



HAL
open science

“Faire appel à la charité” : Trove, les journaux australiens et la foule des internautes

Marie-Louise Ayres, Mathieu Andro

► To cite this version:

Marie-Louise Ayres, Mathieu Andro. “Faire appel à la charité” : Trove, les journaux australiens et la foule des internautes. *IFLA Journal*, 2013, 10 p. hal-01094598

HAL Id: hal-01094598

<https://hal.science/hal-01094598>

Submitted on 17 Dec 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

“Faire appel à la charité” : Trove, les journaux australiens et la foule des internautes

French translation of the Original Paper: “‘Singing for their supper’: Trove, Australian newspapers, and the crowd”.

Translated by: Mathieu Andro, Institut National de la Recherche Agronomique, Versailles, France.

Marie-Louise Ayres

Resource Sharing Division, National Library of Australia, Canberra, Australia

E-mail address: mayres@nla.gov.au



This is a French translation of “‘Singing for their supper’: Trove, Australian newspapers, and the crowd.” Copyright © 2013 by **Marie-Louise Ayres**. This work is made available under the terms of the Creative Commons Attribution 3.0 Unported License:

<http://creativecommons.org/licenses/by/3.0/>

Abstract:

S'appuyant sur 20 ans d'expérience de coopération nationale dans le cadre du Australian Newspapers Plan (ANPlan), la Bibliothèque nationale d'Australie communique des journaux numérisés au public australien depuis 2007, et a implémenté cette diffusion sur le site Trove (www.trove.nla.gov.au) en 2009. Trove a remporté un nombre important de récompenses pour le caractère participatif de son fonctionnement vis-à-vis de ses usagers et en particulier pour la possibilité qui leur est offerte de corriger le texte ocrisé, c'est-à-dire, imparfaitement généré par un logiciel de reconnaissance optique de caractères (OCR). En juillet 2013, Trove a éclipsé tous les autres services en ligne de la Bibliothèque, et la partie de Trove consacrée aux journaux domine largement l'ensemble du projet, les australiens ayant corrigé plus de 100 millions de lignes de textes - soit l'équivalent de 270 années de travail provenant du crowdsourcing. A l'issue d'une expérience de cinq ans, cette forme de participation des usagers est devenue une offre de service mature de la Bibliothèque, et l'évaluation de ses impacts et de son avenir est, à présent, possible et opportune. Cet article résume ce que nous savons sur les motivations de ceux qui s'engagent bénévolement dans cette activité, leurs modes de collaboration, la profondeur avec laquelle ce mode d'engagement pénètre dans le fonctionnement de la Bibliothèque, et sur ce que les statistiques indiquent concernant le point où nous en sommes dans la courbe de croissance de la participation des usagers. L'article examine aussi l'impact du succès rencontré par le crowdsourcing sur la justification que la Bibliothèque présente elle-même pour offrir ce type de services (s'agit-il de chercher à traiter une tâche gigantesque ou plutôt de construire et de maintenir une communauté élargie des amis de la bibliothèque ?), les opportunités et les risques liés au fait d'avoir considérablement augmenté le nombre d'australien passionnément engagés au service de la Bibliothèque, et la manière dont cette expérience positive est en train de changer la manière de penser de la bibliothèque à propos de ses services et de la participation du public. Pour ceux qui envisagent de développer un service de numérisation de journaux, l'article donne un point de vue pragmatique et de terrain sur les choix fondamentaux et les questions que les bibliothèques doivent se poser.

Keywords: Journaux, numérisation, crowdsourcing, bénévolat, communauté.

1 INTRODUCTION

Le programme Trove (<http://trove.nla.gov.au>) de la Bibliothèque nationale d'Australie a été reconnu comme leader international pour la facilitation de l'accès du public au patrimoine documentaire. Bien que Trove donne accès à bien d'autres documents que des journaux numérisésⁱ, c'est ce composant du service qui a le plus capté l'attention du public, et qui représente plus des trois quarts des visites sur le site. De même, alors que Trove offre une gamme variée de fonctionnalités d'engagement de l'utilisateur, et que l'utilisation de chacune ne cesse de croître, ce sont les fonctionnalités de correction des textes de journaux de Trove qui ont attiré le plus haut niveau d'engagement de l'utilisateur. Le public australien et international y corrige les journaux depuis plus de 5 ans. Le personnel de la Bibliothèque et des chercheurs universitaires sont allés au-delà de l'émerveillement au sujet de la popularité du service pour s'interroger plus profondément sur ce qui est en train de se passer dans le domaine de la correction participative de textes. De récentes recherches ont confirmé plusieurs de nos «intuitions» sur ce que signifie cette popularité, invalidé d'autres prévisions, et nous a fait réfléchir sur l'avenir de la correction de texte en particulier, et où nous devrions nous diriger plus généralement par rapport à l'engagement de nos utilisateurs en ligne.

2 CONTEXTE

Le service des journaux australiens numérisé a été lancé, en tant qu'entité, en 2007 et intégré dans Trove en 2009. Néanmoins, des précédents au projet ont existé beaucoup plus tôt. L'Australie a une longue histoire de coopération nationale en matière de bibliothèque. Un exemple typique est le Plan de journaux australiens (ANPLAN), qui a été créé en 1992ⁱⁱ. Le plan a mis en place une structure de coopération et de collaboration pour la collecte, la gestion, la conservation et l'accès aux journaux australiens, et ouvert la voie à une approche nationale pour la numérisation des journaux. En 2005, la Bibliothèque a été associée à une demande de financement de l'Australian Research Council afin de numériser les journaux australiens en soutien à la recherche australienne en sciences humaines et socialesⁱⁱⁱ. Ce projet - qui comprenait une réflexion précoce sur le potentiel de chercheurs universitaires à ajouter de la valeur au contenu numérisé - a échoué, mais il s'est avéré utile lorsque la Bibliothèque, peu après, décida de commencer un projet pilote de numérisation des journaux, et de se concentrer sur le public le plus large possible en termes de participation des utilisateurs.

Ces premières réflexions sur l'engagement des utilisateurs eux-mêmes ne se posent pas dans le vide. La Bibliothèque a été une pionnière en matière de numérisation de collections et dans le développement de services nationaux de découverte du patrimoine documentaire, avec Picture Australia, Music Australia, le Register of Australian Archives and Manuscripts, Australia Dancing, Australian Research Online et la PANDORA web archive, tous désormais intégrés dans Trove. Ces services - développés entre 1997 et 2008 – ont tous précédé le Web 2.0 et les fonctionnalités des réseaux sociaux que nous considérons aujourd'hui pour acquises. Toutefois, certains avaient déjà intégré des options pour la participation des utilisateurs. Ainsi, Picture Australia avait établi une relation avec Flickr dès 2006, permettant aux membres du public d'ajouter leurs propres photographies de la vie australienne au sein d'une collection déterminée de Flickr qui a ensuite été moissonnée par Picture Australia. Dans la même année, Australia Dancing proposait un nouveau service «Participez» qui invitait les utilisateurs enregistrés à faire profiter le projet de leurs connaissances. Le projet a été un grand succès (plus de 180 000 images ont été transférées sur Trove ou sur les plateformes qui l'ont précédées et sont désormais toutes accessibles via Trove). Alors que la Bibliothèque avait de grands espoirs pour ce dernier projet - estimant qu'un tel service serait repris en charge par la

petite mais très dévouée communauté des chercheurs sur la danse - la promesse de «Participez» n'a finalement pas été réalisée.

Le service des journaux australiens a donc été lancé dans un environnement dans lequel la Bibliothèque a longtemps réfléchi sur les façons de collaborer avec les utilisateurs en ligne, après avoir expérimenté plusieurs modes de fonctionnement, et avoir connu quelques succès mais aussi quelques échecs. La fonction de correction de journal intégrée dans le service de journaux numérisés reflète cette volonté de collaborer avec nos utilisateurs, tout en reconnaissant qu'une approche de type crowdsourcing afin de corriger le texte imparfait des journaux numérisés offrait un potentiel susceptible d'améliorer les possibilités de recherches dans Trove et que le travail humain nécessaire à cette correction manuelle ne pourrait jamais être réalisé sur la base des seules ressources financières de la Bibliothèque.

3 L'ENGAGEMENT DE L'UTILISATEUR DE TROVE

Cinq ans après l'ouverture au public de Trove, son succès est une source de fierté et de joie pour la Bibliothèque nationale, et les statistiques de consultation sont impressionnantes :

- Trove rassemble plus de 62% des pages vues sur l'ensemble des sites de la Bibliothèque nationale d'Australie, éclipsant même l'utilisation du catalogue de la Bibliothèque, son site web, et le service des bibliothèques d'Australie utilisé par plus de 1000 bibliothèques ;
- environ 75% des visiteurs de Trove arrivent sur le site après avoir trouvé une ressource dans Google, ce qui souligne l'importance d'exposer et de référencer le contenu de Trove aux grands moteurs de recherche ;
- une moyenne de plus de 60 000 visiteurs uniques par jour consultent Trove, avec des pointes allant jusqu'à 80 000 visites sur un seul jour ;
- le nombre total de visites a doublé au cours des deux dernières années, avec 1,8 millions de visites en juin 2013.

Cela signifie que en seulement quelques années, Trove est devenu le moyen le plus courant par lequel les australiens (et le monde) rencontrent la Bibliothèque nationale d'Australie, et que la grande majorité de ces visiteurs arrivent sur Trove via un moteur de recherche, plutôt qu'en commençant leur navigation directement sur Trove, via le site de la Bibliothèque nationale d'Australie.

Beaucoup des fonctionnalités de participation des utilisateurs de Trove sont très populaires. Plus de 100 000 utilisateurs se sont enregistrés à ce jour, plus de 2 millions de tags et près de 60 000 commentaires ont été ajoutés aux ressources de Trove. La fonction participative qui connaît la croissance la plus rapide est celle de listes de Trove, qui permettent aux utilisateurs de constituer leurs propres collections au sein de Trove ou d'autres ressources web et qui peuvent être rendues publiques ou privées : près de 40 000 de ces listes ont été créées sur une grande variété de sujets.

La correction participative de textes se tient largement en tête de toutes les autres fonctions permettant l'engagement de l'utilisateur. En Juillet 2013, plus de 100 millions de lignes de texte du journal avaient été corrigées par des membres du public. Nous avons estimé que cela équivalait à plus de 425 000 heures de bénévolat, soit 270 années de travail d'un australien moyen. Chiffré au plus faible prix de rémunération pour la Bibliothèque, cela signifie que les correcteurs de texte bénévoles ont versé plus de 17 millions de dollars australiens [moins de

12 millions d'Euros] au service^{iv}. Ces chiffres sont exaltants et il est intéressant de noter que malgré le succès de l'Australie dans ce domaine, peu de services de journaux numérisés offrent encore aux utilisateurs la possibilité d'améliorer la qualité du texte OCRisé des journaux au profit de tous^v.

Des statistiques comme celles-ci soulèvent inévitablement de nombreuses questions, et dans la dernière année, les membres du personnel de la bibliothèque et des chercheurs indépendants ont cherché à en connaître l'origine afin d'essayer de mieux comprendre les utilisateurs, et en particulier ceux qui participent à l'activité de correction de texte. Cette recherche a été menée dans trois grands domaines. Le personnel de la Bibliothèque a utilisé Google analytics et les logs de Trove pour essayer de mieux comprendre nos utilisateurs et leurs habitudes participatives. La Bibliothèque a également commandé une évaluation indépendante de la satisfaction des utilisateurs de Trove, qui révèle plus d'informations sur la démographie des utilisateurs, ce qu'ils utilisent dans le service, ce qu'ils apprécient dans Trove, ce qu'ils aiment et ce qu'ils aimeraient voir changer^{vi}. Pendant ce temps, les chercheurs universitaires étudient ce qui motive les correcteurs bénévoles de textes, et cherchent à démêler les diverses «significations» qui se cachent derrière l'engagement crowdsourcing avec l'aide de l'expérience de Trove.

4 LES UTILISATEURS ENGAGES : QUI SONT-ILS ?

Alors que de nombreux sites de réseaux sociaux recueillent de grandes quantités d'informations sur leurs utilisateurs, Trove ne le fait pas. L'inscription pour les utilisateurs de Trove est facultative, et seule une minorité d'utilisateurs sont identifiés. Même lorsque les usagers s'enregistrent, nous recueillons un minimum d'informations - et aucune donnée sociologique - sur eux. Bien que nous puissions suivre le comportement participatif sur le site de l'utilisateur inscrit, nous ne construisons pas de «profils» des utilisateurs individuels. Nous ne connaissons pas leurs noms, âges, sexes, états matrimoniaux, statuts d'emploi, niveaux d'éducation, tranches de revenu, adresses, etc. Nous ne suivons pas et ne déduisons pas ce qui intéresse les utilisateurs individuels (soit par leur utilisation de Trove, soit par l'analyse des sites d'où ils proviennent ou iront immédiatement après leur passage sur Trove), et dans ces conditions, ne leur proposons aucune suggestion de site, ni n'offrons de services du type « les autres utilisateurs comme vous ont trouvé cela utile». La question de savoir si - dans le modèle de l'information poussée vers l'internaute plutôt que dans celui de l'internaute qui va rechercher celle qui lui convient - nous devrions envisager de le faire, est une question pour un autre jour, et qui soulève de nombreux dilemmes éthiques et professionnelles.

Cela ne veut pas dire que nous n'ayons aucune information sur notre base d'utilisateurs. Le faible pourcentage d'utilisateurs qui utilisent notre service d'aide et de contact^{vii} - et pourraient donc être suffisamment engagés pour solliciter une assistance personnelle - peuvent choisir de nous donner quelques informations sur eux-mêmes : ils peuvent choisir parmi une liste de types d'utilisateurs, nous communiquer leur code postal et leur pays. Une récente évaluation de la satisfaction des utilisateurs Trove nous a donné une plus claire image de notre base d'utilisateurs. En rassemblant toutes ces sources, nous avons été en mesure de conclure que :

- environ 40% de toutes les visites Trove sont des utilisateurs internationaux (y compris quand les moteurs et les robots sont exclus des statistiques). La plupart de ces utilisateurs sont issus du Royaume-Uni, des Etats-Unis et de la Nouvelle-Zélande. Une proportion faible d'utilisateurs internationaux proviennent de pays asiatiques ;

- Trove a atteint une portée véritablement nationale, la proportion d'utilisateurs de chaque état et territoire australien, et la proportion d'utilisateurs dans les régions métropolitaines, régionales et rurales d'Australie sont fortement en adéquation avec la distribution réelle de la population ;
- 70% des utilisateurs de Trove sont des femmes ;
- 65% des utilisateurs sont âgés de 50 ans ou plus, 34% sont âgés de 60 ans ou plus, seulement 17% des utilisateurs ont moins de 40 ans ;
- 60% des utilisateurs sont des employés, un quart sont à la retraite, en adéquation forte avec la participation de la main-d'œuvre australienne ;
- près de la moitié des utilisateurs de Trove gagnent plus de 40 000 dollars australiens [environ 27 000 euros] par an alors qu'ils ne sont que 28% de la population générale, 34% gagnent plus que 60 000 dollars australiens [environ 41 000 euros] ;
- seulement 1% des personnes ayant répondu à l'enquête sont des autochtones aborigènes (alors qu'ils représentent 2% de la population), et 4% seulement parlent des langues autres que l'anglais à la maison (alors que c'est le cas pour 15% de la population générale) ;
- 45% des utilisateurs de Trove sont titulaires ou poursuivent des études en vue de l'obtention d'un diplôme de troisième cycle (contre 2-3% de la population générale), 85% ont eu un diplôme de l'enseignement supérieur ;
- près de la moitié des utilisateurs de Trove considèrent l'histoire de la famille, la généalogie ou l'histoire locale comme leur principale raison d'utiliser Trove ;
- 40% des utilisateurs de Trove déclarent utiliser le service au moins chaque semaine, 19% prétendent utiliser le site quotidiennement ;
- les utilisateurs de Trove qui corrigent texte sont plus susceptibles d'être des historiens de la famille ou des généalogistes, à la retraite et des utilisateurs à long terme (plus de 1 an) de Trove.

En résumé, cela signifie que l'utilisateur «typique» de Trove est un diplômé, bien payé, parlant anglais, une femme employée âgée de cinquante ans ou plus, intéressée par la généalogie ou l'histoire locale, et qui visite le site de Trove très fréquemment. Les utilisateurs de la partie journaux de Trove sont plus âgés que la moyenne des utilisateurs de Trove, seulement 13% des utilisateurs de journaux ont moins de 40 ans. Les correcteurs sont également plus âgés, et sont, des utilisateurs fréquents et à long terme du service.

Les enquêtes récemment menées par Frederick Zarndt, Brian Geiger et Alyssa Pacey ont constaté que les utilisateurs de la Cambridge Public Library et des services journaux de la Californian Digital Library partageaient un certain nombre de caractéristiques communes^{viii}. La majorité des utilisateurs se sont identifiés comme des généalogistes, et une grande majorité utilise le site à des fins d'histoire familiale (85% pour Cambridge, 61% pour la Californie). Environ 40% des utilisateurs ont affirmé qu'ils visitent les sites au moins chaque semaine, pour une durée de 60-70 minutes. En moyenne, 77% de tous les utilisateurs de ces services sont âgés de 50 ans et plus, et 45% sont âgés de 60 ans et plus.

Les utilisateurs «typiques» de ces trois services semblent remarquablement similaires (bien que le contenu et les services de Trove soient plus larges que les exemples cités). Bien sûr, les « types » ou les « moyennes » masquent des variations significatives dans les caractéristiques des utilisateurs, les raisons de l'utilisation du service, quelles sont les parties du service qui sont les plus utilisées, si l'utilisateur se sent «connecté» ou non, et ce que l'utilisateur aimerait voir modifier ou développer à l'avenir. Alors qu'au moins la moitié des utilisateurs de Trove utilisent le service pour l'histoire de la famille, 15% sont de premier cycle ou étudiants de

troisième cycle ou des personnes faisant une recherche universitaire à la faculté ; ces utilisateurs ne se sentent pas «connecté» à la communauté Trove, mais décrivent souvent Trove comme permettant de «révolutionner» leur recherche. Une proportion significative d'utilisateurs de Trove sont en fait des bibliothécaires qui font des recherches pour aider leurs usagers. Le défi de la bibliothèque pour l'avenir est d'essayer d'atteindre autant de points d'équilibre que possible, du contenu et des services qui font appel à un large éventail de types d'utilisations et utilisateurs.

Cependant, si nous nous concentrons sur la communauté autour de la correction de texte, les recherches australiennes et internationales semblent claires : la plupart des utilisateurs de journaux numérisés sont des utilisateurs plus âgés dont l'objectif principal est l'histoire de la famille ou l'histoire locale. Les fonctions de correction de textes de Trove font appel aux plus âgés, et souvent aux retraités, public principalement intéressé par l'histoire familiale et locale.

Alors que la Bibliothèque nationale et d'autres fournisseurs de journaux numérisés peuvent se sentir fiers du grand bond en avant leurs programmes représentent, et la Bibliothèque est certainement ravie d'avoir touché de si nombreux australiens dispersés si largement à travers le continent, il ya quelques questions gênantes pour la bibliothèque, et des questions que d'autres bibliothèques qui envisagent un tel programme pourraient également se poser.

Les utilisateurs inévitablement sont âgés. Est-il possible que l'on assiste actuellement à une «vague» d'intérêt pour la généalogie qui va décroître avec l'arrivée des prochaines générations ? Ou est-il possible que les recherches généalogiques continueront à intéresser les plus de 50 ans lorsque certaines populations auront atteint cet âge, et qu'ils deviendront, alors, des utilisateurs plus avides du service de journaux numérisés ? Y a-t-il des scénarii que nous pouvons prévoir dans lequel il y aura des augmentations substantielles dans l'utilisation de ces services de journaux numérisés par ceux qui sont actuellement à l'école ou à l'université, ou encore ceux qui se trouvent en début de vie ou à mi-vie ? Devons-nous nous contenter de faire appel principalement à une population âgée, diplômée, et aisée ? Devons-nous accepter que les types de contenus que les bibliothèques fournissent actuellement, et que les services que nous mettons à disposition en ligne soient plus attrayants pour ce groupe démographique - ou devons-nous réfléchir à des types de contenu, et à des modèles de participation des utilisateurs qui permettraient d'étendre notre audience et d'atteindre de nouveaux publics, afin d'être socialement d'avantage au service de l'ensemble de la population ?

5 CE QUI MOTIVE LES UTILISATEURS

Une des façons de réfléchir à cette question est de considérer ce qui motive nos utilisateurs actuels ; dans ce contexte, ce qui motive nos correcteurs de texte. Sultana Lubna Alam et John Campbell, deux universitaires de l'Université de Canberra, se sont penchés sur la communauté de correcteurs participatifs de Trove dans le cadre d'une sur les motivations autour du crowdsourcing menée en 2011 et 2012^{ix}. Ils ont élaboré un modèle de motivations pour la correction de textes (basé sur un certain nombre de types de motivations dans plusieurs domaines) englobant des motivations intrinsèques et extrinsèques. Les motivations intrinsèques ont été classées comme fondées sur l'égoïsme (par exemple, les intérêts de recherche personnelle), le caractère communautaire (par exemple, l'altruisme et les aspects communautaires ou collectifs), et la jouissance (par exemple le coté distrayant, amusant, ou agréable). Les motivations extrinsèques ont été principalement identifiées comme des motivations sociales (par exemple, la reconnaissance, les récompenses, l'attribution et de la propriété). Les correcteurs de textes de Trove ont été analysés comme principalement

intrinsèquement motivés. La reconnaissance et les récompenses - présentes dans la communauté de Trove mais relativement modérées - ne sont pas les principales motivations. Au lieu de cela, les correcteurs de textes de Trove sont principalement motivés par leurs intérêts personnels de recherche, par le sentiment d'être impliqués dans quelque chose de «plus grand qu'eux» et «de valeur durable», et par un très fort sentiment redevabilité ou de charité. Ils jouissent de leur autonomie, de trouver la tâche de correction reposante et agréable, ont un fort sentiment de confiance à la fois dans et par la Bibliothèque, et le sentiment que leur travail est apprécié. Alam et Campbell notent que les études sur la durée des travailleurs volontaires du crowdsourcing fournissent des indications sur le dynamisme de leur motivation. Ils ont depuis mené d'autres recherches sur le management des communautés crowdsourcing, et les motivations des institutions - telles que la Bibliothèque nationale d'Australie - qui s'investissent dans ces formes d'engagement^x.

6 LES MODES DE PARTICIPATION DES UTILISATEURS

Paul Hagon, un graphiste web senior à la Bibliothèque nationale d'Australie, a utilisé les logs de transactions du site web de Trove afin d'étudier les caractéristiques de l'engagement des utilisateurs^{xi}. Hagon a utilisé Google Analytics pour montrer que Trove a un grand nombre de visiteurs réguliers, et que la durée moyenne d'une visite est de près de neuf minutes tandis qu'elle ne que de trois minutes pour le catalogue et d'une minute pour le site Web de la Bibliothèque. Les visiteurs australiens restent encore plus longtemps que les autres sur le site, avec une moyenne de douze à quatorze minutes et la visualisation d'une moyenne de quatorze pages par visite. Le taux de retour sur Trove et de la durée moyenne des visites sont très élevés par rapport aux standards du Web, ce qui démontre que le contenu de Trove est très attrayant pour les visiteurs australiens. Cela nous donne une impression très large de la participation des utilisateurs du service, mais ces informations sont moins utiles pour comprendre les modèles de dispositifs de participation des utilisateurs, en particulier pour ce qui concerne la correction de texte. Pour collecter ces informations, Hagon a utilisé les nombreuses données contenues dans les journaux de transactions de Trove.

Il a établi que les corrections de texte par les utilisateurs enregistrés représentent environ 85% du total des corrections, les 15% de corrections restantes étant l'œuvre d'utilisateurs non enregistrés, ou enregistrés mais non identifiés. Des modes de participation ne peuvent être établies que pour les utilisateurs enregistrés. La forte corrélation entre l'enregistrement des utilisateurs et la correction de textes semble indiquer que les correcteurs de texte sentent qu'ils «appartiennent» au service, qu'ils ont noué une relation avec la bibliothèque et son service. La forte corrélation signifie également que nous pouvons avoir un degré raisonnable de certitude que les activités de nos utilisateurs enregistrés agissent comme un bon indicateur de schémas de participation des utilisateurs à travers le tableau de bord.

La recherche de Hagon a établi qu'un très petit nombre de personnes - le top 100 - ont entrepris 43% de toutes les corrections. Cela signifie que les 100 meilleurs correcteurs ont, entre eux, corrigé plus de 41 millions de lignes de texte. Un grand groupe de personnes - le top 1000 - ont réalisé 81% de toutes les corrections, et 96% de toutes les corrections ont été faites par les 5000 meilleurs correcteurs inscrits.

50% de tous les correcteurs ont corrigé à moins de 100 lignes (selon l'estimation de la Bibliothèque d'environ 15 secondes par ligne corrigée, cela signifie que la moitié de tous les correcteurs ont dépensé moins d'une demi heure à la correction participative de textes) et 75% des utilisateurs ont corrigé moins de 500 lignes (autour de 2 heures de leur temps). Nos super-

correcteurs - ceux qui ont corrigé plus de 17 millions de lignes de texte - ne représentent que 0,01% de nos utilisateurs. Leurs contributions de plus de 4000 heures chacun les placent dans une «classe» spécifique de l'engagement. Par exemple, John Warren, top texte correcteur de Trove, a récemment été identifié et interviewé dans la Newsletter des Amis de la Bibliothèque nationale d'Australie^{xii}. M. Warren est un retraité qui passe désormais entre 6 et 8 heures par jour à corriger des textes, il considère cela comme son travail. Il s'est équipé d'un grand écran pour faciliter le zoom, et en même temps que M. Warren concentre son travail sur des éléments liés à l'histoire de sa famille et des communautés dans lesquelles ses ancêtres ont vécu, il en profite pour corriger systématiquement le texte intégral de chaque article qu'il consulte.

Les résultats de cette recherche ont été assez surprenants, et vont certainement provoquer une remise en cause de notre stratégie autour de l'engagement de l'utilisateur. Même si nous savions que certains correcteurs étaient beaucoup plus engagés que d'autres, nous n'avions pas une bonne compréhension du nombre relativement faible de personnes responsables de la majorité des corrections. Pour mettre cela en perspective, la Bibliothèque nationale d'Australie compte environ 80 bénévoles sur place qui travaillent soit mis en avant, face à notre public (surtout en tant que guides d'exposition) soit dans les coulisses des magasins. Les «super-correcteurs» ne sont donc pas beaucoup plus nombreux que le groupe de bénévoles sur place. De même, les 1000 personnes qui ont fait plus de 80% des corrections représentent à peu près la moitié du nombre des amis de la Bibliothèque nationale. Les 5000 ne sont pas très différents de 3 163 lecteurs du Magazine de la bibliothèque. D'une certaine façon, considérer des groupes de cette taille est beaucoup plus facile à manager que la prise en compte d'énormes populations ou chiffres tels que 100 millions de lignes de texte corrigées, ou 100 000 utilisateurs inscrits. Nous nous soucions de tous nos utilisateurs - peu importe leur fréquence et l'intensité de leurs contributions – mais si nous considérons un nombre plus «humain» d'utilisateurs intensément engagés, cela peut libérer notre réflexion et nous permettre d'envisager de nouveaux moyens par lesquels nous pouvons interagir avec eux.

7 QUEL AVENIR POUR LA CORRECTION DE TEXTES ?

Nous avons aussi récemment reconnu que bien que, parallèlement à l'augmentation du contenu sur Trove, le nombre de visites sur le site web de Trove et celui des corrections de textes aient continué de croître, le taux de croissance annuel de corrections au cours des deux dernières années n'avait pas été proportionnel au rythme de croissance des contenus. Cela nous donne à penser que nous avons peut-être atteint un point de masse critique, un point à partir duquel nous aurions peu de chances de recueillir de nouveaux nombreux correcteurs de texte motivés, et que le «marché» des volontaires australiens prêts à donner de leur temps précieux à cet effort – du moins sur la base de notre fonctionnalité actuelle de correction de texte - pouvait être saturé. Est-ce important ? S'il n'existait pas des solutions logicielles d'amélioration de la reconnaissance de caractères à l'horizon, cette information serait certainement d'importance, dans la mesure où une masse sans cesse croissante de contenus n'aurait pas été optimisée pour la recherche, et nous aurait certainement amenés à chercher à développer de nouvelles façons de collaborer avec un public plus large pour atteindre les mêmes résultats^{xiii}. Mais si nous sommes en mesure d'appliquer des solutions logicielles afin de résoudre le problème de l'imparfaite reconnaissance optique de caractères (et il semble probable que ce sera possible dans les cinq prochaines années), est-ce que ce sera grave si nous n'avons plus ce travail à proposer à nos bénévoles numériques ?

8 QUE SE PASSE-T-IL SI ON RENVERSE LA QUESTION ?

Je soutiens qu'une stabilisation de l'engagement des utilisateurs doit être considérée avec importance si vous êtes la Bibliothèque nationale d'un pays avec une petite population dispersée sur un très vaste territoire, et cherchant à atteindre et à engager des communautés distantes.

A l'origine, les raisons pour lesquelles la Bibliothèque avait offert la possibilité de correction participative du texte de journaux étaient que nous pourrions potentiellement exploiter le travail de la foule d'internautes afin d'améliorer la qualité de l'OCR et les résultats de la recherche sur les journaux numérisés, mais aussi notre volonté de collaborer avec les utilisateurs selon de nouveaux modes dans un environnement en ligne. En conséquence, nous avons aujourd'hui un petit nombre d'australien très fortement engagés dans le travail de la Bibliothèque nationale, un plus grand nombre de personnes fortement engagées, et un nombre très élevé qui sont suffisamment engagées pour contribuer en petites quantités sur leur temps libre, dans l'intérêt du bien public, à travers Trove. Beaucoup de ces personnes n'avaient aucun lien antérieur avec la Bibliothèque nationale, et notre récente enquête d'usagers nous suggère qu'elles avaient aussi des liens faibles avec leurs bibliothèques locales, du moins pour ce qui concerne les utilisateurs régionaux et ruraux. Ils représentent effectivement un nouveau public - bien qu'il soit très semblable à notre public traditionnel sur place d'un point de vue démographique et sociologique - et représentent une base de soutien plus large pour la Bibliothèque nationale, et pour toutes les bibliothèques et les organismes qui contribuent à Trove. Je pense que le maintien de l'engagement de nos utilisateurs actuels et la croissance de notre audience font maintenant partis des principaux objectifs de la Bibliothèque.

Si nous pouvions utiliser un logiciel permettant d'améliorer la qualité du texte de nos journaux, quelle valeur ajoutée les fonctionnalités participatives permettraient de faire en sorte que notre vaste et croissante audience de personnes âgées et des retraités australiens cherchant à se renseigner sur leur famille et sur les histoires de leur communauté puisse nous consacrer, ainsi qu'à leurs compatriotes australiens, un peu de leur temps ? La transcription de manuscrits numérisés, semblable au service Hive^{xiv} des Archives nationales d'Australie ? Plus de valeur ajoutée d'initiatives autour de groupes d'intérêts spéciaux sur des thèmes aussi divers que l'histoire ancienne du climat australien et le patrimoine insulaire de la mer du sud de l'Australie ? La possibilité d'ajouter des géo-tags à un grand nombre de ressources de manière ludique, facile et agréable ? La possibilité pour les australiens âgés de constituer leurs propres expositions en ligne afin que les histoires qu'ils proposent puissent être présentées de manière intéressantes et visuellement attrayantes ? Tous ces éléments sont esquissés à court, moyen et long terme à la Bibliothèque nationale, et un certain nombre sont déjà mises en œuvre par nos services frères, Europeana, la Digital Public Library of America, et Digital NZ.

Mais que dire des publics que nous n'atteignons pas - les jeunes, les moins riches, les moins diplômés, les australiens autochtones, et la forte proportion de la population australienne pour qui l'anglais n'est pas la langue principale parlée à la maison ? Étant donné que le public Trove actuel ressemble remarquablement à la profession actuelle de la bibliothèque australienne^{xv}, toucher ces utilisateurs potentiels sera encore plus difficile. C'est, cependant, une aspiration qui - en tant que profession chargée de gérer le patrimoine documentaire et de le rendre disponible pour tous ceux qui souhaitent apprendre et en profiter - nous devons nous efforcer d'atteindre.

ⁱ Trove donne accès à plus de 350 millions de ressources, dont un tiers sont librement accessibles en ligne, un tiers sont des ressources sous licences accessibles aux bibliothèques abonnées, et un tiers nécessite des démarches supplémentaires pour y accéder. Trove donne accès à des livres, des revues, des produits de la recherche, des photos, des manuscrits, des cartes, des enregistrements sonores, des images animées, des sites web archivés et des objets.

ⁱⁱ Pamela Gatenby, Directrice générale adjointe de la gestion des collections de la Bibliothèque nationale d'Australie a présenté un document sur l'histoire de ANPLAN au congrès 2008 de l'IFLA, qui s'était également tenu à Singapour. Son document est disponible sur le site Web de la Bibliothèque: <http://www.nla.gov.au/content/theaustralian-journal-plan-anplan>

ⁱⁱⁱ Peu avant le programme de numérisation JISC de journaux au Royaume-Uni, qui a été réalisé en partenariat public-privé.

^{iv} D'autres façons d'estimer la valeur de la correction de texte sont disponibles. Frederick Zarndt et Brian Geiger ont utilisé une hypothèse de coût de sous-traitance de 50 cents pour 1000 caractères corrigés et sont arrivés à un chiffre de 1,4 millions de dollars pour les 70 millions de lignes de textes corrigés sur Trove dans leur excellente étude à propos de la correction de texte de journaux, présentée en novembre 2012 au Digital Library Federation Forum et accessible sur : <http://www.diglib.org/forums/2012forum/no-tempest-in-my-teapot-analysis-of-crowdsourced-data-and-user-experiences-at-the-california-digital-newspaper-collection>

^v Comme cela a été récemment indiqué par Rose Holley, ancien directeur de Trove et qui travaille actuellement sur un projet similaire aux Archives nationales d'Australie. Voir: <http://roseholley.blogspot.fr/2013/04/crowdsourcing-text-correction-and.html>

^{vi} Dirigé par Gundabluey Pty Ltd en Mai et Juin 2013. La bibliothèque prévoit d'achever l'analyse complète des résultats d'ici la fin 2013. L'enquête a recueilli de nombreuses réponses qualitatives à partir de 28 utilisateurs, et les réponses quantitatives détaillées sur 1086 utilisateurs. La taille de l'échantillon était suffisante pour donner à la bibliothèque un niveau élevé de confiance dans le résultat.

^{vii} L'équipe de Trove répond à environ 3000 demandes par an. En outre, plus de 500 suggestions de numérisations de titres de journaux, ou des questions à propos de titres en cours de numérisation reviennent à l'équipe de numérisation des journaux.

^{viii} Ces résultats ont été présentés à la conférence de l'American Library Association en juin-juillet 2013. Les diapositives sont disponibles sur : <http://www.slideshare.net/cowboyMontana/what-motivates-library-crowdsourcing-volunteers-20130630-ala-lita>

^{ix} Alam, Sultana Lubna and Campbell, John, 'Crowdsourcing Motivations in a not-for-profit GLAM context: the Australian Newspapers Digitisation Program', présenté à la 23ème Australasian Conference on Information Systems, Geelong, 3-5 décembre 2012, and disponible sur : <http://trove.nla.gov.au/version/188319518>

^x Alam, Sultana Lubna and Campbell, John, 'Role of relational mechanisms in crowdsourcing governance: an interpretive analysis' (document qui sera présenté à la dix-neuvième Americas Conference on Information Systems in Chicago, August 15-17, 2013), et 'Dynamic changes in organizational motivations to crowdsourcing for GLAMs' (article proposé au programme de la trente-quatrième International Conference on Information Systems, qui se tiendra à Milan les 15-18 décembre 2013).

^{xi} La recherche menée par Paul Hagon est disponible sur le site : <http://www.nla.gov.au/our-publications/staff-papers/trove-crowdsourcing-behaviour>. Beaucoup d'autres documents de l'équipe de Trove sont également disponibles sur le site.

^{xii} Sylvia Marchant, 'John Warren, champion text corrector', Friends Newsletter, June 2013.

^{xiii} La Bibliothèque nationale de Finlande, par exemple, en partenariat avec Microtask a développé Digitalkoot, en intégrant une approche de gamification ou de ludification pour corriger le texte de journaux numérisés (<http://blog.microtask.com/2011/02/digitalkoot-crowdsourcing-finnish-cultural-heritage/>). Deux jeux, Mole Bridge et Mole Hunt, invitent les utilisateurs à jouer en ligne afin d'atteindre de bons niveaux de certitude sur des mots provenant de l'OCR. L'effort fait appel aux sentiments patriotiques («Commencez à sauvegarder ... la culture finlandaise ici»), mais semble susceptible d'avoir précisément fait appel à un public de moins de 50 ans qui est moins engagé avec Trove ou dans d'autres services de journaux numérisés. Une étude de cette approche tout à fait différente du problème d'OCR serait donc complémentaire de celles mentionnées dans le présent document.

^{xiv} La communauté Hive des Archives nationales d'Australie transcrit des documents gouvernementaux. Plus d'informations sur le service sont disponibles à l'adresse : <http://transcribe.naa.gov.au>

^{xv} Comme l'utilisateur «typique» de Trove, l'auteur de ce papier est de sexe féminin, âgé de cinquante ans, a fait des études postuniversitaires et jouit d'un revenu confortable... mais ce n'est pas une historienne de la famille.