



HAL
open science

**Transcription du Journal de Vincent (1757-1773) par
crowdsourcing Réflexions empruntées 1 sur un nouvel
[outil]/[écueil] du juriste historien**

Aurette Levasseur

► **To cite this version:**

Aurette Levasseur. Transcription du Journal de Vincent (1757-1773) par crowdsourcing Réflexions empruntées 1 sur un nouvel [outil]/[écueil] du juriste historien. 13 en droit, 2018. hal-03163867

HAL Id: hal-03163867

<https://hal.science/hal-03163867>

Submitted on 9 Mar 2021

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.

Transcription du *Journal de Vincent* (1757-1773) par crowdsourcing Réflexions empruntées¹ sur un nouvel [outil]/[écueil] du juriste historien

Aurelle Levasseur
CERAL

Qui rechignerait à retrouver la folie de ses vingt ans, même si ce n'est que le temps d'un article ? La revue *Treizendroit*, enfantée par certains des plus brillants et intrépides doctorants de l'UFR DSPS², invite les chercheurs établis à faire fi des pondérations académiques, et à produire un travail au goût de rock'n roll, que les aînés liraient avec des grimaces horrifiées. Cet article ne manquera pas d'en provoquer, puisqu'il part du postulat que n'importe qui peut participer à la recherche scientifique, par un processus appelé « crowdsourcing », dès lors que celui-ci est mené avec rigueur et méthode. La réflexion sera nourrie par la présentation d'un cas d'étude : une hypothétique transcription collaborative d'un manuscrit d'administration publique du XVIII^e siècle, le *Journal de Vincent*.

Pour comprendre comment un historien du droit peut être tenté par le crowdsourcing, il faut bien prendre la mesure de sa frustration. Au cours de sa carrière, il n'aura le temps d'extraire qu'un ou deux manuscrits du passé pour les offrir à sa communauté scientifique. Il s'agit en effet d'un travail titanesque qui implique de retranscrire la source, de l'analyser et d'en faire une édition savante. Un historien tournera donc le dos à quantité de trésors, qu'il laissera croupir dans l'un des 3800 kilomètres linéaires d'archives françaises, le cœur aussi gros que celui de l'archiviste qui désespère lui aussi de voir ses fonds exploités. Les plus éminents chercheurs ont beau appeler aussi fort qu'ils peuvent à relancer l'édition des sources historiques, rien n'y fait³ : le temps de l'universitaire se rétrécit toujours plus, en un mouvement contraire à celui de la toile. L'archiviste a trouvé dans le web une petite consolation, et ce sont désormais plus de 380 millions de pages numérisées issues des archives qui sont accessibles en ligne⁴. Le chercheur y a surtout gagné des idées *a priori* farfelues, comme d'utiliser l'outil numérique pour faire travailler à son profit et celui de la science ces amateurs d'histoire qu'il côtoie dans les centres d'archives. Ce processus qui consiste à inviter des inconnus non experts à participer à une recherche scientifique est appelé « crowdsourcing ». Le mot est devenu d'usage courant depuis un article du journaliste Jeff Howe paru en juin 2006 dans le magazine *Wired*. Il est lui-même issu d'un autre néologisme, « sourcing », né dans le vocabulaire managérial des années 1960 pour exprimer l'externalisation des tâches qu'une structure ne peut réaliser en interne faute de ressources⁵.

Comme souvent, ce mot tout neuf maquille une réalité ancienne. Le crowdsourcing existe depuis le développement des sciences au XIX^e siècle, peut-être même plus tôt encore : les

-
- 1 Les données et les réflexions présentées ici ont été très largement empruntées au groupe de travail Crowdsourcing du consortium CAHIER de Huma-Num, ainsi qu'aux intervenants au colloque international *Le crowdsourcing pour partager, enrichir et publier des sources patrimoniales* qui s'est tenu à l'Université d'Angers du 18 au 20 octobre 2017, organisé par Cécile Meynard, Florence Alibert, Elisabeth Greslou, Thomas Lebarbé et Valérie Neveu. Nous avons tenté d'attribuer au mieux la paternité des propos tenus mais l'entreprise n'a pas toujours été possible car certaines réflexions sont issues de discussions informelles et d'autres ont été énoncées par la presque totalité des participants.
 - 2 Faculté de droit, sociologie et sciences politiques de l'Université Paris 13.
 - 3 GENET J.-P., « Le politique (l'Etat, le pouvoir, les forces) », *Les historiens français à l'oeuvre (1995-2000)*, J.-F. SIRINELLI, P. CAUCHY, C. GAUVARD (dir.), Presses Universitaires de France, Paris, 2000, p. 161-184 (p. 164) ; HÉBERT M., « Consensus et représentation en Europe occidentale, XIII^e-XVII^e siècles », *Consensus et représentation*, J.-P. GENET, D. LE PAGE et O. MATTÉONI (dir.), Publications de la Sorbonne/École française de Rome, Paris/Rome, 2017, p. 11-39 (p. 35).
 - 4 *Rapport annuel 2015. Des archives en France* [en ligne], p. 5. Disponible sur : <<http://francearchives.fr/article/37979>> (consulté le 06.11.2017).
 - 5 DE CHAMPS E., « Construire une communauté au service du patrimoine manuscrit : les enjeux du crowdsourcing », communication à *Le crowdsourcing pour partager*, coll. cit. ; BURGER-HELMCHEN T., PÉNIN J., « Crowdsourcing : définition, enjeux, typologie », *Management et avenir*, n° 41, 2011, p. 254-269.

observatoires recueillait et traitait les données transmises par les astronomes amateurs. Les grands dictionnaires de langue allemande et anglaise, comme le *Deutsches Wörterbuch* des frères Grimm et surtout l'*Oxford English Dictionary*, furent rédigés grâce aux fiches de citations envoyées par des dépouilleurs volontaires⁶. Si le processus n'est donc pas neuf, il est en revanche métamorphosé par le numérique qui permet sa massification et son institutionnalisation. La perspective a de quoi révolter l'académisme français tant elle va à l'encontre de sa culture, qui s'est forgée en intégrant le principe d'une nette séparation des sociétés civile et scientifique, ce dont témoigne l'existence de deux ministères différents pour la recherche et la culture. Des éléments explicatifs de cette spécificité française (la frontière est généralement moins marquée dans les autres pays européens) peuvent être recherchés dans une culture catholique à tendance centralisatrice qui marginalise le mécénat, dans un élitisme fondé sur les diplômes avant que de l'être sur les compétences, peut-être aussi dans une méfiance traditionnelle envers le *crowd* dont le terme français renvoie à une « foule » irrationnelle et dangereuse. Quoi qu'il en soit, la frontière entre les deux mondes s'amincit aujourd'hui, fragilisée par de multiples facteurs qui ne sont pas que scientifiques ou techniques. Le crowdsourcing en effet se coule à merveille dans un mouvement idéologique et politique qui se veut libéral et neo-démocrate, du moins dans ses discours. L'Etat y est arbitre plutôt que maître, les relations horizontales plus que verticales, la démocratie y est participative et la société « apprenante », délégitimant le regard surplombant et dominateur de l'expert. Cette idéologie placée dans le sillage du crowdsourcing⁷ est une raison supplémentaire de défiance pour le monde académique, qui est témoin des effets désastreux qu'une application trop maladroite a produit dans l'enseignement primaire et secondaire. Toutefois, ce dogme a aussi facilité les premières expérimentations objectives de crowdsourcing, menées par les acteurs de la médiation culturelle, en diminuant les résistances institutionnelles qu'ils pouvaient jusque-là rencontrer⁸. Les centres d'archives et les bibliothèques ont ainsi pu devenir des pionniers du crowdsourcing français, qu'ils utilisent non seulement pour remplir leur mission de médiation culturelle mais aussi pour proposer de nouveaux instruments de recherche au monde académique⁹.

Cette dichotomie entre les deux objectifs possibles du crowdsourcing (médiation culturelle et recherche) est à placer au fondement des réflexions méthodologiques sur cet outil. Si l'objectif principal du crowdsourcing est la médiation, alors la science est un outil qui permet d'améliorer la foule. A l'inverse, si l'objectif principal est la recherche, la foule est l'outil qui permet de faire avancer la science. L'objectif principal n'est pas exclusif de l'objectif secondaire, qui est

6 DE CHAMPS E., « Construire une communauté », communication à *Le crowdsourcing pour partager*, coll. cit. ; PRATS LÓPEZ M., *Managing Citizen Science in the Humanities : The Challenge of ensuring quality*, thèse inédite de l'Université d'Amsterdam, 2017, p. 1.

7 Mathieu Andro, qui a soutenu une thèse de doctorat en sciences de l'information et de la communication sur le crowdsourcing en bibliothèques, a montré que trop souvent les données produites par les crowdsourcers n'étaient jamais réintégrées ni utilisées. Le crowdsourcing se réduit alors à un outil politique ou de marketing (ANDRO M., « Le crowdsourcing dans les bibliothèques », communication à *Le crowdsourcing pour partager*, coll. cit.). Nous n'entrerons pas ici dans les débats éthiques engendrés par le développement du crowdsourcing (travail déguisé, ubérisation de la recherche, etc.) dont il faut toutefois évoquer l'importance.

8 MOIREZ P., « Crowdsourcing, bibliothèques et patrimoine écrit : le projet expérimental Correct à la Bnf », communication à *Le crowdsourcing pour partager*, coll. cit.

9 Mentionnons les projets *Anne-Duportal* et *Jean-Louis Bagot* des Archives municipales de Saint-Brieuc (COUTENTIN Y., « Le participatif, un changement de posture », communication à *Le crowdsourcing pour partager*, coll. cit.) ; le projet Correct mené par la Bibliothèque nationale de France (MOIREZ P., « Construire et enrichir ensemble les savoirs : crowdsourcing et patrimoine numérisé », *Communs du savoir et bibliothèques*, Editions du Cercle de la Librairie, 2017, p. 129-139) ; la transcription collaborative des testaments de guerre des Poilus (Archives nationales et Ecole nationale des Chartes) ; le relevé collaboratif des registres de l'Assemblée et des tables du Conseil général des Ponts et Chaussées des Archives Nationales (DOOM V., RODRIGUEZ S., « Relevé collaboratif des registres de l'Assemblée et des tables du Conseil général des ponts et chaussées : à la découverte d'un trésor archivistique (18^e siècle-1965) », communication à *Le crowdsourcing pour partager*, coll. cit.). D'après Edouard Bouyé, « Le numérique au service de la République des lettres », communication à *Le crowdsourcing pour partager*, coll. cit., 30% des centres d'archives départementales emploient désormais des techniques de crowdsourcing.

généralement incident. Le crowdsourcing-médiation produit des données réutilisables par la recherche, tandis que le crowdsourcing-recherche implique de faire de la médiation, même si c'est *a minima*, pour inciter les volontaires à participer et pour les former. Même si la loi Fioraso a rappelé au chercheur que la médiation était partie intégrante de son métier, il n'en demeure pas moins qu'il sera davantage amené à faire du crowdsourcing-recherche, et donc à réfléchir sur les spécificités méthodologiques de celui-ci. Les premiers retours d'expériences en ce domaine ont dévoilé l'importance d'une méthode qui place au cœur de la réflexion non pas les spécificités du crowdsourcing (II), mais le *corpus* du chercheur – dans notre cas d'étude le *Journal de Vincent* (I).

I.- Présentation de la source et de son intérêt scientifique pour l'histoire du droit

Derrière la cote BIB Ms 66 des Archives départementales de l'Ain se trouve un manuscrit de 1048 pages, connu sous le titre de *Journal de Vincent*¹⁰. Ce manuscrit, en excellent état, est composé de multiples cahiers dont les pages de garde ont été retirées (à l'exception de la première et de la dixième) pour former un ensemble uniforme. Le tout a été récemment relié en mi-cuir par le service d'archive.



Archives départementales de l'Ain, BIB Ms 66
Photographies A. Levasseur

L'auteur de ces cahiers est Louis-Dominique Vincent, avocat et premier syndic du Tiers-Etat de l'assemblée de Bresse du 1^{er} décembre 1757 au 22 février 1773. La province de Bresse, rattachée à la France par le traité de Lyon de 1601, était pays d'Etat, c'est-à-dire un territoire qui jouissait d'une certaine autonomie vis-à-vis des institutions royales et en particulier du droit de consentir à l'impôt. L'exercice de ses libertés était assuré par une assemblée composée de députés des trois ordres. En tant que syndic du Tiers, Louis-Dominique Vincent participait à l'assemblée, en exécutait les décisions et assurait la permanence de la représentation de la province de Bresse entre deux sessions. Il était aussi un rouage et un relai de l'Etat royal dans sa province car il travaillait en étroite relation avec l'Intendant et même, pour certains domaines de compétence, sous sa tutelle¹¹.

La charge de syndic était devenue très polyvalente au XVIII^e siècle, et c'est pourquoi le

¹⁰ L'existence de ce manuscrit a été portée à notre connaissance par Jérôme Dupasquier, archiviste responsable des archives privées des Archives départementales de l'Ain, que nous remercions vivement pour cette découverte.

¹¹ LEGAY M.-L., « Les syndics généraux des Etats provinciaux, officiers mixtes de l'Etat moderne (France, XVI^e-XVIII^e siècles) », *Histoire, économie et société*, 23, 2004, p. 489-501.

Journal de Vincent offre un large aperçu de l'administration de la Bresse, à laquelle Louis-Dominique Vincent se consacra sans compter. Il constitue donc une source intéressante pour l'histoire du droit administratif, en particulier celle des biens et des services publics¹². On ne résiste pas à retranscrire un petit passage, qui évoquera aux juristes publicistes l'un des grands arrêts du Tribunal des Conflits, rendu plus d'un siècle et demi plus tard :

« Le 18 août [1764], les syndics des trois ordres se sont assemblés pour examiner une requête donnée par M. de Crémeaux, seigneur de Chasey, qui demande une indemnité, de ce que par la construction du Pont de Chasey, il perd un droit de bac. L'avis a été qu'il n'était dû aucune indemnité parce que la concession d'un droit pour tenir un bac n'est pas un titre lucratif mais une prestation, qui se paye pour l'entretien du bac [...]. Cependant comme M. de Crémeaux fait réellement une perte considérable, les syndics lui ont proposé de lui céder le logis de Chasey et la dépendance moyennant la somme de 5000 livres. M. de Crémeaux trouvera une indemnité dans la médiocrité du prix. Ce parti a été accepté par M. de Crémeaux et approuvé par M. l'Intendant.»¹³

On y trouve aussi des éléments sur l'histoire de la justice et du droit pénal¹⁴, sur l'histoire de la norme¹⁵, et bien entendu sur celle du droit fiscal puisque le consentement à l'impôt et sa répartition étaient la première compétence historique des Etats, qui l'est restée en volume dans le *Journal*¹⁶. Malgré son intérêt indéniable pour l'histoire du droit, le *Journal de Vincent* ne sera probablement pas transcrit ni édité, du moins par les techniques traditionnelles. L'éparpillement des objets d'intérêt rebutera tous les chercheurs, qui se limiteront à glaner dans le manuscrit ce qui les intéresse. Le *Journal de Vincent* est donc condamné à dormir encore longtemps sur son étagère, sauf à expérimenter une transcription par crowdsourcing.

II.- La méthode du crowdsourcing appliquée au *Journal de Vincent*

La « foule » invitée à participer à une recherche n'est pas un objet naturel mais elle est construite par le chercheur, comme l'est son sujet ou son *corpus*. Construire une foule consiste soit à trouver une communauté préexistante et la modeler en fonction des tâches à lui faire réaliser, soit à la créer de toutes pièces¹⁷ (A). Le chercheur organise ensuite le travail en tenant compte des vecteurs de motivation de sa foule, qui lui sont propres (B). *In fine*, il produit des arbitrages en fonction du niveau de qualité espéré des données produites (C).

A.- La construction de la foule

12 Traité entre la Bresse et le Bugey réglant la répartition des frais de déplacement du pont de Chasey, alors en construction, en cas de changement du lit de la rivière (*Journal de Vincent* [JV], p. 149-150, « dans le cas où la rivière quitterait absolument le Pont et se formerait un nouveau lit dans le Bugey ; Je lui dis qu'il est vrai que l'on parla de ce cas, et que l'on convint, que la rivière restant toujours commune, le transport du pont s'il était jugé nécessaire serait aussi commun ») ; destruction d'immeubles privés qui « offusquent » l'Hôtel de la Province (p. 37) ou qui gênent le passage du carrosse public (p. 446) ; résolution amiable des problèmes inhérents aux réquisitions et expropriations pour travaux publics alors que la province n'a pas le pouvoir d'acquiescer un fond (p. 871) ; intérêt local ou provincial d'un pont (p. 355) ; marchés de travaux publics (p. 201), etc.

13 L'orthographe a été modernisée (p. 449).

14 Question du monopole étatique de la force (p. 141 et 143 et 349) ; de la délinquance politique (p. 224) ; de la faiblesse des épices qui asphyxie le présidial (p. 165).

15 Nombreuses références aux usages de Bresse ; réflexion « si par [une] voye amiable l'on peut mettre la règle » (p. 263) ; il insiste pour que soit établie une règle parce que « dans les matières d'administration publique, il ne doit rien y avoir d'arbitraire » (p. 310) ; délai imposé par l'intendant dans l'application d'une délibération de l'assemblée générale en raison des résistances (p. 725, « l'affaire n'était pas encore assez digérée »).

16 Vincent mentionne un mémoire qu'il a rédigé pour démontrer que la maxime « En Bresse, les tailles sont conventionnelles » est sans fondement (p. 46) ; il rappelle les règles d'enregistrement de l'impôt dans les pays d'Etat (p. 179) ; il programme la suppression des corvées (p. 919).

17 Le terme de « foule » est un peu trompeur car l'essentiel du travail est généralement assuré par une poignée de gros contributeurs. Il n'en demeure pas moins que c'est bien une foule qui doit être mobilisée et construite pour permettre à ces gros contributeurs d'émerger.

Dans un monde idéal, le *Journal de Vincent* serait transcrit par des étudiants en master d'histoire ou d'histoire du droit, sur plusieurs années universitaires. Cette foule captive mais intéressée, « invitée » dans le cadre d'une activité pédagogique à effectuer des tâches scientifiques, produit un travail d'autant plus sérieux qu'il est noté. De plus, la validation scientifique de leur production coïncide avec la correction des travaux qu'un chargé d'enseignement se doit de réaliser, ce qui constitue une économie de temps non négligeable¹⁸. Malheureusement, disposer d'une telle foule est un luxe rare. A défaut, le chercheur essaiera de constituer sa foule à partir d'une communauté préexistante, plutôt que de la créer *ab initio*. Le terme « communauté » est ici entendu dans un sens étroit, comme un groupe de personnes associées, qui ont déjà l'habitude d'interagir et de travailler ensemble. Les sociétés savantes, les clubs de généalogie et les associations d'histoire locale sont des communautés, contrairement aux usagers d'un centre d'archives, d'une bibliothèque ou d'un outil numérique. Ainsi, même si la Bibliothèque nationale de France perçoit les grands usagers de sa bibliothèque numérique Gallica comme une communauté (les « gallicanautes »), ils n'en constituent pas une dans l'acception que nous avons retenue ici. Le chercheur qui parvient à créer sa foule dans une communauté préexistante s'épargne le travail de médiation et surtout de communication qu'exige la création d'une foule *ab initio*, une entreprise qui est déjà difficile et hasardeuse pour des institutions disposant de forts moyens de communication et d'une visibilité nationale¹⁹, et il réduit considérablement ses risques d'échec.

En l'espèce, une analyse préliminaire a permis de s'assurer de l'existence d'un réseau associatif actif susceptible d'être intéressé par une transcription du *Journal de Vincent*. Depuis 1966, une fédération de sauvegarde du *Patrimoine des Pays de l'Ain* accueille cent-vingt-cinq associations et possède un catalogue de publications de plus de six cents titres, en grande partie sur l'histoire locale. Cette fédération reconnue d'utilité publique en 2008 soutient le réseau associatif et a aussi une mission d'inventaire, de publication, d'animation et de sensibilisation. Si les membres du *Patrimoine des Pays de l'Ain* et ses associations seront probablement peu intéressés par les problématiques d'histoire du droit développées par le *Journal*, d'autres points ne manqueront pas d'éveiller leur intérêt.

Le *Journal* permet notamment de mieux connaître Louis-Dominique Vincent, un influent personnage bressan dont il reste à reconstituer en détail la vie et la carrière. Né en 1716 d'un notaire pauvre, il devint avocat en 1740, puis conseiller à la province en 1749 avant d'en devenir premier syndic de 1757 à 1773. Devenu avocat au Parlement, il reçut des lettres de noblesse en 1778 et décéda en 1790²⁰. Il laissa trois fils, dont Charles Humbert Vincent qui fut officier du génie et se lia d'amitié avec Toussaint Louverture. Louis-Dominique Vincent, « notre aigle, l'oracle de notre jurisprudence » d'après l'un de ses contemporains²¹, a rédigé des centaines de lettres et au moins une dizaine de mémoires qu'il reste à retrouver. Si son *Journal* est un instrument de travail qui ne dit presque rien de la vie privée de son auteur, il dévoile en filigrane une partie de son histoire. Elle débute le jour même de son installation comme premier syndic puis se déroule pendant quinze ans, au cours desquels l'affaiblissement progressif du corps se perçoit, l'infirmité qui le gagne et limite

18 GALLERON I., WILLIAMS G., « Les étudiants comme crowdsourcers : enjeux pédagogiques et scientifiques », communication à *Le crowdsourcing pour partager*, coll. cit.

19 La création d'une foule *ab initio* est le pari fait par les membres du projet *Testaments de guerre de Poilus* porté par une équipe de chercheurs de l'Université de Cergy-Pontoise, de l'Université Paris 8, du Centre Jean-Mabillon de l'École nationale des Chartes, d'archivistes des Archives nationales et des Archives départementales des Yvelines. La préparation de l'appel à la foule a duré du printemps 2017 au début de l'année 2018. L'appel à contribution officiel a eu lieu le 20 janvier 2018. Le travail de transcription passera par une plateforme collaborative construite pour le projet.

20 A.N., O/1/130, ff. 137-141, relevé par SCHNEIDER C., « Le colonel Vincent, officier du génie à Saint-Domingue », *Annales historiques de la Révolution française*, n° 329, 2002, p. 101-122.

21 D'après le manuscrit de Jérôme DE LANDE, *Anecdotes de Bresse*, p. 249, cité par Edmond RÉVÉREND DU MESNIL, *L'Armorial historique de Bresse, Bugey, Dombes, Pays de Gex, Valromey et Franc-Lyonnais*, Lyon, Vingtrinier, 1872, p. 698.

son action jusqu'à sa démission pour maladie « sur la vue » en mai 1772²². Louis-Dominique Vincent termine toutefois sa triennalité échue au 1^{er} janvier 1773 et le *Journal* s'achève l'après-midi du 22 février, qu'il passe avec son successeur élu le matin même, à réviser les liasses des affaires de la province avant de les lui transmettre définitivement.

Le *Journal* est aussi un trésor pour l'histoire économique, politique et sociale de la Bresse. Louis-Dominique Vincent s'intéressa aux filatures de coton, protégea le commerce des vins bressans contre ceux de Mâcon et, surtout, s'efforça de développer l'horlogerie²³. Il fut le principal interlocuteur et amendeur des cartes géographiques de Cassini (qu'il refusait de payer) et de Seguin²⁴. Il arbitra les questions touchant les frontières de la Bresse, demanda à ce que la Dombe y soit intégrée et s'opposa aux projets de rapprochement avec Lyon²⁵. Il s'intéressa à la santé publique des Bressans en embauchant des médecins au moment des épidémies, en payant des cours d'accouchement aux chirurgiens, en décidant d'acheter le secret du remède pour les morsures de vipère²⁶. Il travailla à l'établissement des chemins de communication et à l'introduction de la poste²⁷, fut l'un des premiers archivistes de la province²⁸, etc. Tous ces développements et d'autres encore qui intéressent fortement le patrimoine bressan laissent espérer que la motivation d'une foule composée d'amateurs d'histoire locale sera suffisante pour permettre une transcription complète du *Journal*, même s'il contient aussi de nombreux paragraphes plus arides sur la fiscalité. Si le risque d'échec de l'entreprise de crowdsourcing n'est pas nul, il paraît néanmoins assez limité, et peut l'être encore en travaillant sur les autres vecteurs de la motivation de la foule.

B.- Les vecteurs secondaires de motivation de la foule

Les vecteurs secondaires de motivation résident dans une bonne adéquation entre le travail proposé et les capacités de la foule pour l'accomplir, dans l'ergonomie de l'espace de travail et dans la récompense finale. S'ils sont mal adaptés, le travail de la foule sera inexistant ou inutilisable.

La nécessité de créer une adéquation entre le travail proposé et les capacités de la foule construite contraint le chercheur à adapter ses ambitions scientifiques²⁹. En l'espèce, le *Journal de Vincent* peut faire l'objet soit d'une transcription simple, aboutissant à la création d'un fichier en mode texte, soit d'une transcription complexe par l'adoption de la *Text Encoding Initiative* (TEI). Ce standard d'édition numérique est fondé sur le principe de l'édition structurée et du balisage prescriptif, et implique de réaliser un encodage par le langage XML-TEI. Une transcription du *Journal* par TEI suppose que la foule non seulement retranscrive le texte mais que de surcroît elle y intègre des méta-données avec une granularité plus ou moins fine, en utilisant des balises portant sur les noms de personnes et de lieux, sur les sujets abordés, les dates, les sommes d'argent, les ratures, les notes ou les dessins marginaux, etc. D'un point de vue scientifique, la transcription par TEI est bien entendu préférable. Il s'agit toutefois d'un travail aride, qui implique de surcroît que les transcripateurs ne soient pas effrayés par l'utilisation d'un langage numérique. Le choix d'une transcription complexe accentue les risques de démotivation, voire de rejet et elle conduit à des taux d'erreurs plus élevés qui peuvent rendre le travail obtenu inutilisable et, quoi qu'il en soit, qui

22 JV, p. 1012.

23 JV, p. 467, 474, 498, 572, 892, 905, etc.

24 JV, p. 29, 63, 66, 289, 557, 966, etc.

25 JV, p. 363, 970, etc.

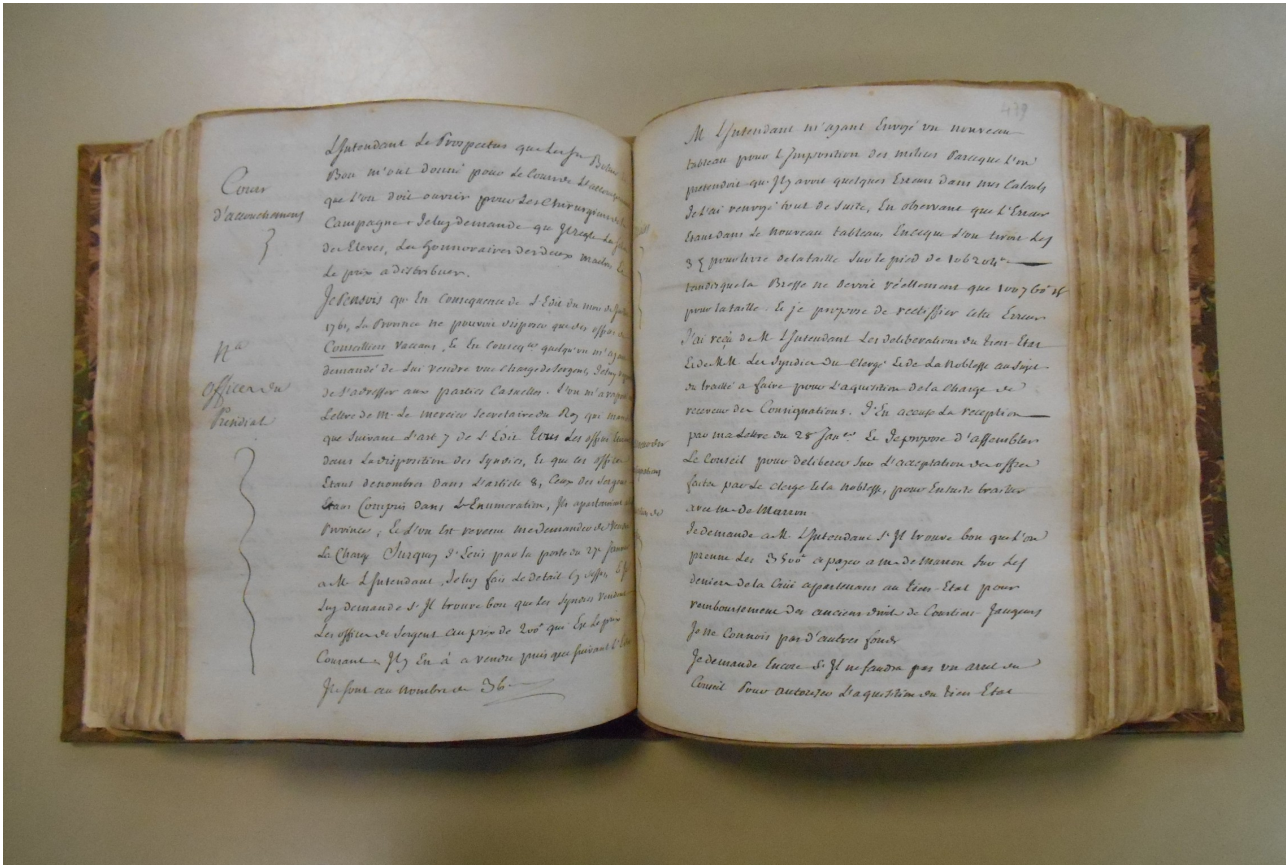
26 JV, p. 947, 960, 1035, etc.

27 JV, p. 35, 36, 47, 57, 66, 67, 111, 136, etc.

28 JV, p. 59, 91, 93, etc.

29 Les chercheurs anglais ont élaboré une typologie distinguant les microtasks (tagging, petites transcriptions), macrotasks (transcriptions plus longues, intégration de méta-données) et metatask (réflexion, recherches sur un personnage par exemple). Lire en particulier les travaux de Mia Ridge (*Crowdsourcing Our Cultural Heritage*, Farnham, Ashgate, 2014 ; « Crowdsourcing in cultural heritage : balancing productivity and engagement », communication à *Le crowdsourcing pour partager*, coll. cit.).

augmentent le coût humain de la vérification³⁰. En l'espèce, le texte clair et fluide du *Journal*, avec peu de ratures et de notes marginales et pas de caviardage, ainsi que le caractère immature de la foule sollicitée rendent préférable une approche par transcription simple³¹. Celle-ci pourra être menée assez facilement, car l'écriture est belle et régulière, déchiffirable sans compétences paléographiques.



Archives départementales de l'Ain, BIB Ms 66
Photographie A. Levasseur

L'ergonomie de l'espace numérique de travail est également déterminante pour la réussite d'un crowdsourcing. Certains projets, portés par des institutions nationales ou soutenus par des subventions importantes, ont pu développer leur propre logiciel de transcription. Si l'interface est alors parfaitement adaptée à la tâche et à la foule, elle est coûteuse et, surtout, elle ne résiste pas toujours à l'obsolescence technique, qui s'accroît avec le temps sans qu'il soit toujours possible

30 L'équipe de EURED, dont l'objectif est de rassembler des expériences de lecture des européens, de l'invention de l'imprimerie à aujourd'hui, n'a pas pu reprendre la base de données UK-RED lancée en 2006 pour un projet identique centré sur le Royaume-Uni, en raison du fort taux d'ambiguïtés, d'erreurs de saisies, de fautes de frappes, d'incompréhension du contenu des champs, etc. (VIGNALE F., « Crowdsourcing et base de données : le cas de EURED », communication à *Le crowdsourcing pour partager*, coll. cit.). Le projet *Testaments de Poilus* des Archives nationales et de l'École des Chartes s'oriente vers une édition par TEI avec une granularité assez fine, mais il permettra aussi au crowdsourcier de se limiter à une transcription simple.

31 Une communauté de crowdsourcing constituée depuis longtemps est capable de mener des tâches de haut niveau scientifique. L'un des projets les plus illustratifs est *Itinera Nova*, développé depuis 2009 par les archives municipales de Louvain, qui a pour objectif de numériser et de transcrire les 1127 registres de la cour échevinale (1362-1795). Les volontaires, au nombre de 40, forment une communauté capable de transcrire des actes du XVI^e siècle, impliquant une maîtrise de la paléographie et du moyen flamand. Ils participent à l'ensemble du processus, de la transcription à la recherche de financements, en passant par la communication autour du projet et le brassage de leur propre bière (CEUNEN M., DIERICK A., « *Itinera Nova*. Tools, People and History », communication à *Le crowdsourcing pour partager*, coll. cit.). Leur travaux sont en ligne sur www.itineranova.be.

d'effectuer les mises à niveau nécessaires³². L'interface de travail peut aussi être externalisée, ce qui règle les problèmes du financement et de l'obsolescence mais soulève d'autres questions. Des clients en open source, légers ou lourds, permettent aujourd'hui de faire un travail de transcription collaborative³³. Ils sont pour la plupart gratuits, assurent eux-même l'évolution technique de leur plateforme, et ont cet avantage d'attirer des transcripateurs supplémentaires si le projet est rendu public. Mentionnons *Wikisource*, *Pybossa*, *From the Page*, *Transkribus*, *Scribe* de *Zooniverse*, bientôt peut-être *PHuN*, etc., chacun possédant ses avantages et ses défauts propres. L'interface de Wikisource par exemple est trop peu intuitive pour motiver une foule de néophytes en numérique³⁴. L'utilisation de l'une de ces interfaces nécessite d'éclaircir en amont la question des droits de propriété sur le *corpus* numérisé qui y est transféré ainsi que sur les données produites puis agrégées, qui sont propres à chaque plateforme. De surcroît, et cela vaut pour tout type de plateforme, l'utilisation d'un espace numérique nécessite toujours un accompagnement de la foule avec la mise en place d'un tutoriel ou de séances de formation, y compris pour une foule composée de webnatifs.

L'octroi de récompenses est un dernier point à évoquer lorsqu'on réfléchit aux vecteurs de motivation d'une foule. La question est délicate car une récompense mal adaptée est un facteur de pollution des tâches. Toute forme de récompense matérielle (argent, goodies, etc.) est à exclure car, outre les questions éthiques qu'elle engendre, elle détourne la foule du véritable objectif de l'entreprise, qui est de créer les données les plus pures possibles³⁵. Les récompenses immatérielles, comme l'octroi de points de réputation, sont moins risquées mais se doivent d'être parfaitement maîtrisées. D'autres formes de récompenses sont presque sans risque. L'organisation de temps de dialogue et de moments de convivialité autour du projet est particulièrement appréciée. La motivation de la foule augmente aussi lorsque le projet peut être terminé dans un temps raisonnable (récompense liée au plaisir du travail achevé) et lorsqu'il porte sur un *corpus* inédit (la transcription d'un texte déjà connu voire déjà édité diminue considérablement la motivation³⁶). En l'espèce, le *Journal de Vincent* remplit ces deux dernières conditions.

C.- Qualité et validation scientifique du travail réalisé par crowdsourcing

Un travail de qualité suppose une phase finale de vérification. Dans l'idéal, elle est assurée de bout en bout par l'expert qui valide *in fine* le caractère scientifique des données créées. Toutefois, les expériences ont montré que le temps économisé par l'expert en amont du projet grâce au crowdsourcing est alors dépensé en aval car le temps de vérification et correction équivaut à celui dont il aurait eu besoin pour réaliser la tâche lui-même (avec cette nuance toutefois que les corrections et le temps passé à la vérification diminuent fortement à mesure que la communauté gagne en maturité). Un tel système de validation scientifique est compatible avec l'objectif du crowdsourcing-médiation mais va à l'encontre du crowdsourcing-recherche, du moins tel qu'il est envisagé dans le projet de transcription du *Journal de Vincent*.

Une autre option envisageable n'exige qu'une intervention minimale de l'expert, concentrée

32 L'obsolescence technique de la plateforme propriétaire est aujourd'hui le problème principal de *Transcribe Bentham*, un projet pionnier dans l'histoire du crowdsourcing. Depuis 2010, cinquante-deux pages du philosophe anglais Jeremy Bentham sont transcrites en moyenne par semaine (SEAWARD L., « Quiet but Powerful : sustaining the Transcribe Bentham volunteer community », communication à *Le crowdsourcing pour partager*, coll. cit.).

33 Le client lourd est un logiciel qui doit être téléchargé puis installé par l'utilisateur sur son matériel, tandis que le client léger fonctionne directement depuis une plateforme internet.

34 MOIREZ P., MOREUX J.-P., JOSSE I., *Etat de l'art en matière de crowdsourcing dans les bibliothèques numériques* [en ligne], février 2013, p. 17. Disponible sur : <http://www.bnf.fr/documents/crowdsourcing_rapport.pdf> (consulté le 06.11.2017).

35 Les résultats d'*Amazon Mechanical Turk* sont paradigmatiques du biais engendré par le paiement d'un salaire dans un système fondé sur le crowdsourcing (LAFOURCADE M., « Promenade dans un réseau lexico-sémantique obtenu par crowdsourcing : 10 ans du projet JeuxDeMots », communication à *Le crowdsourcing pour partager*, coll. cit.).

36 GALLERON I., WILLIAMS G., « Les étudiants comme crowdsourcers », com. cit.

en début de projet pour rectifier les erreurs primaires. La majeure partie de la vérification et de la correction est ensuite laissée à la foule par un système de double ou triple relecture avant une validation automatique par le client numérique³⁷. L'inconvénient de cette option réside dans le caractère non scientifique du travail achevé, ce qui paraîtra de prime abord comme un défaut rédhibitoire au chercheur tant il est contraire à sa culture académique. Mais, pour la recherche scientifique, un texte retranscrit même s'il comporte des erreurs est préférable à un manuscrit oublié dans les archives, d'autant que les scories peuvent être progressivement rectifiées, et qu'un système de renvoi aux pages numérisées permet aux chercheurs de revenir à l'original en cas de doute. L'essentiel est de communiquer clairement sur la nature du travail réalisé, par exemple par l'apposition d'un label « crowdsourcing » sur les données produites.

L'avenir seul nous dira si le crowdsourcing-recherche est amené à se développer ou à disparaître. Aujourd'hui, il importe de l'extraire de sa gangue idéologique et de ne pas le diaboliser. Il ne faut pas non plus faire preuve d'un trop grand angélisme. Le crowdsourcing n'est pas une solution miracle, il n'est même pas vraiment gratuit puisqu'il se contente de déplacer les coûts³⁸. Pour l'instant, il ne faut y voir qu'un outil supplémentaire proposé par la révolution numérique, qui peut rapidement devenir un écueil tant il semble délicat à maîtriser. Il importe donc de l'encadrer par une méthode solide pour s'assurer de l'adéquation entre la démarche et les objectifs, entre la tâche et les compétences de la communauté, et enfin entre ceux derniers éléments et l'interface de travail. Surtout, le chercheur devra garder à l'esprit que, pour les personnes qui composeront sa foule, l'entreprise sera certes une aventure scientifique, mais avant tout une aventure humaine.

37 Il existe aussi des systèmes de validations mixtes, dans lesquels le conflit est identifié par la machine, puis transmis à l'humain qui corrige.

38 ANDRO M., « Le crowdsourcing dans les bibliothèques », *com. cit.*