

Une réflexion sur l'édition à l'époque du numérique

Il est évident que les formes de production, de circulation et de légitimation des contenus ont subi dans les dernières années des changements majeurs en raison du développement et de la diffusion des outils numériques. Le web a contribué à donner une importance centrale à l'écriture dans notre société : tout devient écriture. Dans ce contexte, il est urgent de mettre en place des modèles stables qui nous permettent de comprendre quel est aujourd'hui le rôle de l'édition.

On peut identifier deux tendances contradictoires dans les manières de produire et faire circuler des contenus : celle typique des grandes multinationales du web - on pourrait l'appeler « édition Gafam » en utilisant l'acronyme qui renvoie aux cinq géants Google, Amazon, Facebook, Apple et Microsoft, mais qui désignerait par extension l'ensemble des grandes entreprises qui dominent le marché numérique - et celle des spécialistes, des professionnels de l'édition et des chercheurs - qu'on nommera ici « édition savante ». L'*édition Gafam* est caractérisée par le fait d'être basée sur les données davantage que sur les documents : les entreprises commerciales sont engagées dans la captation de grandes masses d'informations, peu importe leur structure et leur sens. Ces informations ne doivent pas nécessairement être validées, car c'est leur masse qui est précieuse d'un point de vue économique. De la qualité de l'information, la valeur s'est ainsi déplacée vers la quantité d'information, sur la base de laquelle s'est construit leur modèle économique. Dans cette optique de *big data* ou de données massives, il est fondamental d'attirer et de rendre captif un nombre très élevé d'utilisateurs. Pour cela, l'*édition Gafam* concentre ses efforts sur l'expérience de l'utilisateur en offrant des services avant tout simples et performants (rapide), quitte à verser dans la superficialité, au profit de l'attractivité. Son fonctionnement doit être opaque, jusqu'à occulter les aspects techniques, afin que l'utilisateur ne puisse les questionner - l'exemple le plus clair est celui des interfaces Apple. Les usagers s'habituent à ne pas savoir comment la machine fonctionne, à ne pas voir les fichiers, à ne pas avoir accès à la configuration des logiciels. L'édition Gafam doit être efficace et répondre aux pratiques les plus communes. Elle évolue donc rapidement pour rencontrer les besoins et les désirs d'un large public.

À l'opposé se trouve l'édition savante dont le premier objectif est, normalement, d'être structurée et validée par des dispositifs de légitimation clairs et bien fondés. L'édition savante est donc riche, complexe, et produit avant tout des documents - et non seulement des données - qui déclarent de façon transparente leur structure et leur fonctionnement. L'édition savante est ainsi plus difficile à appréhender pour l'utilisateur non averti, ce qui l'expose au risque d'être moins efficace en termes de rapidité, et d'être donc moins consultée. Elle demande en effet une forte conscience technique de l'utilisateur, conscience qui est aussi celle des enjeux politiques liés à la circulation des contenus. L'édition savante requiert nécessairement plus de ressources, tant dans sa production que sa réception, générant une temporalité problématique, car en décalage avec

des exigences qui émergent des pratiques communes. Le grand problème de cette forme d'édition est donc qu'elle s'éloigne souvent des pratiques réelles - et non seulement des pratiques du grand public, mais aussi des pratiques des chercheurs, qui préfèrent utiliser - comme plusieurs études le démontrent (Rieger, 2009 - <http://firstmonday.org/article/view/2716/2385#p7> ; Jose Van Dijck, "Google Scholar as the co-producer of scholarly knowledge", in Takseva, *Social Software and the Evolution of User Expertise*, 2013) -, tant pour leurs recherches que pour leur diffusion, des outils grand public : par exemple Google Search à la place des moteurs de recherche sémantique destinés spécifiquement aux besoins de la recherche, ou encore Academia.edu ou ResearchGate comme dépôt de publication à la place des dépôts institutionnels et ouverts (Papyrus de l'Université de Montréal ou HAL en France).

La bataille entre le format json et le format xml illustre bien cette opposition : json, d'une part, est un format de balisage à plat, simple, léger, mais très peu structuré et non validé ; xml, d'autre part, est un format très riche, plus complexe, validé, mais produisant parfois des informations très peu exploitées par les utilisateurs (ce qui explique que la plupart des éditeurs/programmeurs, n'en voient pas l'intérêt).

Or le problème dans cette opposition est que les pratiques savantes ont tendance à s'uniformiser par rapport aux pratiques Gafam : d'une part parce que les chercheurs n'ont pas tous les compétences techniques suffisantes et de l'autre parce que les informaticiens qui développent les outils n'ont pas la sensibilité suffisante aux enjeux qui devraient intéresser les chercheurs.

Cet abandon des questions techniques aux informaticiens - surtout en sciences humaines - commence donc à avoir des effets néfastes. Le problème ne vient pas, à mon avis, des géants de l'édition gafam, mais du manque de conscience des chercheurs qui devraient être intéressés par l'édition savante.

La Chaire de recherche du Canada sur les écritures numériques - dont je suis le titulaire - développe une série de projets qui ont pour but de penser l'édition savante en réconciliant les caractéristiques scientifiques de structuration, de richesse et de complexité, et les pratiques réelles des chercheurs. Associer efficacité et richesse suppose peut-être de déplacer ces pratiques ou de les accompagner vers plus de scientificité, grâce à des outils et des protocoles plus adaptés aux différents utilisateurs de la chaîne éditoriale. Pour ce faire l'édition savante doit repenser ses fondements et ses modèles épistémologiques, tant d'un point de vue théorique que pratique.

Un des projets de la Chaire qui va dans ce sens est l'éditeur de texte Stylo. La philosophie de Stylo consiste à remettre dans les mains des chercheurs la gestion du balisage en partant cette fois-ci de leurs compétences, qui sont des compétences sémantiques plutôt que graphiques. Conçu sur le principe d'un éditeur WYSIWYM (what you see is what you mean), Stylo a le potentiel de changer l'ensemble de la chaîne de production des contenus dans le domaine des revues savantes en Sciences Humaines, en proposant une interface uniformisée et

sans perte de données, depuis la rédaction jusqu'à la diffusion, en passant par l'évaluation, la correction et l'édition.

L'éditeur propose aux auteurs un environnement d'écriture très simple, permettant un enrichissement sémantique à travers des fonctionnalités de balisage léger, mais spécifique au travail d'auteur. Le chercheur se voit ainsi donner la possibilité de baliser l'ensemble des informations qu'il produit de façon sémantique (titres, sous-titres, références bibliographiques, notes, index), simplement en sélectionnant les éléments, en leur appliquant des styles préprogrammés, ou selon les compétences de l'auteur, en utilisant le format markdown.

La suite de la chaîne éditoriale (évaluation, édition, diffusion) est assurée au sein du même outil, permettant aux producteurs de contenus – les éditeurs de la revue ou de la plateforme de diffusion, ou encore directement les chercheurs eux-mêmes – d'enrichir le texte en sélectionnant et ajoutant des métadonnées (mots-clés, nom de l'auteur, nom de la revue, etc.) et en les alignant avec des autorités.

Les métadonnées, étant alignées, pourront être facilement mises en relation avec d'autres contenus existants (traduction automatique en plusieurs langues, notamment, en se basant par exemple sur les alignements des vocabulaires RAMEAU de la Bnf et LCSH de la Library of Congress). Grâce à l'alignement avec des autorités, les contenus créés via Stylo pourront par ailleurs être mis en relation avec d'autres contenus – par exemple via un moissonnage qui interroge les autorités - et être mieux indexés par les moteurs de recherche spécialisés.

N'importe quelle plateforme de diffusion peut ainsi exploiter les contenus produits par Stylo. La brique logicielle centrale communiquera via une API avec plusieurs plateformes en fournissant des métadonnées très riches, sans que l'enrichissement soit réalisé par le diffuseur ou l'exploitant final – puisque pris en charge directement par le producteur.

Dans l'esprit des éditeurs WYSIWYM, le formatage graphique du contenu est automatiquement effectué lors de l'export, en s'appuyant sur des modèles programmables et intégrés à la chaîne éditoriale des revues et des plateformes de diffusion (notamment Erudit.ca qui participe au projet pilote).

Ces exports pourront être de plusieurs types selon la diffusion visée :

1. des fichiers XML selon les schémas sélectionnés (erudit, TEI, etc.)
2. des fichiers HTML pour publication directe sur des CMS (grâce à des API)
3. des fichiers print (.pdf) stylés selon des modèles programmables.

L'éditeur pourra être utilisé pour différents types de contenus, notamment :

1. les articles de revues
2. les monographies
3. les mémoires et les thèses

4. les billets de blogue de recherche

De cette manière, l'éditeur stylo essaie d'atteindre les niveaux de richesse et de complexité des contenus typiques de l'édition savante tout en offrant à l'auteur un environnement de travail qui ne demande pas des compétences techniques trop poussées.

Certes, des projets de ce type n'ont de valeurs que s'ils sont accompagnés par de vastes opérations de sensibilisation des publics. La question de la littéracie numérique est devenue particulièrement fondamentale aujourd'hui. Il faut que les chercheurs réalisent que l'acquisition de compétences de base qui leur permettent de savoir produire des contenus de qualité est la condition de la légitimité des contenus qu'ils produisent.

Nous sommes désormais arrivés à des formes de circulation des textes où la valeur de légitimation ne vient plus principalement de l'instance de publication - comme la maison d'édition ou la revue -, mais de la qualité technique du contenu - qui détermine sa visibilité, son indexation, sa bonne place dans l'ensemble des contenus disponibles. En d'autres termes, si dans le modèle de l'édition papier, on devenait auteur au moment où l'on était publié, à l'époque du numérique, on devient auteur parce qu'on produit des contenus correctement structurés.