



Raphaële Bats (dir.)

Construire des pratiques participatives dans les bibliothèques

Presses de l'enssib

Correct : une plate-forme de correction participative et collaborative pour la BnF

Isabelle Josse et Pauline Moirez

DOI : 10.4000/books.pressesenssib.4355

Éditeur : Presses de l'enssib

Lieu d'édition : Villeurbanne

Année d'édition : 2015

Date de mise en ligne : 30 janvier 2019

Collection : La Boîte à outils

ISBN électronique : 9782375460696



<http://books.openedition.org>

Édition imprimée

Date de publication : 1 janvier 2015

Référence électronique

JOSSE, Isabelle ; MOIREZ, Pauline. *Correct : une plate-forme de correction participative et collaborative pour la BnF* In : *Construire des pratiques participatives dans les bibliothèques* [en ligne]. Villeurbanne : Presses de l'enssib, 2015 (généré le 16 février 2021). Disponible sur Internet : <<http://books.openedition.org/pressesenssib/4355>>. ISBN : 9782375460696. DOI : <https://doi.org/10.4000/books.pressesenssib.4355>.

Ce document a été généré automatiquement le 16 février 2021.

Correct : une plate-forme de correction participative et collaborative pour la BnF

Isabelle Josse et Pauline Moirez

- 1 Lorsqu'une institution patrimoniale nationale comme la Bibliothèque nationale de France (BnF) s'engage dans une politique participative, elle a pour atout de pouvoir susciter une adhésion naturelle de contributeurs « citoyens », mais il lui est indispensable d'imaginer des projets d'envergure, à la hauteur du volume des collections concernées et donc des contributions attendues.
- 2 Ainsi, la plate-forme Correct¹ est le fruit d'un projet de Recherche & Développement (R & D), mené de janvier 2012 à juin 2015 et regroupant 9 partenaires publics et privés², dont l'objectif est la conception d'une plate-forme de *crowdsourcing*^{*}, appuyée sur un réseau social, pour la correction du mode texte issu de l'océrisation (OCR)³ des documents numérisés de Gallica, la bibliothèque numérique de la BnF⁴.
- 3 Les enjeux de la participation des usagers sont au centre de la méthodologie expérimentale du projet. Pour le préparer, nous avons réalisé un état de l'art⁵ du *crowdsourcing* dans les bibliothèques, dont la synthèse nous a permis d'identifier des facteurs de réussite et d'échec de projets similaires, en particulier en termes d'interactions avec les usagers. Par ailleurs, parmi les partenaires du projet, la BnF s'est concentrée sur la réalisation des expérimentations, afin de bénéficier d'un retour d'expérience concret sur les pratiques participatives. Cette démarche nous a donné une idée plus précise des services qui pourraient être véritablement intégrés dans la stratégie numérique de la bibliothèque, avec l'enjeu de passer d'une dynamique de projet à une dynamique institutionnelle, sans perdre l'esprit d'innovation indispensable pour la réussite d'un service participatif.
- 4 Nous souhaitons partager nos réflexions sur les moyens d'atteindre une masse critique de participants et de contributions, afin de remplir les objectifs de correction que nous nous sommes fixés, à savoir améliorer le mode texte des documents de Gallica afin de fournir une recherche plus efficace dans la bibliothèque numérique. Pour qu'un projet

de *crowdsourcing* portant sur les 3 millions de documents diffusés par Gallica ait un véritable impact, il est en effet nécessaire qu'un grand nombre de contributeurs participe. Par ailleurs, la plate-forme doit pouvoir trouver une certaine autonomie, afin que les usagers assument eux-mêmes le recrutement et l'animation. Un effort de médiation doit évidemment être apporté, mais il ne peut pas être démesuré. L'ampleur des collections concernées et la granularité extrêmement fine des corrections attendues rendent nécessaire cette autonomisation* des correcteurs, et tout l'enjeu du projet est là : recruter les contributeurs, les motiver à rester et revenir, et faire en sorte qu'ils s'approprient la plate-forme.

Recruter : comment faire venir les contributeurs sur Correct ?

- 5 Nous avons travaillé à identifier les contributeurs potentiels et les raisons qui les inciteraient à contribuer, et nous avons adapté l'ergonomie de la plate-forme et la médiation afin de toucher en priorité ces publics.

Premiers constats sur le recrutement pour des projets participatifs

- 6 Cette expérience de *crowdsourcing* n'est pas la première initiée par la BnF. Déjà en 2009, en partenariat avec Wikimedia France, 1 416 documents ocrésés avaient été ouverts à la correction collaborative sur Wikisource. Le résultat, bien que loin d'être négligeable (fin 2014, 54 documents corrigés par une double vérification, soit 10 300 pages ; 104 documents dont toutes les pages ont été corrigées mais validées par un seul correcteur, soit 30 000 pages), est décevant au regard de l'investissement. Nous avons eu du mal à identifier les communautés susceptibles de s'investir dans ce projet, si bien qu'aucune médiation ciblée n'avait pu être réalisée.
- 7 À la différence de pays anglo-saxons où le sentiment d'appartenance nationale est très puissant, en France la participation la plus active est réalisée à un niveau local : les premiers projets de *crowdsourcing* initiés par des institutions culturelles ont été le fait des services d'archives, qui s'appuient sur des généalogistes et érudits locaux très investis et aux pratiques collaboratives anciennes⁶. C'est donc finalement un défi pour une institution nationale de se lancer sur le chemin du participatif. Malgré tout, la BnF, en tant qu'institution patrimoniale nationale, dispose de l'atout particulier de pouvoir générer une participation « citoyenne », pour contribuer à améliorer l'accès aux documents du patrimoine. Le projet gagne ainsi à mettre en avant les « marques » BnF et Gallica.
- 8 De plus, une enquête sur les usages de Gallica en 2008 avait montré que les utilisateurs ne se sentaient pas légitimes pour corriger des contenus produits par une institution patrimoniale. En 2011, après deux années d'une médiation numérique intense qui a contribué à faire évoluer l'image de Gallica et à fédérer autour d'elle une communauté de Gallicanautes, une nouvelle enquête a montré que ceux-ci sont désormais prêts à participer à l'amélioration et à l'enrichissement des contenus⁷.
- 9 Enfin, nous avons constaté une évolution de la perception de la participation dans les bibliothèques, chez les professionnels, qui auparavant réticents, sont désormais plus enclins à voir la complémentarité entre les données structurées et normalisées

produites par les bibliothécaires et des données sociales moins cadrées mais répondant à des besoins nouveaux des usagers (recherche plein texte, granularité de description plus fine).

Identifier les leviers de motivation des publics-cibles

- 10 En s'appuyant sur ces constats, nous nous sommes naturellement adressés à des publics connus et auprès desquels il nous était facile de communiquer (usagers de la BnF, Gallicanautes). Nous avons également cherché à nous assurer une audience plus large auprès d'autres communautés utilisatrices de plates-formes participatives (blogueurs, généalogistes, contributeurs actifs sur les forums). Enfin, nous avons mené une réflexion spécifique autour des publics pouvant être bénéficiaires des résultats de la plate-forme et, notamment, les personnes mal- et non-voyantes.
- 11 Les Gallicanautes constituent un public évident, à condition de savoir construire avec eux une interaction d'un nouveau type et de les convaincre de participer. L'amélioration de l'OCR, même si elle bénéficie directement aux usagers de Gallica, ne semblait pas pouvoir être un facteur de motivation suffisant pour mobiliser de nombreux contributeurs. Aussi la sélection des premiers documents s'est faite au regard des statistiques d'usage de Gallica. La thématique des sciences occultes, figurant dans la liste des documents les plus téléchargés, a fait l'objet d'un des premiers corpus. D'autres thématiques se sont dégagées au regard des usages constatés sur le Web : les blogs culinaires utilisant des recettes anciennes de manuels de cuisine trouvés dans Gallica ont inspiré la collection « Cuisine et gastronomie » ; les discussions sur des forums spécialisés autour des documents relatifs à la guerre de 14-18 ont inspiré la collection « Témoins de la Grande Guerre » ; la présence active sur le Web de communautés dites *Steampunk* a suscité la sélection de « Romans d'anticipation ». Enfin, nous avons imaginé d'autres collections susceptibles de rassembler un large public : les contes d'ici et d'ailleurs, les romans populaires.
- 12 Pour cela, nous avons privilégié une communication sur les réseaux sociaux (blog, Facebook, Twitter), notamment à travers la médiation numérique de Gallica. Nous pouvons ainsi atteindre directement des communautés actives sur le Web, dont les pratiques participatives sont préexistantes, ou qui interagissent déjà avec Gallica sur les réseaux sociaux. Nous réalisons également une médiation continue en publiant des billets de blog autour de chaque corpus sélectionné, et en s'appuyant sur la propagation virale permise par les réseaux sociaux, Twitter en particulier. Ainsi lors de la mise en ligne de documents sur les « Témoins de la Grande Guerre », nous avons pu noter une augmentation de plus de 20 % des inscrits en quelques jours (70 personnes supplémentaires).
- 13 Par ailleurs, dans la conception même de la plate-forme, nous avons souhaité permettre à tous types de contributeurs de participer. Alors que les usagers doivent s'identifier pour bénéficier des fonctionnalités du réseau social, nous avons souhaité permettre également des contributions anonymes, et accueillir autant des personnes intervenant ponctuellement que des utilisateurs assidus. Enfin, le projet de recherche porte une forte exigence d'accessibilité numérique pour permettre aux utilisateurs en situation de handicap d'avoir accès aux documents. En effet, ces publics, mal- ou non-voyants, peuvent être les premiers bénéficiaires des documents corrigés (texte lisible par les synthèses vocales, réédition adaptée aux personnes non-voyantes). Nous avons réalisé

une interface vocalisée pour permettre aux personnes déficientes visuelles de participer à la correction collaborative en signalant les erreurs rencontrées. Ce prototype a fait l'objet d'une expérimentation auprès de 24 utilisateurs non- et mal-voyants. Les résultats nous ont confirmé la faisabilité et l'intérêt d'un accès à la correction collaborative de ces utilisateurs.

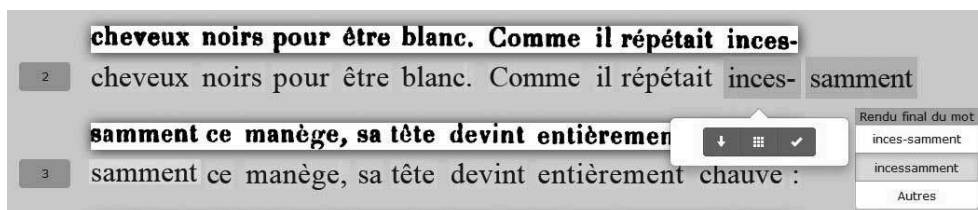
Motiver : comment faire rester les contributeurs ?

- 14 Une fois que les contributeurs se sont rendus sur Correct, il s'agit d'emporter leur adhésion et de les amener à revenir régulièrement, en leur fournissant une ergonomie de correction agréable, et en permettant l'échange et la collaboration entre eux.

Créer une interface ergonomique construite avec les usagers

- 15 L'ergonomie d'un projet participatif est un élément fondamental de motivation des usagers, qui peut les amener à adhérer à la plate-forme et à contribuer régulièrement. Des ergonomes d'Orange et de Paris 8 ont travaillé sur les interfaces de l'éditeur de correction de Correct, dont chaque fonctionnalité a été pensée pour être la plus intuitive et la plus simple d'utilisation possible. Pour éprouver les choix d'usages et d'ergonomie, et pour tester les fonctionnalités, nous avons réalisé des expérimentations des prototypes développés.
- 16 La première expérimentation, en avril 2013, avait pour objectif l'observation de l'utilisation du premier prototype de l'éditeur de correction. Les 24 participants ont exprimé leur satisfaction sur la simplicité de l'interface de correction et sur l'intérêt du projet lui-même. Les différents modes de correction (ligne-à-ligne et vis-à-vis texte et image) ont été fortement appréciés car ils proposent des points de vue sur le document adaptés aux différentes étapes (correction, relecture, validation).

Schéma : Un exemple de fonctionnalité : la gestion des césures



Légende : le survol d'une césure signale les deux parties du mot pour vérifier qu'elle a bien été prise en compte, un clic permet de préciser s'il s'agit d'une césure simple ou d'un mot composé.

Source : Réseau Correct.

- 17 L'objectif de la deuxième expérimentation, en juin 2013, était de tester la collaboration, par l'intégration de l'interface de correction au sein d'un réseau social. L'alliance d'un éditeur de correction et d'un réseau social constitue un défi technique qui justifie à lui seul un projet de R & D : interfaçage des deux modules techniquement distincts et travail collaboratif en temps réel, où plusieurs utilisateurs peuvent corriger la même page du même document en même temps, ce qui nécessite ensuite une fusion des corrections (validation des corrections similaires, mise en doute des corrections différentes, processus de levée de doute).

- 18 Pendant 2 heures, nous avons simulé la mise en réseau au travers d'un jeu de rôle réunissant 10 personnes sur place et 22 à distance. Le réseau social a été apprécié pour sa convivialité, pour les interactions et pour la possibilité d'initier des projets. Cependant, les testeurs ont trouvé que la dimension collaborative n'était pas assez mise en valeur : absence de suivi des corrections et faible visibilité de l'avancement de son travail et de celui des autres.
- 19 Les résultats de ces deux expérimentations ont alimenté les adaptations que nous avons apportées aux interfaces. Une dernière expérimentation a enfin été réalisée dans les conditions réelles en ligne. Nous avons invité les usagers de Gallica et les autres communautés cibles à tester Correct pendant 6 mois (de novembre 2014 à juin 2015) et mis en place une étude d'usage pour mesurer cette expérience, évaluer les outils et interfaces, et recueillir les attentes des utilisateurs.
- 20 Enfin, si la dynamique d'innovation de partenaires aux enjeux divers et aux compétences variées a permis de donner vie à des hypothèses d'usage qui auraient été impossibles avec nos moyens propres, en revanche, nous n'avons pas toujours eu la main sur les choix ergonomiques, et n'avons pas toujours pu faire jouer les enjeux de motivation identifiés (*gamification*, microtâches, sélection des documents par des présentations variées, etc.).

Susciter l'émulation par la collaboration

- 21 L'interface de correction s'appuie donc sur un réseau social dédié, afin de créer du lien entre les contributeurs, et donc de renforcer leur sentiment d'appropriation de la plate-forme.
- 22 Cependant, le choix de fonder ces relations entre usagers sur la collaboration et non sur la compétition, sur une entraide vertueuse, correspond aux usages des contributeurs identifiés (généalogistes, personnes non-voyantes, étudiants ou chercheurs). Nous souhaitons que les usagers s'engagent sur une production réellement participative, qu'ils puissent corriger ensemble les mêmes documents, et non qu'ils avancent parallèlement les uns aux autres.
- 23 Le réseau social permet ainsi de créer des groupes de correction, où les contributeurs se rassemblent autour d'un projet commun, et invitent d'autres correcteurs à les rejoindre. Ces groupes mobilisent des forces autour de la correction de documents, et l'on a plus de chances de voir ces documents entièrement corrigés plutôt que d'avoir un éclatement des efforts de correction sur une multitude de documents différents dont aucun ne serait jamais terminé. Des fonctionnalités de discussion (messages, commentaires, *chat*) permettent aux membres d'un groupe d'échanger entre eux, de poser en temps réel des questions en cas de problème, ou de se répartir les différentes pages d'un ouvrage, bref d'organiser plus efficacement leur activité de correction.
- 24 Par ailleurs, les usagers sont davantage incités à contribuer s'ils constatent que leurs efforts sont utiles. Les indicateurs mis en place sur chaque document et sur chaque page (nombre de contributeurs, de pages corrigées et validées, de corrections) permettent aux correcteurs de voir qu'ils ne sont pas seuls à contribuer, et de constater l'état d'avancement des travaux.

Passer le relais : comment faire en sorte que les contributeurs s'emparent de la plate-forme ?

- 25 L'ampleur du projet nécessite d'aller plus loin dans l'autonomisation des contributeurs. Correct se veut au croisement du participatif et du collaboratif, ce qui constitue une véritable innovation en termes d'usages participatifs pour une institution culturelle. Au-delà du défi technique, c'est un pari théorique : le collaboratif comme garant de la dynamique et de la qualité de la participation ; mais pour que le réseau fonctionne, il faudra que l'animation et la mobilisation viennent des utilisateurs eux-mêmes et que l'institution passe le relais.
- 26 La BnF propose ainsi un cadre accompagnant la mise en place de projets collaboratifs mais va s'effacer devant les contributeurs qui se réunissent par affinité ou autour d'un projet commun. Nous souhaitons que les usagers s'emparent de la plate-forme, et qu'au-delà du rôle de correcteur, ils endossent d'autres rôles qui recouvrent toutes les facettes de la contribution : animateur, formateur, ambassadeur. Pour simplifier la prise de possession de la plate-forme par les usagers, il n'y a pas de hiérarchie entre les utilisateurs, tout contributeur est un correcteur et peut devenir animateur d'un groupe dès qu'il le désire. Les usagers peuvent progressivement acquérir des expertises de correction ou d'animation, ils deviennent des formateurs, et contribuent à assurer la qualité des contributions, le bon fonctionnement du réseau, l'émergence de nouveaux experts. Enfin, ils se positionnent comme « ambassadeurs » de la plate-forme auprès de nouveaux utilisateurs, ils peuvent développer le recrutement et donc la fréquentation, en touchant les communautés de l'intérieur.

Premiers résultats

- 27 Quatre mois après le lancement de Correct, en mars 2015, les premiers résultats enregistrés ont conforté notre démarche de recrutement et de motivation des contributeurs : 531 correcteurs inscrits, 62 documents mis à disposition, 1 318 049 mots corrigés sur 5 753 pages. Certains documents ont de très bonnes audiences (49 correcteurs pour un même document), d'autres ont été entièrement corrigés. Enfin, on a pu identifier 37 utilisateurs très actifs et, parmi eux, 10 personnes réalisent 50 % du total des corrections.
- 28 L'enquête réalisée en ligne auprès des utilisateurs de Correct trois mois après son ouverture a confirmé certaines intuitions sur la motivation des contributeurs : 75 % affirment vouloir participer à l'amélioration de Gallica ; 75 % d'entre eux ont choisi les documents à corriger en fonction de leur sujet et 38 % ont préféré des ouvrages où peu de correcteurs étaient déjà intervenus.
- 29 Même s'ils restent encore rares, nous avons constaté plusieurs exemples d'appropriation de la plate-forme par ses usagers. Ainsi un animateur est allé recruter des correcteurs au sein du réseau social Lecteur.com dont il était un membre actif. Ce groupe, composé de 13 personnes, s'entraide et s'organise grâce aux outils du réseau social.

NOTES

1. < <http://www.reseau-correct.fr/> >.
2. Orange, BnF, Jamespot, Urbilog, i2S, ISEP, INSA Lyon, Universités Lyon 1 et Paris 8.
3. Reconnaissance optique de caractères (Optical Character Recognition) : système informatique permettant la génération d'un fichier texte à partir d'images numérisées.
4. < <http://gallica.bnf.fr/> >.
5. Pauline Moirez, Jean Philippe Moreux, Isabelle Josse, *État de l'art en matière de crowdsourcing dans les bibliothèques numériques*, février 2013. [En ligne] : < http://www.bnf.fr/documents/crowdsourcing_rapport.pdf >.
6. Pauline Moirez, « Archives participatives », in Muriel Amar, Véronique Mesguich (dir.), *Bibliothèques 2.0 à l'heure des médias sociaux*, Paris, Éditions du Cercle de la Librairie, 2012 (coll. Bibliothèques), pp.187-197. [En ligne] : < http://archivesic.ccsd.cnrs.fr/sic_00725420 >.
7. Il y avait une question sur les évolutions attendues de l'interface de Gallica ; parmi les réponses proposées, « Échanger autour des documents de Gallica », « Annoter les documents de Gallica », etc., ont reçu plus de 30 % d'avis positifs.