Wikipedia, un trouble-fête de l'édition scientifique

Lionel Barbe

▶ **To cite this version:**

Lionel Barbe. Wikipedia, un trouble-fête de l'édition scientifique. Hermès, La Revue - Cognition, communication, politique, 2010, Sciences.com : libre accès et science ouverte, 1 (57), pp.69-74. 10.4267/2042/38640 . halshs-00643121

HAL Id: halshs-00643121
https://shs.hal.science/halshs-00643121

Submitted on 21 Nov 2011

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.

Manuscrit auteur publié dans Hermès 57 Sciences.com octobre 2010 (p 69/74)

Lionel Barbe
Modyco (équipe info-com)
Université Paris X - Nanterre

WIKIPEDIA, UN TROUBLE-FÊTE DE L'ÉDITION SCIENTIFIQUE

Souvent critiquée, l'encyclopédie en ligne Wikipedia n'en est pas moins devenue en quelques années un outil de diffusion d'informations scientifiques de plus en plus utilisé dans le monde entier, par les élèves mais également par les chercheurs. En 2005, 17 % des chercheurs publiant dans la revue *Nature* reconnaissaient en effet consulter Wikipedia, et 10 % y contribuer (Giles, 2005). Les producteurs classiques de connaissances s'affichent même parfois : dans la section francophone, par exemple, plus de 400 contributeurs se sont auto-recensés dans la catégorie « Enseignants et chercheurs », 43 dans la catégorie « Juristes » et presque 500 dans la catégorie « Ingénieurs ». Ces chiffres ne concernent que des Wikipédiens actifs volontairement recensés et ils sont donc très partiels car de nombreux contributeurs ne précisent pas leur profession. Mais cela pose néanmoins la question du rapport délicat entre production des savoirs dans le monde académique et dans Wikipedia.

Une « illusion technophile » pourtant bien réelle

Pour reprendre le bon mot de Piotr Konieczny (2007), chercheur et contributeur de Wikipedia, « le problème avec Wikipedia, c'est que ça marche en réalité, mais pas en théorie ». Construit autour d'un système d'édition plus complexe et plus opaque qu'il n'y paraît, l'encyclopédie Wiki occupe une place de choix dans les résultats des moteurs de recherche, malgré les imperfections, grâce à un souci d'innovation constant et aux millions de contributions des « wikipédiens ».

Wikipedia aurait pu n'être qu'un feu de paille, mais l'encyclopédie offre désormais 15 millions d'articles en plus de 270 langues, créés et mis à jour par près de 100 000 contributeurs bénévoles. Le succès de Wikipedia dépend sans doute des algorithmes de classement de Google puisque l'omniprésent moteur de recherche classe systématiquement en première position Wikipedia pour des mots-clés aussi divers que

« relativité générale », « physique quantique », « théorie de l'information » et même « sciences de l'information et de la communication ». Google ne peut être accusé de favoritisme, puisque, comme on le sait, il privilégie les sites commerciaux, ce qui n'est assurément pas le cas de Wikipedia, site indépendant, gratuit et sans publicité. De plus, depuis 2007, Google a développé un projet concurrent d'encyclopédie, baptisé Google Knol. Le succès de Wikipedia dans les moteurs de recherche est donc lié à l'extrême « citabilité » (interne comme externe) de ses articles, qui explique son succès auprès de l'algorithme Google.

On aurait pu également ranger l'entreprise au rang des vulgates, de grande ampleur certes mais sans portée dans le monde du savoir académique. Or, en 2005, un article dans la revue *Nature* donnait un sceau de respectabilité à l'entreprise, puisque la comparaison avec l'encyclopédie commerciale *Britannica* montrait que le taux d'erreurs décelées dans les deux encyclopédies était proche. (Giles, 2005)

Le paradoxe est que les articles de Wikipedia répondent mal aux exigences éditoriales scientifiques : les auteurs se cachent souvent derrière un pseudonyme qui fait écran aux références statutaires ; la validation ne se fait pas entre pairs ; le principe commun aux contributeurs n'est pas l'objectivité mais un mythique *neutral point of view* (« NPOV ») aussi emblématique que problématique.

De la nature des savoirs sur Wikipedia

Wikipedia se résume dans un principe à trois processus : une écriture collaborative, un système de contrôle, une dynamique communautaire.

L'écriture collaborative est donc l'une des bases de l'encyclopédie libre. N'importe qui peut créer un article, peut l'amender, peut le détruire. Seule la collaboration des wikipédiens encadre le processus. L'IBM Watson Research Center (Viegas et Wattenberg, 2004) a montré l'efficience de ce modèle, évoquant un « effet piranha » pour parler de la dynamique d'incrémentation sémantique d'un article grâce aux participations des contributeurs. L'efficience peut être relativisée pourtant (Barbe, 2006), la variabilité de « l'effet piranha » étant très grande. Certes, l'article intitulé « Pomme » a débuté par « La pomme est un fruit » et a connu un grand succès collaboratif. Mais il n'en va pas de même pour des sujets plus spécialisés, en particulier pour les sujets scientifiques.

Une analyse méticuleuse de la version francophone montre que, dans plusieurs cas d'articles ayant obtenu le label « articles de qualité », le nombre de contributeurs à l'origine de l'essentiel du contenu est réduit : deux à cinq contributeurs principaux, en tout cas moins que les centaines de contributeurs évoqués. Si les contributeurs sont parfois nombreux, ils n'interviennent pas tous au même titre, la plupart se contentant de modifications mineures ; ils sont donc moins des auteurs que des correcteurs, des maquettistes, des « référenceurs », etc. Plus un sujet est spécialisé, moins le nombre de contributeurs intervenant sur le fond de l'article est important. Selon nos calculs (*ibid*), 5 % environ des contributeurs sont à l'origine de 90 % du contenu. La genèse de la plupart des articles de l'encyclopédie en ligne est donc due à quelques centaines de wikipédiens.

L'encyclopédie est dite libre, mais son fonctionnement est en partie hiérarchique. Au cœur du système de contrôle du contenu se trouvent en effet les « administrateurs ». Maintes fois dénoncé pour ses possibles dérives totalitaires (le site wiki-observateur.org s'en fait une spécialité), ce statut, acquis dans le cadre d'un système de vote à la fiabilité aléatoire, confère un pouvoir impor-

tant sur les choix éditoriaux de Wikipedia. La mesure la plus décriée et néanmoins nécessaire est celle du ban temporaire ou définitif d'internautes indélicats. La mise en application de cette mesure a même abouti au bannissement d'une personnalité, Grishka Bogdanov, soupçonnée de modifier à son avantage l'article qui lui était consacré. Plusieurs cas d'exclusions de chercheurs ou d'universitaires pour la rédaction de pages traitant de leur spécialité scientifique ont également été signalés (Christian Magnan, de l'Institut d'astrophysique de Paris, s'est plaint d'avoir été « mis à la porte », par exemple).

Dans la plupart des cas, les administrateurs ou les gros contributeurs appuient leur plainte sur un défaut de *neutral point of view* pour justifier leurs interventions, donnant lieu à de conventionnelles « controverses de neutralité » (Jacquemin *et al.*, 2008). Bien souvent, les *ego* s'affrontent (Auray *et al.*, 2009) et l'entente est difficile entre experts reconnus d'un domaine et administrateurs néophytes, souvent anonymes et pas toujours compétents en la matière. Ce sont pourtant eux qui ont la plupart du temps le dernier mot à l'intérieur du comité d'arbitrage de l'encyclopédie (Jacquemin *et al.*, 2008) ou à l'extérieur, en poussant notamment à la faute de forme leurs opposants (Auray *et al.*, 2009). Finalement, le système basé sur les administrateurs et le *neutral point of view* a pour effet de gommer les contenus qui ne sont pas pré-existants ou avérés, tout en laissant une marge de manœuvre très mince à ceux qui sont par ailleurs justement responsables de leur validation ou de leur création. Ce paradoxe apparent masque une réalité : Wikipedia est un projet agrégatif et normatif, mais ce n'est pas un espace d'édition de connaissances de première main. Reste le problème de l'auto-application de cette règle par les administrateurs eux-mêmes, qui sont, rappelons-le, à l'origine d'une grande partie du contenu validé.

Les motivations des gros contributeurs de Wikipedia ne sont pas non plus clairement identifiées. Si le rôle de la dynamique communautaire a pu être démontré, des études ont également établi qu'un positionnement identitaire est souvent recherché par les wikipédiens assidus (Anthony, 2005). Certains chercheurs se focalisent davantage sur le rôle des dynamiques sociales et communicationnelles entre individus que sur le contenu des articles, envisagé dans ce cas comme le résultat du tissage de liens sociaux et communautaires. Cette approche est soutenue par le fait que le contenu des articles ne représente qu'un quart du total des mots écrits sur le site, ainsi qu'on le découvre sur la page « Statistiques spéciales » de l'encyclopédie en ligne.

Cependant, si les gros contributeurs ont joué un rôle majeur pendant la phase d'extension rapide de Wikipedia, en créant notamment les entrées correspondantes à des ensembles de mots, il semble que l'enjeu réside davantage aujourd'hui dans les disciplines plus spécialisées, en particulier les disciplines scientifiques. Nous avons vu que plus un article est spécialisé, moins le nombre de contributeurs est important. On constate également en parcourant les pages utilisateurs que les contributeurs sont plus fréquemment des experts du domaine ou des personnes disposant de connaissances approfondies. Il s'agit donc de contributeurs spécialisés, qui n'ont pas une optique de croissance généralisée de l'encyclopédie mais davantage un but de communication sur un ou deux domaines particuliers, ou de promotion de ces domaines auprès des communautés d'experts qui leur sont proches (Barbe, 2008). Il semble donc qu'il faille invoquer deux dynamiques contributives distinctes. Une dynamique grand public de construction commune d'une base de connaissances générales et une dynamique de communication d'informations spécialisées à destination d'une population plus restreinte d'initiés, concernant l'information scientifique et technique en particulier.

L'analyse des arcanes de Wikipedia montre que le contenu final de l'encyclopédie est issu d'un processus

de sélection et invite à fortement relativiser la liberté théorique de publication. C'est d'autant plus probant en matière de sciences où les experts et chercheurs interviennent de plus en plus fréquemment sur le contenu, lorsqu'ils ne sont pas eux-mêmes à l'origine de celui-ci. Par ailleurs, il est nécessaire de distinguer d'une part les contributeurs généralistes, dont l'activité consiste essentiellement en un effort de synthèse et de remise en forme de contenus pré-existants, et d'autre part les auteurs d'articles spécialisés. Néanmoins, il faut garder à l'esprit que l'encyclopédie libre n'a pas vocation à accueillir du contenu inédit, même venant d'experts, puisqu'un des principes fondateurs consiste à demander aux contributeurs de citer des sources pour appuyer leurs propos. Le niveau de qualité des sources constitue donc un élément fondamental pour évaluer la solidité des articles scientifiques de Wikipedia.

Les analyses bibliométriques sur Wikipedia sont peu nombreuses. Une étude empirique de la *Wee Kim Wee School of Communication and Information* de Singapour (Luyt et Tan, 2010) conclut à une forte prédominance des sources officielles ou à des contenus trouvés sur Internet dans les citations, au détriment des publications scientifiques. Cependant, cette étude est basée uniquement sur les articles traitant de l'histoire des pays dans la Wikipedia américaine. Une autre étude (Arup, 2008), cette fois systématique, s'est intéressée à la représentation des grandes revues scientifiques américaines dans les citations de Wikipedia. Cette étude démontre une forte corrélation entre les niveaux de citations des revues dans le *Journal Citation Reports* et dans Wikipedia. Autrement dit, les revues qui bénéficient des plus forts facteurs d'impacts en science sont aussi celles qui sont le plus souvent citées sur Wikipedia (*Nature* et *Science* en premier lieu). Autre point, selon l'étude, la croissance des citations scientifiques dans Wikipedia entre novembre 2006 et avril 2007 a été très importante, passant de 20 000 à plus de 30 000 uniquement pour les 80 revues les plus citées.

Ces deux études ne sont pas contradictoires, elles montrent que dans de nombreux cas, comme les articles sur l'histoire des pays, les citations de publications académiques sont trop peu nombreuses sur Wikipedia par rapport à des sources provenant de médias grand public plus accessibles pour les non-spécialistes d'un domaine. Mais la forte augmentation des sources scientifiques et leur bonne représentativité s'inscrivent logiquement dans le tournant qualitatif entrepris par l'encyclopédie depuis 2006 avec, en particulier, le développement du label « article de qualité » et l'apparition d'un nouveau label « bon article ». Cependant, ces articles labellisés ne représentent qu'une très petite partie du total, soit environ 1 500 articles sur 900 000 dans la Wikipedia francophone. Pour qu'un article soit déclaré « de qualité », il faut qu'une demande soit faite, puis qu'un vote ait lieu, sachant que le label requiert nombre d'autres éléments que les citations, comme par exemple un plan détaillé, des illustrations, etc. Ainsi, Wikipedia pourrait favoriser la visibilité de certaines publications scientifiques sur le Web (Black, 2008).

Quelle place pour Wikipedia et ses concurrents vis-à-vis de l'édition « scientifique » ?

Il convient d'abord de situer Wikipedia face à ses deux concurrents les plus crédibles en matière d'édition de connaissances qui sont Citizendium et Google Knol. Lancés respectivement en 2006 et 2007, ces deux sites ont vocation à faire concurrence à Wikipedia en exploitant son talon d'Achille : l'anonymat des auteurs. Dans les sites concurrents, il est nécessaire de s'identifier. Mais si le modèle de Citizendium est basé sur la validation des articles par des experts, celui de Google

Knol, qui propose par ailleurs de rémunérer les auteurs en fonction de l'audience de leurs articles, semble beaucoup moins convaincant en matière de fiabilité des contenus. En effet, l'identification sur Google Knol est réduite au dépôt d'un e-mail avec nom et prénom sans aucune autre vérification. L'édition d'un Knol, unité de savoir en langage « googlien », est donc possible en quelques minutes et de façon automatique, et ce sans comité de lecture, bien que Google affirme étudier cette possibilité.

Face au mythe du *neutral point of view* de Wikipedia, Google Knol propose donc la responsabilité des auteurs, retournant ainsi à un modèle plus traditionnel d'édition personnelle, permettant cependant les commentaires, à l'image d'un blog. Une option permet aux auteurs d'ouvrir à la collaboration la rédaction de leur texte. L'indice de popularité d'un article est mis en avant, augmentant la visibilité dudit article sur le site. Cette organisation pyramidale de l'audience semble favoriser les sujets polémiques et la promotion de produits ou de services. Les conseils pratiques sont également largement représentés sur Google Knol : résolution de problèmes informatiques, réussir en bourse, etc. En conclusion, Google Knol rencontre deux principaux écueils. Le premier est sa croissance très lente ; seulement quelques centaines d'articles publiés en français et quelques milliers en anglais. Le second concerne l'absence d'un système de contrôle de la qualité des contenus. Paradoxalement, la facilité de publication sur Google Knol est mise en avant par certains auteurs d'articles de Knol[1] qui trouvent trop restrictif le système de contrôle *a posteriori* de Wikipedia. On constate par ailleurs que bien peu d'articles de Knol utilisent des citations.

Le modèle de Citizendium est plus efficient au niveau scientifique car il est basé sur des groupes de travail d'experts réunis par thèmes ou disciplines. La validation par un comité de lecture est obligatoire pour la publication d'articles et la qualité est donc en principe garantie. Cependant, Citizendium reste confidentiel sur Internet et une comparaison qualitative serait limitée par le très faible nombre d'article de Citizendium par rapport à Wikipedia.

Pour l'instant, la place de Wikipedia comme porte d'entrée de la connaissance sur Internet ne semble pas remise en cause au regard du développement lent de ses concurrents. Une intéressante étude d'Alireza Noruzi (2009) montre qu'entre 2004 et 2009 le nombre de citations contenant le mot Wikipedia dans le *Web of Science* de l'ISI est de 263, toutes disciplines et langues confondues, soit 26 fois supérieur au nombre de citations de l'*Encyclopaedia Universalis*. L'étude montre également qu'après avoir atteint en 2007 le nombre maximum de 80 citations, ce même nombre a ensuite diminué en 2008 et 2009, à seulement 21 citations. Les articles de l'encyclopédie libre seraient donc de moins en moins cités dans des publications scientifiques et l'on peut penser que les polémiques liées à la fiabilité de son contenu n'y sont pas pour rien. Wikipedia, tout en s'imposant au grand public, serait donc de moins en moins reconnue comme une source fiable d'édition scientifique, source qu'elle n'a *a priori* pas vocation à devenir.

Par ailleurs, l'irruption du modèle de l'accès ouvert a engendré un remaniement du secteur de l'édition scientifique et créée des espaces nouveaux de publications et de références dédiés aux sciences (Lardy, 2009 ; Bosc, 2008). Le risque de remplacement ou du moins de parasitage du monde de l'édition scientifique par Wikipedia semble s'éloigner, car le problème de la fiabilité du système demeure, et en particulier celui de sa vulnérabilité face aux attaques volontaires pour le mettre en défaut (Gourdain *et al.*, 2007). Et la controverse sur sa fiabilité semble devoir durer. Cependant, ayant atteint un plateau quantitatif (la croissance du nombre d'articles n'augmente plus), l'encyclopédie libre pourrait passer à un modèle mixte en validant des versions définitives d'articles qui ne seraient plus modi-

fiables. C'est déjà le cas de plusieurs centaines d'articles, objets de conflits d'édition et protégés temporairement ou définitivement en écriture. Cette cristallisation progressive du contenu de Wikipedia aurait pour avantage de permettre la mise en place de garanties qualitatives et de pallier aux faiblesses qualitatives de son système éditorial, pourtant à l'origine de sa croissance.

NOTE

1. Voir<http://knol.google.com/k/pourquoi-knol-est-il-un-grand-prématuré-et-wikipedia-un-grand-vieillard#>.

RÉFÉRENCES BIBLIOGRAPHIQUES

ANTHONY, D., SMITH, S. W. ; WILLIAMSON, T., *Explaining Quality in Internet Collective Goods : Zealots and Good Samaritans in the Case of Wikipedia*, Hanover, Dartmouth College, 2005.

AURAY, N., HURAULT-PLANTEL, M., JACQUEMIN, B., POUDAT, C., « La négociation des points de vue : une cartographie sociale des conflits et des querelles dans Wikipedia francophone », *Réseaux*, vol. 27, n° 153, 2009.

ARUP, N., *Scientific Citations in Wikipedia*, 2008. En ligne sur
<arxiv.org>.

BARBE, L., « Experts, professionnels et profanes : jeux d'acteurs dans la co-construction des informations et savoirs sur le Web participatif et collaboratif », *Actes de la conférence internationale Isko*, 2009.

BARBE, L., « Wikipedia et Agoravox : des nouveaux modèles éditoriaux ? », *Actes de la conférence "Document numérique et société" (DocSoc)*, ADBS Éditions. 2006.

BLACK, E. W., « Wikipedia and academic peer review : Wikipedia as a recognised medium for scholarly publication », *Online Information Review*, 2008.

BOSC, H., « Le droit des chercheurs à mettre leurs résultats de recherche en libre accès : appropriation des archives ouvertes par différentes communautés dans le monde », *Terminal*, numéro spécial « Propriété intellectuelle », 2008.

GILES, J., « Internet encyclopaedias go head to head », *Nature*, nº 438, 2005, p. 900-901.

GOURDAIN, P., O'KELLY, F., ROMAN-AMAT, B., SOULAS, D., DROSTE ZU HÜLSHOFF, T. VON, *La Révolution Wikipedia. Les encyclopédies vont-elles mourir ?*, préface de Pierre Assouline, Paris, Mille et Une Nuits, 2007.

JACQUEMIN, B., LAUF, A., POUDAT, C., HURAULT-PLANTEL, M., AURAY, N., *La Fiabilité des informations sur le Web : le cas Wikipedia*, 5e Conférence en recherche d'information et applications (Coria), Trégastel, 2008.

KONIECZNY, P., « Wikis and Wikipedia as a teaching tool », *International Journal of Instructional Technology and Distance Learning*, 2007.

LARDY, J.-P., « Multiplication des modèles économiques d'accès aux publications scientifiques », *Archimag*, hors-série, nº 35, 2009.

LUYT, B., TAN, D., « Improving Wikipedia's credibility : References and citations in a sample of history articles », *Journal of the American Society for Information Science and Technology*, janvier 2010.

NORUZI, A., « Wikipedia popularity from a citation analysis point of view », *Webology*, vol. 6, n° 2, juin 2009.

RAYNAUD, D., *Sociologie des controverses scientifiques*, Paris, PUF, 2003.

VIEGAS, F., WATTENBERG, M., *History Flow Visualizing Dynamic, Evolving Documents and the Interactions of Multiple Collaborating Authors : a Preliminary Report*, IBM Watson Research Center, 2004.