ÉTUDES DE COMMUNICATION
langages, information, médiations

Études de communication

61 | 2023
Les pratiques informationnelles : dynamiques conceptuelles, questionnements méthodologiques

Blogger seul ou à plusieurs sur *Hypotheses.org* : une étude à partir de l'entropie des carnets

Blogging alone or with others on the *Hypotheses.org* platform: a study based on entropy

Elsa Poupardin et Camille Prime-Claverie

OpenEdition Journals

Édition électronique
URL : https://journals.openedition.org/edc/16451
DOI : 10.4000/edc.16451
ISSN : 2101-0366

Éditeur
Université de Lille

Édition imprimée
Date de publication : 18 décembre 2023
Pagination : 231-252
ISBN : 978-2-917562-30-7
ISSN : 1270-6841

Référence électronique
Elsa Poupardin et Camille Prime-Claverie, « Blogger seul ou à plusieurs sur *Hypotheses.org* : une étude à partir de l'entropie des carnets », *Études de communication* [En ligne], 61 | 2023, mis en ligne le 02 janvier 2025, consulté le 30 janvier 2025. URL : http://journals.openedition.org/edc/16451 ; DOI : https://doi.org/10.4000/edc.16451

Le texte et les autres éléments (illustrations, fichiers annexes importés), sont « Tous droits réservés », sauf mention contraire.

Blogger seul ou à plusieurs sur *Hypotheses.org* : une étude à partir de l'entropie des carnets

Blogging alone or with others on the Hypotheses.org *platform: a study based on entropy*

Elsa Poupardin

URP 441 CERILAC (Centre d'études et de recherches interdisciplinaires en lettres, arts et cinéma), Université Paris Cité
elsa.poupardin@u-paris.fr

Camille Prime-Claverie

EA 7339 DICEN (Dispositifs d'information et de communication à l'ère numérique), Université Paris Nanterre
cclaveri@parisnanterre.fr

Nous proposons une typologie originale des 3 540 « carnets » tenus sur la plateforme *Hypotheses.org* depuis sa création et ce, à partir de l'observation des pratiques d'écriture et de la manière dont ils sont investis par un ou plusieurs contributeurs affiliés au monde de la recherche. La variété des modes de participation sur la plateforme (intervention unique, mono-carnet, collaborations à différents blogs, etc.) et leur évolution montrent combien la co-écriture occupe une place importante dans le blogging scientifique.

◇◇◇◇◇◇◇◇◇◇◇◇

Mots clefs : blogs académiques, pratique d'écriture, participation, collaboration, co-autorialité, entropie de l'information

We propose an original typology of the 3540 «research notebooks» hosted on the Hypotheses.org *academic blog platform since its creation, through analysis of contributors' writing practices. The variety of modes of participation on the platform (single contribution, single notebook, collaborations across different blogs, etc.), as well as the evolution of these modes of participation over time, show the increasing importance of co-writing in scientific blogging.*

◇◇◇◇◇◇◇◇◇◇◇◇

Keywords: academic blogs, writing practices, participation, collaboration, co-authorship, information entropy

La diversité des blogs scientifiques qui sont apparus dans les années 2000, leur succès et leur longévité, ont généré de nombreuses recherches. L'utopie initiale d'une communication sans remédiation entre scientifiques et grand public a progressivement été remise en question du fait de l'évolution des différents modèles de plateforme de *blogging*, financés par des éditeurs privés ou des institutions scientifiques (Poupardin et Prime-Claverie, 2023). L'intérêt s'est alors resserré sur le *blogging* académique, sous-groupe du *blogging* scientifique, caractérisé par le fait que ses auteurs appartiennent au monde de la recherche et non pas à celui du journalisme ou de la médiation. Les études ont principalement cherché à catégoriser cette production, le plus souvent à partir des pratiques d'écriture de leurs auteurs pour déterminer le profil de ces derniers, leur motivation ou évaluer les retombées de ces textes. Certaines ont considéré les types de contenus textuels (propos intimes, amateurs, experts, etc.) qu'ils proposent (Morelli, 2007) ou leurs modes d'énonciations (Cardon et Delaunay-Téterel, 2006). D'autres ont choisi de prendre en compte le type de lectorat visé par les blogueurs (Gardiner *et al.*, 2018) ou celui effectivement touché (Jarreau et Porter, 2017) pour effectuer des regroupements pertinents. Des sélections ont pu également être opérées en tenant compte de la forme du blog (présence de liens externes, de commentaires ou de menus) (Herring *et al.*, 2004), du style d'écriture que les chercheurs adoptent pour séduire leur lectorat (Zou et Hyland, 2020a ; 2020b) ou la nature des motivations avancées par les auteurs pour justifier leur investissement sur différentes plateformes européennes (Kjellberg, 2010 ; Marht et Puschmann, 2014).

Le nombre de co-auteurs d'un blog est une variable rarement prise en compte pour caractériser ces écrits. En partie parce que le blog reste considéré comme un objet éminemment personnel. Le site français de Wikipédia ne suggère que le 29 mars 2020 sur sa page consacrée au blog que ce dernier peut être collectif et ses billets publiés par « son/ses propriétaire(s) ou son/ses webmaster(s) ». Ce sont principalement les écrits à la première personne qui focalisent l'attention, parce que susceptibles de faire apparaître un nouveau type d'écriture scientifique relevant du journal intime (Deseilligny, 2010). La sous-estimation de l'écriture collective sur un blog est relevée dès 2006 (Broudoux, 2006) et plusieurs enquêtes signalent que le nombre d'auteurs du blog a un

impact sur la nature des contenus proposés. Certaines études s'intéressent aux liens existants entre le nombre d'auteurs d'un blog et la longueur et qualité des écrits qu'il propose, ainsi qu'à la longueur et la qualité de ses écrits (Hearst et Dumais, 2009 ; Brown, 2021), mais ne permettent pas d'avoir une vue globale des pratiques. Cependant, les chevauchements importants de celles-ci ne permettent pas d'établir une typologie solide.

Nous nous pencherons dans cet article sur les blogs publiés sur la plateforme *Hypotheses.org* créée par le Centre pour l'édition électronique ouverte (CLÉO) en 2008. Accessible par le portail OpenEdition, elle héberge des blogs sur différents domaines des sciences humaines et sociales, en français, mais également en allemand, anglais, espagnol.

Nous proposerons ici une étude de la « co-présence » des auteurs sur les carnets. Cette « co-présence », qui se manifeste par la présence de textes signés par différents auteurs sur un même carnet, ne s'apparente pas à la co-signature de publication. Elle est bien moins engageante puisqu'elle ne nécessite pas le partage de données ou de cadre théorique. Elle n'est pas non plus assimilable à la co-présence au sein d'une revue ou d'un livre, d'une part, parce qu'elle est moins rentable a priori académiquement, puisque les écrits dans un blog ne sont pas reconnus au même titre que des publications scientifiques classiques ; d'autre part, parce qu'elle relève d'un autre investissement dans la recherche (Poupardin et Faury, 2018 ; Gunthert, 2010).

Ce partage d'un espace d'écriture ouvert est relativement spécifique au *blogging* et nous partons ici de l'hypothèse qu'il dénote des relations qui peuvent être de nature variée entre les auteurs ou tout au moins annonce un lien en devenir. Mesurer la répartition des interventions des différents auteurs au sein d'un blog donné permettra de déterminer comment la parole est partagée. Pour cela, nous utiliserons la mesure d'entropie largement utilisée dans d'autres domaines, comme l'écologie, pour mesurer la biodiversité d'un système. Nous dresserons à partir de l'entropie une catégorisation de ces carnets.

À partir des modalités de co-écriture des carnets, nous essayerons de déterminer le type de contenus qu'ils peuvent proposer. Nous établirons ensuite, à partir de la catégorisation des carnets et du niveau d'investissement des auteurs sur la plateforme, une typologie de ces derniers afin de préciser les dynamiques d'écriture.

1. Les recherches sur *Hypotheses.org*

La plateforme *Hypotheses.org* permet à tous les acteurs de la sphère scientifique pouvant justifier d'une affiliation académique en France

(chercheurs en poste ou doctorants, ingénieurs de recherche, etc.) d'ouvrir des carnets de recherche. Elle assure la cohérence scientifique de l'ensemble grâce à la mise en place de procédures de validation et la présence d'un conseil scientifique. Elle héberge aujourd'hui la quasi-totalité des blogs de chercheurs en SHS français et occupe une place à part dans le paysage de la recherche. Elle est tout à la fois : outil de valorisation des travaux réalisés, lieu d'échange ou de débats auprès de ses pairs et média de communication vers le grand public.

Le terme « carnet de recherche », « concept sans doute plus respectable » (Blanchard, 2016, 8) a remplacé celui de blog. La plateforme fournit conseils et suivi grâce à une liste de diffusion, une FAQ et des formations d'utilisateurs dans différents cadres.

> « [OpenEdition] a pris au sérieux la question de la formation, de l'animation et de la curation, en consacrant du personnel et des moyens à cet accompagnement des utilisateurs et cette valorisation des contenus ; enfin, il a créé de la confiance en inscrivant *Hypotheses.org* dans un écosystème de publication académique grandissant et respecté. » (Blanchard, 2016, 8).

Développé sous WordPress, le *back-office* prévoit le travail collectif en proposant la gestion des rôles des fonctions éditoriales (administrateur, éditeur, auteur, abonné), en laissant une grande marge de manœuvre pour les choix esthétiques et de mise en forme. Dès que le carnet est accepté sur la plateforme, les billets sont directement mis en ligne par leur auteur. Une fois ouvert, le carnet n'est pas figé. Son « créateur » garde toujours la possibilité de modifier son apparence et sa structure, de supprimer des billets ou même de le retirer complètement de la plateforme avec éventuellement une migration vers un autre serveur, même si cette solution est déconseillée par l'équipe d'OpenEdition (Casanova, 2022). La plateforme est ouverte à tous les types d'écriture, ce qui permet la digression, l'exposition des « à-côtés d'une activité de recherche » aussi bien que la mise en visibilité de l'activité de recherche dans son actualité (Mayeur, 2019).

La recherche sur les carnets d'*Hypotheses.org* s'est heurtée à la croissance rapide du site : 55 carnets en 2010, 940 cinq ans après et 1 400 en 2016 (Dacos, 2019). Leur diversité gêne également les différentes tentatives de description. Les études se cantonnent donc souvent à un « type » de carnets (carnets de chercheur ou carnets de thèse) ou se limitent à considérer un groupe de billets. Quand elles essaient de rendre compte de la diversité de pratiques des auteurs, elles doivent le plus souvent s'arrêter aux déclaratifs des auteurs sur leur pratique (Gunthert, 2010) et

ne parviennent pas à quantifier leur importance respective au sein de la plateforme. Faute de pouvoir appréhender la plateforme dans sa totalité, la possible évolution des profils des auteurs ou de leur comportement, les différences disciplinaires ou langagières sont mises de côté.

Rares sont les études purement quantitatives. On peut cependant recenser ici deux d'entre elles. La première a tenté de comprendre le fonctionnement des échanges entre carnetiers à partir d'une étude des liens hypertextes présents sur chaque page d'accueil des carnets (Debaz, 2010). La deuxième s'est penchée sur la circulation des lecteurs entre les carnets eux-mêmes, à partir de l'étude des liens entrants sur les carnets (Mercklé, 2019). Les cartographies réalisées sont peu exploitables pour le moment. Elles permettent d'identifier les carnets qui sont des lieux de passages importants mais les liens mis à jour ne signifient pas qu'il existe vraiment des échanges d'un ordre ou d'un autre entre les auteurs.

2. Présentation du corpus

OpenEdition a mis à notre disposition deux corpus. Le premier, livré en juillet 2019, comprend tous les billets des blogs référencés au catalogue des carnets de recherche à cette date. Le second, livré en juillet 2020, ajoute les nouveaux billets édités dans tous les carnets au catalogue depuis juillet 2019. Ont été fournis le texte intégral de chaque billet, un login d'auteur, la langue utilisée, la date de publication et le nom du carnet sur lequel le billet a été publié. Pour étudier nos corpus, nous avons récupéré les métadonnées sur les carnets (date de mise en ligne, d'entrée au catalogue, type de carnet, adresse et pays de production, ISSN, etc.), informations disponibles sur le catalogue d'*Hypotheses.org* et obtenues par *scraping*.

Nous avons au final traité 389 089 billets écrits par 12 893 auteurs différents (c'est-à-dire avec des logins différents) sur 3 540 blogs. Ce corpus est une photographie de la plateforme à un instant donné. Nous pensons cependant qu'il offre une bonne visibilité de la nature et de la progression du *blogging* académique en SHS, en France tout particulièrement, progression quantitative du nombre d'auteurs mais également qualitative en considérant les types de carnets ouverts et les thématiques traitées.

Certains carnets créés sur la plateforme sont absents de notre corpus, soit parce qu'ils ne respectent pas encore les trois conditions d'entrée au catalogue (posséder au moins trois billets, témoigner d'une certaine structuration et disposer d'une page « Crédits »), soit parce qu'ils ont disparu avant la formation de nos corpus. Mais une comparaison avec

les bases de données officielles[1] pour les carnets disposant d'un ISSN permet de considérer ces « trous » comme négligeables.

Les logins des auteurs sont très souvent proches de leur véritable nom. Dans quelques cas (Rédaction, Administrateur, etc.), ils sont utilisés indifféremment par différentes personnes dans différents carnets et ne peuvent être distingués. Ces logins ne seront pas pris en compte dans la typologie des auteurs. De la même manière, nous ne considérons pas les pseudonymes partagés (par exemple, le login « Alexandre Hobeika et Étienne Ollion ») dans cette même typologie[2]. La longévité de la plateforme nous incite à penser que certains chercheurs ont pu utiliser plusieurs logins au cours du temps ou selon les carnets. Il n'en reste pas moins que la mise à jour des logins d'auteur est répercutée sur chaque billet et que nous pouvons assez finement suivre la production de chaque login.

Nous avons utilisé pour cette étude une approche quantitative de traitement de corpus en utilisant la suite Jupyter/Anaconda pour le traitement des données. Cela permet de conserver une trace des traitements successifs effectués sur les données et de reproduire éventuellement la méthodologie sur des corpus remis à jour ou redimensionnés.

3. Une première approche par les types de carnets

À partir de ce corpus, nous avons étudié l'évolution dans le temps de la publication sur la plateforme ; nous sommes parties pour tous nos calculs de la date de mise en ligne du carnet au catalogue. La croissance est continue et rapide entre 2008 et 2020 (OpenEdition team, 2021).

Dans cette étude, nous choisissons de garder toutes les langues et de faire une analyse macro de la plateforme. Les auteurs sur la plateforme *Hypotheses.org* peuvent écrire dans un seul carnet (88,6 % des auteurs), deux carnets (7,8 %), trois carnets (2 %) ou plus (1,6 %).

Les responsables peuvent choisir, s'ils le désirent, à la création de leur carnet de lui attribuer un type. Ce dernier est à choisir parmi les 17 types proposés par OpenEdition : carnet de thèse, de vulgarisation, de chercheurs, de projet de recherche, de master, etc. Les types évoluent dans le temps et certains ont été ajoutés au fil de l'eau. Le contenu qu'ils sont censés receler n'est pas décrit. Ils proposent donc une variété importante de style d'écrits et de participations, le créateur d'un carnet pouvant interpréter les intitulés à sa guise. Ainsi un carnet de master peut être

1. — https://portal.issn.org/ et le site de la BNF.

2. — Nous avons repéré 14 auteurs problématiques, dont nous avons supprimé les billets. Le nouveau corpus contient 380 299 billets publiés sur 3 521 carnets différents et 12 879 auteurs.

écrit par un seul enseignant ou des promotions successives d'étudiants, un carnet de structure de recherche peut être alimenté par un *community manager* isolé aussi bien que par une équipe rédactionnelle nombreuse.

De ces types, nous ne pouvons déduire ni la taille ni la forme des billets. En utilisant une nouvelle variable (l'entropie), nous essayerons de voir s'il est possible d'anticiper plus finement le contenu de ces carnets.

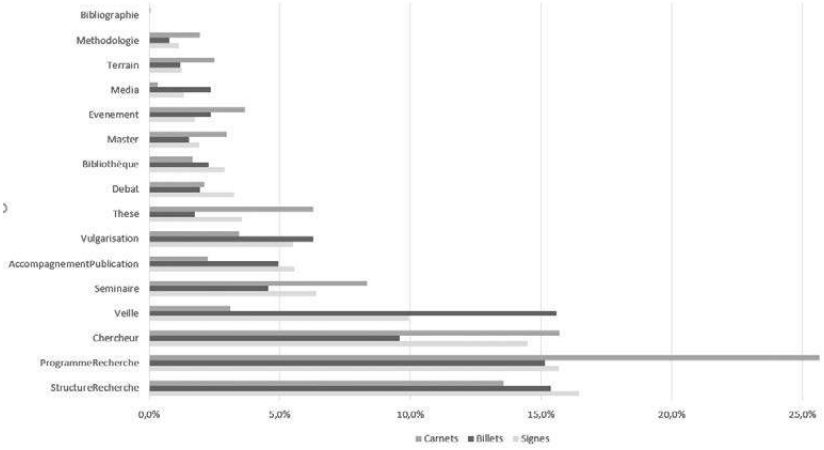


Figure 1 : Effectif et production des carnets par type (nombre de carnets, nombre de billets, nombre de signes) sur *Hypotheses.org*

La figure 1 ci-dessus permet de voir comment la production en nombre de carnets, de signes et de billets se répartit sur chacun des types de carnets et de mesurer leur importance relative sur la plateforme. Par exemple, les carnets de thèse représentent 3,5 % de la production en nombre de signes pendant que les carnets des structures de recherche pèsent de leur côté 16,5 %. Certains types de carnets semblent générer des billets plus longs ou plus courts en moyenne que les autres. Les carnets de chercheurs représentent 9,6 % des billets mais 14,5 % du contenu en nombre de signes. À l'inverse, les carnets de veille représentent 15,6 % des billets de blog, mais seulement 10 % du contenu en nombre de signes.

Le type « carnet des programmes de recherche » est ainsi placé en seconde position pour la production en nombre de signes, non pas parce que ces carnets produisent beaucoup de contenus, mais parce que ce type regroupe le plus grand nombre de carnets (25 %). Quand nous considérons les types, il faut également prendre en considération la dimension temporelle du corpus. Les carnets des programmes de

recherche sont ouverts régulièrement mais actifs uniquement pendant la durée du projet alors que les carnets de structure de recherche peuvent être à l'inverse alimentés sur de longues périodes. Ce n'est alors pas tant la nature des billets qui change mais tout simplement la durée de vie d'un type de carnet.

La grande diversité des carnets classés dans un type ne permet pas d'aller plus loin. Tout au plus pouvons-nous constater que le nombre de signes moyen par carnet diminue pour tous les types à partir de 2017.

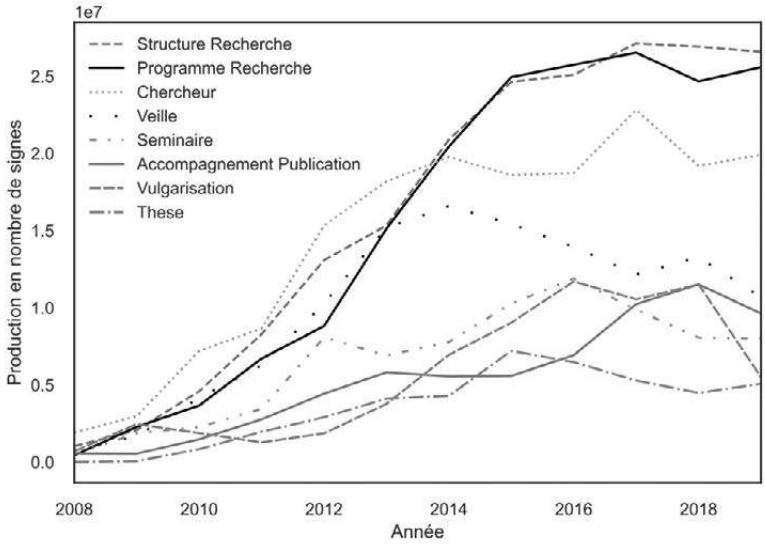


Figure 2 : Évolution de la production (en signes) pour les 8 types de carnets les plus prolifiques

Il est difficile de tirer des conclusions et des tendances nettes de l'évolution de la production dans chaque type de blogs en ce qui concerne l'écriture à plusieurs mains (cf. Figure 2). Ces derniers, nous l'avons vu, ont évolué au cours du temps, de nouveaux types ayant été proposés au fil de l'eau. L'importance croissante des carnets « programme de recherche » et « structure de recherche », hébergeant a priori plus de contributions collectives, semble claire, mais nous assistons dans le même temps à la baisse de la production des carnets de séminaire, qui sont également des carnets censés être alimentés par plusieurs auteurs.

Un autre fait notable est que les carnets dont les « types » suggèrent qu'ils sont écrits à plusieurs (structure de recherche, programme de recherche) ont une production croissante en nombre de signes alors même que le nombre d'auteurs par carnet qui publient dans ces types diminue. Cela marque sans doute le fait que ces carnets sont progressivement pris en main par des auteurs principaux, sans doute des *community managers*.

Un examen de l'évolution du nombre d'auteurs considérés selon le type de carnet dans lequel ils publient nous apporte deux informations. Une majorité des carnets sur *Hypotheses.org* est écrite par plusieurs auteurs : seulement 1 757 carnets sont rédigés par un seul auteur/login. Cette moitié des carnets (50,3 %) qui proposent des billets rédigés par plusieurs auteurs ouvre leur espace rédactionnel de manière très différente. Il n'est pas possible de les décrire plus précisément sans utiliser une nouvelle variable pour évaluer la dispersion des interventions.

4. Une mesure du partage de la parole sur les carnets : calcul de l'entropie

Nous proposons d'utiliser ici l'entropie pour mesurer le degré de diversité des participations d'un carnet. Cette valeur nous permettra de qualifier la nature d'un carnet en distinguant, par exemple, ceux dans lesquels les contributions sont équilibrées entre différents auteurs, de ceux dans lesquels la parole n'est partagée que par une ou deux personnes.

En mathématique de l'information, l'entropie de Shannon (1948) est une mesure qui permet d'apprécier l'apport informationnel d'un corpus. Celle-ci dépend de la probabilité d'apparition des termes le composant. « Plus un système est composé d'un grand nombre d'éléments différents, plus sa quantité d'informations est grande, car plus grande est son improbabilité de le constituer tel qu'il est en assemblant au hasard ses constituants » (Atlan, 1979). Comme en physique, l'entropie mesure donc le désordre, l'hétérogénéité. L'entropie d'un système se calcule par la formule suivante : (avec i variant de 1 à S) $H = -\sum (Ni/N) \, Log \, (Ni/N)$ où N est le nombre d'éléments du système, S le nombre de modalités (valeurs différentes) que peuvent prendre les éléments et Ni l'effectif de chaque modalité i.

Cette mesure calculée pour chacun des carnets permettra de considérer la répartition du nombre de billets écrits par chacun des auteurs, avec N le nombre total de billets du carnet, S le nombre d'auteurs y contribuant, et Ni le nombre de billets écrits par le i^e auteur.

Ainsi, si tous les billets d'un carnet sont écrits par un seul et même auteur, son entropie sera nulle. À l'inverse, si chaque billet est écrit par

un auteur différent, son entropie sera alors maximale. Il faut noter que l'entropie est une grandeur non normalisée et peut en théorie tendre vers l'infini si, par exemple, un carnet contient un nombre infini de billets, tous écrits par un auteur différent. Dans notre corpus, le carnet le plus prolifique, « Archivalia[3] » contient 43 022 billets. En théorie, son entropie maximale pourrait atteindre 15,4 si tous ses billets étaient écrits par des contributeurs différents. Dans les faits, son entropie est égale à 1,25 du fait qu'une très grande partie des billets est écrite par un seul auteur.

Lorsque les billets sont strictement répartis entre les auteurs, l'entropie progresse selon une croissance logarithmique en base 2. Elle est égale à 1 pour 2 auteurs, à 1,58 pour 3 auteurs, 2 pour 4 auteurs, 2,32 pour 5 auteurs, 3 pour 8 auteurs, 4 pour 16 auteurs, 5 pour 32 auteurs, etc. Lorsque les contributions ne sont pas équilibrées entre les auteurs, les valeurs d'entropie sont plus faibles.

Nous donnons ci-dessous quelques exemples de valeur d'entropie pour des carnets de notre corpus. Celui qui présente la plus forte valeur d'entropie, égale à 7,89, est « *Mondes sociaux* : Magazine des sciences humaines et sociales toulousaines[4] ». Le carnet ouvert en 2013 se présente sous la forme d'un blog de publications. Dans notre corpus, il propose 569 contributions (billets) écrites par 359 auteurs différents. Le plus prolifique a écrit 33 billets, 5 auteurs plus de 10 et la plupart (336 auteurs) 1 ou 2 billets seulement. Un autre exemple de carnet présentant une forte valeur d'entropie (4,33) est « Contrepoints[5] », un carnet de master ouvert en novembre 2019 avec 46 contributions rédigées par 24 auteurs (1 à 3 contributions par auteur).

D'une manière générale, nous pouvons dire que l'ensemble des carnets présentant une forte valeur d'entropie sont des carnets collectifs avec un grand nombre de contributeurs différents et avec un degré de participation assez bien réparti. À l'inverse, les carnets de faible entropie (inférieure à 0,5) sont des carnets dont la parole est particulièrement concentrée par un auteur. « Freakonometrics[6] », carnet présentant des applications d'économétrie et ouvert en 2010 par Arthur Charpentier, en est une bonne illustration. Il a pour valeur d'entropie 0,006. Toutes les contributions (1 838), sauf une, sont écrites par son rédacteur en chef.

Les valeurs d'entropie intermédiaires sont parfois plus difficiles à interpréter, car elles peuvent traduire des réalités différentes. Une valeur d'entropie proche de 1 peut refléter soit des contributions équilibrées entre deux

3. — https://archivalia.hypotheses.org/
4. — https://sms.hypotheses.org/
5. — https://mastersts.hypotheses.org
6. — https://freakonometrics.hypotheses.org

auteurs, soit un carnet comprenant un plus grand nombre d'auteurs, avec un contributeur principal concentrant autour de 80 % de la participation, puis un auteur secondaire contribuant à hauteur de 10 % et le reste partagé entre un plus grand nombre d'auteurs. Le carnet « Forum interdisciplinaire des sciences numériques du texte[7] » ouvert en novembre 2018 propose ainsi un carnet collaboratif entre deux chercheurs (entropie 0,97).

5. Catégorisation des carnets basée sur l'entropie

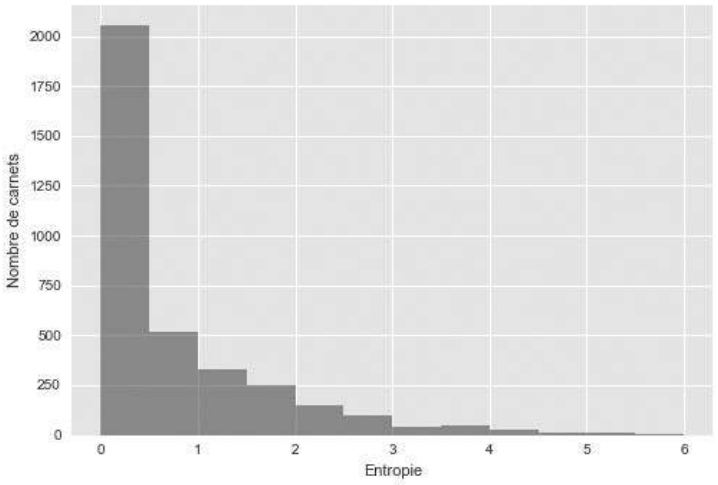


Figure 3 : Distribution des carnets selon leur valeur d'entropie

La figure 3 propose la répartition des carnets en fonction de leur valeur d'entropie. Nous choisissons de fixer les deux seuils de 0,5 et 2 pour classer les carnets. Ces seuils ont été choisis d'un point de vue théorique de l'entropie et après analyse qualitative des carnets, car ils semblaient les plus à même de donner des classes cohérentes de blogs, même si certains d'entre eux sortent de la norme.

Nous considérerons donc trois catégories de carnets : ceux ayant une entropie inférieure à 0,5 qui sont plutôt des carnets presque mono-signataires ou « personnels » dont la plupart des contributions sont produites par un signataire principal ; ceux ayant une entropie comprise entre 0,5

7. — http://infoditex.hypotheses.org

et 2 et qui sont soit le fruit d'une collaboration entre un faible nombre d'auteurs (2 à 4), soit rassemblent un plus grand nombre de personnes autour de quelques auteurs très productifs comme « Archivalia » cité plus haut. Nous les appellerons carnets « collaboratifs ».

Enfin, les carnets ayant une entropie supérieure à 2 seront appelés carnets « collectifs ». Ces derniers sont souvent de véritables publications qui possèdent une équipe de rédaction, parfois même avec des procédures de soumission avec relecture. Des blogs pédagogiques (carnet de masters) ou des blogs de réflexion thématiques se trouvent aussi dans cette catégorie.

Catégorie de carnets	Nombre de carnets	En pourcentage
Carnets personnels	2 057	58,11 %
Carnets collaboratifs	1 101	31,1 %
Carnets collectifs	382	10,79 %

Tableau 1 : Répartition des carnets par catégorie (calculé selon l'entropie)

La plateforme héberge en majorité des carnets personnels (58,1 %) qui, pour 85 % d'entre eux, ne proposent aucune contribution extérieure (soit 49 % des carnets du corpus). Les carnets collaboratifs représentant le deuxième groupe en termes de nombre de carnets sur la plateforme (31,1 %) et les carnets collectifs arrivent en troisième position (19,79 % des carnets).

En termes de production, les carnets personnels éditent 42,7 % des contenus de la plateforme en nombre de signes, les carnets collaboratifs 31,1 % et les carnets collectifs 25,4 %. Sans surprise, si les carnets personnels représentent la majorité des carnets de la plateforme, les carnets à plusieurs mains (carnets collaboratifs et collectifs) éditent la majorité des contenus en nombre de billets et nombre de signes (57,2 %). Les carnets collectifs ne représentent que 10,79 % des carnets au catalogue, mais éditent 20,1 % des billets et 25,4 % des contenus. Leurs billets sont plus longs en moyenne que ceux des autres catégories comme le montre la Figure 4.

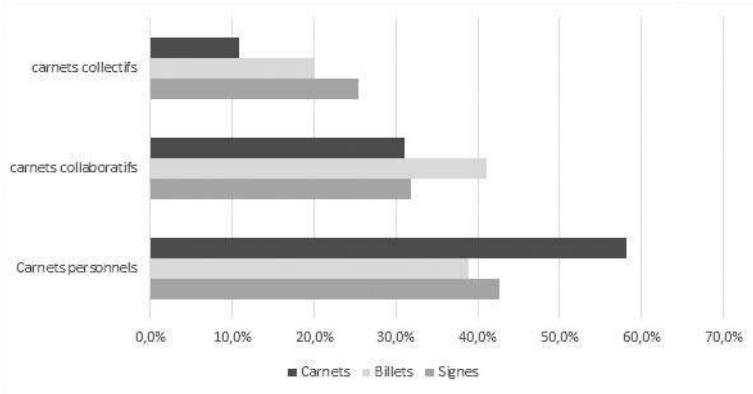


Figure 4 : Effectif et production des carnets par catégorie (nombre de carnets, nombre de billets, nombre de signes)

La production moyenne sur les carnets personnels paraît plus faible en fréquence mais avec une taille moyenne des billets assez longue. Les carnets collaboratifs se distinguent par des billets nettement plus courts que les deux autres catégories et une production relativement importante. Nous pourrions, à ce stade, faire un lien entre la forme et le fond. Les billets publiés dans les carnets collaboratifs semblent, parce qu'ils sont plus courts, s'apparenter plus souvent à des articles de veille ou d'annonce d'événements. Les articles plus longs pourraient représenter des articles de fond nécessitant plus de travail et s'apparentant plus à une publication scientifique produite pour une revue en ligne.

Comme le montre le tableau 2, il y a bien entendu un lien entre les types de carnets et les catégories affectées selon l'entropie, un test de Khi-deux[8] sur ces variables le prouve, avec un risque d'erreur inférieur à 0,001 % de rejeter l'hypothèse d'indépendance. Les carnets de « chercheurs » sont plus souvent des carnets personnels (82,5 %) que les carnets de « structure de recherche » (41,3 %). Les carnets « de thèse » sont presque toujours des espaces à une voix. Réaliser des croisements systématiques entre les catégories que nous avons mises à jour et les « types » proposés par la plateforme permet de repérer les fonctionnalités très différentes de carnets qui appartiennent pourtant aux mêmes types. Ainsi certains carnets dont le type suggère la multi-auctorialité (carnet « de structure de recherche », carnet « de programme de recherche ») et qui ont une entropie proche inférieure à 0,5 sont probablement tenus par des *community managers*.

8. — Pour ce test, les types « carnet de média » et « carnet de bibliothèque », dont les effectifs théoriques sont inférieurs, à 5 sont regroupés sous la catégorie « Autre ». Les carnets sans type déclaré ne sont pas comptabilisés dans les tableaux.

	Carnets personnels		Carnets collaboratifs		Carnets collectifs	
Type/ Catégorie	Effectifs observés	Effectifs théoriques	Effectifs observés	Effectifs théoriques	Effectifs observés	Effectifs théoriques
Bibliothèque	24	33,71	11	17,50	22	5,79
Chercheur	456	327,08	81	169,78	16	56,14
Débat	33	42,59	27	22,10	12	7,31
Événement	75	76,89	40	39,91	15	13,20
Master	46	61,51	26	31,93	32	10,56
Méthodologie	33	40,22	23	20,88	12	6,90
Programme de recherche	472	535,27	366	277,85	67	91,88
Séminaire	158	173,89	103	90,26	33	29,85
Structure de recherche	197	281,54	202	146,14	77	48,33
Terrain	53	50,27	27	26,10	5	8,63
Thèse	213	130,12	6	67,54	1	22,34
Veille	62	63,88	31	33,16	15	10,96
Vulgarisation	71	70,38	32	36,53	16	12,08
Accompagnement à la publication	39	44,95	29	23,33	8	7,72
Autre	8	7,69	3	3,99	2	1,32

Tableau 2 : Calcul du Khi-deux : effectifs observés/effectifs théoriques

6. Typologie des auteurs basée sur leur participation à différentes catégories de carnets

Un des objectifs de cet article est de proposer une typologie originale des carnetiers sur *Hypotheses.org*, basée sur l'importance de leur investissement sur la plateforme et sur la nature des carnets qu'ils investissent.

Une première évaluation quantitative montre que les auteurs qui publient sur un carnet collectif représentent 52,7 % des carnetiers, ceux qui publient dans un carnet collaboratif 30,9 % et ceux sur un carnet personnel 16,4 %. Du point de vue du nombre d'auteurs engagés, l'écriture sur *Hypotheses.org* semble donc être une pratique sociale principalement collective. Distinguer des auteurs uniquement sur le fait qu'ils publient dans telle ou telle catégorie de carnets n'a pas grand sens, car l'implication des uns et des autres est très variable, aussi bien en nombre de billets produits au cours du temps qu'en termes de longueur des articles proposés.

Les auteurs peuvent également alimenter plusieurs carnets appartenant à des catégories différentes. Deux variables supplémentaires nous semblent donc importantes à prendre en considération pour pouvoir typologiser les comportements des auteurs sur la plateforme : celle de la dispersion des participations sur la plateforme et celle du nombre des interventions.

En fait, à partir des analyses de présence des auteurs dans les carnets de la plateforme, nous distinguerons trois groupes d'auteurs aux comportements distincts.

Le premier rassemble les auteurs « invités » (4 615 auteurs). Ceux-ci se caractérisent par le fait qu'ils ne contribuent qu'une seule fois (un seul billet) sur la plateforme dans notre corpus. Ils ne sont ni animateurs ni créateurs des carnets dans lesquels ils publient. Cela ne préjuge en rien de leurs futures interventions. Nous considérons qu'ils interviennent à la demande des animateurs principaux des carnets et qu'ils ne discutent pas du contenu éditorial du carnet.

Le deuxième groupe réunit les auteurs « occasionnels » (3 485 auteurs). Ces derniers n'ont écrit qu'entre 2 et 5 billets sur la période considérée. Ils peuvent être fidèles à un carnet ou s'exprimer sur plusieurs. Là encore nous pouvons estimer leur intervention sur l'orientation éditoriale du carnet comme faible, puisque leur investissement sur la plateforme est relativement faible.

Le troisième groupe regroupe les auteurs « investis » (1 179 auteurs). Ils publient 6 articles ou plus sur la plateforme et constituent un groupe relativement hétérogène. Nous les considérerons donc en tenant compte de la catégorie de carnets sur laquelle ils publient.

Un premier sous-groupe est constitué d'auteurs « solitaires » qui contribuent uniquement à l'animation de leur(s) carnet(s) personnel(s). Ils sollicitent parfois très ponctuellement la participation d'autres auteurs, mais la grande majorité d'entre eux (85 %) publient sur des carnets à entropie nulle en alimentant leurs carnets de manière totalement indépendante sans aucun lien apparent avec d'autres membres.

Un second sous-groupe réunit les auteurs « partenaires » (3 183 auteurs) qui n'alimentent que des carnets partagés. Les partenaires peuvent ne publier qu'un billet par carnet ou être relativement fidèles à plusieurs carnets et y intervenir régulièrement ou occasionnellement.

Enfin, un troisième sous-groupe (417 auteurs) rassemble les auteurs « polymorphes ». Ces derniers sont caractérisés par le fait qu'ils publient sur un ou plusieurs carnets personnels et qu'ils s'engagent également sur des carnets partagés (collaboratifs ou collectifs). Les auteurs polymorphes sont donc présents à la fois sur des carnets à forte et à faible entropie.

Les trois groupes principaux d'auteurs, en nombre, sont plus ou moins équivalents (36 % d'invités, 27 % d'auteurs occasionnels et 37,1 % d'auteurs investis). Les auteurs « investis » cependant, à eux seuls, proposent 91 % de la production d'*Hypotheses.org*.

Chez les auteurs « investis », la répartition des sous-groupes est déséquilibrée. Les auteurs « partenaires » constituent le groupe majoritaire, suivent les auteurs « solitaires » puis les auteurs « polymorphes » qui ne représentent que 3,2 % des auteurs.

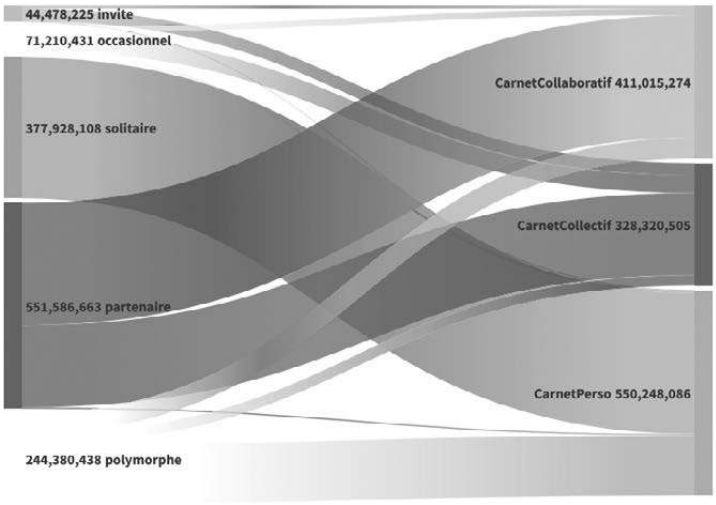


Figure 5 : Répartition de la production selon le type d'auteurs dans les catégories de carnets (nombre de signes)

Sur la plateforme *Hypotheses.org*, le nombre de billets dans les carnets partagés (collaboratifs et collectifs) est plus important que celui des carnets personnels, nous l'avons vu dans la figure 4. Ces billets sont principalement écrits par des auteurs « partenaires » (cf. Figure 5). Les auteurs « invités » et « occasionnels » interviennent majoritairement dans les carnets collectifs alors que les auteurs « polymorphes » alimentent avant tout leurs carnets personnels.

Les auteurs « polymorphes » sont les plus actifs sur la plateforme : ils proposent 18,9 % des contenus (pour 3,2 % des auteurs). La majorité de leur production se trouve dans les carnets personnels qu'ils complètent avec des écrits dans les carnets collaboratifs et collectifs. Un auteur « polymorphe » moyen, rapporté à un auteur « partenaire », publie presque trois fois plus. Les auteurs « partenaires » (24,7 % des auteurs)

produisent 42 % du contenu. Nous calculons de la même manière un ratio de 1,5 rapporté à un auteur « solitaire » moyen. Un auteur moyen solitaire publie de son côté deux fois plus qu'un auteur partenaire. Les « solitaires » (7 %) proposent 30 % des articles.

Les invités, qui représentent 36 % des auteurs, ne produisent que 3,4 % du contenu d'*Hypotheses.org*. Tout comme les auteurs occasionnels (27 % des auteurs), qui publient finalement assez peu (5,5 % du contenu). Ils interviennent majoritairement dans les carnets collectifs, qui s'apparentent souvent à des publications. Nous avons vu que dans ces carnets, les billets étaient plus longs en moyenne que dans les autres catégories. Il s'agit donc de participation qui s'apparente à une contribution dans un article de revue.

Les auteurs « occasionnels », qui publient moins de 6 billets sur la plateforme, et les invités, qui sont principalement des contributeurs de carnets collaboratifs et collectifs, publient en fait peu de contenu, 8,9 %, alors qu'ils représentent 63 % des auteurs.

Conclusion

Une analyse de l'évolution dans le temps des catégories de blogs (personnel, collectif, collaboratif) montre que le poids des carnets collaboratifs, relativement aux autres, continue de croître en nombre de signes et de billets. Ceci montre, selon nous, que sur *Hypotheses.org*, le *blogging* s'envisage de plus en plus comme une pratique d'écriture à plusieurs qui s'éloigne de ce qui constituait le *blogging* à sa naissance, les carnets s'apparentant de plus en plus à des sites web ordinaires, vitrines de la communication de l'actualité de la recherche (annonces de séminaires, de colloques).

L'analyse nous a permis de différencier et quantifier les différentes pratiques des auteurs en considérant l'importance et les modalités de leur investissement sur les différents carnets. Conformément aux lois rencontrées en sciences de l'information (Lotka, 1926), nous observons que la production des contributeurs est très dispersée.

Les auteurs les plus investis dans la plateforme en termes de production peuvent être décomposés en trois groupes : les partenaires, les solitaires et les polymorphes. Ces trois groupes se comportent relativement différemment, certains privilégiant l'écriture à plusieurs, d'autres les espaces personnels, les troisièmes s'essayant à tout. En règle générale, les auteurs qui animent un carnet personnel ont une activité supérieure à celle des auteurs publiant principalement sur des carnets partagés.

Nous avons montré que les pratiques très différentes hébergées par la plateforme *Hypotheses.org* sont quantifiables. Cette photographie, réalisée à un instant donné, peut être précisée en prenant en compte des périodes plus réduites. Les pratiques des auteurs ne sont pas figées : un auteur « solitaire » peut devenir « polymorphe » avec le temps. Il semblerait intéressant d'essayer de déterminer ce qui conditionne une carrière d'auteurs polymorphes ou d'auteurs partenaires.

Nous pouvons aussi sans doute distinguer les pratiques des auteurs selon un gradient de participation. La coopération se signale quand les auteurs participent à une œuvre commune, sans remettre en cause les objectifs communs, décidés en amont. Parler de collaboration nécessite la présence d'un investissement plus lourd. Nous avions choisi ces termes et distingué blogs collectifs et collaboratifs parce que « l'activité de collaboration suppose la possibilité de discuter des règles » (Broudoux, 2014). En ce sens, la plupart des interactions entre les auteurs sur *Hypotheses.org* relève plus de la coopération ou de la simple contribution que de la collaboration. C'est évident chez les auteurs « invités », « occasionnels » ou « isolés », c'est plus délicat à discerner dans le cas des auteurs « polymorphes » et « partenaires » dont l'investissement est variable. Une analyse plus fine par sous-corpus serait nécessaire pour en savoir plus.

Nous pouvons en tout cas, à ce stade, repérer deux niveaux de coopérations des auteurs : celle intra-carnet signalée par la co-présence au sein d'un même carnet ; celle inter-carnet, plus invisible, qui consiste à participer à la construction de la plateforme elle-même. La question des conditions de travail du chercheur[9], des outils méthodologiques[10] ou des raisons de s'investir dans le blog[11] par exemple sont au cœur de ces deux niveaux de coopération. Ces carnets peuvent concrétiser des groupes informels ou des relations institutionnelles aussi bien que générer une véritable écriture en réseau dans une communauté en formation.

En 2013, Marin Dacos rappelait déjà combien l'importance des infrastructures était sous-évaluée dans la recherche en SHS (Dacos, 2013), les plateformes conversationnelles jouant selon lui un rôle déterminant dans l'émergence de nouvelles idées. Finalement, la plateforme *Hypotheses.org*, comme infrastructure de recherche, influence les dynamiques collaboratives entre communautés et aide à la structuration de réseaux préexistants. Retracer l'évolution des carnets partagés sur certaines thématiques, par exemple sur la question des outils de recherche ou

9. — https://reflexivites.hypotheses.org/a-propos
10. — http://outiquanti.hypotheses.org
11. — https://books.openedition.org/oep/174?lang=fr

sur les humanités numériques, permettrait de vérifier cette hypothèse. L'étude des co-présences sur les carnets au cours du temps aiderait à mieux comprendre comment les réseaux se sont constitués. Enfin, sortir de l'écosystème d'*Hypotheses.org* et étudier la vie de ces communautés repérées sur la plateforme affinerait la mesure du rôle et du poids de cette plateforme de *blogging*.

Bibliographie

Atlan, H. (1979). *Entre le cristal et la fumée. Essai sur l'organisation du vivant*. Seuil.

Blanchard, A. (2016). Quand la culture scientifique s'affranchit sur le web : l'exemple des blogs de science en français (2003-2014). P. Poirrier (dir.), *Histoire de la culture scientifique en France. Institutions et acteurs*. Éditions universitaires de Dijon (EUD), 1-16. https://shs.hal.science/halshs-01242707/document

Broudoux, E. (2014). Participer, coopérer et collaborer en milieu scientifique… *Les Cahiers de la SFSIC*, 9, 178.

Broudoux, E. (2006). Autorités énonciatives et espaces de publication et de référencement [Billet]. *UrfistInfo*. https://urfistinfo.hypotheses.org/324

Brown, S. A. (2021). The Role of the Editor of an Academic Publication Blog. *Journal of Scholarly Publishing*, 52(4), 199-211. https://doi.org/10.3138/jsp.52.4.0

Cardon, D. et Delaunay-Téterel, H. (2006). La production de soi comme technique relationnelle. *Réseaux*, 138 (4), 15-71. https://doi.org/10.3917/res.138.0015

Casanova, N. (2022, 12 janvier). Supprimer son carnet : Enjeux et modalités [Billet]. *La Maison des carnets*. https://maisondescarnets.hypotheses.org/5723

Dacos, M. (2019, 8 janvier). *Hypothèses* 2008-2018. De la coupe aux lèvres, et des lèvres à la recherche sur les usages. 10 ans déjà, 20 ans bientôt. *Zenodo*. https://zenodo.org/record/2533845

Dacos, M. (2013, 27 octobre). Un angle mort ? Les infrastructures pour les SHS en général et pour les carnets de recherche en particulier [Billet]. *Blogo-numericus*. https://bn.hypotheses.org/11261

Debaz, J. (2010, 8 décembre). Hypothesosphère [Billet]. *Socio-informatique et argumentation*. https://socioargu.hypotheses.org/1921

Deseilligny, O. (2010). Le blog intime au croisement des genres de l'écriture de soi. *Itinéraires. Littérature, textes, cultures*, 2, 73-82. https://doi.org/10.4000/itineraires.1985

Gardiner, A., Sullivan, M. et Grand, A. (2018). Who Are You Writing for? Differences in Response to Blog Design Between Scientists and Nonscientists. *Science Communication*, 40(1), 109-123. https://doi.org/10.1177/1075547017747608

Gunthert, A. (2010). Why Blog ? M. Dacos (dir.), *Read/Write Book. Le livre inscriptible*. OpenEdition Press, 167-171. http://books.openedition.org/oep/174

Herring, S. C., Scheidt, L. A., Bonus, S. et Wright, E. (2004). Bridging the gap : A genre analysis of Weblogs. *Proceedings of the 37th Annual Hawaii International Conference on System Sciences, 5-8 January 2004*. Institute of Electrical and Electronics Engineers, 11. https://doi.org/10.1109/HICSS.2004.1265271

Jarreau, P. B. et Porter, L. (2017). Science in the Social Media Age:

Profiles of Science Blog Readers. *Journalism & Mass Communication Quarterly*, 95(1), 142–168. https://doi.org/10.1177/1077699016685558

Kjellberg, S. (2010). I am a blogging researcher : Motivations for blogging in a scholarly context. *First Monday*, 15(8). https://doi.org/10.5210/fm.v15i8.2962

Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, 16(12), 317-323.

Marht, M. et Puschmann, C. (2014). Science blogging: An exploratory study of motives, styles, and audience reactions. *Journal of Science Communication*, 13(3). https://jcom.sissa.it/archive/13/03/JCOM_1303_2014_A05

Mayeur, I. (2019). *Recherches en cours. Discours de savoir numériques et actualité scientifique dans les carnets de recherche de la plateforme* Hypothèses. Thèse de doctorat, Université de Liège. http://hdl.handle.net/2268/232256

Mercklé, P. (2019, août). Hypothèses 2. *10 ans d'*Hypotheses. https://halshs.archives-ouvertes.fr/halshs-02270049

Morelli, P. (2007). Blogs et médias, quels rapports aujourd'hui : essai de typologie. *Actes du colloque EUTIC Enjeux et usages des Technologies de l'information et de la communication, Novembre 2007, Athènes (Grèce)*, 169-178.

OpenEdition team. (2021, 7 avril). *Hypothèses* : histoire d'une collaboration internationale [Billet]. *L'Édition électronique ouverte*. https://leo.hypotheses.org/17463

Poupardin, E. et Prime-Claverie, C. (2023). Carnetiers et blogueurs de sciences : analyser des modèles et des communautés en construction. A. Mahé et I. Mayeur (dir.), *Communication scientifique et science ouverte : opportunités, tensions et paradoxes. Actes du colloque « Document numérique et société »*. De Boeck Supérieur, 27-46.

Poupardin, E. et Faury, M. (2018). *Hypotheses* : l'inscription d'une pratique de communication dans l'activité de recherche. *Revue française des sciences de l'information et de la communication*, 15. https://doi.org/10.4000/rfsic.4877

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379-423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Zou, H. et Hyland, K. (2020a). Academic blogging : Writers' views on interacting with readers. *Iberica*, 39, 267-294.

Zou, H. et Hyland, K. (2020b). "Think about how fascinating this is": Engagement in academic blogs across disciplines. *Journal of English for Academic Purposes*, 43, https://doi.org/10.1016/j.jeap.2019.100809