, devenus un outil éditorial, les formats informatiques jouent un rôle central dans la définition et la forme des produits éditoriaux tels qu'ils sont conçus et utilisés.

Cette situation ne concerne pas exclusivement les formats numériques. Déjà, dans la production éditoriale sur supports papier, l'environnement technique de production des différents formats conditionnaient en partie les produits éditoriaux proposés au public. L'exemple le plus simple de ce réseau de contraintes techniques est le nombre de pages de certains ouvrages. Ainsi, plusieurs collections de poche comme « Que sais-je » ou « Repères » ont un format de 128 pages qui correspondait à l'origine à un processus déterminé de fabrication (l'assemblage de plusieurs cahiers qui imposaient un certain nombre de pages). Pour le même type de raisons, les cahiers d'écoliers font 96 pages et non 100.

Cette contrainte sur le nombre de pages n'est qu'un exemple : les grammages de papier, la possibilité ou non d'intégrer des illustrations ou de la couleur, les choix typographiques, tous ces éléments contraignent la forme de l'objet éditorial mais

également la production du contenu. La maîtrise de ces contraintes techniques fait partie de l'expertise professionnelle des acteurs du livre. Ils les ont intégrées à leurs réflexions et processus de travail, que ce soit au niveau du contenu ou de la fabrication.

Avec l'édition numérique, plusieurs de ces dimensions techniques sont remises en question. Il ne s'agit pas seulement d'une évolution, mais bien d'un bouleversement du cadre de contraintes dans lequel les activités éditoriales se déploient. Ce bouleversement porte évidemment en lui autant d'opportunités (moins de limitations matérielles) que de nouvelles contraintes (complexité technique, instabilité permanente, etc.). Avec les formats informatiques, l'édition numérique est ainsi confrontée à un choix complexe quant à la forme qu'elle souhaite donner à ces productions. Ce choix, qui se concrétise par des options techniques, est celui du niveau de contrôle dont l'éditeur souhaite disposer sur la mise en page de ses ouvrages. En effet, en passant aux formats numériques, ces derniers offrent à leurs lecteurs des possibilités de modification nettement plus importantes que le support papier (pour lequel elles se limitaient à l'annotation ou au cornage de page). Il est par exemple techniquement possible de proposer de modifier plusieurs paramètres de mise en page tels que la taille des caractères, la police, la couleur du fond de la page, la façon dont les pages s'affichent ou s'enchaînent, etc. Ces possibilités constituent à la fois une réelle valeur ajoutée pour la lecture numérique mais également une perte de contrôle de l'éditeur sur l'expérience du lecteur.

Du point de vue informatique, cette perte de contrôle peut conduire au choix par l'éditeur du *fixed layout* (mise en page fixe), qui consiste à proposer les ouvrages sous une forme homothétique de leur mise en forme sur support papier, en contraignant techniquement l'affichage des pages, en termes d'organisation notamment. Il s'agit d'une approche principalement utilisée pour les ouvrages illustrés. Elle permet à l'éditeur de s'assurer que l'affichage du livre numérique sera optimal lorsqu'il sera lu sur un dispositif ayant les caractéristiques prévues (taille d'écran, résolution...) lors de sa conception. Cette catégorie de formats dits homothétiques comprend notamment le format pdf, mais également le format doc de Word. Dans ces différents formats, le gabarit de la page est défini au départ et la mise en forme se fait dans ce cadre. Tout l'enjeu réside dans la capacité de ces formats à être lus sur des dispositifs ne correspondant pas à ce gabarit de départ. Ainsi, les liseuses ou tablettes proposent des formats différents de l'une à l'autre mais également généralement plus petits que ceux des écrans d'ordinateurs. Dans ce contexte, la difficulté pour les éditeurs consiste à

choisir de construire le livre numérique sur la base d'un format de page qui ne correspondra qu'à une partie des dispositifs.

À l'inverse, une autre approche, baptisée *reflowable*, consiste à optimiser l'adaptation de l'ouvrage au dispositif utilisé. L'objectif est de permettre un rendu pertinent quel que soit le dispositif de lecture ou les paramètres utilisés par le lecteur. Dans cette approche, principalement utilisée pour des textes à la mise en page relativement simple, l'ouvrage s'adaptera au mieux aux différentes configurations. Le travail consistera donc principalement à veiller au respect des standards des formats et à l'interprétation optimale du code de l'ouvrage.

Il faut préciser que cette question de l'homothétie appliquée au livre numérique n'est pas seulement éditoriale. Elle questionne la diversité des aspects de l'édition numérique dont, entre autres, le régime fiscal qui y correspond. Le taux de TVA réduit appliqué au livre numérique « homothétique » a suscité de nombreux débats sur la possibilité d'une définition objective du livre numérique, dans toute la diversité de création qu'il offre aujourd'hui [Bon, 2010].

Livre-Web, livre-application

Pour diffuser ou commercialiser des livres numériques, l'approche a longtemps été celle d'un fichier informatique autonome (c'est-à-dire contenant l'ensemble des éléments nécessaires) et interprétable par des logiciels permettant la lecture, suivant le modèle déjà déployé dans d'autres secteurs comme la musique (avec le format mp3 par exemple) ou la vidéo (avec, il est vrai, une plus grande diversité des formats utilisés).

Pour mettre en œuvre ce modèle, les acteurs du livre numérique ont développé le standard ePub, porté par l'organisme professionnel IDPF (International Digital Publishing Forum). Ce format a l'avantage pour les acteurs du livre de présenter des caractéristiques proches du livre papier : un objet unique, produit en amont de la filière et distribué ensuite dans l'ensemble des réseaux de commercialisation. L'objectif était de retrouver un fonctionnement similaire avec un livre numérique incarné par un fichier informatique au format unique et diffusable dans l'ensemble des réseaux moyennant l'intégration éventuelle de systèmes spécifiques de gestion des droits.

D'autres modèles de développement sont aujourd'hui intégrés par la filière du livre avec des approches très différentes. La première approche consiste à prendre en compte l'évolution du format ePub, dont la version 3.0, adoptée en 2011, offre des possibilités bien plus importantes d'intégration des contenus multimédias. D'un point de vue

fonctionnel, le rapprochement entre le livre numérique et le Web est de plus en plus net. En effet, un fichier au format ePub est en réalité un ensemble de fichiers HTML et CSS regroupés dans un « conteneur » ePub (comme pour une archive zip). Ainsi le format ePub 3.0 intègre-t-il HTML5 et CSS 3.0. Ce rapprochement s'observe à un niveau plus institutionnel avec l'intégration en 2017 de l'IDPF au sein du W3C, en charge de la normalisation des standards du Web. Cette évolution nourrit une réflexion académique et professionnelle autour de la place respective du livre numérique et du Web et d'une possible fusion, faisant du livre numérique un site Web comme un autre.

Une autre approche mise en œuvre depuis l'arrivée de l'iPad en 2010 consiste à considérer les livres numériques comme des applications, au même titre que des jeux ou des utilitaires. Cette logique est notamment à l'œuvre dans le domaine de l'édition numérique pour la jeunesse où la gestion des illustrations et des interactions est centrale. De ce fait, des ouvrages numériques illustrés ou destinés à la jeunesse prennent la forme d'applications et intègrent donc l'ensemble des éléments caractéristiques, y compris le modèle économique, des applications ou des logiciels.

Dans les deux approches, livre-Web et livre-application, l'ensemble des aspects de l'édition numérique sont questionnés. Encore une fois, il ne s'agit pas d'un simple choix technique, mais d'orientations qui couvrent toutes les dimensions de l'édition, des choix éditoriaux aux circuits de diffusion en passant par les modèles économiques et le cadre juridique.

La chaîne de production numérique

Papier et numérique

Parler de la chaîne de production numérique implique nécessairement d'évoquer le rapport entre papier et numérique, et cela pour deux raisons. En premier lieu parce que l'un des enjeux majeurs de l'édition consiste désormais à rendre disponibles en format numérique des contenus qui ont été conçus et réalisés pour le papier. En second lieu, et parallèlement, parce que la totalité des contenus qui sont aujourd'hui destinés à la diffusion papier est produite dans un environnement et avec des outils numériques. Le papier et le numérique, loin d'être, comme on pourrait le croire, en opposition, sont donc tous deux au fondement du monde de l'édition actuelle. Les rapports qu'ils entretiennent n'en sont pas moins complexes : si d'un côté ces deux formes de gestion des contenus

présentent des caractéristiques différentes, elles s'imitent de l'autre et se transforment réciproquement.

Nous pouvons en effet identifier deux mouvements complémentaires caractérisant les rapports entre papier et numérique. Tout d'abord, il est facile de constater combien les environnements numériques essaient d'imiter et de reproduire les caractéristiques et les structures propres au papier. Que l'on pense, par exemple, à l'idée de « page ». Sur écran, le concept de page n'a aucune nécessité. Pourtant, cette notion est loin d'avoir disparu des pratiques d'écriture et de lecture sur écran. Dans la plupart des logiciels de traitement de texte (LibreOffice, Word, Pages, pour les plus courants), on écrit sur des pages, séparées les unes des autres : l'image même de la page est graphiquement reproduite à l'écran. De la même façon, certaines applications de lecture sur écran tactile proposent la possibilité de « tourner » les pages en imitant le geste que l'on accomplit sur un livre papier. Même lorsque les contenus ne sont pas encadrés par une forme graphique rappelant la page papier, la notion a tendance à persister : c'est ainsi que nous consultons des « pages » web.

Ce phénomène a été baptisé par Bolter et Grusin [1998] *remediation* (traduit en français par « remédiation » ou « remédiatisation »). La *remediation* désigne ce mouvement selon lequel un support médiatique reprend et imite des caractéristiques typiques d'un autre médium : dans ce cas, le numérique remédie le papier. Il ne faut pas penser, en revanche, que ce phénomène s'opère dans une seule direction, à savoir du support plus « ancien » au support plus récent : s'il est sans doute vrai que le numérique remédie le papier, il est aussi vrai que le papier remédie le numérique. Plus précisément, notre conception contemporaine de la publication sur papier est profondément affectée par nos expériences dans les environnements numériques. Par exemple, notre façon de lire a changé, tout comme notre rapport aux références, aux citations, aux renvois... Le papier participe dorénavant d'un environnement de circulation des contenus plus vaste, qui inclut nécessairement le Web, les applications mobiles, les écrans, etc.

La numérisation

Commençons par analyser le mouvement d'adaptation de matériaux conçus et diffusés pour le papier dans l'environnement numérique. Ce processus est appelé « numérisation ». Au sens propre, la numérisation est l'opération qui consiste à transformer un signal analogique en un signal numérique. Ce processus implique une discrétisation du signal analogique – qui est continu – réalisée grâce à un échantillonnage.

On peut numériser n'importe quel type d'information : du son, des images, des vidéos ou, évidemment, des textes. Prenons l'exemple du son : concrètement, numériser le son revient à sélectionner des échantillons de l'onde sonore – c'est-à-dire que l'on ne considère pas l'ensemble de l'onde sonore, qui est continue, mais seulement les changements qui se produisent à des intervalles déterminés. Plus court est l'intervalle choisi, plus précis sera l'échantillonnage et plus haute sera la qualité de la numérisation. Le son que l'on obtient de cette manière est essentiellement de qualité inférieure à l'analogique, car il ne rend pas compte de la continuité du son d'origine, mais seulement d'un nombre fini – bien qu'élevé – d'échantillons. Mais le processus de discrétisation permet une simplification de l'enregistrement qui est réduit à une série de chiffres entiers, plus précisément de 0 et de 1.

Le même processus permet de transformer une page imprimée en un fichier numérique. La page imprimée – analogique et continue – est discrétisée *via* un échantillonnage et transformée d'abord en image, ensuite en texte, pour être finalement diffusée dans l'environnement numérique. Ce processus peut être décrit en trois phases :

- 1) Le scanneur transforme la page papier en image, en échantillonnant la continuité de sa trame en une série discrète de pixels. Plutôt qu'un objet continu – comme la feuille de papier –, l'image numérique est une série de carrés disposés l'un après l'autre, qui prennent chacun une valeur de couleur. On discrétise donc la page en un certain nombre de pixels : cela détermine la résolution de l'image, résolution qui est mesurée en ppi (*pixels per inch*). Si nous réglons le scanneur à 300 ppi, par exemple, ce qui représente une résolution moyenne, nous aurons 300 pixels pour chaque pouce (2,54 cm) de papier. On discrétise aussi les couleurs, en établissant combien de valeurs de couleurs différentes un pixel peut prendre. Dans le cas le plus simple, il n'y a que deux valeurs possibles : le pixel peut être soit blanc, soit noir. Toutes les couleurs sont réduites à ces deux possibilités – ce qui est souvent suffisant dans le cadre de la numérisation des textes. La haute résolution de couleurs permet de spécifier des milliards de couleurs (le HDMI 1.3 par exemple arrive à 281 500 milliards).
- 2) Une fois que la page a été transformée en image, il est préférable de transformer cette image en texte. En effet, une image permet au lecteur de lire le texte, mais elle ne permet aucun des avantages propres aux outils numériques comme la recherche de texte, le copier-coller, la mise en page, etc. Il est donc préférable d'avoir recours à un logiciel pouvant reconnaître dans l'image les caractères et transformer ainsi

l'image en texte. Ce processus est appelé « reconnaissance optique de caractères » (ROC) ou « océrisation » (de l'anglais OCR, *optical character recognition*). Les logiciels de reconnaissance optique de caractères permettent de transformer l'image en texte avec une certaine marge d'erreur qui dépend de la qualité du support original et de sa lisibilité, ainsi que de la qualité de l'image numérisée. Ces logiciels ont aussi recours à une phase de post-traitement pendant laquelle un algorithme analyse le texte obtenu linguistiquement (orthographe, syntaxe) et contextuellement pour limiter le nombre d'erreurs.

- 3) Le texte numérique peut ensuite être structuré pour rendre compte de la richesse d'information présente dans le support papier – ou même pour augmenter le nombre d'informations potentiellement utiles. Ainsi, on peut vouloir garder des éléments de mise en forme (police de caractère, taille, couleurs, etc.), mais on peut aussi vouloir ajouter des métadonnées – titre de l'ouvrage, nom de l'auteur, mots-clés, etc. –, ou encore baliser sémantiquement le texte pour indiquer qu'une série de mots correspond au titre d'un ouvrage cité, ou qu'une série de caractères (par exemple « Athènes ») correspond au nom d'une ville. Pour ce faire, il faudra bien évidemment choisir un format – xml, pdf, epub, etc. – et parfois rajouter manuellement les informations que la machine n'est pas capable d'extraire automatiquement du document original. Cette troisième phase ne concerne pas seulement les contenus qui ne sont disponibles qu'en format papier, mais aussi tous les contenus produits aujourd'hui et destinés à une double diffusion papier et numérique – par exemple un article de revue qui sera imprimé et en même temps diffusé par un diffuseur numérique (comme OpenEdition en France, Érudit au Québec ou Muse aux États Unis), ou encore un roman qui sera imprimé, mais aussi diffusé en format ePub ou Kindle.

Ce processus de numérisation ne va pas sans poser problème. En premier lieu, il requiert des choix techniques – concernant la résolution, le type de structuration et de format, les informations à baliser, etc. – qui sont loin d'être neutres. En deuxième lieu, il implique un changement des possibilités de circulation du contenu qui met en crise des dispositifs de gestion des droits et des modèles économiques caractérisant le modèle papier : en effet, est-il encore pertinent d'essayer de limiter la circulation d'un fichier qui peut être multiplié à volonté sans aucun coût ? Peut-on encore faire fonctionner le modèle de rémunération des auteurs qui s'est stabilisé à partir du XVIII^e siècle ? Des entreprises de numérisation massive comme celle de Google Books ont montré à quel point ces

questions sont sensibles. En troisième lieu, le passage du papier au numérique provoque une décontextualisation du document originaire qui en rend l'interprétation problématique. Bruno Bachimont [2007] a démontré que dans le document papier, le contenu et son inscription matérielle sont une même chose : le livre est à la fois le support et le contenu. Dans le cadre des ressources numériques, le contenu se distingue de son inscription : un même contenu peut être interprété de façon différente par la machine. Ainsi, le même texte peut être visualisé dans une taille ou dans une mise en page différente – ce qui rend problématique, par exemple, la citation d'un passage précis – ou peut même être interprété comme une image ou comme une vidéo. C'est pour ces raisons que, loin d'être une opération technique neutre, la numérisation demande une réflexion importante : chaque forme de numérisation porte des valeurs, des interprétations de ce qu'est le contenu, de son sens, de son contexte, de la manière dont il doit être lu et compris, de ses possibilités de circulation et de réception...

Tableau 3. Principaux programmes de numérisation

Projet	Fondateur	Date de création	Types de documents numérisés	Nombre de documents numérisés
Gutenberg	Michael Hart	1971	Livres du domaine public	53 000 en 2017
Google Books	Google	2004 (initialement sous le nom Google Print)	Tous types de livres	25 millions en 2015
Internet Archive	Brewster Kahle	1996	Livres, films, enregistrements audio, images, logiciels et pages web archivées (Wayback Machine)	11 millions de livres et textes, 4 millions d'enregistrements audio, 3 millions de vidéos, 1 million d'images, 100 000 logiciels et 279 milliards de pages web en 2018
Gallica	Bibliothèque nationale de France	1997	Livres, manuscrits, revues, photographies, cartes géographiques, enregistrements sonores, etc., d'origine française	Plus de 4 millions de documents en 2018
HathiTrust	Committee on Institutional Cooperation (CIC), aujourd'hui Big Ten Academic Alliance (BTAA)	2008	Mise en commun des contenus numérisés de plus de 120 partenaires, notamment Google Books, Internet Archive et plusieurs universités américaines et européennes	15 millions de volume en 2017
Canadiana	Conseil des Arts du Canada	1978 (initialement sous le nom d'Institut canadien de microreproductions historiques)	Livres, revues, journaux, publications gouvernementales, photographies, cartes géographiques, enregistrements audio et vidéo, etc. d'origine canadienne	180 000 documents, ou 70 millions de page en 2018
Europeana	Commission européenne	2008	Mise en commun des contenus numérisés de plus d'une centaine d'institutions	Plus de 54 millions de documents en 2017

American Memory	Library of Congress	1994	Archives écrites, sonores et vidéo sur l'histoire américaine	5 millions de document en 2000
--------------------	------------------------	------	---	-----------------------------------

Linéarité et non-linéarité

Considérons à présent les contenus produits d'emblée en version numérique. On pourrait dire qu'à partir de la fin du XX^e siècle, la totalité des contenus qui circulent est produite dans un environnement numérique – y compris ceux pensés pour une impression papier. Les premiers logiciels de traitement de texte apparaissent dans les années 1960 et leur utilisation se diffuse rapidement auprès des auteurs [Kirschenbaum, 2016]. Microsoft Word est lancé en 1983 et conquiert rapidement le marché en s'imposant comme l'un des outils d'écriture les plus utilisés [Liebowitz, 2001]. Dans les années 2000, il est rare de trouver des auteurs qui n'écrivent pas leurs textes en format numérique. Il existe cependant des différences entre les contenus pensés pour l'impression sur papier – même s'ils sont produits avec des outils numériques – et les contenus dont la vocation est d'être diffusés sur support numérique. L'une de ces différences – probablement la plus importante – est le niveau de linéarité ou de non-linéarité du contenu.

Il serait excessivement simplificateur d'affirmer que le support papier est linéaire et que le support numérique est non linéaire : encore une fois, les rapports entre papier et numérique sont plus complexes et ne peuvent être réduits à une série d'oppositions.

Christian Vandendorpe, dans son ouvrage *Du papyrus à l'hypertexte*, préfère ainsi opposer la linéarité à la « tabularité » : la tabularité est en effet « la possibilité pour le lecteur d'accéder à des données visuelles dans l'ordre qu'il choisit » [Vandendorpe, 1999, p. 39]. Vandendorpe montre que cette caractéristique est présente dans toute forme d'écriture (le lecteur peut regarder où il veut, sauter des passages, etc.), mais qu'elle se radicalise dans un support comme le codex – beaucoup moins linéaire que le papyrus – et à plus forte raison dans l'hypertexte.

La notion d'hypertexte – qui désigne un texte électronique relié à une série d'autres textes *via* des hyperliens – souligne l'aspect de non-linéarité propre au support électronique : dès que le lecteur trouve un lien, il peut abandonner la page et passer à une autre, en créant ainsi des parcours de lecture différents. Or cette notion n'est évidemment pas exclusive aux supports numériques, bien que ces derniers en simplifient la mise en place technique. Concrètement, lorsque le lecteur feuillette les pages d'un dictionnaire pour trouver un terme, il accomplit une opération qui relève de la même non-linéarité que celle d'un hypertexte. Par ailleurs, l'idée de relier physiquement les textes entre eux a été formulée à plusieurs reprises dans l'histoire de la pensée, indépendamment des outils numériques. Ainsi, en 1910, Paul Otlet créait le Mondaneum, un répertoire

bibliographique dans lequel les documents étaient reliés mécaniquement les uns avec les autres [Wright, 2014]. En 1945, Vannevar Bush [1945] proposait un bureau mécanique (le Memex) qui devait permettre, grâce à un dispositif mécanique de leviers, de créer des liens entre différents documents stockés dans le même bureau sous forme de microfilms.

S'il est vrai qu'une grande culture de la non-linéarité précède l'apparition et la diffusion des supports numériques, il est aussi vrai, *a contrario*, que le support numérique n'est pas exclusivement utilisé pour produire des documents non linéaires. Certes, le Web peut effectivement être considéré comme un grand hypertexte, mais d'autres formes d'écriture numérique – par exemple les éditions homothétiques avec des formats tels que le pdf ou l'ePub dont nous avons parlé – reproduisent fidèlement la linéarité du support papier, parfois même en la radicalisant. En effet, dans un livrel, il est particulièrement difficile de sauter des pages et de passer d'une partie du texte à une autre. Il est aussi difficile de se faire une idée globale du texte – opération paradoxalement plus aisée avec un support papier.

Les éditions augmentées

Concentrons-nous désormais sur les expériences d'édition numériques fortement non linéaires, en particulier sur l'édition augmentée. Le livre augmenté, ou enrichi, est une forme encore loin d'être institutionnalisée, et dont les contours sont difficiles à saisir [Tréhondart, 2016]. On pourrait le définir de façon minimale comme un livre qui, par la structure et l'agencement de ses contenus, ne peut être imprimé sans perdre des éléments importants. Mais concrètement, il existe des livres fortement hypertextuels – sans parcours de lecture linéaire de référence –, ou des livres qui contiennent des vidéos, des contenus interactifs ou dynamiques, etc. Les livres enrichis peuvent être accompagnés par des versions papier ou numériques homothétiques, mais les deux versions présentent des différences fondamentales. Considérons quelques exemples.

La collection « Parcours numériques » publiée aux Presses de l'université de Montréal propose deux éditions de chaque livre : une édition papier et une édition numérique en ligne. L'édition papier est évidemment linéaire, elle est aussi caractérisée par l'absence d'appareil critique (notes, références). Cette édition est également disponible en format numérique homothétique (pdf et ePub), correspondant parfaitement à la version papier. Une seconde version est proposée en ligne, en HTML. Cette version contient l'ensemble du texte de l'édition papier, augmenté d'un appareil critique (notes et références), de vidéos qui illustrent les contenus du livre, de bibliographies dynamiques

(mises à jour périodiquement), de liens vers des pages d'approfondissement ou vers d'autres contenus... Dans le domaine de la littérature, la maison d'édition Publie.net propose des livres numériques de différents types : certains ne comportent que des éléments textuels qui pourraient être reportés sans aucune perte dans une édition papier (laquelle est d'ailleurs souvent disponible), tandis que d'autres profitent abondamment des possibilités du format numérique. Notamment, la revue numérique de création *D'ici là*, dirigée par Pierre Ménard, comporte plusieurs éléments hypertextuels, des bandes sonores, des vidéos, etc., qui ne pourraient pas exister sur support papier.

L'Alliance for Networking Visual Culture a créé en 2009 Scalar, une plate-forme de production de contenus enrichis qui permet justement de réaliser des livres augmentés. Le principe de Scalar consiste à remplacer la notion de texte linéaire par un dispositif reliant de multiples documents entre eux. Au lieu de créer des pages, on y crée des documents : textes, images, vidéos, sons, références, etc. Chaque document peut être décrit et mis en relation avec tous les autres. Cela permet la constitution de différents parcours de lecture que chaque lecteur pourra actualiser selon ses intérêts. Chaque lien entre un document et un autre est un parcours de lecture potentiel. Depuis 2009, plusieurs livres enrichis ont été produits avec Scalar, certains sous la supervision d'une maison d'édition (par exemple un numéro spécial de la revue *American Literature* publiée par Duke University Press, voir <http://scalar.usc.edu/showcase/new-media-and-american-literature/>), mais la plupart édités directement par des chercheurs ou des créateurs.

Tableau 4. Les pure player et les éditions augmentées

Nom	Éditeur	Date de création	Description	Ressource
Zones	La Découverte	2007	Publication en libre accès d'ouvrages politiquement engagés, dont les versions en ligne, accessibles en HTML, sont enrichies de dossiers et de modules complémentaires.	www.editions-zones.fr/
—	Publie.net	2008	Coopérative d'auteurs fondée par François Bon, publiant des livres papier et des livres numériques sans DRM de littérature contemporaine.	www.publie.net/
Scalar	Alliance for networking visual culture (ANVC)	2013	Plate-forme <i>open source</i> d'édition universitaire se conformant aux principes du Web sémantique. Permet l'intégration de contenus multimédias, d'annotations, de commentaires des lecteurs, etc.	http://scalar.usc.edu/scalar/features/
OpenEdition	Centre pour l'édition électronique ouverte (Cléo – UMS 3287)	2013	Portail regroupant quatre plates-formes de publications scientifiques en sciences humaines (Revue.org, OpenEdition Books, Calenda et Hypothèses). OpenEdition propose un modèle Freemium et soutient le libre accès.	http://books.openedition.org/
Parcours numérique	Presses de l'université de Montréal	2014	Essais théoriques d'horizons divers (philosophie, humanités numériques, sémiologie, anthropologie, arts visuels, littérature, etc.) portant sur les technologies, pratiques et cultures numériques.	www.parcoursnumeriques-pum.ca/
Collection SHS	Terra-HN	2015	Ouvrages de sciences humaines. <i>Actualité de l'habitat temporaire</i> est la	www.reseau-terra.eu/spip.php ? rubrique224

			première expérience d'édition augmentée de la collection.	
Humanités numériques plurielles	Terra-HN	2016	Ouvrages dans le champ des humanités numériques portant sur les rapports entre technologies numériques et société actuelle	www.reseau-terra.eu/spip.php ? rubrique304
Open Road Media		2009	Une des premières plates-formes éditoriales américaines à expérimenter le modèle <i>pure player</i> .	https://openroadmedia.com/
StoryLab		2011	Publication de livres numériques courts de littérature contemporaine. La maison d'édition publie cependant des livres papier depuis 2017.	www.storylab.fr/
<i>Le Livre 010101 (1971-2015)</i>		2015	Livre en ligne de Marie Lebert portant sur l'histoire du livre numérique. On y retrouve une centaine de résumés, de chronologies et d'entretiens. [Lebert, 2015]	www.010101book.net/fr/
Living Books about History		2016	Anthologies numériques à caractère savant. Les sujets sont divers, mais adoptent tous une perspective historique ; on y trouve autant l'histoire d'Internet et du Web que l'histoire des représentations des monstres et des miracles au Moyen Âge.	http://livingbooksabouthistory.ch/fr/

Le statut des contenus

Dans ce contexte complexe se pose la question cruciale du statut des contenus. Qu'est-ce qu'un livre à l'époque du numérique ? Si l'on résume les différents cas de figure que l'on vient de présenter, nous comptons : 1) des livres papier numérisés ensuite ; 2) des livres produits en même temps en versions papier et numérique homothétique ; 3) des livres produits en deux versions différentes, l'une papier et l'autre numérique augmentée ; 4) des livres produits exclusivement en version numérique.

Or, selon la loi française, un livre numérique est une « œuvre de l'esprit créée par un ou plusieurs auteurs [...] à la fois commercialisée sous sa forme numérique et publiée sous forme imprimée ou [...], par son contenu et sa composition, susceptible d'être imprimé, à l'exception des éléments accessoires propres à l'édition numérique » (loi n° 2011-590 du 26 mai 2011 relative au prix du livre numérique).

Cette définition ne pose aucun problème pour les livres papier numérisés. En revanche, en ce qui concerne les livres originairement numériques – en particulier les livres augmentés – il est difficile d'établir la frontière entre ce que la loi considère comme un livre et ce qui est exclu de cette définition. Par ailleurs, le fait de fonder la notion de livre numérique sur le modèle papier est profondément problématique, car les modèles de production, de diffusion et de circulation du livre papier sont aujourd'hui fondamentalement affectés par le numérique : la notion même d'homothétie devient donc floue. En même temps, sans une définition précise de ce qu'est le livre numérique, comment gérer la question des droits ? Comment comprendre et interpréter le contexte d'un contenu ? Comment garantir sa stabilité et sa citabilité ? Par exemple, un livre réalisé sur Scalar est-il stable, et donc citable ? Ou bien est-il destiné à changer, compromettant ainsi le principe de citation ? Puisqu'il se situe au milieu d'autres contenus sur le Web et que sa visibilité est fondamentalement garantie par les moteurs de recherche généralistes (comme Google), comment le distinguer de ces autres contenus ? Comment le penser comme une unité autonome ?

Ces questions montrent que les modèles éditoriaux ne sont pas stabilisés dans ces premières décennies du XXI^e siècle, où commence à s'opérer un mouvement d'institutionnalisation de l'édition numérique. L'édition évolue rapidement et s'éloigne des repères institutionnels qui ont été posés entre le XV^e et le XVIII^e siècle : il est donc nécessaire d'en créer de nouveaux.

Potentialités du numérique

Les métadonnées

La question des métadonnées pour l'édition n'émerge pas avec l'édition numérique. Elle est depuis longtemps un enjeu pour le livre papier, au départ pour permettre la gestion de la distribution, des commandes et du référencement.

Dans ce but, les acteurs du livre ont depuis des années mis en place des standards de métadonnées dédiés à leur *back office*, c'est-à-dire à la gestion en interne des flux physiques des ouvrages. Cette normalisation a pris plusieurs formes, de l'ISBN (International standard book number), identifiant international d'un ouvrage dans une édition donnée, à la norme ONIX (norme d'échanges de données relatives aux ouvrages, basée sur XML, surtout utilisée par les éditeurs, les distributeurs et les libraires).

L'ensemble de ces standards de métadonnées avaient comme objectif la description et l'identification d'objets physiques, manipulables.

Avec l'édition numérique, le rôle des métadonnées associées aux livres numériques évolue.

Ainsi, le rôle des métadonnées pour l'édition numérique devient un point-clé dans la structuration des offres en ligne, les mécanismes de recommandation et les services d'exploration et de découverte proposés aux usagers. Ces enjeux de référencement et d'identification ne se limitent pas aux seuls livres numériques, les livres papier proposés par des librairies en ligne étant également concernés. Toutefois, dans le cas des livres numériques, c'est la totalité des circuits de distribution qui sont directement dépendants des métadonnées associées à un ouvrage. En effet, pour les livres numériques, les interfaces d'exploration et de recherche des offres proposées reposent entièrement sur un traitement des métadonnées disponibles. Que ce soit pour répondre à une requête précise ou pour organiser une forme de proximité entre des ouvrages, les informations disponibles pour les différents algorithmes proviennent essentiellement des métadonnées associées.

Dans ce contexte, le choix, la précision ou la fiabilité des métadonnées constituent non seulement un ensemble décisif pour la logistique de suivi ou de localisation mais également un élément essentiel du référencement et donc de la visibilité des ouvrages. Ces observations renvoient naturellement à l'ambivalence des métadonnées, levier pertinent et efficace de recommandation et de médiation mais aussi biais potentiel pour une exploitation détournée permettant de créer une fausse notoriété ou une manipulation

du référencement dans les interfaces d'exploration et de recherche. L'enjeu de ces métadonnées est d'autant plus important que celles-ci sont produites de plus en plus en amont de la filière, au moment de la production des titres et qu'elles influencent donc l'ensemble de la circulation et de la diffusion des ouvrages numériques.

Les potentialités de l'édition numérique relatives aux métadonnées concernent également deux autres aspects. Le premier réside dans la circulation et la manipulation des métadonnées des ouvrages tout au long des filières. La dématérialisation des ouvrages qu'induit l'édition numérique autorise une diffusion bien plus facile des livres. La circulation des livres numériques au sein des différentes filières de production et de distribution peut exploiter les capacités de traitement automatisé autorisées par les outils informatiques. Ainsi la disponibilité permanente d'un livre numérique, sur un grand nombre de sites, marchands ou non, nécessite que l'ensemble des métadonnées qui y sont associées puisse être traitées de façon homogène et standardisée par l'ensemble des acteurs.

Mais la circulation des métadonnées entre les différents acteurs concernés n'est aujourd'hui pas totalement optimisée, comme le montrent par exemple les formats différents utilisés par les acteurs de l'édition et des bibliothèques.

La seconde potentialité réside dans la possibilité de créer dynamiquement des métadonnées liées au contenu de l'ouvrage mais également à l'usage qui peut en être fait. Pour les livres imprimés, les métadonnées résultaient quasiment exclusivement d'un traitement intellectuel des ouvrages qui étaient décrits par les éditeurs ou plus précisément encore par les bibliothécaires. Cette production d'informations relevait d'aspects objectifs (titre, auteur, année, nombre de pages, dimensions...), mais également d'interprétations (genre éditorial, côte, sujets ou thématiques traités, résumé...). Pour l'ensemble de ces opérations d'identification, ou de catalogage pour les bibliothèques, une forme de subjectivité est liée, en partie, au caractère non calculatoire des ouvrages imprimés, c'est-à-dire l'impossibilité de traiter de façon informatisée l'ouvrage et son contenu pour en extraire automatiquement des informations permettant de le décrire.

Avec le livre numérique, il devient possible d'appliquer des outils de recherche ou d'exploration au texte intégral de l'ouvrage. Cette approche modifie considérablement le rôle des professionnels en charge de la production de ces métadonnées, en termes de valeur ajoutée et de pratiques professionnelles.

Le livre numérique rend techniquement possible l'extraction automatique à partir des ouvrages d'un ensemble de métadonnées qui étaient jusqu'alors difficilement

accessibles, comme par exemple le champ lexical, la durée de lecture en fonction du nombre de mots, les passages repérés ou surlignés ou encore les notes prises par d'autres lecteurs.

À ces nouveaux types de métadonnées vient s'ajouter le champ infiniment vaste des données produites par les usages. Ces dernières représentent un changement considérable entre le livre papier et le livre numérique. Pour les ouvrages sur support papier, les données d'usages étaient restreintes, ne concernant souvent que des informations relatives à leur achat (nombre, lieux et dates des ventes...), parfois enrichies par les revendeurs de quelques informations issues des cartes de fidélité, l'exploitation de ces données se résumant généralement à la production de classements divers en fonction des ventes sur des périodes données.

Avec le livre numérique, les données d'usages deviennent bien plus disponibles et nombreuses. Elles peuvent toucher des niveaux de détails jamais atteints dans l'observation et la mesure des pratiques de lecture, comme le temps passé à lire, la part des lecteurs ayant terminé l'ouvrage, l'endroit auquel ils se sont arrêtés, etc. Ces données deviennent ainsi un levier puissant de construction de la recommandation et d'analyse des pratiques de lecture ouvrage par ouvrage.

Le XML et d'autres langages de balisage

Nous avons déjà évoqué l'importance des formats informatiques dans le développement de l'édition numérique. Au-delà de leur rôle dans la construction des écosystèmes du livre numérique, les formats utilisés sont des vecteurs importants dans l'introduction d'une approche sémantique de l'exploitation des ouvrages.

Les langages utilisés pour l'édition numérique sont en effet tous basés sur des logiques de balisage (comme le HTML) qui peuvent présenter des éléments sémantiques. Ainsi, en introduisant un balisage sémantique dans les ouvrages numériques, il devient possible de proposer aux usagers des modalités de lecture innovantes, fondées sur cette structure. Il devient par exemple envisageable d'offrir pour un même ouvrage des niveaux de lecture et d'approfondissement différents, en affichant selon le choix du lecteur certains éléments, balisés comme correspondant à des niveaux de détail définis. Cette logique de construction modulaire et sémantique est une perspective de développement intéressante, mais se heurte à des contraintes économiques et juridiques complexes.

Le balisage XML, par l'introduction d'éléments sémantiques standardisables et manipulables de façon algorithmique, ouvre la voie à une forme nouvelle de construction éditoriale fondée sur la modularité. En combinant les formats numériques et le balisage sémantique porté par le XML, il devient en effet possible et nécessaire de redéfinir une unité de base pour la production, le traitement et la distribution éditoriale. Dans l'édition sur support papier, cette unité de base est évidemment le livre : c'est à partir et autour de celui-ci que l'ensemble de l'écosystème éditorial se met en place. Avec l'édition numérique, il devient possible de modifier cette unité de base pour la ramener à une partie, à un chapitre ou à un élément précis comme une figure ou une illustration.

Cette prise en compte des potentialités offertes par le XML pour l'édition [Prost, 2011] passe également par d'autres dimensions de l'activité éditoriale. Nous en pointerons deux : la structuration fine et stable des ouvrages d'une collection et l'intégration de logiques de chaîne de publication multisupports.

Pour le premier point, la structuration d'un document par balisage XML suppose la formalisation en amont de la *définition* des documents au travers de DTD (Document type definition) ou de schéma XML. Dans ces structures, on peut ainsi identifier, directement dans le document, les différents niveaux de titres, l'auteur, les noms de lieux, la date de mise à jour... Cette formalisation consiste à lister et définir l'ensemble des balises exploitables par les fichiers XML. Cela suppose de définir une structure logique commune à un ensemble d'ouvrages ou d'utiliser des DTD existantes comme TEI ou DocBook.

Dans tous les cas, ce travail peut être conséquent et nécessite des compétences spécifiques parfois absentes des maisons d'édition.

Le second point découle en partie du premier. Un des objectifs de la mise en place d'une structuration XML est le déploiement d'une chaîne de publication multisupports. En raison de la multiplication des supports et des formats disponibles ou à venir pour les livres numériques, il est intéressant de concevoir et d'exploiter une chaîne de publication numérique qui permette à partir d'un unique fichier une exploitation dans des formats et sur des supports variés. Pour mettre en œuvre des projets de ce type, il convient non seulement de disposer des compétences spécifiques nécessaires, mais surtout de passer à une logique de structuration des processus de production éditoriale qui ne vise plus à produire un objet mais à structurer un contenu. Cela suppose d'appréhender des problématiques nouvelles comme l'accessibilité ou la préservation et la conservation.

Face à ces nouvelles problématiques, la maîtrise des logiques sous-jacentes au XML est indispensable.

La fouille de texte et d'autres possibilités

L'édition numérique, en passant d'un support papier à un support informatique, ouvre le champ du traitement algorithmique des contenus éditoriaux. La fouille de textes est, par exemple, un enjeu important pour l'édition scientifique. Elle consiste en un traitement informatique d'ensembles de textes et de données, parfois conséquents, pour en extraire des données ou informations nouvelles.

Cela permet par exemple de repérer des éléments précis de certains textes (noms de personnes, de lieux, etc.) pour pouvoir y associer d'autres informations provenant de sites ou de bases de données. Ce type d'approche suppose une expertise technique avancée et l'existence de bases de données structurées et fiables.

L'objectif de la fouille de texte est de pouvoir s'appuyer sur des corpus textuels importants afin d'en extraire automatiquement des éléments susceptibles d'être utilisés par des services divers proposés aux usagers. Ainsi, l'analyse d'un champ lexical, de références bibliographiques ou de formes stylistiques permet de concevoir des outils de recommandation avancés optimisant la navigation dans des corpus littéraires ou scientifiques. De la même manière, la production automatisée de résumés ou de synthèses, la production d'indicateurs statistiques sur les contenus permettent de construire des métadonnées intéressantes rattachées aux textes eux-mêmes.

Le développement de l'autoédition (literacy numérique)

Parmi les potentialités ouvertes par le développement de l'édition numérique, l'autoédition représente un changement conséquent et déjà largement mesurable des modèles d'édition. Il s'agit d'une production éditoriale réalisée et pilotée par le ou les auteurs, parfois en dehors d'une maison d'édition. Ce type de production éditoriale n'est pas propre à l'édition numérique, elle existe pour l'édition papier sous la forme de la publication à compte d'auteur – l'auteur fait réaliser, à ses frais, le travail de préparation éditoriale, de fabrication, de diffusion et de distribution par la maison d'édition –, ou de l'autoédition – l'auteur s'édite sans recours à une maison d'édition. Ce type d'édition restait toutefois marginal sur support papier en raison des obstacles de production et de distribution. En effet, une publication autoritative (c'est-à-dire directement prise en

charge par l'auteur) imprimée suppose un investissement parfois non négligeable et une diffusion souvent réduite. Les volumes concernés et les zones géographiques couvertes restaient donc limités.

Avec l'édition numérique et dès les années 2000 (notons par exemple la création de la société états-unienne Lulu.com en 2002), l'autoédition numérique connaît une croissance importante, d'abord aux États-Unis et plus récemment en France. Cette autoédition prend plusieurs formes qu'il convient de distinguer.

Un premier type d'autoédition numérique consiste pour un auteur à produire un contenu éditorial de façon autonome et à le proposer en ligne, de façon gratuite, avec une rémunération publicitaire ou sur une plate-forme commerciale de diffusion. Un second type se rapproche plus des éditions à compte d'auteur avec l'achat d'une prestation de fabrication numérique sur la base d'un tapuscrit (un manuscrit numérique) pour ensuite proposer l'ouvrage à la vente.

Ce paysage de l'autoédition se structure lentement depuis les années 2000 avec l'arrivée sur ce marché de plusieurs acteurs (Amazon, Lulu.com, SmashWords, Librinova...) qui proposent des palettes de services très variés (correction, impression à la demande...). Il est sujet à une croissance extrêmement rapide, avec un nombre d'ISBN attribués à des titres autoédités multiplié quasiment par 10 entre 2007 et 2015.

Depuis le lancement de son service Kindle Direct en 2007, Amazon a pris une place importante sur le marché de l'autoédition numérique. De plus, les ouvrages autoédités représentaient en 2016 40 % des titres proposés en version numérique sur sa plate-forme et 25 % de son chiffre d'affaires sur ce marché. Il s'agit donc d'une forme de production éditoriale non négligeable et qui questionne plusieurs dimensions de l'activité d'édition.

En premier lieu, la croissance de l'autoédition a fait émerger plusieurs succès commerciaux, comme *Fifty Shades of Grey* ou *Les gens heureux lisent et boivent du café*. Les succès commerciaux issus de l'autoédition présentent des caractéristiques communes : ils s'appuient au départ sur un investissement fort des auteurs dans la promotion de leur ouvrage *via* les sites de fan-fiction et les communautés qui s'y créent et commentent les textes, ou *via* un travail de promotion auprès de blogueurs ou de médias plus classiques. Que ce soit auprès d'une communauté qui préexiste au livre (autour d'un blog ou sur les réseaux sociaux) ou en la fédérant progressivement, la prise en charge de la promotion et de la publicité est portée de façon très active par ces auteurs. Cette dynamique permet, dans ces quelques cas de succès, de se positionner dans les meilleures ventes dès la sortie de l'ouvrage sur les plates-formes comme Amazon, iTunes ou Kobo.

L'affichage dans les classements renforce la visibilité de l'ouvrage et facilite l'enclenchement d'un cercle vertueux. Il s'agit d'un phénomène classique dans l'économie des industries culturelles, mais qui se construit pour l'autoédition sur des plates-formes produisant elles-mêmes, parfois automatiquement, leurs propres outils de classement.

Ces succès d'autoédition donnent ensuite lieu à des formes de valorisation bien plus traditionnelles avec en général la publication de l'ouvrage ou de ses suites dans des maisons d'éditions « classiques », qui accompagnent un éventuel développement à l'international. Dans ce cas, l'autoédition joue un rôle particulier, puisqu'elle permet aux éditeurs de choisir des ouvrages déjà confrontés aux appréciations des lecteurs sans avoir à assumer le risque important d'un premier roman.

Enfin, le marché de l'autoédition connaît une structuration oligopolistique, avec des niveaux de concentration du marché très élevés, phénomène bien connu dans l'édition « classique ».

Ce qui change avec l'autoédition numérique, c'est la nature des acteurs qui concentrent les parts de marché. Les plates-formes en ligne proposant des services d'autoédition sont en majorité la propriété de nouveaux entrants dans le domaine de l'édition, voire plus globalement du livre. Ces plates-formes, peu nombreuses, concentrent un très grand nombre de titres autoédités, qui peuvent être utilisés comme leviers d'attractivité et de captation d'une attention exploitée ensuite à d'autres fins, comme la vente de matériel, l'abonnement ou la promotion de divers produits.

Avec l'autoédition numérique, plusieurs facettes de l'activité éditoriale sont questionnées. La première est le rôle et la valeur ajoutée des acteurs traditionnels de l'édition. Dans la logique de l'autoédition, la sélection opérée auparavant par les éditeurs parmi les multiples manuscrits et fondée sur l'appréciation des textes est remplacée par la mise à disposition d'un très grand nombre de titres accompagnée d'outils algorithmiques de sélection et de recommandation. Basés sur la maîtrise des données d'usages des lecteurs utilisant la plate-forme (Amazon) ou sur des critiques et avis produits par les lecteurs (Babelio), cette médiation automatisée permet d'offrir au lecteur une forme individualisée et automatisée de recommandation. Le glissement s'opère d'une autorité de la maison d'édition, fondée sur une expertise sur le contenu, vers une forme d'autorité algorithmique, fondée sur la maîtrise technique de la recommandation et un nombre de titres très important. Il est intéressant de noter que, dans le cas des livres autoédités

mentionnés, la valeur symbolique des maisons d'édition traditionnelles les ayant par la suite édités demeure un paramètre important.

Formes numériques de production des contenus

L'extension du domaine de l'édition

Après ce premier panorama des changements déterminés par le numérique, on peut donc à nouveau se poser la question : qu'est-ce que l'édition aujourd'hui ? Les frontières de ce que l'on peut considérer comme « édité » semblent se déplacer. L'édition ne peut plus se limiter au travail réalisé par des maisons d'édition : la fonction éditoriale se manifeste de façon évidente dans d'autres formes de production des contenus. Nous avons décrit ce phénomène lors de notre analyse de l'éditorialisation, et nous avons pu le remarquer à nouveau à travers le cas des livres enrichis qui mettent en crise la définition de livre telle qu'elle s'est cristallisée à partir du XVIII^e siècle. Or, plusieurs autres formes de production des contenus semblent contribuer également à remettre en question les frontières de l'édition. Il est difficile, sinon impossible, d'en donner un aperçu exhaustif en raison de leur instabilité, aussi choisirons-nous celles qui nous semblent les plus représentatives et les plus durables.

Les blogs

Le blog est une forme éditoriale qui commence à s'affirmer à la fin des années 1990. Le terme « blog » est la contraction du mot anglais « weblog », qui pourrait être traduit comme « journal web » ou « registre web ». Les premières formes de weblog sont en effet des pages web qui compilent des listes de liens vers d'autres sites, accompagnés de courts commentaires. Le premier d'entre eux serait celui de Jorn Barger qui utilise, pour la première fois en décembre 1997, le mot « weblog » pour faire référence à son activité d'écriture sur son site Robot Wisdom [Blood, 2002] : Barger commence à écrire de façon assidue (presque quotidienne) et ordonne ses textes en ordre antichronologique. Ce sont là les deux premières caractéristiques du blog : une écriture régulière et un ordre antichronologique.

Depuis cette première apparition, la forme blog évolue, notamment avec l'apparition des premiers CMS (Content management system) qui permettent une appropriation de cette forme par le grand public. Cela favorise le développement d'une écriture diaristique et autobiographique qui peut être définie comme une « écriture de

soi ». En 2000, c'est cette expression qu'utilise Philippe Lejeune [2000], qui relève et étudie l'émergence des pratiques d'écriture en ligne. À cette époque, des plates-formes qui facilitent la création de blogs sans réclamer de grandes compétences informatiques commencent à apparaître : les plus connues sont Blogger (1999) ou, en France, Skyblog (2002). En 2003 apparaît la première version de WordPress, le CMS le plus représentatif de la pratique du blog – et probablement le plus utilisé, avec près de 24 millions de sites fonctionnant sur cette base en 2015 [Feldman, 2015]. WordPress se démarque par sa facilité d'installation et d'utilisation, tout en donnant la possibilité aux usagers de créer leur propre site sur le serveur de WordPress, sans avoir à posséder d'hébergement personnel. Ce CMS propose par ailleurs une organisation qui permet de rompre avec l'ordre antichronologique que les blogs proposaient jusque là.

À partir des années 2000, le blog se répand rapidement. Le billet (post) de blog devient un véritable format textuel. Cette forme commence à être utilisée à de multiples fins, de l'écriture diaristique [Lejeune, 2000] à l'écriture académique [Rak, 2005]. En France, le Centre pour l'édition électronique ouverte (Cléo) lance en 2009 Hypothèses.org, une plate-forme de carnets de recherche en sciences humaines et sociales : en 2017, ce portail accueillait presque 2 000 sites. Or plusieurs questions se posent : comment définir les frontières du format blog ? Peut-on considérer ces contenus comme édités ?

L'hétérogénéité du matériel textuel regroupé sous le terme blog complique l'édification de frontières nettes propres à ce format. Par ailleurs, plusieurs auteurs refusent que l'on décrive leur activité d'écriture comme un blog. Il semble en effet indéniable que pour des plates-formes complexes, par exemple Le Tiers Livre (letierslivre.net), portail d'écriture de l'écrivain François Bon, ou Liminaire de Pierre Ménard (liminaire.fr), le terme blog semble inadéquat. Si l'on peut donc proposer une définition restreinte du blog – conçu comme un ensemble de textes écrits à la première personne, postés régulièrement sur un site et ordonnés en ordre antichronologique –, ce format tend à éclater dans la réalité des pratiques, et glisser vers des formes plus complexes, ressemblant davantage à de véritables revues en ligne dont il n'y aurait qu'un seul auteur. Que peut-on dire du statut éditorial de ces contenus ? S'agit-il seulement d'une forme d'édition ? Les contenus publiés sur un blog sont bien le fruit d'une instance de production, ils bénéficient d'un processus de circulation et d'une diffusion (elle-même très hétérogène selon les sites) ainsi que d'une forme de légitimation dans certains cas très forte (des blogs ont parfois acquis une crédibilité et une autorité considérables auprès

de publics très larges). Indéniablement, les dispositifs techniques de production de ces contenus, et en particulier les CMS, jouent un rôle fondamental : ils établissent le cadre et les règles de production (une écriture régulière, un certain type de classement et donc un certain type de contenus), assurent la visibilité et l'accessibilité (notamment en rendant indexables les contenus par les moteurs de recherche) et, par leur diffusion, créent une forme de légitimation. Finalement, cette forme d'écriture est donc fortement formatée par la fonction éditoriale que jouent les CMS.

Les wikis

Les wikis sont un dispositif de production de contenus en ligne dont la fonction éditoriale est particulièrement évidente. Ils reposent sur une forme de production collaborative fonctionnant grâce à des « moteurs de wikis » (*wiki software*), un cas particulier de CMS. Le premier wiki, WikiWikiWeb, a été créé en 1995 par Ward Cunningham, mais le plus connu et le plus utilisé reste sans aucun doute Wikipédia. L'encyclopédie participative, dont nous avons déjà évoqué l'importance, a été créée par Jimmy Wales et Larry Sanger en 2001. À partir de 2002, ce wiki fonctionne grâce au moteur MediaWiki. Il est important de bien comprendre la différence entre le moteur de wiki et le wiki lui-même : le premier est le logiciel qui permet la création de contenus, tandis que le second est le projet qui utilise la solution logicielle. En d'autres termes, Wikipédia est un wiki réalisé avec MediaWiki, mais il n'est pas le seul. Plusieurs wikis fonctionnent sur MediaWiki, qui est un CMS gratuit et libre : chacun peut l'installer sur son propre serveur pour créer son propre wiki.

Les wikis et les moteurs comme MediaWiki sont pensés pour favoriser la production collaborative de contenus. Le succès d'un wiki comme Wikipédia dépend ainsi du dispositif qui structure et règle cette collaboration. La qualité des contenus qui y sont publiés est garantie par les règles de production, purement formelles, incarnées dans le fonctionnement même du CMS. Ce dernier joue un rôle essentiel : il guide les interactions des usagers, contrôle et favorise les liens entre les contenus et réalise une vérification algorithmique du respect de ses principes. En outre, l'ergonomie et les éléments graphiques – par exemple l'accessibilité de l'historique, ou de la table des matières de la page, ou du même contenu en d'autres langues – conditionnent non seulement la lecture, mais aussi la contribution sur Wikipédia. C'est ce que Dominique Cardon appelle « la dépendance récursive des pratiques et des règles » [Barbe *et al.*, 2015, p. 19]. On peut affirmer que le CMS MediaWiki et l'ensemble des règles formelles

qui régissent les échanges de la communauté ont une fonction éditoriale. Et si l'un des objectifs de la fonction éditoriale est de garantir la qualité des contenus, la mission d'éditeur des wikis est sans doute réussie : malgré les défauts de certains contenus, la qualité des articles de Wikipédia a été souvent évaluée comme comparable à celle des encyclopédies traditionnelles [voir par exemple Fraser et Temple, 2014]. Ce projet s'affirme ainsi comme un outil de référence non seulement pour le public généraliste, mais aussi pour les chercheurs.

Tableau 5. Les CMS

Nom	Date de création	Usages	Proportion de sites web utilisant le CMS	Parts de marché
Wordpress	2003	Blogs, sites vitrine, e-commerces, sites d'actualités, etc.	27,7 %	58,9 %
Joomla!	2005	E-commerces, sites d'actualités, sites vitrines, sites web gouvernementaux, etc.	3,3 %	7,1 %
Drupal	2001	Blogs, sites web gouvernementaux, e-commerces, sites vitrines, etc.	2,2 %	4,7 %
TYPO3	2000	Sites web commerciaux, sites vitrines, blogs, etc.	0,7 %	1,4 %
Squarespace	2004	Sites vitrines, e-commerces, blogs.	0,6 %	1,2 %
Wix	2006	Sites vitrines, e-commerces, blogs.	0,3 %	0,7 %
Concrete5	2003	Blogs, sites web gouvernementaux, e-commerces, sites vitrines, etc.	0,1 %	0,2 %
Spip	2001	Blogs, sites vitrines, revues scientifiques. Particulièrement populaire au sein de la communauté française.	0,1 %	0,2 %
MediaWiki	2002	Publication de wikis (Wikipédia, Wikimedia, Wiktionary, Wikia, WikiHow, Wikileaks, etc.)	0,1 %	0,1 %
WebGUI	2001	Sites web commerciaux, sites gouvernementaux, sites éducatifs, etc.	< 0,1 %	< 0,1 %

Lodel	2014	Édition de revues scientifiques	—	—
-------	------	---------------------------------	---	---

Les réseaux sociaux

Parmi les formes de production de contenus à l'intérêt éditorial certain, il faut citer les réseaux sociaux qui, depuis leur apparition au début des années 2000, n'ont cessé de prendre une importance croissante dans les pratiques. Des plates-formes grand public comme Facebook, Twitter ou Instagram sont des lieux où les usagers écrivent, déposent des documents, republient des informations trouvées ailleurs. Deux aspects propres à cette forme de production de contenus méritent d'être relevés : leur fonction de formatage de l'écriture et leur fonction d'agrégation de contenus existants. À ces deux aspects s'ajoutent les fonctions de diffusion et de validation dont on discutera dans le prochain chapitre.

Les plates-formes des réseaux sociaux ont d'abord une fonction de structuration des contenus très claire : c'est ce que l'on appelle l'« affordance » [Norman, 1988]. Chaque plate-forme pousse les utilisateurs à produire un certain type de contenu : en termes quantitatif (les 140 caractères pour Twitter, par exemple), structurel (lorsque les liens doivent être utilisés d'une certaine manière), rhétorique (comme Facebook qui nous conduit à écrire notre « journal » à la première personne) et thématique. La plate-forme prédéfinit donc ce que l'on peut écrire et canalise les actions de production. De cette manière, il y a dans chaque plate-forme une homogénéité de format qui peut être comparée à celle obtenue grâce à un processus éditorial classique.

En second lieu, ces plates-formes fonctionnent comme de véritables agrégateurs de contenus : les usagers « curent » les documents auxquels ils ont accès sur le Web en les ordonnant selon leurs intérêts ou selon les besoins de leur communauté. De cette manière s'opère un travail de structuration des contenus qui est encore une fois formaté par la plate-forme elle-même. Par leur grand nombre d'utilisateurs (1,18 milliard d'utilisateurs actifs par mois pour Facebook selon les statistiques de l'entreprise, 313 millions pour Twitter), les réseaux sociaux ont acquis un rôle fondamental dans la production, la diffusion et la validation des contenus. Il est donc désormais indispensable, à l'ère numérique, de les compter parmi les acteurs de l'édition.

III / La légitimation des contenus

Les modèles « classiques »

La reconnaissance symbolique de l'édition papier

De quelle manière peut-on donner de la crédibilité à un contenu ? Comment faire en sorte qu'un texte, un document, une information soient considérés comme fiables ? Comment donner de l'autorité à des affirmations ou à des idées ? Voici quelques unes des questions fondamentales auxquelles doit répondre aujourd'hui la fonction éditoriale.

Comme on l'a vu, les maisons d'édition comptent, au moins depuis le XVIII^e siècle, parmi les instances qui remplissent la fonction de légitimation. La légitimité du contenu est fonction de la légitimité symbolique de la maison d'édition qui le publie. Mais comment une maison d'édition peut-elle concrètement être source de légitimation ? Pour répondre à cette question, on peut identifier trois différents types de légitimation : le prestige, la visibilité et l'évaluation scientifique.

1. Le prestige dépend de la réputation qu'une maison d'édition a pu acquérir auprès des lecteurs. Ces derniers connaissent le nom de la maison et associent à ce nom une idée de qualité. Cette valeur symbolique est transférée de la maison d'édition aux ouvrages qu'elle publie. Concrètement, le nom et le logo de la maison confèrent du prestige à l'ouvrage et garantissent sa qualité. Ce type de légitimation est typique des contenus destinés au grand public.
2. La visibilité dépend de la capacité à faire connaître l'existence d'un contenu à un certain nombre de personnes. Les maisons d'édition ont la possibilité de communiquer sur leurs publications et de les rendre visibles. Si cette visibilité ne garantit pas nécessairement la qualité de l'ouvrage, elle participe tout de même de sa légitimation, car elle en détermine l'influence. Dans cette production de visibilité, plusieurs maillons de la chaîne de valorisation du livre peuvent jouer un rôle fondamental, comme les prix littéraires, qui augmentent l'audience des ouvrages et se présentent comme un gage de qualité.
3. L'évaluation scientifique concerne seulement un type particulier de contenus : ceux destinés à un public de spécialistes – notamment d'universitaires –, souvent publiés par des maisons d'édition spécialisées – comme des presses universitaires – ou dans des collections particulières. Chaque ouvrage est en général soumis, avant

publication, à un processus d'évaluation par les pairs : la qualité et la rigueur scientifique du contenu sont analysés et appréciés, souvent en aveugle, par des spécialistes. Ce processus garantit la légitimité de l'ouvrage. Dans ce cas, la maison d'édition est le garant du bon déroulement du processus d'évaluation, les spécialistes auxquels est demandée l'évaluation étant extérieurs à la maison.

Un contenu peut bénéficier de ces trois types de légitimation en même temps : avoir du prestige, de la visibilité et être garanti par une évaluation scientifique. Souvent, ces aspects sont liés : la visibilité peut dépendre du prestige, le prestige du sérieux de l'évaluation scientifique, etc. Mais ces types de légitimation peuvent aussi jouer de manière isolée : on peut prendre l'exemple d'une petite maison d'édition universitaire, à la visibilité très réduite et au prestige parfois limité, ou d'une grande maison d'édition commerciale, se donnant les ventes pour seul objectif, et misant tout sur la visibilité.

Les environnements numériques ont indéniablement produit de nouveaux modèles de légitimation. Ils ont ainsi, en partie, remis en question le monopole que les maisons d'édition avaient conservé pendant plusieurs siècles en matière de légitimation. Mais il serait faux d'affirmer que le numérique a enlevé – ou même affaibli – le pouvoir de légitimation et la valeur symbolique des maisons d'édition. Le fait d'être édité par l'une d'entre elles – en papier ou en format numérique – reste la source principale de légitimité pour un contenu, même si ce n'est plus la seule.

L'auteur se porte bien

Une autre forme de légitimation d'un contenu provient de son auteur – ou, très concrètement, du nom de l'auteur affiché sur la couverture d'un livre.

On peut identifier trois éléments propres à la fonction auctoriale en partie croisés avec ceux de la fonction éditoriale. Tel qu'elle s'est stabilisée à partir du XVIII^e siècle, la fonction auctoriale implique la propriété, l'originalité et enfin l'autorité [Rose, 1993].

1. l'auteur a tout d'abord un certain droit de propriété sur ce qu'il produit. Cette notion s'est affirmée et institutionnalisée à partir du XVIII^e siècle (le Statut d'Anne de 1710 est considéré, on l'a vu, comme la première loi sur le copyright). L'idée de lier l'auctorialité à la propriété a rendu théoriquement possible la naissance de l'édition papier telle que nous la connaissons. Le fait que l'auteur soit propriétaire du contenu – et non du support qui le fait circuler – lui permet de vendre le droit d'exploitation de son bien, notamment à une maison d'édition. Ce principe est à la base du modèle économique de l'édition papier.

2. En second lieu, l'auteur exprime un point de vue particulier et original, car il manifeste, dans ses productions, son individualité. Les notions d'individu et d'originalité deviennent essentielles pour expliquer la propriété intellectuelle : l'auteur est propriétaire du fruit de son originalité, à savoir de ce qu'il est le seul à pouvoir produire. La fonction auctoriale garantit donc aussi la reconnaissance de l'auteur, son style, sa « marque ».
3. En dernier lieu, l'auteur légitime et garantit les contenus qu'il produit. Il en est donc aussi le responsable – y compris d'un point de vue légal. Cette fonction est justement partagée avec l'éditeur, qui augmente et encadre la légitimité de l'auteur. Les deux fonctions de légitimation – auctoriale et éditoriale – sont alors entremêlées. Dans certains cas, l'une prévaut sur l'autre : par exemple, un auteur encore inconnu verra sa légitimité garantie par la fonction éditoriale ; à l'inverse, un auteur très connu pourra faire profiter de sa légitimité à un nouvel éditeur.

Qu'en est-il de ces structures dans les environnements numériques ? Depuis quelques années, on entend souvent parler d'une crise de l'auteur. Cette crise est indéniable, mais il serait faux de l'assimiler à une disparition de la fonction auctoriale. La crise est due à la convergence d'une théorie critique du modèle de l'auteur que l'on vient de décrire et de l'émergence de pratiques en environnement numérique qui mettent en question l'unicité de ce modèle. À partir des années 1960, en effet, des théoriciens comme Roland Barthes et Michel Foucault ont annoncé la « mort de l'auteur ». Leur objectif était d'attaquer la notion romantique de l'auteur-créateur, du génie produisant à partir de rien, dans l'originalité absolue. À ce modèle s'oppose une idée de la circulation des textes et des idées qui ne relève jamais d'un individu isolé, mais d'une série complexe de dynamiques culturelles, sociales et politiques. Ce n'est pas l'auteur qui produit à partir de rien des contenus, ce sont les interactions sociales et culturelles qui les font émerger. Or le Web – et en particulier le Web 2.0 – a permis l'émergence de pratiques qui vont tout à fait dans ce sens : le nom de l'auteur s'efface au profit d'une écriture collective, la signature de l'auteur devenant moins importante que celle de la plate-forme. Les affirmations « j'ai lu cela sur Internet » ou « je l'ai trouvé sur Google » sont à cet égard bien plus significatives qu'on pourrait le croire.

Or, s'il est vrai que de nouveaux modèles de légitimation sont en train d'émerger, il n'en reste pas moins que la fonction auctoriale conserve son rôle de légitimation. Les environnements numériques sont caractérisés par une hybridation de formes traditionnelles et de formes classiques. Prenons l'exemple des blogs. Comme l'a

démontré Dominique Cardon [2010], un blogueur peut devoir sa réputation à la popularité de son blog – et dans ce cas c'est le blog qui légitime l'auteur –, mais l'inverse est aussi vrai : un blog peut être lu car le blogueur est déjà connu en tant qu'auteur. Ce type de phénomène est très courant dans les environnements numériques, où les auteurs peuvent démultiplier leur visibilité et donc leur autorité. S'il y a bel et bien une certaine crise du concept d'auteur liée à l'émergence de nouveaux modèles de production et de légitimation de contenus, l'auteur et la fonction auctoriale n'en ont pas pour autant perdu leur propre pouvoir de légitimation.

La reconnaissance par le commerce

Dans les mécanismes de légitimation des contenus, la légitimation commerciale occupe une place particulière. Par légitimation commerciale nous entendons l'ensemble des logiques de légitimation s'appuyant sur des indicateurs de vente mais également les logiques symboliques associées à des acteurs commerciaux. Cet aspect a été abordé précédemment dans la reconnaissance symbolique de l'édition papier, mais nous l'abordons ici sous l'angle des acteurs de la commercialisation, et notamment des librairies.

Considérons tout d'abord la logique de légitimation par les chiffres de vente. Cette logique est bien connue de l'ensemble des acteurs des industries culturelles, édition – numérique ou non – comprise : dans un premier temps, un produit culturel, en raison de sa qualité, de la communication, de l'actualité ou d'autres paramètres connaît un certain succès commercial. Il apparaît de ce fait dans des palmarès de vente qui peuvent jouer un rôle, pour une partie du public, de recommandation. Cette visibilité renforcée induira des ventes supplémentaires et améliorera la place dans les classements (ou par exemple sur les tables des meilleures ventes de certaines librairies). Ces volumes de vente deviennent parfois eux-mêmes des arguments commerciaux (« Déjà *n* exemplaires vendus ! »). Cette mécanique est dans un deuxième temps contrariée par une saturation des publics visés ou l'arrivée d'un nouveau titre profitant d'une dynamique semblable

L'édition numérique modifie cependant ce fonctionnement pour deux raisons : l'émergence d'indicateurs non commerciaux (réputation, visibilité...) et le rôle considérable des classements de ventes dans les pratiques d'achat sur les plates-formes de vente en ligne. L'environnement numérique permet en effet une multiplicité de modalités commerciales, de la vente classique de fichiers numériques ou de droits d'accès à la rémunération par la publicité, en passant par les modèles *freemium* (contenu gratuit et

services payants). Cette multiplicité de modèles empêche de s'appuyer sur un seul type de données commerciales pour établir des classements pertinents. Ce sont ainsi les moteurs de recherche qui, *via* leurs algorithmes, agglomèrent les données les plus diverses et proposent aux utilisateurs les classements les plus pertinents, dans lesquels les ventes réalisées ou les droits d'accès vendus ne représentent qu'un aspect parmi d'autres. Les stratégies d'optimisation de référencement de l'édition numérique se rapprochent alors des stratégies de construction de réputation généralement utilisées pour les sites Web.

À un second niveau, celui des classements sur les sites de vente, l'édition numérique est confrontée à une problématique nouvelle. En effet, comme dans l'ensemble des industries culturelles numériques, les plates-formes de vente occupent une place croissante dans la vente de l'édition numérique, comme la place prépondérante d'Amazon sur ce marché dans de nombreux pays le montre. Or, ces plates-formes présentent, entre autres, deux caractéristiques qui modifient la place des données commerciales dans les mécanismes de légitimation : le très grand nombre de titres proposés, largement supérieur à toutes les offres existantes, et l'absence de recommandation non algorithmique. Ces deux éléments concourent à l'exploitation renforcée par les plates-formes de vente des classements fondés sur les ventes comme outils de recommandation automatisés.

Cet enjeu des classements associé à la souplesse permise par le numérique dans l'évolution du prix de vente a conduit au développement de stratégies de prix et de choix de catégorisation des titres qui visent à faire rentrer des ouvrages dans les classements en utilisant des prix bas et en choisissant des catégories moins concurrentielles que d'autres. Une fois le titre entré dans le classement des meilleures ventes d'une catégorie, sa notoriété croissante permet de faire remonter les prix pour retrouver un niveau de rentabilité satisfaisant. Ces tactiques sont évidemment risquées, les titres n'atteignant pas un niveau de vente suffisant demeurant à la fois moins visibles et moins rentables.

Le rôle des bibliothèques

Les bibliothèques, du fait de leur poids symbolique important, peuvent jouer un rôle dans les logiques de légitimation des contenus. Pour l'édition papier, l'acquisition par les bibliothèques peut conférer une légitimité ou une reconnaissance à des titres parfois peu présents dans les circuits commerciaux habituels. Pour l'édition numérique cette *légitimité bibliothéconomique* s'exerce de plusieurs façons.

L'édition numérique ouvre deux nouvelles voies de légitimation par les bibliothèques. La première passe par le repérage, voire l'intégration dans leurs catalogues et portails, d'ouvrages numériques librement accessibles en ligne. Cette démarche joue un rôle de légitimation différent selon les ouvrages concernés. Pour les classiques, déjà largement référencés, il ne s'agit pas d'une légitimation des œuvres – qui n'en n'ont pas besoin – mais du travail d'édition numérique réalisé sur les textes et fichiers. Pour d'autres types d'ouvrages proposés librement en ligne, la présence sur le portail ou le catalogue d'une bibliothèque confère de fait une légitimité, ne serait-ce que d'avoir été considéré et sélectionné par les bibliothécaires. L'autre voie de légitimation par des bibliothèques passe par l'organisation par ces dernières d'activités et d'ateliers de création ou d'écriture. Cette approche, déjà mise en œuvre dans des bibliothèques publiques américaines, consiste à considérer le public comme producteur de contenus qui viendront potentiellement enrichir le fonds. Cette démarche s'inscrit dans une dynamique plus large à la croisée des modes de production (*fablab*, *maker spaces*...) et de l'implication des publics dans la production de connaissances. Il s'agira par exemple d'ateliers visant à recueillir les récits de vies d'habitants du quartier d'une bibliothèque, récits qui viendront enrichir la documentation déjà proposée.

Dans ces deux cas, l'enjeu pour la bibliothèque est de se positionner par rapport à la question de la légitimation. En reconnaissant certains contenus (issus de ses publics ou du Web), elle joue un rôle qui la rapproche d'un éditeur. En effet, de la même façon qu'un éditeur légitime un contenu éditorial en acceptant de le publier sous son nom, la bibliothèque joue un rôle éditorial en acceptant de faire figurer dans ses espaces numériques des contenus pas forcément issus d'une chaîne de production éditoriale traditionnelle. Ces démarches s'accompagnent de plusieurs questionnements notamment sur la légitimité professionnelle des bibliothécaires en matière d'édition, mais également sur la neutralité des acteurs publics dans ce type de processus.

Validation collective

Parmi les multiples enjeux de l'édition numérique, celui de la validation des contenus occupe une place centrale. En effet, la dématérialisation et la plus grande facilité technique de production des contenus impliquent des mécanismes innovants de validation qui amènent, en creux, à questionner la valeur ajoutée par les acteurs traditionnels de l'édition dans ces mécanismes. Le chapitre précédent a traité de la

question des modèles « classiques », nous aborderons ici les mécanismes de validation collective.

Contenus grand public et savants et formes de légitimation

Les contenus à destination du grand public ou des publics savants ou académiques s'appuient sur des logiques de légitimation et de validation apparemment distinctes. Ainsi, l'édition savante, notamment de revues, repose souvent sur la validation par les pairs : les éditeurs ou les revues recueillent des propositions de chercheurs, les soumettent à l'évaluation d'autres chercheurs reconnus et décident ou non, au terme d'échanges éventuels, de leur publication. Il s'agit d'un mécanisme circulaire et collectif, dans la mesure où la validation est réalisée au sein d'un réseau de reconnaissance et de légitimité partagée entre les chercheurs. Dans le cas des contenus grand public, la qualité du travail de repérage, de sélection et d'accompagnement des œuvres est reconnue à l'éditeur par les lecteurs. Cette reconnaissance devient alors un outil de validation pour les productions suivantes. Malgré ces différences, on assiste dans les deux cas à un processus de validation où une communauté confère à un acteur central, éditeur ou revue, une légitimité à valider des productions nouvelles.

La numérisation de l'édition et la mise en réseau des acteurs remet en cause cette centralisation du processus de validation collective, c'est-à-dire le fait que l'ensemble de ces mouvements de validation se rapportent à un acteur central du dispositif. On passe de modèles de validation *a priori* et centralisés à des modèles *a posteriori*, plus diffus. Ces nouveaux processus s'appuient sur deux caractéristiques principales de l'édition numérique : la capacité de mise à disposition large et peu coûteuse des contenus, et la possibilité de dynamiques collectives fondées sur les communautés.

La première caractéristique, la facilité de mise à disposition de contenus éditoriaux, modifie en profondeur la problématique de la validation collective. En modifiant les modalités de publication, l'édition numérique autorise un basculement des mécanismes de validation vers une légitimation *a posteriori* des contenus par des communautés en ligne constituées autour des contenus. Dans cette logique, à l'œuvre dans l'autoédition ou la *fan fiction* par exemple, la validation collective s'exerce sur un contenu déjà disponible auquel une communauté de lecteurs va réagir et construire progressivement une validité. La structure de coûts qui accompagne la publication numérique (notamment un coût de mise en ligne très faible) renforce la place des plates-formes de publication évoquées précédemment. La validation en amont, traditionnellement opérée par les éditeurs, est

remplacée par une validation collective en aval, après publication sur des plates-formes proposant de nombreuses publications et réunissant de larges publics. Cette mise en relation, ne s'appuyant sur aucune validation éditoriale, doit s'appuyer sur d'autres outils permettant à l'ensemble des lecteurs de participer à cette validation. Les plates-formes organisent une validation collective des productions éditoriales *via* des systèmes de vote, d'évaluation, de commentaires... qui alimentent des algorithmes de recommandation à destination de leurs usagers. Il s'agit d'une forme *horizontale* de validation collective dans laquelle le poids relatif de chaque usager est en principe identique.

C'est notamment le cas sur des plates-formes comme Amazon, sur lesquelles les commentaires et les traces d'achat et de navigation produites par les usagers servent de base au système de recommandation. Ces recommandations, en rendant visibles certains contenus éditoriaux au détriment d'autres, procèdent d'une forme de *validation* collective. Ainsi, c'est l'agrégation des achats et des consultations des internautes sur ces plates-formes qui rendront certaines productions éditoriales visibles dans les classements.

En construisant leurs mécanismes de recommandation sur ces logiques, les plates-formes s'exposent à deux risques principaux. Le premier est celui de l'émergence de concurrents. Ce risque, présent sur tous les marchés, prend pour ces plates-formes une forme spécifique qui doit être soulignée. En effet, des plates-formes comme Amazon, en proposant des mécanismes algorithmiques de validation ou de recommandation, fondent leur valeur ajoutée sur une maîtrise technologique bien plus que sur une réelle expertise éditoriale. Cette maîtrise technologique s'appuie également sur une base considérable d'usagers, mais elle reste concurrentielle par d'autres acteurs disposant de moyens techniques ou de bases d'usagers plus importantes même si ils sont largement extérieurs au secteur de l'édition. C'est l'enjeu des rapports de forces entre ces plates-formes et les moteurs de recherche notamment.

Le second risque est celui d'une concentration excessive des navigations et des achats sur un nombre trop restreints de contenus. En effet, les logiques algorithmiques de recommandation fondées sur les achats précédents vont tendanciellement limiter la distribution des consultations sur un nombre de plus en plus limité de titres, entraînant dans le même mouvement une concentration des commentaires et des évaluations des usagers. Cette dynamique risque de limiter les effets de cette validation collective à un nombre restreint de titres et donc de ne pas satisfaire les usagers sortant de ce cœur de sélection. C'est pourquoi les recommandations algorithmiques intègrent un part de découverte, difficilement évaluable mais qui doit maintenir un délicat équilibre entre une

proposition de proximité pertinente et une proposition originale élargissant l'offre proposée à l'utilisateur. En revanche, dans l'édition académique, les plates-formes de publication numérique, dépôts institutionnels ou archives ouvertes ne proposent pas d'outils d'évaluation. Le dépôt et la mise à disposition des contenus éditoriaux académiques sont facilités et leur accessibilité renforcée, mais la validation collective se déploie en d'autres lieux, notamment dans les revues.

La deuxième caractéristique de l'édition numérique qui joue un rôle dans les problématiques de validation est la possibilité de mise en relation de larges communautés. Il existe de multiples plates-formes qui ne proposent pas de contenus éditoriaux numériques mais sur lesquelles des communautés de lecteurs peuvent commenter et évaluer des productions éditoriales. On retrouve la même logique sur les plates-formes d'écriture où des communautés de lecteurs/auteurs réagissent et accompagnent les processus d'écriture. Il s'agit dans les deux cas de formes de validation collectives, la légitimation passant par la visibilité et les interactions en ligne, sur réseaux ou des plates-formes. La production de contenus savants est également concernée par cette dynamique avec le développement d'espaces de publication académiques hors revues traditionnelles qui proposent des outils d'échange et de discussion, à défaut d'outils formels d'évaluation. L'articulation entre la validation traditionnelle, par les revues académiques, et celle pouvant intervenir sur des plates-formes de publication en ligne est un enjeu important dans la recherche de processus d'évaluation de la recherche actuelle.

Ainsi, si la validation collective des contenus est une pratique ancienne dans le monde académique, elle constitue une nouveauté importante pour les contenus grand public. Mais dans les deux cas le modèle traditionnel de validation et de légitimation est remis en cause par la capacité des lecteurs/auteurs à échanger sur des plates-formes en ligne, instaurant de fait une validation collective hors éditeurs.

Réseaux sociaux académiques, épijournaux et mégarevues

Le monde de la publication académique a connu au cours des dernières décennies des évolutions considérables. Il n'est pas possible de les aborder toutes ici tant ce secteur a été bouleversé par l'édition numérique. Nous nous limiterons ici aux nouveaux dispositifs de validation collective. La validation dans le monde académique consiste à reconnaître, *via* des acteurs reconnus et des méthodes éprouvées, la validité d'une

proposition scientifique émise par un ou des chercheurs. Dans le contexte du numérique, cette mécanique de validation se heurte toutefois à de multiples écueils.

Le premier problème est l'inflation constante de la publication scientifique au cours des dernières décennies. Le nombre de publications à évaluer ne cesse de croître, et de nouvelles revues émergent régulièrement, leur fiabilité scientifique étant difficile à évaluer. Tous ces éléments posent la question de la pertinence du modèle traditionnel de validation et des propositions alternatives que la communauté académique peut mettre en place.

Plusieurs propositions de nouvelles formes de publication académique ont vu le jour autour du mouvement, maintenant ancien, de l'*open access*. Sans entrer dans le détail des différents voies ouvertes par ces nouveaux modèles (principalement les voies *dorée* et *verte*), il est important d'identifier les logiques de validation qu'elles impliquent. Les épirevues et épijournaux constituent une forme intéressante. Elles proposent un modèle qui se place en surplomb des archives ouvertes. Il s'agit concrètement d'opérer une sélection, donc une éditorialisation, des contenus académiques déjà disponibles dans les archives ouvertes et les dépôts institutionnels. Ces épirevues s'appuient donc sur une infrastructure – les archives ouvertes – mais également sur des processus de validation traditionnels en ayant recours à des comités de lecture et à une évaluation par les pairs. La validation y reste donc encore du ressort des chercheurs, avec les mêmes enjeux de légitimité, mais participe à l'activité d'un type nouveau de publications scientifiques, en accès libre.

Les mégarevues constituent une autre proposition intéressante. PLOS ONE, qui a publié en 2016 plus de 20 000 articles, en est un exemple. Dans ces mégarevues, l'évaluation est toujours réalisée par les pairs, mais ne concerne que les aspects méthodologiques des articles soumis. L'ensemble des coûts relatifs à la publication sont pris en charge par les chercheurs, suivant la voie dite « dorée » d'un financement en amont de la publication scientifique. Il s'agit bien d'un modèle de publication nouveau, qui autorise la publication d'un nombre colossal d'articles, inenvisageable sur support papier. Ici encore, la question de la validation par les pairs, même réduite à la seule dimension méthodologique, reste incontournable, elle s'appuie encore sur la légitimité académique évoquée plus haut.

La masse considérable de publications scientifiques disponibles en accès libre a ainsi induit un glissement des mécanismes de validation et de légitimation des revues académiques vers des réseaux sociaux spécialisés comme ResearchGate ou Academia.

L'objectif du chercheur n'est plus seulement d'obtenir une forme de reconnaissance académique par la validation de ses travaux, mais également de bénéficier d'une visibilité numérique maximale. Les réseaux sociaux académiques viennent ainsi répondre à un besoin de la part des chercheurs de valorisation de leurs travaux, mais aussi d'eux-mêmes.

Validation algorithmique

Qu'est-ce qu'un algorithme ?

Dans les environnements numériques, les algorithmes s'imposent comme une forme de validation complètement nouvelle par rapport au modèle papier. Désormais, la technologie aurait le pouvoir de valider un contenu et de se porter garante de sa qualité ou de sa pertinence. L'exemple le plus significatif de cette fonction de validation assumée par les algorithmes est Google Search.

Mais il faut d'abord définir ce qu'est un algorithme. Il s'agit d'un ensemble ordonné et fini d'instructions telles que, après chaque instruction, on peut connaître l'instruction suivante. En d'autres termes, l'algorithme est une procédure formalisée qui permet d'obtenir un résultat. Cette procédure a pour caractéristique d'être computable – c'est-à-dire qu'elle peut être réalisée par une machine. Prenons un exemple : si l'on possède un sac de billes de toutes les couleurs et que l'on souhaite isoler les billes rouges, on pourrait imaginer l'algorithme suivant : 1) prend une bille dans le sac ; 2) si la bille est rouge, mets-la à droite – si la bille n'est pas rouge, mets-la à gauche ; 3) s'il y a encore des billes, retourne à la première instruction – s'il n'y a plus de billes, arrête-toi.

Cet algorithme est exprimé ici en langage naturel, mais il peut très facilement être exprimé avec un langage informatique et exécuté par un ordinateur. La machine saura exactement ce qu'elle doit faire à chaque étape du processus. Progressivement, les billes seront séparées – les rouges à droite et les autres à gauche – et, lorsqu'il ne restera plus aucune bille, la machine saura qu'elle doit s'arrêter. On peut construire des algorithmes très complexes, qui permettent de réaliser des opérations variées : trier, classer, calculer, mais aussi, par exemple, ajouter un effet particulier à une image, réduire la taille d'un fichier audio...

La « magie » de l'algorithme

Dans les environnements numériques – en particulier sur le Web –, les algorithmes jouent un rôle fondamental dans la sélection, le tri et la hiérarchisation des contenus. Que l'on pense notamment à PageRank, l'algorithme à la base du moteur de recherche de Google, ou à l'algorithme d'Amazon, qui structure les suggestions d'achat sur la page de chaque usager, ou encore à EdgeRank, l'algorithme qui ordonne le « mur » des usagers de Facebook. Les algorithmes nous recommandent des contenus, en rendent certains visibles et d'autres invisibles. Le fait que le classement et la hiérarchisation soient le fruit du travail automatisé d'une machine pourrait laisser penser que le résultat proposé est objectif. D'ailleurs, la rhétorique adoptée par la plupart des sociétés propriétaires de ces algorithmes affirme justement leur neutralité et leur objectivité.

Lorsque l'on effectue une recherche sur Amazon, par exemple, le moteur nous propose de classer les résultats en ordre de « pertinence », donnant pour acquis que la définition de « pertinence » utilisée par l'algorithme est neutre. Google Search affirme la même chose, en nous proposant les résultats les plus « pertinents » et en choisissant une dizaine de pages « pertinentes » parmi les millions de pages existantes. La force de légitimation de cet algorithme est stupéfiante : concrètement, Google est devenu l'une des instances d'autorité les plus influentes de l'histoire. Pour preuve, 96 % des usagers ne regardent pas au-delà de la première page des résultats proposés [Harkless, 2012] : Google a ainsi acquis le pouvoir de choisir, parmi des millions de contenus, la dizaine de contenus « pertinents ». Ces dix premiers contenus sont légitimés par Google de la même manière qu'un livre imprimé est légitimé par une maison d'édition. Sauf que l'omniprésence de Google – le moteur de recherche le plus utilisé – et la quantité de contenus triés rendent la puissance de légitimation de son moteur incomparablement plus forte que celle de n'importe quelle maison d'édition.

La plupart du temps, la structure exacte de ces algorithmes est inconnue : ce sont des algorithmes propriétaires, dont les spécifications ne sont pas rendues publiques par les sociétés qui les développent. En outre, même les principes de base sur lesquels ils sont fondés – qui sont normalement déclarés par leurs propriétaires – sont inconnus de la quasi totalité des usagers. En conséquence, les résultats obtenus semblent être le fruit d'une « magie », d'un processus presque surnaturel et incompréhensible qu'il faut accepter parce qu'il se présente comme objectif.

La capacité de légitimation est renforcée justement à cause de cette impossibilité d'interroger les principes sur lesquels s'appuie la garantie de qualité et de pertinence.

Paradoxalement, le comportement du lecteur change totalement selon qu'il consulte un quotidien – dont il sait très bien que le choix des informations, leur hiérarchisation, ainsi que la façon de les présenter, dépend du point de vue de l'équipe éditoriale – ou une page web – dont il ne questionne que trop peu la pertinence, comme si le point de vue de l'algorithme était neutre ou complètement objectif.

Les valeurs des algorithmes : le cas de PageRank

L'objectivité des algorithmes n'est pourtant qu'un leurre : ils proposent en réalité de véritables visions du monde. En particulier, leur façon de classer, de trier et de hiérarchiser est toujours fondée sur une conception particulière de ce qui est « pertinent » et de ce qui est donc crédible et légitime.

Analysons ainsi les idées sur lesquelles est fondé PageRank, l'algorithme le plus important pour le moteur de recherche Google. Même si le véritable algorithme est propriétaire – nous n'avons par conséquent pas accès à son code – et que PageRank n'est pas le seul algorithme utilisé par Google pour classer les résultats, nous connaissons ses principes, qui ont été illustrés par ses créateurs dans un article [Brin et Page, 1998] et qui sont par ailleurs toujours mis en avant dans la communication de l'entreprise. PageRank classe les pages à partir des liens entrants : plus le nombre de pages pointant vers un document est grand, plus ce document est considéré comme important et donc pertinent. Bien évidemment, ce principe de classement n'est pas neutre, il renvoie à une idée particulière de la légitimité des contenus, qui s'oppose à un autre principe de légitimation : l'autorité. L'idée qu'un article est d'autant plus important qu'il est cité a été formalisée par Eugène Garfield en 1964 lors de la création du Science Citation Index. Un tel modèle n'a donc pas été inventé par les créateurs de Google, puisqu'il est à la base, depuis plusieurs décennies, du système de classement des contenus propre au milieu académique. L'objectif d'Eugène Garfield était de rendre plus démocratique le classement des contenus scientifiques et de remplacer le modèle d'autorité selon lequel l'importance – et la crédibilité – d'un contenu est déterminée par la réputation de l'auteur. Bien évidemment, le modèle de Garfield n'est pas parfait et produit parfois des aberrations : il ne tient notamment pas compte des raisons pour lesquelles un article est cité. Il n'est en effet pas rare qu'un contenu soit cité justement parce qu'il est faux. En d'autres termes, cette vision du monde tend à mesurer l'impact et la visibilité d'un contenu, plus que sa qualité. Elle fait ainsi correspondre légitimité et visibilité. Par ailleurs, le fait qu'un article soit bien classé entraîne d'autres citations, mettant ainsi en

place une boucle qui rend de plus en plus visible un groupe restreint de contenus – c’est ce que l’on appelle le *Matthew effect* [Merton, 1968].

Les principes du Science Citation Index ont été représentés en un modèle mathématique sur lequel s’est ensuite appuyé le PageRank : en d’autres termes, ces idées ont été transformées en une formule. Cette transformation est une interprétation, donc une opération idéologique qui modélise l’idée de départ, supposant une sélection et une systématisation de certains principes. Cette systématisation n’est pas nécessaire ni neutre. En revanche, Google tient beaucoup à affirmer que l’algorithme seul produit le classement, ce qui garantit l’objectivité des résultats de la recherche. Selon la rhétorique de l’entreprise, l’indexation de Google est « naturelle » puisqu’elle est le fruit du travail de l’algorithme sans être retouchée par l’intervention humaine. On revient à l’idée d’objectivité des algorithmes qui, comme nous l’avons souligné, n’est pas sans poser problème.

Google produit, de fait, de l’autorité : la quasi totalité des contenus auxquels nous accédons a été sélectionnée par ce moteur de recherche et notre niveau de confiance en son choix est très élevé, puisque nous nous fions aux premiers résultats qu’il nous propose. De fait, il a acquis un pouvoir de légitimation qui dépasse celui de n’importe quelle maison d’édition ou de n’importe quel auteur. Nous devons donc désormais l’apprécier comme un acteur majeur de l’édition, en prenant garde toutefois de ne pas considérer son classement comme un résultat objectif ou « naturel », mais bien comme une vision singulière du monde.

Tableau 6. Les algorithmes et moteurs de recherche du Web

Nom	Propriétaire	Créateurs (date de création)	Description
Yahoo! Search	Yahoo inc.	(1995)	À ses débuts, Yahoo! Search est moins un moteur de recherche algorithmique qu'un annuaire web animé par l'ambition d'un classement humain, manuel, du Web. À partir de 2000, Yahoo! Search est alimenté par le moteur de recherche d'Inktomi, société rachetée par Yahoo en 2002, puis par celui de Google jusqu'en 2004, date à laquelle Yahoo commence à utiliser son propre algorithme. Cependant, depuis 2009, Yahoo est de nouveau alimenté par un fournisseur externe, Microsoft Bing.
PageRank	Google	Larry Page et Sergueï Brin (1996)	L'algorithme célèbre du moteur de recherche de Google, fondé sur un principe méritocratique de réputation des sites web. S'inspirant du Science Citation Index d'Eugène Garfield (1964), il mesure en l'autorité d'un site web en calculant les liens entrants comme des votes, ces votes ayant plus de poids lorsqu'ils proviennent eux-mêmes d'un site réputé.
Hilltop Algorithm	Google (rachat en février 2003)	Krishna Bharat et George A. Mihăilă (1999)	Développé à l'Université de Toronto en 1999, l'algorithme Hilltop cherche à identifier les résultats les plus pertinents lorsqu'un internaute effectue une recherche vaste (ou peu détaillée) en attribuant une valeur d'autorité aux pages web. Le classement du Hilltop est réalisé au moyen d'un index de « documents d'experts », qui sont des pages web identifiées comme expertes sur un certain sujet. Plus les relations hypertextuelles entre une page web et sa page experte associée sont proches, plus la page en question sera qualifiée d'autoritaire (« <i>authoritative</i> »). L'algorithme Hilltop est utilisé par Google depuis 2003.
HITS Algorithme		Jon Kleinberg (1999)	L'algorithme HITS (Hyperlink-induced topic search) classe les pages web en leur attribuant deux scores définis par récursion mutuelle : 1) leur valeur de référencement (<i>hub value</i>), qui équivaut à la valeur de leurs liens renvoyant vers d'autres pages ; 2) leur valeur d'autorité (<i>authority value</i>), qui correspond à la valeur du contenu de la page, calculée par la somme des <i>hub values</i> des liens entrants. L'article de Kleinberg dans lequel il esquisse le fonctionnement du HITS constituera une source d'influence pour Page et

			Brin dans la conception du PageRank.
TrustRank		Hector Garcia-Molina, Zoltán Gyöngyi et Jan Pedersen (2004)	Exposé lors d'une conférence de la Very Large Data Bases Endowment Inc. à Toronto, TrustRank est une technique d'analyse semi-automatique des liens visant à distinguer les pages web pertinentes des pages employant une stratégie de spam de mots-clés et de liens sortants. En ce sens, le TrustRank s'inscrit dans la même lignée que le PageRank en ce qu'il utilise le réseau des liens comme mesure de qualité des pages web, mais la nécessité d'un contrôle humain de l'algorithme vise à combattre avec davantage de précision certaines tactiques du Search Engine Optimization (SEO ou « optimisation pour les moteurs de recherche », désignant les techniques d'optimisation de l'indexation et du référencement).
Bing	Microsoft	(2009)	Bing est l'aboutissement d'une longue succession de moteurs de recherche créés par Microsoft, à savoir MSN Search (1998), Windows Live Search (2006) et Live Search (2007). Peu d'informations sur l'algorithme de Bing sont rendues publiques, mais Microsoft a été accusé en 2010 de copier les résultats de recherche de Google.
EdgeRank	Facebook	Serkan Piantino (v. 2010)	Succédant à <i>Turning Knobs</i> , l'algorithme qui avait accompagné l'introduction du <i>News Feed</i> en 2006, EdgeRank procède à une hiérarchisation des informations selon une métrique d'affinités. Chaque fois qu'un utilisateur interagit sur Facebook, il produit un <i>edge</i> qui reçoit ensuite un score calculé à partir de trois variables : 1) U_e , ou la proximité relationnelle entre l'utilisateur-acteur et l'utilisateur-lecteur (le score d'affinité), elle-même déterminée par une série de paramètres (fréquence de visite du profil, fréquence des conversations entretenues sur le site, etc.) ; 2) W_e , qui correspond au poids de l' <i>edge</i> , c'est-à-dire à la nature même de l'action de l'utilisateur (publication, <i>like</i> , commentaires, partage, etc.) ; 3) D_e , le vieillissement (<i>decay</i>) du <i>edge</i> en fonction du temps passé depuis sa création. En somme, le score d'un <i>edge</i> peut être calculé grâce à la formule suivante : $\Sigma U_e W_e D_e$; plus le score est élevé, plus l'action prise par l'utilisateur-acteur a des chances d'apparaître sur le fil d'actualité de l'utilisateur-lecteur. Cependant, le terme d'EdgeRank est aujourd'hui tombé en désuétude : devant la massification des usages, la commercialisation croissante du réseau et la multiplication des appareils

mobiles, ces trois variables ont été submergées par près de 100 000 facteurs de poids (types de publication, personnalisation des paramètres par l'utilisateur, publicités, appareils utilisés, vitesse de connexion, etc.).

IV / La circulation des contenus

Le Web

L'importance du Web

Le terme « numérique » a désormais acquis une signification culturelle très large [Doueïhi, 2011]. Il n'est plus utilisé seulement comme un adjectif, mais aussi comme un substantif : le numérique. Cela signifie que ce terme ne réfère plus seulement à des outils ou à des technologies particulières, mais aussi à un ensemble de transformations sociales, culturelles, politiques, économiques et anthropologiques déclenchées par le développement technologique, mais ne dépendant plus exclusivement de ce dernier. En particulier, lorsque l'on parle des changements que les technologies numériques ont déterminés dans le champ de l'édition, on peut faire référence à une multiplicité d'outils et d'instruments informatiques, mais on parle aussi de changements institutionnels, sociaux, politiques, etc. Cependant, on ne peut pas nier que la naissance et la diffusion du Web a eu un impact particulièrement fort sur le monde de la production, de la circulation et de la légitimation des contenus. Le Web peut être considéré comme la première détermination des caractéristiques de ce que nous appelons le « numérique ». En particulier, le changement majeur qu'il a apporté, entraînant ensuite des transformations dans l'ensemble des autres domaines, est la facilitation de la circulation et de l'accessibilité des contenus.

L'idée du Web : faciliter la circulation

En 1989, Tim Berners-Lee, un informaticien anglais qui travaillait au CERN de Genève, remarque un problème majeur dans la gestion des informations au sein du centre de recherche : les documents ne sont pas assez accessibles et il est très difficile de trouver les informations dont on a besoin, malgré le fait qu'elles soient potentiellement disponibles. Le problème identifié par Tim Berners-Lee concerne la circulation de l'information : les employés du CERN changent souvent et les nouveaux arrivants ont du mal à récupérer les informations dont ils ont besoin, car la seule manière d'y parvenir est d'en discuter avec les autres. Parce que les documents concernant le centre sont en perpétuelle évolution, ils sont difficilement organisables en une structure figée (comme le

permettrait un livre imprimé). C'est en cherchant la solution à ce problème que l'informaticien jette les bases, en 1989, du *World Wide Web* [Berners-Lee, 1989].

L'idée fondamentale est de trouver une manière de rendre accessibles l'ensemble des documents en réseau et de les raccorder entre eux *via* un système de liens. Concrètement, cela signifie : 1) attribuer à chaque document un nom unique sur tout le réseau (le principe des URL, pour *uniform resource locator*) ; 2) établir un format universel pour les documents (le HTML, pour *hypertext markup language*) ; 3) définir un protocole de transmission de ces documents *via* Internet (le HTTP, pour *hypertext transfert protocol*).

Ces trois principes permettent la création d'une nouvelle forme de circulation des contenus : les documents deviennent accessibles à tous ceux qui disposent d'une connexion Internet. À partir de 1990, formater un document en HTML et le déposer sur un serveur signifie le rendre public, le « publier ». La fonction de diffusion des contenus propre à l'édition a trouvé une nouvelle expression. Par ailleurs, en raison de la généralisation de la connexion Internet, la diffusion *via* le Web produit une accessibilité incomparablement plus élevée que celle pouvant être offerte par une maison d'édition à travers l'impression.

Du Web 1.0 au Web 3.0

L'idée initiale du Web était de mettre à disposition des documents. Les documents formatés en HTML sont déposés sur des ordinateurs connectés (des serveurs) et deviennent des « pages » accessibles. Cette première époque du Web, qui va de sa naissance jusqu'à la fin des années 1990, correspond à ce qu'on appelle le Web 1.0, ou Web statique. Le modèle du Web 1.0 est assez proche de celui de l'édition papier, car il est fondé sur une idée du document comme une ressource stable – exactement comme une page imprimée. Il y a d'une part un producteur du document et d'autre part des lecteurs de ce document : les rôles sont clairement établis.

À partir de la fin des années 1990, le Web commence à s'orienter vers une nouvelle forme d'organisation des contenus : alors qu'il se contentait de mettre en relation une série de documents, il commence à inclure des relations avec des individus. Le Web est alors fait de documents et de personnes. C'est ce que Darcy DiNucci [1999] appelle le Web 2.0 et qu'on qualifiera par la suite de « Web social ». Les contenus deviennent dynamiques, car les usagers commencent à pouvoir ajouter eux-mêmes des informations : commentaires, recommandations, images, informations sur leurs profils, etc.

TripAdvisor, qui permet aux usagers d'évaluer restaurants et hôtels, apparaît en 2000. Wikipédia, l'encyclopédie participative, est lancée en 2001. Quant à Facebook et Twitter, ils apparaissent en 2006. Le rôle des « lecteurs » change radicalement : le lecteur devient contributeur. En même temps, les contenus perdent leur stabilité : ils sont susceptibles de changer à tout moment. Ce sont deux transformations majeures par rapport au modèle de l'édition papier.

Le Web 3.0, ou Web sémantique, ajoute aux relations entre documents et personnes d'autres relations avec les machines. Les informations, à partir du moment où elles sont correctement structurées et balisées, peuvent ainsi être utilisées, comprises et réagencées non seulement par les lecteurs, mais aussi par des algorithmes. Les machines deviennent à leur tour des instances éditoriales, capables d'organiser les contenus.

Encadré 1. L'histoire d'Internet et du Web

1969	Première connexion du ARPANET, réseau physique créé à des fins militaires. Le réseau fera l'objet d'une présentation officielle en 1972 à l'International Computer Communication Conference, date à laquelle soixante terminaux seront connectés à l'ARPANET.
1970	Sous la supervision de Stephen Crocker, le Network Working Group crée le premier protocole de transmission de données d'ARPANET, le Network control protocol (NCP).
1973	Création d'un second protocole, le Transmission control protocol (TCP), par Vinton Cerf et Robert Kahn.
1982	Devant le nombre croissant d'hôtes se connectant au réseau, Paul Mockapetris crée le Domain Name System (DNS, protocole de correspondance entre IP et nom de domaine).
1983	Réorganisation du TCP en un nouveau protocole, le Transmission control protocol/Internet protocol (TCP/IP). ARPANET adopte le protocole TCP le 1 ^{er} janvier 1983.
1986	La National scientific foundation (NSF) adopte le protocole TCP/IP et finance la construction d'une dorsale (câble transocéanique) traversant les États-Unis.
1990	S'inspirant de la notion d'hypertexte de Ted Nelson, Tim Berners-Lee développe le World Wide Web à partir du langage HTML (Hypertext markup language) et du protocole HTTP (Hypertext transfer protocol).
1992	Développement du navigateur Mosaic, dont hériteront entre autres Netscape et Internet Explorer. Premier navigateur à afficher directement les images, Mosaic contribue au gain exponentiel de popularité du World Wide Web.
1994	Tim Berners-Lee fonde le World Wide Web Consortium (W3C), qui assure la standardisation des technologies web et l'évolution de ses différents protocoles.
1995	L'introduction de Netscape (NSCP) sur les marchés financiers marque le début de la bulle internet (« dot-com bubble ») et un gain d'intérêt marqué du secteur économique à l'égard du Web. Après que l'indice NASDAQ a quintuplé en cinq ans, la bulle éclate vers mars 2000.
1998	Fondation de Google par Larry Page et Sergueï Brin.
2001	Suite à l'échec relatif de Nupedia, projet d'encyclopédie numérique dont la rédaction et la validation des articles devaient être assurées par des experts, Jimmy Wales et Larry Singer créent Wikipédia avec l'intégration de l'application libre wiki.
2003	Dale Dougherty propose le terme « Web 2.0 », parfois référencé sous les noms de « Web social » ou « Web participatif », pour désigner une nouvelle phase interactive du Web— interactivité qui était d'emblée envisagée par Berners-Lee à la création du World Wide Web.
2004	Création de Facebook, alors réservé aux étudiants de Harvard, par Mark Zuckerberg. Le réseau social sera rendu public en 2006.
2005	Création de YouTube par Steve Chen, Chad Hurley et Jawed Karim. Le site web sera racheté par Google en 2006.
2006	Création de Twitter par Jack Dorsey, Evan Williams, Biz Stone et Noah Glass.

Les limites de la circulation des contenus numériques

L'expansion du Web a rendu possible une circulation des contenus potentiellement sans limite. La généralisation de la connexion Internet fait qu'un document, une fois mis en ligne, devient accessible à des millions de personnes. Le rêve qui était à la base du projet de Tim Berners-Lee semble s'être réalisé. Il faut souligner combien ce rêve s'insère parfaitement dans la tradition de l'édition, dont l'une des principales fonctions est la circulation des contenus. L'invention de la presse à caractères mobiles au XV^e siècle avait déjà permis de multiplier la puissance de diffusion des contenus. Les manuscrits, pour circuler, devaient être copiés manuellement : le coût et le temps requis par cette opération en limitaient très fortement la circulation. Par ailleurs, la circulation était freinée par la difficulté physique de faire voyager le support : un livre arrivait en effet seulement là vers où on le transportait.

À première vue, la circulation sur le Web semble avoir abattu toutes ces frontières : la diffusion des contenus est désormais ubiquitaire, les fichiers peuvent être multipliés et copiés gratuitement, voyager à la vitesse de la lumière, etc. Ce discours ne résiste pourtant pas totalement à une analyse plus attentive. On compte au moins quatre limites principales à la circulation des contenus dans les environnements connectés : les barrières techniques et culturelles, les barrières géopolitiques, les barrières linguistiques et, enfin, la différence fondamentale entre accessibilité et visibilité.

1. Les barrières techniques et culturelles se manifestent principalement à travers les phénomènes de « fracture numérique » (*digital divide*) et de « fracture culturelle » (*cultural divide*). La fracture numérique renvoie à la division entre les zones géographiques ayant accès à une connexion Internet et celles qui n'en ont pas. Cette fracture peut se manifester entre différents pays ou zones géographiques (l'Afrique centrale a par exemple un accès très modeste à la connexion par rapport à l'Europe occidentale et l'Amérique du Nord), mais aussi à l'intérieur d'un même pays – on sait par exemple que les campagnes sont moins connectées que les villes. Cela crée une véritable différence sociale et culturelle entre les différentes parties de la population mondiale.
2. À cela s'ajoutent les barrières géopolitiques, qui correspondent aux politiques de censure et de limitation de l'accès à l'information propres à certains pays, tels que la Chine. Mais même dans des régions connectées et non soumises à la censure, on peut identifier une fracture culturelle : le fossé qui sépare les personnes dotées d'une culture numérique suffisante pour se servir des médias électroniques et ceux à qui cette culture fait défaut. Ce problème est réel et particulièrement critique à une

époque où plusieurs informations ne sont plus disponibles que sur un support informatique accessible *via* une connexion.

3. Les barrières linguistiques semblent constituer un fait banal, et pourtant il est important de souligner combien la circulation des contenus continue à s'opérer à l'intérieur de communautés linguistiques bien délimitées. Pour pallier cette difficulté, on assiste à une prédominance de l'anglais comme *lingua franca* – avec les conséquences politiques et culturelles que peuvent entraîner une telle suprématie. C'est pourquoi la recherche sur les traductions – en particulier sur les traductions automatiques – est devenue fondamentale. Cependant, les progrès récents réalisés dans le domaine de la traduction automatique sur base statistique par Google Translate, et l'association de cette approche à des algorithmes fondés sur l'apprentissage machine, n'ont fait que laisser dans les mains du géant de la Silicon Valley la gestion de cette limite et la possibilité de la dépasser. Si bien que la circulation des contenus a tendance à acquérir une structure fortement centralisée, au lieu de favoriser la décentralisation et la dissémination des documents propre au projet initial du Web.
4. Si ces deux premières barrières limitent l'accessibilité des contenus – en empêchant les usagers d'avoir, dans le premier cas, un accès matériel et, dans le second cas, un accès cognitif aux ressources –, une dernière barrière reste à prendre en compte : celle de la visibilité. Car en raison de la quantité considérable de contenus existants (et donc potentiellement accessibles), un nombre très limité de documents est réellement visible. En effet, pour être visible, un contenu doit être référencé, indexé ou recommandé sur d'autres plates-formes. Dans le cas contraire, il est noyé dans la masse d'informations existantes et demeure invisible. Mais quels sont les instances qui rendent un contenu visible – et qui, de fait, opèrent véritablement une fonction de diffusion ? Concrètement, il s'agit des moteurs de recherche – Google en premier lieu – et des réseaux sociaux – comme Twitter et Facebook. En d'autres termes, nous sommes encore une fois face au risque d'une forte centralisation de la fonction éditoriale.

Les librairies et les maisons d'édition face à la circulation GAFAM

La constitution d'un oligopole à franges

La structuration progressive de l'écosystème numérique a fait émerger une forte concentration des acteurs. Cette dynamique oligopolistique de concentration des activités numériques autour d'un nombre restreint de plates-formes est observable dès le début des années 2000 avec pour certaines plates-formes des externalités positives (retombées positives alimentant leur succès) et qui leur permet d'occuper progressivement des positions fortes sur leurs secteurs d'activités. C'est ainsi que le paysage des moteurs de recherche est passé de quelques dizaines d'acteurs au début du Web à un marché structuré dans sa quasi-totalité autour de trois acteurs (Google, Yahoo, Bing), hors les exceptions notables de la Chine (Baidu) et de la Russie (Yandex). Cette évolution est observable dans un grand nombre de secteurs d'activité avec comme corollaire une concentration des moyens et donc de la capacité à proposer des services innovants et coûteux en terme d'infrastructures. Ces nouveaux services permettant de capter de nouveaux usages, ce qui renforce davantage ces positions. La situation de ces marchés peut ainsi être qualifiée d'oligopole à franges, c'est-à-dire un marché structuré principalement autour d'un nombre réduit d'acteurs captant la majeure partie du marché et entourés d'un grand nombre d'acteurs de taille bien plus réduite se partageant une part très limitée du marché.

Cette tectonique des usages numériques a abouti à l'émergence d'acteurs largement dominants regroupés sous l'acronyme GAFAM pour Google, Apple, Facebook, Amazon et Microsoft. Ces acteurs occupent aujourd'hui une place prépondérante sur leur marché. Google dispose ainsi depuis plusieurs années d'une part de marché mondiale dans les recherches en ligne supérieure à 75 %. Son système d'exploitation Android a capté plus de 80 % de la part de marché des OS mobiles et son système de messagerie Gmail est également largement en tête des services de messagerie. Le cas d'Apple est un peu particulier : en effet, depuis l'arrivée de Samsung et d'Android sur le marché des smartphones, Apple a perdu sa première position en termes de ventes de smartphones et de part du parc des systèmes d'exploitation utilisés. L'entreprise reste toutefois largement dominante en termes de valeur avec environ 90 % des profits sur le marché des smartphones en 2016. Pour ces deux acteurs, la part du parc des OS installés induit également la part dans le marché des contenus associés (iTunes Store, App Store ou

Google Play). Ainsi, iBooks représente en 2015 11 % du marché américain du livre numérique en volume et 12 % en valeur.

Sur le marché des réseaux sociaux la position de Facebook est également largement dominante avec, en 2017, plus de 50 % du trafic sur les réseaux sociaux et une base d'utilisateurs actifs nettement majoritaire avec plus de 1,8 milliard d'usagers en janvier 2017. À la même époque, la part de marché de Facebook sur les réseaux sociaux était de 18 %, devant What's App à 11 %, également propriété de Facebook.

La place dominante de Microsoft dans le domaine des systèmes d'exploitation et bien connue, mais pour l'édition numérique, c'est Amazon qui est l'acteur le plus pertinent à observer. En effet, avec 74 % du volume des ventes de livres numériques aux États-Unis en 2015, Amazon est nettement en position de leader sur ce marché. Au-delà du livre numérique, la part estimée d'Amazon dans le commerce en ligne aux États-Unis est estimée à 43 % en 2016. Les industries culturelles connaissent depuis longtemps ce type de marché, les niveaux de concentration atteints dans le domaine de la musique enregistrée ou de l'édition en témoignent. Elles se retrouvent donc confrontées pour leurs activités numériques à une structuration du marché qui questionne leur propre place dans les filières.

L'édition académique a également été touchée par ces changements de modèles. On peut distinguer deux périodes. Au cours de la première période la logique de plateformes et de structuration oligopolistique du marché joua à plein, avec l'émergence d'acteurs atteignant des positions très fortes et rassemblant des volumes de ressources éditoriales de plus en plus importants. Cette situation s'accompagna d'une augmentation des budgets d'acquisition des bibliothèques (principale clientèle de ce type de documentation) pour les ressources numériques. Une seconde période s'est ensuite ouverte, en partie en réaction à l'augmentation des tarifs, avec le développement d'un modèle tourné vers un accès ouvert aux productions éditoriales scientifiques, s'appuyant sur les outils du Web permettant cette désintermédiation. Pour l'ensemble des secteurs éditoriaux, les conséquences principales de cette situation doivent être analysées de façon précise tant elle conditionne les modèles susceptibles d'être déployés dans les différents filières et notamment dans celle de l'édition.

Poids des modèles, évolution et comparaison internationale

La première conséquence de l'arrivée des GAFAM dans le monde du livre est une évolution forte des modèles de circulation. Nous en détaillerons trois : le modèle éditorial, le modèle de la publication Web et celui de l'édition académique.

Le modèle éditorial. L'activité éditoriale s'est structurée depuis longtemps dans un modèle d'affaires et d'organisation de filière stabilisé dans lequel les éditeurs occupent une place centrale [Chartron, 2016]. Du fait de sa position très en amont dans la filière, l'éditeur dispose de plusieurs leviers de pilotage. Ainsi, en sélectionnant les contenus qui seront édités, il joue un rôle clé dans l'alimentation du circuit de distribution. Il maîtrise également, notamment en France, les prix de vente et la répartition de la remontée des recettes du fait de la présence dans les principaux groupes d'édition d'acteurs en charge de la distribution. Ce dernier aspect est un élément central dans la construction des groupes d'édition (il a d'ailleurs joué un rôle important dans les mouvements de concentration qu'a connus l'édition française depuis les années 1980). Cette intégration des plates-formes de distribution est en effet un facteur important puisqu'elle permet de fixer les conditions de vente et de remise qui influent directement sur les marges des libraires et des éditeurs, surtout sur un marché à prix unique comme en France.

Ce modèle est largement remis en question avec l'arrivée des GAFAM et le passage à l'édition numérique. C'est l'arrivée d'Amazon et la croissance très rapide de sa part de marché dans la vente de livres en ligne qui a tout d'abord modifié les rapports de forces. En se positionnant dès sa création sur le créneau de la vente en ligne de livres imprimés, Amazon a développé l'offre de titres de loin la plus importante aujourd'hui et des capacités logistiques de haut niveau. Au cours de son développement mondial, l'entreprise a également constitué une base de clientèle de plusieurs centaines de millions de personnes. En atteignant ce poids sur le marché du livre physique, Amazon a inversé en partie le rapport de forces entre l'amont de la filière (l'édition) et l'aval (la distribution). Cette inversion est très importante, car elle ne se limite pas aux ouvrages sur support papier mais touche également le livre numérique. Sur ce marché, Amazon occupe une place de leader, ayant largement participé au développement du marché aux États-Unis avec la commercialisation d'une liseuse, la Kindle. Lancée en 2007 aux États-Unis et en 2009 en France, elle est régulièrement mise à jour et constitue pour Amazon un vecteur central dans la construction de son écosystème du livre numérique.

Avec le développement du marché du livre numérique, les GAFAM bouleversent encore un peu plus le modèle éditorial. L'ensemble de ces firmes ont investi d'une façon ou d'une autre dans le marché du livre numérique, à des niveaux et avec des stratégies différentes. Google a, dès 2004, lancé Google Print (qui deviendra Google Books ou Google Livres), programme de numérisation de masse d'ouvrages essentiellement issus de bibliothèques partenaires. S'il a donné lieu à plusieurs actions en justice, Google Livre constitue aujourd'hui un corpus de quelque 25 millions d'ouvrages. Apple a ouvert son iBooks Store simultanément au lancement de l'iPad, en 2010. Comme l'entreprise l'avait fait pour la musique, les applications ou les jeux, l'objectif est de proposer des contenus pouvant alimenter un dispositif numérique et ainsi encourager son achat. Microsoft, qui s'était associé en 2012 à Barnes & Nobles et à sa liseuse Nook avant de mettre fin au partenariat en 2014, envisageait à nouveau en 2017 de proposer des livres numériques à la vente *via* une librairie intégrée à son système d'exploitation. Enfin, Amazon a encore fait évoluer son offre numérique en lançant en 2014 Kindle Unlimited, qui propose par abonnement mensuel l'accès, sous conditions, à plusieurs centaines de milliers de livres numériques.

Ainsi, à l'exception de Facebook, les principaux acteurs du numérique et les plus importantes plates-formes en ligne ont tous une activité en lien avec le livre numérique. Pourtant, celui-ci occupe une place particulière dans leurs modèles d'affaires. En effet, pour l'ensemble de ces exemples, le livre numérique ne constitue pas une source directe de valeur ajoutée. Il joue un rôle de « produit d'appel » dont la valeur sera réalisée au travers d'autres services ou offres. Dans le cas d'Apple, c'est la vente de matériels qui constitue, de loin, la principale source de revenus. Pour Google, la logique est différente : en proposant un corpus de cette taille, c'est une captation de l'attention qui est visée et donc sa valorisation *via* les mécanismes publicitaires qui produisent l'énorme majorité des bénéfices de Google (il convient d'ajouter que les ouvrages numérisés constituent également un outil précieux dans le développement des services de traduction automatique sur lesquels travaille la firme depuis de nombreuses années). La part du livre numérique dans l'activité d'Amazon est compliquée à mesurer, mais il est probable qu'il joue également un rôle dans l'accroissement de sa base de clients, notamment à travers son service premium.

Cette analyse du rapport à l'édition des GAFAM souligne bien les enjeux actuels de l'édition numérique, s'agissant des rapports de forces entre éditeurs et plates-formes,

mais aussi de la place du livre, en l'occurrence numérique, dans la stratégie de ces dernières.

Le modèle de la publication Web. Nous proposons de rassembler sous cette terminologie un ensemble de modèles d'affaires qui sous-tendent en partie l'activité de publication numérique prise dans une acception large, couvrant aussi bien des textes que des contenus audiovisuels ou multimédia. Dans l'ensemble de ces modèles le principe est le même, à savoir une dissociation nette entre les circuits de distribution/diffusion et les circuits de financement. Dans les modèles traditionnels de publication, les mécanismes de remontée des recettes sont bien connus. Soit il s'agit d'une logique de paiement à l'acte qui amorce une remontée financière tout au long de la filière (par exemple la vente d'un ouvrage correspond à la perception par le libraire, le distributeur, l'éditeur, l'auteur... d'un montant calculé comme une part du prix de vente). Soit, et de façon non exclusive, il s'agit d'une logique de fructification de l'attention suscitée par le produit (le fameux « temps de cerveau disponible »). Dans cette seconde logique, des annonceurs achètent au diffuseur une partie de cette attention *via* un support publicitaire dont le prix est calculé en fonction de l'audience captée par le produit (c'est le cas par exemple lorsqu'un annonceur achète un espace publicitaire sur une chaîne de télévision). Ce faisant, ils alimentent également une remontée des recettes qui sera partagée entre les différents acteurs de la filière. Ce type de marché, dit biface, combine une source de revenus issue de ventes à l'utilisateur final et un financement publicitaire.

Sur le Web, les modèles utilisés pour la publication se sont largement construits sur un seul versant, celui de la publicité. Cette approche a longtemps été la principale solution de financement pour les publications en ligne : la vidéo est dans cette logique d'accès gratuit avec comme contrepartie des messages publicitaires ajoutés au début ou au cours de la vidéo. La presse en ligne, comme la musique, ont également longtemps exploité ce modèle.

Pour le secteur de l'édition, cette approche ne s'est jamais réellement déployée, le poids symbolique du livre freinant sûrement la juxtaposition de messages publicitaires et de contenus éditoriaux. De ce fait, la publication sur le Web utilise plusieurs mécanismes de financement : la publicité, le financement amont (subvention par exemple), le *freemium* (un accès gratuit au contenu mais proposant des formats ou services complémentaires payants) ou encore le financement participatif.

Le modèle académique. L'édition académique est constituée de deux marchés différents, celui des ouvrages ou manuels de formation et celui des revues de recherche. Historiquement, les auteurs d'articles parus dans les revues académiques ont très tôt investi la publication en ligne avec la création d'archives ouvertes dédiées à la publication et au partage d'articles. Les acteurs commerciaux ont également proposé très tôt des offres numériques avec un accès à distance aux revues numériques. À l'occasion de ce changement de support, les éditeurs scientifiques ont transposé leur modèle classique, celui de l'abonnement, en proposant aux institutions abonnées un accès en ligne à des « bouquets » de revues numériques. Ces abonnements étaient onéreux mais, ramenés au nombre de titres, présentaient encore un intérêt pour les bibliothèques, même si cela s'est amenuisé avec la hausse importante des tarifs imposée par les grands éditeurs scientifiques.

Le modèle de l'abonnement utilisé pour les publications académiques (revues ou ouvrages), papier d'abord et numérique ensuite, demeure cependant particulièrement intéressant.

Du point de vue économique, l'abonnement, ou plus précisément la licence d'accès, correspond à un modèle dans lequel le tarif est partiellement décorrélié de l'usage. En effet, le paiement par une institution académique ou une bibliothèque universitaire d'une licence pour des publications scientifiques ouvre l'accès à l'ensemble des usagers de cette institution. Quels que soient les accès effectifs à ces ressources, le montant restera globalement forfaitaire. Ce type d'offre présente deux caractéristiques : d'une part, le paiement d'un abonnement forfaitaire pour accéder à une offre de contenus ; et, d'autre part une collection de titres ou de contenus extrêmement importante.

Ce modèle de licences s'est également développé dans les secteurs de la musique avec Deezer en 2007 et Spotify en 2008, dans le secteur du jeu vidéo dès 2003 avec Steam ou dans celui de l'audiovisuel avec Netflix en 2010. Il est également présent dans l'édition grand public via l'offre Kindle Unlimited d'Amazon depuis 2014.

Reconnaissance des formats et marché

Le numérique, questionne également les modèles éditoriaux traditionnels du point de vue des formats – nous désignons ici par « format » des formes éditoriales et non des formats informatiques de fichiers. Avec l'arrivée des nouveaux formats numériques, les libraires et éditeurs sont amenés à revoir leur pratiques dans la construction de gammes de produits, et ce à deux niveaux. À un premier niveau, on assiste à un certain

nivellement des livres numériques vers un type unique de format(age) éditorial, indépendamment du fichier, qui ne propose pas au lecteur d'expériences de lecture différenciées. Il n'existe pas, pour le livre numérique, de logique de gamme ou de formats comme le support papier l'autorise : les éditions « grand format », « poche », « limitée », « luxe »... n'y ont en effet pas de sens, la perception qu'a le lecteur d'un livre numérique étant principalement conditionnée par le dispositif de lecture dont il dispose. Il n'est donc pas possible de faire varier la forme de l'ouvrage numérique pour en proposer différentes déclinaisons.

Un des leviers disponible pour repenser cette construction de gamme est de segmenter l'offre selon le niveau de services accompagnant l'ouvrage numérique. Ces services associés peuvent notamment consister en l'ajout de contenus illustratifs, d'approfondissement, etc. Cela revient en réalité à produire deux versions (ou éditions) d'un ouvrage : une version complète et une version partielle proposant uniquement le texte, sans les enrichissements. Les services complémentaires peuvent également être liés à la portabilité – c'est-à-dire la possibilité de lire le livre numérique sur plusieurs supports différents (smartphone, tablette, ordinateur, livre lu...) – ou à l'accessibilité de l'ouvrage numérique. Il s'agit d'une approche déjà utilisée dans certains domaines avec au départ un accès à une version en ligne de l'ouvrage ou du texte, et, en montant en gamme, l'accès à d'autres versions, souvent autonomes, sous forme de fichiers à utiliser sur tablette ou liseuse. Cette montée en gamme peut également correspondre à des accès simultanés ou multiples à l'ouvrage, permettant par exemple une lecture suivie sur plusieurs dispositifs.

Cette difficulté de segmentation du marché pour l'édition numérique est liée, à un deuxième niveau, à la question de la fixation des prix de vente. Depuis les débuts de l'édition numérique, la question du prix du livre numérique s'est posée, du point de vue économique et législatif. Du point de vue économique la difficulté consiste pour l'éditeur à définir un prix de vente qui intègre la nouvelle structure de coûts (plus de coûts d'impression mais des coûts de développement), le nouveau circuit de distribution, la disparition des gammes (moyenne entre grand format et poche) et l'acceptation par les acheteurs. Cette équation est d'autant plus complexe à résoudre que certaines plateformes de distribution comme Amazon encouragent une uniformité des prix (9,99 \$ pour un ouvrage), comme Apple l'a fait pour la musique (0,99 \$ pour un titre). Si la question de la fixation des prix n'est pas complètement nouvelle pour les éditeurs, la souplesse potentielle dans la variation des niveaux de prix est par contre une vraie nouveauté.

Contrairement à l'édition papier, l'édition numérique dispose d'un nouveau levier marketing ou commercial *via* la possibilité de faire varier le prix du livre numérique tout au long du cycle de vie. Cette possibilité (proposer un prix de lancement, remonter ensuite au tarif normal ou proposer des variations liées à l'actualité par exemple) est déjà largement utilisée sur le marché des applications mobiles mais nécessite, de la part des éditeurs, des compétences nouvelles dans l'analyse en temps réel des volumes de ventes déclencher ou mesurer les effets des fluctuations tarifaires.

Les bibliothécaires face à la « grande bibliothèque numérique »

L'édition et le livre sont au cœur de l'idée de bibliothèque. Les ouvrages sur support papier en ont conditionné le modèle même : ses espaces physiques, construits autour des rayonnages, l'identité de ses professionnels ou encore son inscription dans la cité. Les développements du Web et du livre numérique amènent les bibliothèques vers un renouvellement de leur rapport à leur objet, de leurs pratiques et de leur définition en tant qu'institution.

Le Web comme grande bibliothèque

Depuis les débuts du Web, le parallèle avec les bibliothèques n'a cessé d'être développé et analysé. Il s'appuie sur deux caractéristiques communes au Web et aux bibliothèques : la facilité d'accès à un coût très faible et la mise à disposition de vastes collections de ressources accessibles à tous.

Ce rapprochement entre Web et bibliothèques se fonde également sur un aspect que nous avons déjà évoqué précédemment, le paiement forfaitaire décorrélé de l'usage effectif des ressources proposées. Ce modèle de licence ou d'abonnement, qui se diffuse aujourd'hui à un nombre croissant de secteurs d'activité en ligne, est historiquement celui de la bibliothèque. L'analogie entre les modèles va même plus loin. Dans les deux cas, s'il est proposé à l'utilisateur d'accéder gratuitement à une vaste collection de ressources, les services supplémentaires – comme le prêt, la suppression des publicités ou l'accès hors ligne (pour la musique par exemple) – seront payants ou nécessiteront une démarche d'identification.

Cette évolution importante des possibilités d'accès à des ressources, *via* la numérisation et ces nouveaux modèles, fait évoluer le rôle des bibliothèques et questionne la spécificité de leur modèle. Le rapprochement entre les modèles de la

bibliothèque et des industries culturelles est renforcé par la numérisation progressive des contenus proposés par les deux types d'acteurs. Cette numérisation conduit à des pratiques de recherche, de consultation et d'accès qui passent par les mêmes dispositifs numériques. La consultation en ligne des ressources proposées par la bibliothèque se fait *via* le même terminal que pour consulter librement des ressources publiées en ligne ou commercialisée.

Cette proximité entre les différents fournisseurs, rassemblés du point de vue de l'utilisateur sur le même dispositif, renforce la mise en concurrence de l'ensemble des acteurs, bibliothèques comprises, dans l'offre de contenus numériques. Cette mise en concurrence ne se limite pas à des questions de coûts mais couvre l'ensemble des aspects de l'expérience utilisateur : qualité des interfaces, facilité d'usage ou adéquation avec son propre écosystème numérique. En effet, avec le passage au support numérique, l'ensemble des acteurs du livre entre dans un nouvel écosystème, en ligne, dans lequel chacun doit réinventer ses rôles, modèles et valeurs ajoutées. Les bibliothèques n'échappent pas à ces enjeux. Ainsi, plus encore que les médias qui le précédèrent dans le développement numérique, le livre joue un rôle majeur dans l'évolution des bibliothèques. Cette évolution est observée et mesurée depuis plusieurs années pour les bibliothèques universitaires et académiques, en raison principalement du développement des offres numériques des revues spécialisées. Le développement des pratiques et la croissance de l'offre de livres numériques entraînent les bibliothèques de lecture publique vers les mêmes questionnements.

Car il ne s'agit pas pour le livre numérique d'un « simple » changement de support comme le passage de la VHS au DVD ou du disque vinyle au CD. Il s'agit d'un changement à la fois de modèle (l'acquisition, la sélection, l'indexation, la mise à disposition ou la conservation changent radicalement) mais aussi de place dans les écosystèmes de l'édition et plus largement dans les pratiques des usagers. En intégrant le livre numérique à ses collections la bibliothèque se positionne dans une offre éditoriale numérique bien plus vaste, accessible en ligne. Si ce rôle de repérage et de signalement de ressources en ligne n'est pas nouveau ou lié uniquement au livre numérique, il est toutefois aujourd'hui un enjeu fort de la visibilité des bibliothèques. En effet, depuis les premiers temps du Web les modalités d'accès aux ressources publiées en ligne ont largement évolué. D'une identification des différents acteurs, publics ou privés, *via* leur site web les usages sont passés à un accès direct aux ressources *via* les moteurs de recherche. La visibilité institutionnelle des bibliothèques au travers de leur site web n'est

donc plus suffisante, celui-ci ne constituant plus un point de passage obligé pour accéder aux collections. L'enjeu, et les bibliothèques ont fortement investi ce champ, est donc la visibilité, la « trouvabilité » de l'offre des bibliothèques sur le chemin de navigation des internautes.

Cette nouvelle donne implique deux enjeux différents. D'une part il est primordial pour les professionnels des bibliothèques de maîtriser les techniques d'intégration des ressources des bibliothèques dans les résultats des moteurs de recherche. Cette intégration suppose une ouverture des catalogues au Web pour permettre aux moteurs de recherche d'indexer efficacement, titre à titre, l'offre des bibliothèques. D'autre part, la diffusion des métadonnées des catalogues aux moteurs de recherche pose la double question de la visibilité des bibliothèques en tant qu'institution et de leur ancrage territorial.

Bibliothèques et algorithmes

Avec le passage sur support numérique de l'offre documentaire des bibliothèques une autre dimension du rôle des bibliothèques a également été modifiée. En effet, il appartenait jusqu'à présent aux bibliothèques d'établir l'ensemble des relations entre les documents de leur fonds. Cette mise en réseau s'incarne toujours aujourd'hui dans les pratiques de catalogage et de cotation des documents. En attribuant à un document une côte et donc un emplacement, physique ou virtuel, dans une collection, les bibliothécaires construisent un maillage sémantique qui relie les éléments de la collection sur la base des métadonnées extraites ou produites autour des documents. Il s'agit d'une expertise professionnelle au cœur des métiers des bibliothèques qui consiste à attribuer, *via* une notice par exemple, un ensemble d'attributs, libres ou prédéfinis. Ces attributs, côte, mots-clés, thématiques... permettent de rassembler au sein d'une organisation tangible (les étagères d'une bibliothèque) ou virtuelle (la page de résultats du catalogue) des ressources documentaires.

Cette construction sémantique, à l'échelle de la collection ou d'un sous-ensemble, constitue une valeur ajoutée unique des bibliothèques. Elle est établie sur un temps long et s'appuie sur des outils et des méthodes spécifiques et éprouvées. Cet aspect du travail des bibliothèques ne leur est pas exclusif, il se retrouve dans de très nombreux secteurs d'activité, de la librairie au supermarché. La particularité réside pour les bibliothèques dans deux aspects : le des documents traités, et le caractère intellectuel des ses traitements. En effet, la production de métadonnées nécessite en bibliothèques une connaissance des domaines traités, une interprétation éventuelle des informations

disponibles (titre, éditeur...), etc. Cette expertise limitait jusqu'à présent l'automatisation d'une partie de ces tâches et a conduit les bibliothécaires à développer des pratiques renforcées de mutualisation et de partage.

Le glissement vers le numérique des documents peut être appréhendé au travers de trois aspects particuliers de ce type d'activité des bibliothèques. Tout d'abord, comme nous l'avons évoqué, les bibliothécaires, en tant que professionnels de l'information, disposent de compétences élevées dans la production de métadonnées, la manipulation de données structurées et l'utilisation d'outils puissants d'interrogation de corpus. La difficulté est de positionner la bibliothèque, soit comme point d'entrée vers des collections provenant en partie du Web (sur lesquelles le niveau de stabilité et de structuration des données n'est pas garanti) ou vers des ressources provenant de bases connues et identifiées (archives ouvertes par exemple) ; soit comme fournisseur de ressources, par exemple de données bibliographiques, qui seront utilisées et reprises par d'autres points d'accès comme les moteurs de recherche. Dans ce dernier cas, c'est la perception de la bibliothèque par les usagers qui posera question.

Ensuite, les documents numériques autorisent une forme nouvelle de traitement algorithmique qui s'appuie sur les documents directement et non sur les métadonnées. C'est déjà le cas pour les textes, aisément manipulables par des outils informatiques, mais cette capacité de traitement automatisé et algorithmique se déploie également pour les autres formats comme la vidéo ou la musique. Cette capacité offerte par le numérique de confier la production des métadonnées (des index dans la plupart des cas) à un traitement logiciel fait glisser l'expertise du traitement documentaire des professionnels de l'information aux spécialistes des outils informatiques. Cette approche n'est évidemment pas exempte de difficultés et de limites. Le bruit, par exemple, dans les résultats d'une recherche, c'est à dire la propension du système à proposer des résultats non pertinents par rapport à une requête, est important (il suffit d'observer le nombre de résultats d'une requête sur un moteur de recherche) mais la masse de documents traités et la capacité de l'algorithme à ordonner les résultats selon des critères dits de « pertinence » suffisent apparemment à répondre aux attentes des usagers. Il s'agit d'une évolution importante car elle place l'algorithme comme principal outil de traitement et d'interrogation des corpus documentaires, sans avoir recours à un traitement et une production humaine de métadonnées.

Enfin, la place centrale prise par les algorithmes dans l'indexation et l'interrogation des corpus documentaires a conduit les bibliothèques à déployer deux approches. D'une

part une évolution des interfaces d'interrogation des catalogues qui, en se rapprochant des moteurs de recherche, amène à masquer la complexité et la finesse des métadonnées exploitables. Cette approche pose en creux la question du *retour sur investissement* du travail conséquent de production manuelle de métadonnées. Cette interrogation est d'autant plus forte qu'on assiste à une forme d'externalisation des activités d'indexation (lorsque les ressources sont acquises avec des métadonnées déjà établies) et des outils d'interrogation (comme avec les outils de découverte, qui permettent une interrogation unique de multiples réservoirs de contenus). D'autre part, les bibliothèques ont travaillé à la diffusion et à l'intégration de leurs propres métadonnées dans les index des moteurs de recherche. Ce faisant elles se placent ainsi dans le *champ de vision* des usagers lorsqu'ils utilisent un moteur de recherche pour leurs recherches documentaires.

Le dernier aspect de la médiation algorithmique à l'œuvre sur le Web est l'exploitation par les algorithmes des données d'usage issues des pratiques des internautes. Cela correspond à des logiques de recommandation bien connues sur les réseaux sociaux mais également utilisées par des moteurs de recherche. Il s'agit d'exploiter les données d'usage provenant des navigations de l'internaute, des liens qu'il suit ou encore de ses achats. À partir de ces données, il devient possible de construire un maillage de proximité entre des pages web ou des produits comme nous l'évoquions pour les bibliothèques. À la différence principale que cette proximité est construite de façon automatisée et individualisée.

Dans cette confrontation des bibliothèques aux algorithmes d'indexation, d'interrogation et de recommandation celles-ci ne sont pas dépourvues d'atouts. Leur statut d'acteur public leur permet en effet de proposer une véritable transparence concernant les algorithmes qu'utilisent leurs outils. Par rapport aux moteurs de recherche, cette capacité de rendre public le fonctionnement de leurs outils représente une réelle valeur ajoutée. Elle permet de mettre en avant le caractère neutre du traitement documentaire et de l'accès proposés en bibliothèques, cette neutralité et cette ouverture pouvant aujourd'hui rencontrer un écho favorable auprès des usagers. Cette neutralité se retrouve également dans le fait que l'indexation, et donc les proximités entre les ouvrages, sont établies de manière stable, en amont et indépendamment des usages. Cela induit une médiation et une recommandation basée uniquement sur une logique sémantique et non sur les intérêts commerciaux de la mise en avant d'un document ou d'un site Web en particulier.

Ce déploiement des potentialités d'indexation algorithmiques offertes par le format numérique ouvrent, entre autres, deux champs de réflexion pour les bibliothèques. Le premier champ concerne l'exploitation des données d'usage pour des services de recommandation ou de médiation. Cette question est complexe pour les bibliothèques. Elle s'inscrit dans une tension délicate entre d'une part la garantie d'une forme relative d'anonymat, c'est-à-dire la garantie que les données d'usage ne seront exploitées que pour permettre le fonctionnement normal de l'établissement (prêts, retours, l'identification auprès des fournisseurs de ressources en ligne...) et d'autre part la mise en œuvre de services exploitant ces données, de la forme « ce qui ont emprunté ça ont aussi emprunté ça ». Cette réflexion doit prendre en compte plusieurs aspects : le cadre juridique qui définit les conditions de conservation et d'utilisation des données personnelles en bibliothèque, l'attente des usagers de retrouver dans leur bibliothèque des services similaires à ceux qu'ils utilisent ailleurs sur le Web ou encore l'importance qu'ils accordent à la protection de leurs données personnelles.

Le deuxième champ est celui de l'exploitation des potentialités offertes par la dématérialisation des collections pour proposer des interfaces d'exploration innovantes. Avec le livre papier, les contraintes physiques liées au rangement de volumes dans l'espace réel amènent les bibliothèques à choisir des modalités d'organisation qui privilégient la proximité. Ainsi, dans la plupart des bibliothèques, la disposition des espaces et des documents donne aux usagers une indication sur l'organisation, généralement thématique, des collections. Les classifications utilisées en bibliothèques, comme la classification de Dewey, définissent l'arborescence de ces thématiques. Cette classification choisie par la bibliothèque est la seule proposée aux usagers. Celle-ci placera tel ouvrage à côté de tel autre, imposant ainsi une modalité unique de découverte et d'exploration.

L'édition numérique ouvre de nouvelles possibilités dans ce domaine. Affranchie des contraintes physiques, les collections peuvent être structurées autour d'une multitude d'axes différents et ainsi être abordées selon des critères très différents (par exemple le nombre d'emprunts, la date d'acquisition, la longueur ou la mention d'une personne ou d'un lieu). Ces possibilités de structuration autour de critères différents, choisis potentiellement par l'utilisateur, créent de nouvelles proximités entre les ouvrages. Elles permettent une exploration et une sérendipité bien plus dynamiques que sur papier, à condition que les interfaces proposées le permettent. Les listes proposées par les catalogues, bien que pouvant souvent être triées selon de multiples critères, restent encore

trop limitées pour permettre la même facilité de découverte que la déambulation dans les espaces physiques des bibliothèques.

Communautés et circulation

En intégrant des produits éditoriaux numériques dans leurs collections, les bibliothèques déploient progressivement une offre documentaire adressée non plus à un territoire mais à une communauté d'utilisateurs. Cette communauté, caractérisée jusqu'alors par une proximité géographique avec les établissements peut s'affranchir des contraintes de déplacement pour se rassembler autour d'une thématique ou d'un corpus particulier (un exemple de ces communautés est celle rassemblée autour des ressources proposées dans Gallica par la BNF). En utilisant les réseaux sociaux ou leur propre site web, les bibliothèques construisent des communautés d'utilisateurs en ligne, qui ne s'inscrivent plus dans un territoire défini.

Plus largement, l'édition numérique amène les bibliothèques à interroger le maillage territorial existant à l'aune d'une offre dématérialisée accessible de n'importe où. Dans certains établissements des aspects administratifs ou réglementaires limitent l'inscription des utilisateurs à ceux issus d'une zone géographique précise, ce n'est pas le cas partout et cela n'apporte pas d'éléments de réponse quant à la pertinence de ce type de maillage d'un territoire virtuel. En s'affranchissant des contraintes logistiques et matérielles d'accès physique aux ouvrages, l'édition numérique questionne le sens des communautés d'utilisateurs construites autour des bibliothèques. Elle ouvre des opportunités réelles de toucher un public plus large, disséminé sur des territoires plus vastes, parfois éloigné ou empêché dans son accès à la bibliothèque comme bâtiment. Ces nouvelles communautés d'utilisateurs amènent une nouvelle conception du rôle de la bibliothèque comme espace public, numérique, et comme espace de sociabilisation.

Conclusion

Permanence des fonctions

Dans ce livre, nous avons montré comment le développement des technologies numériques affecte le monde de l'édition, sur le plan des formes de production, de légitimation et de diffusion des contenus. Malgré l'importance de ces changements et leur impact sur l'ensemble de l'écosystème éditorial, il est nécessaire de constater que la fonction éditoriale ne disparaît pas et qu'elle ne semble aucunement vouée à la disparition. Si le Web et certains de ses grands acteurs ont profondément changé les enjeux de l'édition, ils n'ont pu le faire qu'en se chargeant, d'une manière ou d'une autre, de la fonction éditoriale.

Ce constat a deux implications : d'une part, les dispositifs de médiation qui permettent aux contenus d'exister et d'être accessibles continuent d'opérer et d'avoir un rôle fondamental dans nos sociétés ; d'autre part, il faut interroger les formes de ces dispositifs et les manières de les déterminer. En effet, si la fonction éditoriale continue d'exister et d'avoir un rôle fondamental, il est indéniable que les instances et les institutions qui s'en chargent sont en train de changer. En particulier, les maisons d'édition, qui ont été pendant plusieurs siècles les dépositaires presque uniques de ce savoir-faire, ne sont plus les seuls acteurs à remplir la tâche de produire, légitimer et diffuser les contenus. La question qu'il devient urgent de se poser est donc : qui sont les instances éditoriales d'aujourd'hui ? Et comment les institutions chargées de cette fonction dans les sociétés prénumériques peuvent-elles continuer à avoir une place dans les sociétés numériques ?

Au constat de la persistance de la fonction éditoriale s'ajoute le fait que l'édition numérique n'est pas qu'une question de maisons d'édition ni d'éditeurs – au moins dans le sens restreint qu'on a donné à ce terme dans les derniers siècles. Il s'agit donc d'un changement qui ne concerne pas seulement un métier ou une profession, mais qui bouleverse un nombre très élevé d'acteurs en reconfigurant la construction de l'ensemble de notre culture collective.

Enjeux et défis pour les acteurs

Depuis l'apparition du Web, nous avons assisté à l'émergence de plusieurs nouveaux acteurs qui assument un rôle fondamental dans le processus de production,

légitimation et diffusion des contenus. En particulier, nous avons souligné l'importance des entreprises rassemblées sous l'acronyme GAFAM. Google, Amazon, Facebook, Apple, Microsoft, mais aussi, plus récemment, AirBnB, Uber, Netflix... se sont imposées dans le panorama des instances gérant la production du savoir.

Le premier défi pose l'époque numérique à l'édition est donc de repenser, et possiblement redessiner, les sources d'autorité et de pouvoir. Qui sont les dépositaires du savoir ? Qui a le pouvoir de légitimer des contenus et de les rendre accessibles ? Quelles sont les valeurs et les idéologies portées par ces institutions ?

Ce défi crée indéniablement un malaise dans les institutions qui détenaient ce pouvoir avant le développement des technologies numériques : les maisons d'édition, mais aussi les universités, les institutions littéraires, les librairies, les bibliothèques, peuvent se sentir dépossédées d'une fonction qui a été pendant longtemps entre leurs mains. La première réaction pourrait être – et est souvent – de se retrancher dans des positions de résistance aveugle, en niant par exemple les effets de ces changements et en revendiquant la nécessité de revenir à des formes éditoriales prénumériques ou en méprisant les nouvelles instances éditoriales qui s'imposent dans le panorama de l'industrie culturelle.

Or notre conviction est que cette position n'est pas viable. Il est au contraire nécessaire de reconsidérer le rôle des acteurs ainsi que de repenser les formes que peut prendre aujourd'hui la fonction éditoriale.

À partir de ces considérations, on peut identifier trois défis majeurs que les acteurs de l'édition doivent impérativement relever : 1) les formes d'autorité et de pouvoir doivent être analysées et reconsidérées ; 2) la structuration des contenus requiert de nouveaux savoir-faire et ouvre de nouveaux champs d'action que les acteurs de l'édition doivent s'approprier ; 3) les métiers des acteurs prénumériques de l'édition sont appelés à se renouveler pour pouvoir s'adapter à la réalité des pratiques éditoriales actuelles.

Perspectives ouvertes

Ces défis ont déjà été relevés par un certain nombre d'acteurs du numérique : c'est ce qui leur garantit leur place dans le domaine de l'édition. Les GAFAM, mais aussi des initiatives comme Wikipédia ou un certain nombre de *pure players* ont construit leur légitimité en répondant à ces trois défis. Ils ont en effet tout d'abord repensé les formes d'autorité – on l'a vu par exemple à propos de la validation algorithmique de PageRank, de la validation formelle de Wikipédia ou de la production de visibilité par les réseaux

sociaux. Ils ont proposé de nouvelles manières de structurer les contenus, notamment en travaillant sur les métadonnées et sur les différentes manières d'exposer ces contenus. Ils ont finalement pensé de nouveaux métiers, fondés sur de nouvelles compétences et savoir-faire pertinents.

Si les acteurs traditionnels de l'édition veulent assurer leur rôle dans la production, la légitimation et la diffusion des contenus, il est nécessaire qu'ils répondent à leur tour à ces questions – ce que, sans doute, certains ont déjà commencé à faire. En premier lieu, il faut accepter que les rapports de pouvoir et les dispositifs d'autorité ont changé. Les institutions traditionnelles ne sont plus, de fait, les seuls dépositaires du savoir. Cela implique aussi d'accepter des changements dans les rapports institutionnels : pour prendre un seul exemple, Google est devenu un acteur qui peut être mis à la même table qu'une bibliothèque nationale ou qu'une université. Pour garder une place dans le groupe de ces acteurs, les institutions doivent se poser la question de la structuration des contenus, processus déjà largement engagé par nombre d'entre elles (en particulier les grands groupes d'édition). Comme le soulignait de façon provocatrice Hugh McGuire dans un billet de blog en 2013, le métier d'un éditeur devrait être de fournir une bonne API pour les livres. Une API (*Application programming interface*) est une interface qui permet de standardiser l'accès à des informations numériques. Par exemple, nous pouvons avoir une base de données et un système qui permet de l'interroger pour récupérer, exploiter et afficher des informations. C'est ce que font tous les services du Web, de Google à Facebook. L'API est justement ce qui structure, rend accessible et légitime un contenu. Et, conclut McGuire, c'est justement ce que devrait faire un éditeur [McGuire, 2013].

Ce qu'il est nécessaire de retenir de cette réflexion est qu'« éditer » ne se réduit plus à mettre en forme pour l'impression et la diffusion sur papier : il s'agit aussi de structurer et de mettre en forme pour le format numérique qui a ses propres particularités. Il ne suffit pas d'essayer de mimer le papier en produisant des versions homothétiques, il est nécessaire d'essayer de comprendre comment l'information peut être agencée, structurée et diffusée dans les environnements numériques. À partir de cet effort, il est nécessaire de s'interroger sur les compétences attachées au métier d'éditeur. Depuis les années 2000, les acteurs de l'édition ont ainsi appris à mobiliser de nouvelles compétences qui ne faisaient pas partie de leurs savoir-faire traditionnels.

Le futur de l'édition dépendra de la capacité de chaque acteur de repenser sa mission et de prendre une place dans le panorama qui se dessine. Nous croyons qu'il est

souhaitable que les acteurs traditionnels relèvent ces défis et qu'ils prennent place parmi les acteurs émergents en évitant les concentrations de pouvoir décrites par Morozov [2011]. Ce qui requiert sans doute que les acteurs de l'édition repensent l'ensemble de leur philosophie.

Repères bibliographiques

ABBATE J., *Inventing the Internet. Inside Technology*, Cambridge, MIT Press, 1999.

BACHIMONT B., « Nouvelles tendances applicatives : de l'indexation à l'éditorialisation », in GROS P. (dir.), *L'Indexation multimédia*, Paris, Lavoisier, 2007.

BARBE L., MERZEAU L. et SCHAFER V. (dir.), *Wikipédia, objet scientifique non identifié*, Nanterre, Presses universitaires de Paris Ouest, 2015.

BERNERS-LEE T., « Information management. A proposal », **CERN**, 1989, www.w3.org/History/1989/proposal.html.

BHARAT K. et MIHAILA G. A., « Hilltop : a search engine based on expert documents », vol. 10, 1999, <http://ftp.cs.toronto.edu/pub/reports/csrg/405/hilltop.html>.

BLOOD R., *The Weblog Handbook. Practical Advice on Creating and Maintaining your Blog*, Cambridge, Perseus Publishing, 2002.

BON F., « Le livre sera homothétique (ou ne sera pas) », 2010, www.tierslivre.net/spip/spip.php?article2003&var_mode=calcul.

BOULÉTREAU V. et HABERT B., « Les formats », in SINATRA M. E. et VITALI-ROSATI M. (dir.), *Pratiques de l'édition numérique*, Montréal, Presses de l'université de Montréal, « Parcours numériques », 2014, p. 125-146.

BRIN S. et PAGE L., « The anatomy of a large-scale hypertextual web search engine », *Computer Networks and ISDN Systems*, vol. 30, n° 1-7, avril 1998, p. 107-117, [https://doi.org/10.1016/S0169-7552\(98\)00110-X](https://doi.org/10.1016/S0169-7552(98)00110-X).

BRIN S., MOTWANI R., PAGE L. et WINOGRAD T., « The PageRank citation ranking. Bringing order to the Web », Technical Report, Stanford InfoLab, 29 janvier 1998, <http://ilpubs.stanford.edu:8090/422/>.

BUSH V., « As we may think », *Atlantic Magazine*, juillet 1945, www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/.

CARDON D., « Dans l'esprit du PageRank. Une enquête sur l'algorithme de Google », *Réseaux*, n° 177, avril 2013, p. 63-95.

— « Du lien au like sur Internet. Deux mesures de la réputation », *Communications*, n° 93, 2013, p. 173-186.

— *La Démocratie Internet. Promesses et limites*, Paris, Seuil, 2010.

— « L'ordre du Web », *Médium*, n° 29, 2011, p. 191-202.

CERUZZI P. E. « Aux origines américaines de l'Internet : projets militaires, intérêts commerciaux, désirs de communauté », *Le Temps des médias*, n° 18, juin 2012, p. 15-28.

CHARTRON G., « Édition et publication des contenus : regard transversal sur la transformation des modèles », in CALDERAN L., LAURENT P., LOWINGER H. et MILLET J. (dir.), *Publier, éditer, éditorialiser. Nouveaux enjeux de la production numérique*, De Boeck Supérieur, « Information & stratégie », 2016, p. 9-36, <https://halshs.archives-ouvertes.fr/halshs-01522295>.

COMPIÈGNE I., *Internet. Histoire, enjeux et perspectives critiques*, Paris, Ellipses, « Infocom », 2007.

COUSIN G., CROZAT D., LIÈVRE M., LOISEAU G. et VIALA L. (dir.), *Actualité de l'habitat temporaire. De l'habitat rêvé à l'habitat contraint*, Marseille, Terra-HN, « SHS », 2016, www.shs.terra-hn-editions.org/Collection/?-Actualite-de-l-habitat-temporaire-1-.

DACOS M. et MOUNIER P., *L'édition électronique*, La Découverte, « Repères », 2010.

DESEILLIGNY O., « Matérialités de l'écriture . Le chercheur et ses outils, du papier à l'écran », *Sciences de la société*, n° 89, octobre 2013, p. 38-53.

DINUCCI D., « Fragmented future », *Print Magazine*, 1999.

DOUEIHI M., *La Grande Conversion numérique suivi de Rêveries d'un promeneur numérique*, Paris, Seuil, « Points », 2011.

Facebook Newsroom, « Company Info », consulté le 27 janvier 2017, <http://newsroom.fb.com/company-info/>.

FELDMAN V., « 540 million active plugins makes WordPress a billion dollar market », Freemius, 8 avril 2015, <https://freemius.com/blog/540-million-active-plugins-makes-wordpress-a-billion-dollar-market/>.

FRASER J. et TEMPLE N. J., « How accurate are Wikipédia articles in health, nutrition, and medicine ? / Les articles de Wikipédia dans les domaines de la santé, de la nutrition et de la médecine sont-ils exacts ? », *Canadian Journal of Information and Library Science*, vol. 38, n° 1, juin 2014, p. 37-52.

GENETTE G., *Seuils*, Paris, Seuil, « Points », 2002 (1987).

GOLDMAN D., « Microsoft and Yahoo : search partners », *CNN*, 29 juillet 2009, http://money.cnn.com/2009/07/29/technology/microsoft_yahoo/.

GOODY J., *La Raison graphique*, Paris, Minuit, 1979 (1977).

GUICHARD É. (dir.), *Écritures. Sur les traces de Jack Goody*, Villeurbanne, Presses de l'Enssib, 2012.

GYÖNGYI Z., GARCIA-MOLINA H. et PEDERSEN J., « Combating Web spam with trustrank », in Collectif, *Proceedings of the 30th VLDB Conference*, Saint-Louis, Morgan Kaufmann, 2004, p. 576-587.

HARKLESS G., « Importance of showing up on the first page of Google », *The Unique Side of Entrepreneurship*, 2012, <http://prograshion.ceopress.com/2012/07/18/importance-of-showing-up-on-the-first-page-of-google/>.

HAUBEN R., « À la recherche des pères fondateurs d'Internet. Pourquoi a-t-on besoin d'une histoire d'Internet ? », *Multitudes*, n° 11, 2003, p. 193-99.

HAUSTEIN S., LARIVIÈRE V. et MONGEON P., « The oligopoly of academic publishers in the digital era », *PLOS ONE* vol. 10, n° 6, juin 2015, <https://doi.org/10.1371/journal.pone.0127502>.

HEYMAN S., « Google Books : a complex and controversial experiment », *New York Times*, 28 octobre 2015, <https://www.nytimes.com/2015/10/29/arts/international/google-books-a-complex-and-controversial-experiment.html>.

HÖFERT A. (dir.), *Miracles, Marvels and Monsters in the Middle Ages*, Infoclio.ch, « Living Books about History », 2016, <http://livingbooksabouthistory.ch/fr/book/miracles-marvels-and-monsters-in-the-middle-ages>.

H J., « Yahoo dumps Google Search technology », *Cnet*, 18 février 2004, <https://www.cnet.com/news/yahoo-dumps-google-search-technology/>.

JACKSON J., « Google : 129 million different books have been published », *PC World*, 6 août 2010, www.pcworld.com/article/202803/google_129_million_different_books_have_been_published.html.

KINCAID J., « EdgeRank : the secret sauce that makes Facebook's news feed tick », *Tech Crunch*, 22 avril 2010, <https://techcrunch.com/2010/04/22/facebook-edgerank/>.

KIRSCHENBAUM M. G., *Track Changes. A Literary History of Word Processing*, Cambridge, Belknap Press-Harvard University Press, 2016.

LEBERT M., *Le livre 010101 (1971-2015)*, 2015, www.010101book.net/fr/.

LEINER B. M. *et al.*, « A brief history of the Internet », *ACM SIGCOMM Computer Communication Review*, vol. 39, n° 5, octobre 2009, p. 22-31, <https://doi.org/10.1145/1629607.1629613>.

LEJEUNE P., « *Cher écran...* » *Journal personnel, ordinateur, Internet*, Paris, Seuil, 2000.

LESSIG L., *Culture libre. Comment les médias utilisent la technologie et la loi pour verrouiller la culture et contrôler la créativité*, Petter Reinholdtsen, 2015 (2004).

LEVREL J., « Wikipédia, un dispositif médiatique de publics participants », *Réseaux*, n° 138, 2006, p. 185-218.

LIEBOWITZ S. J. et MARGOLIS S. E., *Winners, Losers & Microsoft. Competition and Antitrust in High Technology*, deuxième édition, Oakland, Independent Institute, 2001.

MACCOLL J., « Out of the shadows : the case for a national repository of open content », *Insights*, vol 27, n° 2, juillet 2014, p. 205-209, <https://doi.org/10.1629/2048-7754.167>.

MCGEE M., « EdgeRank is dead. Facebook's news feed algorithm now has close to 100 K weight factors », *Marketing Land*, 16 août 2013, <http://marketingland.com/edgerank-is-dead-facebooks-news-feed-algorithm-now-has-close-to-100k-weight-factors-55908>.

MCGUIRE H., « A publisher's job is to provide a good API for books », *Tools of Change for Publishing*, 1^{er} février 2013, <http://toc.oreilly.com/2013/02/a-publishers-job-is-to-provide-a-good-api-for-books.html>.

MERTON R., « The Matthew effect in science », *Science*, vol. 159, n° 3810, janvier 1968, p. 56-63.

MILLE A., « D'internet au Web », in SINATRA M. E. et VITALI-ROSATI M. (dir.), *Pratiques de l'édition numérique*, Montréal, Presses de l'université de Montréal, « Parcours numériques », 2014, p. 7-11.

Ministère de la Culture et de la Communication, « Le secteur du livre. Chiffres clés 2014-2015 », mars 2016, www.culturecommunication.gouv.fr/content/download/137323/1508351/version/1/file/Chiffres-cles_Livre_SLL_2014-2015.pdf.

MOROZOV E., *The Net Delusion. The Dark Side of Internet Freedom*, New York, PublicAffairs, 2011.

— *To Save Everything, Click Here. The Folly of Technological Solutionism*, New York, PublicAffairs, 2013.

NORMAN D. A., *The Design of Everyday Things*, New York, Basic Books, 1988.

O'REILLY T., « What is Web 2.0 », O'Reilly, 30 septembre 2005, www.oreilly.com/pub/a/web2/archive/what-is-web-20.html.

PÉDAUQUE R. T., *Le Document à la lumière du numérique*, Caen, C&F, 2006.

PLANE S., « Les effets d'un instrument d'écriture à l'épreuve de la recherche. Deux ou trois choses que l'on sait ou que l'on ne sait pas sur le traitement de texte », *Repères, recherches en didactique du français langue maternelle*, n° 26-27, 2002, p. 163-186.

POIRIER P. et GENÊT P., « La fonction éditoriale et ses défis », in SINATRA M. E. et VITALI-ROSATI M. (dir.), *Pratiques de l'édition numérique*, Montréal, Presses de l'université de Montréal, « Parcours numériques », 2014, p. 15-29.

RAK J., « The digital queer : weblogs and Internet identity », *Biography* vol. 28, n° 1, juin 2005, p. 166-182.

REBILLARD F., *Le Web 2.0 en perspective. Une analyse socio-économique de l'internet*, Paris, L'Harmattan, « Question contemporaines », 2007.

ROSE M., *Authors and Owners. The Invention of Copyright*, Cambridge, Harvard University Press, 1993.

SCHAFFER V., et SERRES A. (dir.), *Histoires de l'Internet et du Web*, Infoclio.ch, « Living Books about History », 2017, <http://livingbooksabouthistory.ch/fr/book/histories-of-the-internet-and-the-web>.

SCHROEDER S., « Google : Bing's search results are a "cheap imitation" », Mashable, 2 février 2011, http://mashable.com/2011/02/02/google-bing-copying/#mOnnbj_2vEqu.

SINATRA M. E. et VITALI-ROSATI M. (dir.), *Pratiques de l'édition numérique*, Montréal, Presses de l'université de Montréal, « Parcours numériques », 2014, <http://parcoursnumeriques-pum.ca/pratiques>.

TAKŠEVA T. (dir.), *Social Software and the Evolution of User Expertise. Future Trends in Knowledge Creation and Dissemination*, Hershey, Information Science Reference, 2012.

TRÉHONDART N., « Le livre numérique "augmenté" au regard du livre imprimé : positions d'acteurs et modélisations de pratiques », *Les Enjeux de l'information et de la communication*, vol. 15, n° 2, janvier 2016, p. 23-37.

VANDENDORPE C., *Du papyrus à l'hypertexte. Essai sur les mutations du texte et de la lecture*, Paris, La Découverte, 1999, http://litmedmod.ca/sites/default/files/pdf/vandendorpe-papyrusenligne_lr.pdf.

VITALI-ROSATI M., « Qu'est-ce que l'éditorialisation ? », *Sens Public*, 18 mars 2016, <https://papyrus.bib.umontreal.ca/xmlui/handle/1866/16068>.

WRIGHT A., *Cataloging the World. Paul Otlet and the Birth of the Information Age*, New York, Oxford University Press, 2014.

Table des matières

Introduction.....	3
I La fonction éditoriale.....	5
Fonction de choix et de production.....	5
Fonction de légitimation.....	7
Fonction de diffusion.....	7
Instances éditoriales et maisons d'édition.....	8
Les éditeurs face au numérique.....	9
Des enjeux très hétérogènes.....	9
La littérature : nouvelles formes de diffusion.....	11
L'édition jeunesse : des nouvelles productions multimédia.....	12
Les manuels universitaires : nouvelles formes de consultation.....	13
Les encyclopédies : de nouvelles formes de légitimation.....	14
L'édition scientifique et savante : légitimation, modèles économiques, usages et nécessité de structuration.....	16
Édition et droits d'auteur.....	18
Naissance des droits d'auteur et naissance de l'édition.....	18
Formes de droit d'auteur – États-Unis/Europe.....	20
Problèmes actuels.....	21
Plusieurs modèles : du DRM aux creative commons.....	22
Tableau 1. Les différents types de Digital rights management (DRM).....	23
L'éditorialisation.....	23
Une définition de l'éditorialisation : éditer dans l'espace numérique.....	23
Différences par rapport à l'édition.....	25
Un processus ouvert.....	26
II / La production des contenus.....	30
La plate-forme et l'écriture.....	30
Les techniques et la pensée.....	30
Les outils numériques.....	31
Tableau 2. Principaux formats de textes associés aux logiciels qui permettent de les traiter.....	33
Enjeux stratégiques de la structuration des documents.....	33
Les formats de fichiers.....	33

Homothétique, fixed layout, reflowable.....	36
Livre-Web, livre-application.....	38
La chaîne de production numérique.....	39
Papier et numérique.....	39
La numérisation.....	40
Tableau 3. Principaux programmes de numérisation.....	44
Linéarité et non-linéarité.....	46
Les éditions augmentées.....	47
Tableau 4. Les <i>pure player</i> et les éditions augmentées.....	49
Le statut des contenus.....	51
Potentialités du numérique.....	52
Les métadonnées.....	52
Le XML et d'autres langages de balisage.....	54
La fouille de texte et d'autres possibilités.....	56
Le développement de l'autoédition (literacy numérique).....	56
Formes numériques de production des contenus.....	59
L'extension du domaine de l'édition.....	59
Les blogs.....	59
Les wikis.....	61
Tableau 5. Les CMS.....	62
Les réseaux sociaux.....	63
III / La légitimation des contenus.....	64
Les modèles « classiques ».....	64
La reconnaissance symbolique de l'édition papier.....	64
L'auteur se porte bien.....	65
La reconnaissance par le commerce.....	67
Le rôle des bibliothèques.....	68
Validation collective.....	69
Contenus grand public et savants et formes de légitimation.....	70
Réseaux sociaux académiques, épijournaux et mégarevues.....	72
Validation algorithmique.....	74
Qu'est-ce qu'un algorithme ?.....	74
La « magie » de l'algorithme.....	75
Les valeurs des algorithmes : le cas de PageRank.....	76

Tableau 6. Les algorithmes et moteurs de recherche du Web.....	78
IV / La circulation des contenus.....	81
Le Web.....	81
L'importance du Web.....	81
L'idée du Web : faciliter la circulation.....	81
Du Web 1.0 au Web 3.0.....	82
Les limites de la circulation des contenus numériques.....	84
Les librairies et les maisons d'édition face à la circulation GAFAM.....	87
La constitution d'un oligopole à franges.....	87
Poids des modèles, évolution et comparaison internationale.....	89
Reconnaissance des formats et marché.....	92
Les bibliothécaires face à la « grande bibliothèque numérique ».....	94
Le Web comme grande bibliothèque.....	94
Bibliothèques et algorithmes.....	96
Communautés et circulation.....	100
Conclusion.....	101
Permanence des fonctions.....	101
Enjeux et défis pour les acteurs.....	101
Perspectives ouvertes.....	102
Repères bibliographiques.....	105