

Anonymiser ou non les données sensibles d'un corpus numérique en sciences du langage

L'intégrité contextuelle comme solution théorique et méthodologique

Eugénie Pereira Couttolenc

Laboratoire Education Discours et Apprentissages (EDA) – Université Paris Cité

Numéro spécial 2 / 2023 - T3



RÉSUMÉ *La manipulation et la diffusion de données recueillies sur le web s'accompagnent nécessairement de questions juridiques, éthiques, déontologiques et pratiques. De fait, s'il est vrai que le numérique fournit un accès facile à de nombreuses productions authentiques (Demonet, 2018 : 283), il est aussi juste d'affirmer que cet accès n'est pas toujours ouvert (Marcoccia, 2016 : 43). Ainsi, depuis la loi n°2004-801, il est requis de rendre anonyme toutes les informations permettant une identification directe ou indirecte des personnes mentionnées dans une étude (Huygbe, Cailly et Oppenchain, 2018 : 160-162). Or, les difficultés à supprimer les éléments à caractère personnel dans un corpus numérique sont nombreuses : pour éviter toute probabilité de recoupement, doit-on substituer aux données sensibles des étiquettes indiquant [nom], [toponyme], [date], [photo] ? Doit-on appliquer un système de « brouillage » (Op. Cit. : 163) en remplaçant les pseudonymes ? L'investigabilité des énoncés (Merzeau, 2009 : 27 ; Paveau, 2017 : 134) ne rend-elle pas stérile ce type d'entreprise (Marcoccia, 2016) ?*

Cette communication partage une réflexion menée lors de l'exploitation de documents plurisémiotiques constituant le corpus de blogs voyage d'une recherche récente menée en sciences du langage. L'ensemble étudié, composé de 8 pages d'accueil, de 164 billets, de 110 images, de 9 vidéos, de 345 commentaires et de 536 hyperliens, présente ainsi un grand nombre de renseignements et de photographies exposant l'identité des animateur.trice.s des blogs, de leurs proches ou encore celle de leurs lecteur.trice.s. Après avoir rappelé quelques principes clés des textes législatifs parus ces dernières années (Carvalho, 2019 : 307) et exposer les moyens mis en œuvre pour les appliquer, on justifie de notre choix à faire de l'intégrité contextuelle (Nissenbaum, 2010 et 2011 ; Paveau, 2017 : 229-232) le pivot de toutes réflexions éthiques accompagnant le traitement de données prélevées sur le web.

MOTS-CLÉS • *Anonymisation - Données numériques - Intégrité contextuelle – Déontologie - Sciences humaines et sociales*

ABSTRACT *The manipulation and dissemination of data collected on the Web are closely linked to legal, ethical, deontological and practical issues. Indeed, if it is true that digital world allows easy access to many authentic productions (Demonet, 2018: 283), it is also fair to consider that these data are not always open (Marcoccia, 2016: 43). Thus, since Law No. 2004-801, it is required to anonymize all information allowing the direct or indirect identification of the people cited in a study (Huygbe, Cailly and Oppenchain, 2018: 160-162). However, there are many difficulties in deleting personal elements from a digital corpus: to avoid any risk of cross-checking, should sensitive data be replaced by labels indicating [name], [toponym], [date], [photo]? Should we apply a system of "disguise" (Op. Cit.: 163) by*

replacing pseudonyms? Do the criteria for the statements investigability (Merzeau, 2009: 27; Paveau, 2017: 134) make this type of enterprise sterile (Marcoccia, 2016)?

This communication shares a reflection carried out during the discursive analysis of plurimodal documents constituting a corpus of travel blogs. The whole set, composed of 8 home pages, 164 posts, 110 images, 9 videos, 345 comments and 536 hypertext links, thus presents a large amount of information and photographs exposing the identity of the webmasters, their relatives or that of their readers. After recalling some key principles taken from legislative texts published in recent years (Carvalho, 2019: 307) and outlining the means implemented to apply them, we justify our choice to make contextual integrity (Nissenbaum, 2010 and 2011; Paveau, 2017: 229-232) the pivot of all ethical reflections accompanying web data processing.

KEYWORDS • Anonymization - Digital data - Contextual integrity – Ethics - Human and social sciences

Introduction

L'exploitation de corpus numériques plurisémiotiques expose des informations personnelles et des images aux contenus jugés parfois délicats, preuves en sont les noms et les prénoms figurant dans les commentaires publiés sur un forum ou encore les photographies d'enfants exposées pour illustrer la qualité d'un séjour touristique. En amont et au cours de notre étude sur les mises en scène de soi des créateurs et créatrices de contenu animant, entre autres, des blogs voyage, une réflexion à la fois éthique et pratique a accompagné continuellement le traitement de ces données sensibles. Dans la première partie de cet article, une fois introduit le contexte théorico-méthodologique de la recherche, on présente les limites des recommandations basées sur la distinction public et privé avant d'exposer les tentatives infructueuses mises en place pour anonymiser les éléments du corpus. Puis, on expose les raisons qui nous ont poussés à faire le choix d'une démarche déontologique se réclamant de l'intégrité contextuelle, c'est-à-dire d'une réflexion éthique intégrant le contexte de production et de diffusion des données recueillies dans l'appréhension de ces dernières.

Contexte de la recherche

Dans un cadre théorique subordonné à l'analyse du discours (AD), notre questionnement reposait sur les moyens déployés par un individu, a priori inconnu du grand public, pour, d'une part, bénéficier d'une audience dans un champ d'activité sociale donné, en l'occurrence le tourisme, et, d'autre part, pour disposer d'une influence sur les agissements des internautes (Pereira Couttolenc, 2022). En AD, on cherche à comprendre les relations entre les textes (entendus ici au sens large) et la société dans laquelle ils sont produits (Maingueneau, 2014 : 13) et qu'ils façonnent en retour. Pour ce faire, le praticien du discours caractérise usuellement des événements discursifs en les rattachant aux lieux sociaux¹ dans lesquels ils sont apparus (Maingueneau, 2014 : 43) accordant, dès lors, aux conditions d'émergence de ces discours une place importante. Ainsi, pour étudier la présentation de soi (Amossy, 1999 et 2015) telle qu'elle est construite et projetée dans les productions des blogueur.euse.s – celle-ci se façonnant au diapason des représentations que le.la locuteur.trice se fait de son auditoire, de ce dont il.elle parle, du

1 La notion de lieu social est à saisir au sens large à l'instar de l'usage qu'en fait Maingueneau : « [1]a notion de "lieu social" ne doit pas être prise dans un sens trop immédiat : ce lieu peut être une position dans un champ symbolique (politique, religieux...) » (Maingueneau, 2012 : 5).

domaine dans lequel il.elle intervient² et en relation co-constitutive avec la nature des données qui la composent – on a considéré des éléments d’ordre langagier, sémiotique et technique. De fait, méconnu.e.s du grand public, les influenceur.euse.s voyage ne peuvent s’appuyer que sur les dimensions iconiques, techniques, quantitatives et sur le caractère authentique de leurs témoignages pour donner de la crédibilité à leurs interventions.

Composition du corpus

Le corpus de travail a réuni divers documents tels que des sections de présentation des webmestres (« À propos », « Qui-suis-je ? »), des pages d’accueil des blogs, des énoncés, des photographies, des cartes, des dessins, des vidéos, des commentaires (ceux des blogueur.euse.s et ceux postés par leurs lecteur.trice.s) et des hyperliens (soit un ensemble total de 8 pages d’accueil de blogs, 164 billets, 110 images, 9 vidéos, de 345 commentaires et de 536 hyperliens). Certains de ces éléments plurimodaux comportaient du contenu sensible. Sensible, d’une part, du point de vue du cadre prévu par la loi, puisque des photographies de mineurs étaient parfois exposées. Sensible, d’autre part, dans la mesure où ces clichés apparaissaient comme essentiels à la construction de la légitimité à dire de l’auteur.e. C’est notamment le cas de Christine, blogueuse animant le site de Maman Voyage³, qui dispense des conseils pour voyager en famille sur la base de ses expériences en tant que mère de famille et qui met en scène systématiquement ses enfants dans les billets publiés. C’est aussi le cas des informations contenues dans les sections « Qui suis-je ? » et « À propos » qui non seulement concourent à la construction d’un ethos de témoin authentique mais comportent également des renseignements permettant d’identifier les lieux de vie ou encore l’état-civil des lecteur.trice.s des blogs ou de leurs auteur.e.s. Par exemple, l’introduction du site The Daydreameuse⁴ est accompagnée d’une photo portrait de la blogueuse, de la mention de son prénom, de sa ville d’origine, de son parcours universitaire, des villes qu’elle a visitées, des voyages qu’elle a effectués seule, avec une amie, en amoureux ou en famille. De la même manière, si la plupart des commentateur.trice.s ajoutent leurs impressions, leurs demandes ou encore leurs remerciements au texte du billet en utilisant uniquement leur prénom ou en recourant à un pseudonyme, d’autres utilisent leur identité civile. Notre démarche d’exploitation a donc oscillé entre les contraintes légales encadrant le recueil et le traitement des données, les conditions de production et de diffusion de ces éléments publiés à destination du grand public et les éventuelles réponses que ceux-ci pouvaient apporter à la question de recherche. Pour faire face à l’inconfort ressenti face à la publication de noms et de prénoms ou encore à l’utilisation de visuels montrant des visages d’enfants, nous avons entamé une réflexion éthique dont les objectifs visaient non seulement à respecter le cadre de la loi mais aspiraient surtout à nous forger une norme de

2 Dans les travaux d’Amossy (Amossy, 1999 et 2015), la conception de la notion de *présentation de soi* ne recouvre *que* la mise en scène de la personne à partir de ressources langagières (Amossy, 2015 : 7). Néanmoins, l’auteure rejoint volontiers la proposition d’expansion de la notion proposée par Maingueneau qui stipule que l’ethos ne concerne pas uniquement les énoncés oraux, il vaut pour tout discours même écrit (Maingueneau, 2021 [2013] : 89). Dans ce travail, on désigne par le terme de « présentation de soi en ligne » l’application de ce concept aux éléments composant les mises en scène de soi des internautes intervenant dans les réseaux sociaux du web 2.0 (Pereira Couttolenc, 2022 : 90).

3 <https://www.mamanvoyage.com/a-propos/>, consulté le 29 mars 2023.

4 <https://www.thedaydreameuse.com/about/>, consulté le 29 mars 2023.

manipulation des données qui reflète le plus fidèlement possible notre intégrité de chercheuse en sciences humaines et sociales.

Réflexions éthiques et traitements envisagés

Le traitement de documents accessibles sur le web impose aux chercheur.euse.s d'adopter une démarche réflexive (Barats, 2016 ; Marcoccia, 2016) sur les tenants et les aboutissants d'une éthique de la recherche⁵ (Lambert-Chan, 2012 : 105) appliquée au cadre de leurs travaux. Dans cette section, on fait état des exigences légales actuelles et on décrit les chemins méthodologiques peu féconds que nous avons empruntés pour nous y conformer.

Cadre légal et recommandations actuelles

Si on se reporte à l'article 4 du chapitre 1, correspondant à la modification du règlement général sur la protection des données (RGPD) datant du 23 mai 2018, est considérée comme « données à caractère personnel » toute information se rapportant à une personne physique identifiée ou identifiable « directement ou indirectement, par référence à un identifiant, tel qu'un nom, un numéro d'identification, des données de localisation, un identifiant en ligne, ou à un ou plusieurs éléments spécifiques propres à son identité physique, physiologique, génétique, psychique, économique, culturelle ou sociale » (CNIL, consulté le 4 novembre 2022). Ces informations doivent alors faire l'objet d'un traitement licite, loyal et transparent au regard de la personne concernée. De cette formulation assez générale, on retient qu'il est nécessaire de supprimer tout renseignement pouvant conduire à l'identification des participants d'une étude. Rien n'est cependant précisé (ni clairement ni explicitement) en ce qui concerne les contenus publiés volontairement dans l'espace public des blogs et leur traitement dans le cadre d'études scientifiques.⁶

Selon Marcoccia, il serait possible de calquer pour le numérique le guide des bonnes pratiques de traitement des corpus oraux publié en 2006 aux éditions du CNRS (Marcoccia, 2016 : 42). Selon les préconisations du livret, il faudrait alors s'interroger sur le statut juridique du corpus : contient-il des données personnelles ? Est-il le produit

5 « L'éthique de la recherche vise la protection des êtres humains qui prennent part à la recherche scientifique » (Lambert-Chan, 2012 : 105).

6 Le site de la CNIL répertorie 23 entrées concernant le traitement des données émis sur un blog à la date de rédaction de cet article (15 juin 2023) (<https://www.cnil.fr/fr/recherche/blog>, consulté le 15 juin 2023). On note qu'il est possible de consulter « Les obligations du blogueur », « Supprimer ou anonymiser des données personnelles ou des contributions », « Connaître les informations détenues par un blog / forum » etc. En revanche, aucune information ne vient éclairer le chercheur.euse quant aux droits d'exploitation de données volontairement publiées sur internet dans une perspective de large diffusion. Par ailleurs, au chapitre II du « Règlement général sur la protection des données » (RGPD), l'article 5 – traitant des « principes relatifs au traitement des données à caractère personnel » – mentionne au point 1, alinéa b) que les données à caractère personnel doivent être : « collectées pour des finalités déterminées, explicites et légitimes, et ne pas être traitées ultérieurement d'une manière incompatible avec ces finalités ; le traitement ultérieur à des fins archivistiques dans l'intérêt public, à des fins de recherche scientifique ou historique ou à des fins statistiques n'est pas considéré, conformément à l'article 89, paragraphe 1, comme incompatible avec les finalités initiales (limitation des finalités) » (<https://www.cnil.fr/fr/reglement-europeen-protection-donnees/chapitre2#Article5>, consulté le 15 juin 2023).

d'une création intellectuelle ? L'obtention du consentement des enquêtés, solution recommandée par l'institution, apparaît complexe à mettre en place (Nissenbaum, 2011 : 34-35) compte tenu de la nature particulière des terrains d'observation dans le domaine des sciences humaines et sociales (Marcoccia, 2016). L'Association of Internet Researchers (AoIR), pour sa part, considère que plus les messages sont « publics moins est forte l'obligation de protéger l'intimité et la vie privée des internautes et la confidentialité de leurs discours » (Marcoccia, 2016 : 43), précepte qui fait sens mais qui peut ne pas pleinement satisfaire, les études en ligne relevant difficilement d'une situation ou de l'autre. C'est donc vers l'anonymisation des données que notre choix s'est porté.

Anonymiser les données

Dans un premier élan et en s'inspirant d'une proposition de Marcoccia (Marcoccia, 2016 : 45), on a effacé toutes les références permettant, d'une manière ou d'une autre, de déduire des énoncés l'identité des auteur.e.s. Les noms, les prénoms et la totalité des toponymes ont alors été remplacés par des étiquettes telles que [nom B1], [prénom B1] et [nom de villages] comme on peut en faire l'observation dans les captures d'écran 2 et 3 infra.

Titre du post : [NOM B1] – VIVRE À [NOM DE VILLE],
VOYAGER [NOM DE PAYS]

QUI SUIS-JE ?

[photo]
[photo]
[photo]
[photo]

Vous êtes sur le [nom B1], et [prénom de l'auteure], hé bien c'est moi. Ce blog est né en [date]: j'étais sur le point de partir vivre aux [nom de pays], j'allais démissionner de mon job d'éditrice, j'étais enthousiaste à l'idée de ce gros changement, de ce plongeon vers l'inconnu... J'avais envie d'avoir une activité créative en attendant de trouver un « vrai job », et je voulais aussi créer une sorte d'**album-souvenir interactif** : voilà mes motivations initiales pour créer ce blog.

Dès mon arrivée à [nom de ville], on est parti en week-end régulièrement, notre objectif était de partir un week-end par mois. Au départ, on ne devait rester que 2 ans, il n'y avait pas de temps à perdre ! Il n'y avait pas grand chose à lire sur la région, rien en français en tout cas. Le blog était encore tout petit, c'était un véritable terrain de jeu pour moi de raconter, prendre des photos, faire des vidéos après une découverte de la ville, [nom de ville], ou de la région autour, la [nom de région], et puis le reste des [nom de pays]. Le premier road [lien hypertexte vers un article] trip dans l'ouest américain a été une véritable claque visuelle, une expérience unique, festive, extraordinaire, je ne m'attendais pas du tout à ça.

Et puis les 2 ans de vie à [nom de ville] se sont prolongés, les voyages se sont multipliés, les expériences professionnelles, amicales, ou simplement de la vie de tous les jours se sont diversifiées.

Capture d'écran 2. Application de la règle de transcription sur un billet d'introduction de la personne de l'auteur.e.

[pseudonyme – commentateur 1] [date] [répondre]

Bonjour [prénom B1] et bravo pour le blog, une véritable mine d'or d'informations sur [nom de pays] ! Je suis rédactrice pour le magazine de [nom de marque] (entreprise de location de voitures) et je réalise un article sur les meilleurs blogs voyage aux [nom de pays] dont tu fais partie. J'ai essayé de t'envoyer un mail via le formulaire de contact plus haut mais ça n'a pas fonctionné donc j'essaie ici. Je souhaiterais illustrer l'article avec une photo de toi ou du blog, du coup serait-il possible de m'envoyer une photo ou m'indiquer laquelle je peux utiliser avec les crédits à ajouter?

Merci,

1. [pseudonyme – commentateur 1] [Prénom B1 – Réponse commentateur 1] [date]
[répondre]

Hello [pseudonyme – commentateur 1]

Le formulaire fonctionne pourtant... j'ai reçu plusieurs messages.
Vous pouvez prendre une photo de la page About en la créditant du blog + lien.
Merci pour la sélection !

2- [pseudonyme masculin – commentateur 2] [date] [répondre]

Bonjour ,

J'avais lu dans un article il y a quelques années la référence à un site de vente apprécié de [prénom mari de B1] sur lequel on pouvait trouver différents objets très [nom de pays]

...

Je ne retrouve pas l'article possible de me le renvoyer ?

Merci

Capture d'écran 3. Exemples d'anonymisation des données de la zone dédiée aux commentaires.

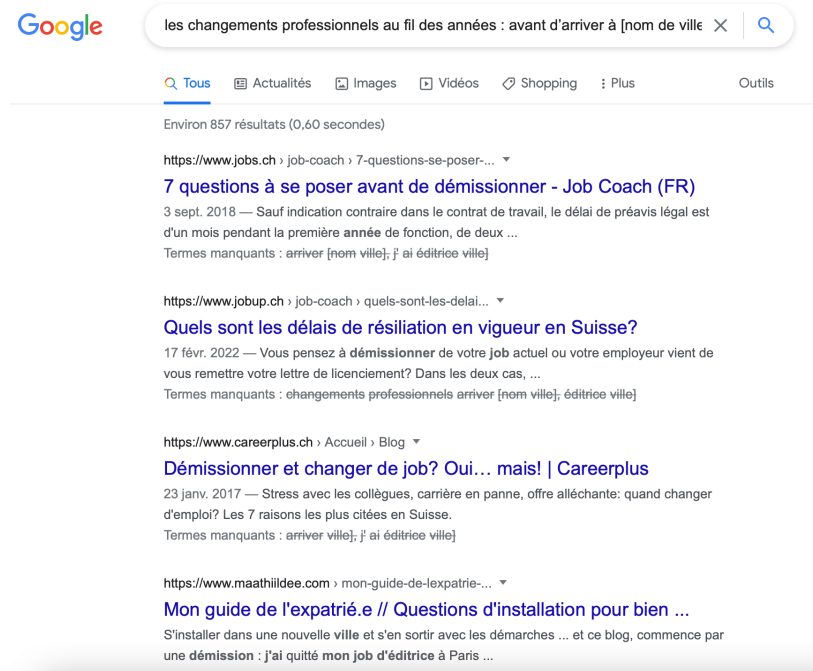
Les clichés représentant personnes ou lieux ont de même été réduits à de simples mentions ; les dates ont été supprimées ; les noms des sites ont été codés. Or, on le constate, dans le cadre d'une étude portant sur la construction d'objets sociaux en discours, ce processus d'étiquetage rend malaisé la lecture et l'appréhension des composantes du corpus. L'anonymisation des données se révèle, dès lors, être un procédé d'exploitation qui questionne le sens que les chercheur.euse.s donnent à leurs travaux en sciences humaines et sociales.

Limites théoriques et pratiques de la démarche

Bien que solution recommandée et légalement irréprochable, retirer des billets de blogs voyage toute référence toponymique, chronologique, civile et photographique se révèle dommageable quant au développement de la question de recherche. De fait, comment analyser la mise en scène de soi des influenceur.euse.s dans le champ discursif du tourisme sans intégrer le nom des destinations, ce qu'elles évoquent ou la manière dont elles sont représentées ? Sociologues, historiens ou anthropologues francophones ou anglophones s'accordent, en effet, pour attester du rôle majeur de la dimension visuelle dans la circulation des imaginaires touristiques et dans la construction du sens du voyage passé et de celui à venir (Boorstin, 1992 : 116 ; Boyer et Viallon, 1994 : 55 ; Urry, 1990). La scène photographiée et diffusée institutionnalise, de fait, l'image des lieux à voir et conditionne le regard que les voyageurs sont incités à poser sur l'emplacement (Boorstin, 1992 : 117 ; Urry, 1992 : 172). De surcroît, n'est-ce pas oblitérer une dimension inhérente de la parole de l'internaute amateur que d'ôter ce qui en établit le caractère véridique ou l'authenticité ? La relation qui se tisse au fil des connexions entre webmestres et lecteur.trice.s est empreinte de la sédimentation des pratiques qui entoure le phénomène de blogging depuis son apparition et se voit nécessairement dotée d'une perception

morale (Lehti, 2011). La filiation des usages impose donc à la voix du.blogueur.euse d'apporter la preuve, a minima iconographique, de la véracité de ses propos (Candel, 2010 : 30), ceci d'autant plus que l'appréciation de la qualité de ses productions est liée à l'image et à la crédibilité qu'il véhicule (Cardon et Delaunay-Téterel, 2006 : 23-24).

Du point de vue de l'efficacité de la pratique et au-delà de la dimension chronophage requise par le traitement manuel – celui-ci n'étant pas exempt d'oubli – la traçabilité des énoncés numériques (Barats, 2016 : 46 ; Merzeau, 2009) a rendu inutile cette démarche de protection. Quel que soit le segment sélectionné, une fois celui-ci introduit comme mot clé dans un moteur de recherche, l'identification du site source est immédiate (Cf. Capture d'écran 4).



Capture d'écran 4. Résultats d'un moteur de recherche à partir du segment anonymisé « les changements professionnels au fil des années : avant d'arriver à [nom de ville] ». Le site source est positionné en quatrième position.

Enfin, à la lecture assidue des blogs et des comptes YouTube de notre corpus, on observe que les influenceur.euse.s voyage ne sont pas en quête d'anonymat mais qu'ils.elles aspirent plutôt à atteindre une forme de reconnaissance (Barats, 2016 : 50 ; Granjon et Denoël, 2010) de leurs capacités créatrices. À la lumière de ces deux derniers éléments, nous avons réorienté notre réflexion vers le déploiement d'une autre solution.

De la pertinence de la notion d'intégrité textuelle pour l'exploitation de données numériques

Confrontée à des obstacles d'ordre pratique et déontologique, nous avons privilégié une approche du terrain respectant l'esprit de la notion d'*intégrité contextuelle* (Nissenbaum, 2011) et qui transcende les limites imposées par la dyade « public-privé ».

L'intégrité contextuelle

Le concept proposé par la philosophe Helen Nissenbaum (Nissenbaum, 2010) suggère au. à la chercheur.euse de s'interroger sur les motivations du. de la locuteur.trice et sur les caractéristiques du cadre communicationnel dans lequel il.elle s'exprime et, par suite, d'en respecter l'intention initiale (Barats, 2016 : 48). L'intérêt d'une telle approche est de considérer que dans chaque contexte social, des normes tacites ou implicites régissent la diffusion de la qualité et du volume d'information qu'il est alors possible de transmettre à des tiers (Ibid.). L'atteinte à la vie privée se produit, en ce cas, lorsque ces normes sont violées (Ibid.). Dans la situation qui nous concerne, les auteur.e.s de blogs voyage créent des contenus qu'ils.elles publient sur leur site et qu'ils.elles relaient sur diverses plateformes telles que YouTube, Pinterest, Facebook ou encore Instagram. De cette diffusion élargie et simultanée de leurs productions via les réseaux sociaux, nous avons déduit que le critère de visibilité importe dans la réalisation de l'acte de publication des blogueur.euse.s.

Application

Compte tenu de leur aspiration à être lus, nous avons donc traité le contenu et les interventions des webmestres de notre corpus en tant que messages intentionnellement conçus et diffusés dans l'espace public du web en vue d'obtenir une audience. À ce titre, nous avons procédé à un travail de citation systématique qui rend hommage à la créativité des individus intégrés à l'étude. Chaque élément (énonciatif, sémiotique ou iconographique), retranscrit dans le manuscrit de notre thèse, est ainsi associé à sa source par la mention du nom du blog et/ou de son adresse internet. À chaque référence faite, un lien hypertextuel propose également au lecteur de consulter en contexte l'extrait présenté. Les potentialités de l'outil informatique sont donc exploitées non seulement en faveur du respect du travail de production de contenu des blogueur.euse.s mais sont également mobilisées pour favoriser, on l'espère, les flux de passage des internautes sur les sites cités. En respectant les principes et l'esprit de cette approche et bien que tous propos soient aisément localisables, on a malgré tout décidé de maintenir, par ailleurs, la protection de l'anonymat des commentateurs via les moyens de brouillage mentionnés *supra*.

Apports de la notion

La pertinence de l'*intégrité contextuelle* réside dans l'interrogation systématique des normes informationnelles qui régulent les données observées. Ce qui semble en effet relever d'une décision déontologique peut, dans d'autres situations concernant le milieu de la blogosphère, ne pas être approprié. Par exemple, les conditions de visibilité sont sensiblement distinctes entre un blog voyage et un blog s'apparentant à un journal

intime. De même, citer des extraits issus d'un forum comme Doctissimo revient à mettre en lumière à ce qui se trouvait dans l'ombre et y était destiné (Barats, 2016). Dès lors, l'intérêt de la notion d'*intégrité contextuelle* est, précisément, de poser le point de départ d'une réflexion qui requiert de la part des chercheur.euse.s de s'interroger sur les conditions de production des messages et sur les intentions des locuteur.trice.s qui en sont à l'origine.

Conclusion

Les situations évoquées et les extraits présentés démontrent qu'il est difficilement réalisable de faire le choix de l'anonymisation des données tout en adoptant un cadre de recherche théorico-méthodologique qualitatif. En revanche, en optant pour une approche relevant de l'intégrité contextuelle, on s'éloigne de la dyade privé/public et des recommandations rarement concrètes et applicables qui en résultent pour traiter les productions publiées en ligne. De même, en se conformant aux préceptes prônés par Nissenbaum, il est alors possible de considérer la singularité du système de normes et de régulations (explicites ou implicites) du partage de l'information propre à chaque contexte social et d'en déduire la manière dont l'information peut-être véhiculée.

Notre cheminement méthodologique illustre les limites résultant de l'application d'un code déontologique généraliste. Certes, il est impératif d'associer la production de tous travaux scientifiques au respect des êtres qui en sont l'objet. Toutefois, au-delà de l'importance de se doter d'un « guide de bonnes pratiques », une réflexion collective doit être menée sur ce signifie être éthique lorsqu'on est chercheur.euse en sciences humaines et sociales.

Références bibliographiques

Amossy, R., *La présentation de soi – Ethos et identité verbale*, P.U.F, Paris, [2010] 2015.

Amossy, R., *Images de soi dans le discours – la construction de l'ethos*, Delachaux et Niestlé, Lausanne, 1999.

Barats, C., *Manuel d'analyse du web*, Armand Colin, Paris, 2016.

Boorstin, D., *The Image: A Guide to Pseudo-Events in America*, Vintage Books Editions, Alfred A. Knopf, New York, [1962] 1992.

Boyer, M., Viallon, P., *La communication touristique*, Que-Sais-Je ? Paris, Presses Universitaires de France, 1994.

Candel, É., « Penser la forme des blogs, entre générique et génétique », In A. Tomiche et P. Zoberman (dir.), *Les blogs – Écritures d'un nouveau genre ?*, L'Harmattan, 23-31, Paris, 2010.

Cardon, D., et Delaunay-Téterel, H., « La production de soi comme technique relationnelle – un essai de typologie des blogs par leurs publics », *Réseaux*, n°138, 15-71, [en ligne], 2006. [<https://www.cairn.info/revue-reseaux1-2006-4-page-15.htm>]

Carvalho, S., « L'éthique de la recherche entre réglementation et réflexivité », *Revue d'anthropologie des connaissances*, n°2, vol. 13, 299-326, [en ligne], 2019. [<https://www.cairn.info/revue-anthropologie-des-connaissances-2019-2-page-299.htm>]

- Demonet, M.-L., « La confiscation des données issues de l'humanisme numérique - un paradoxe résistant », In Ginouvès, V. et Gras, I., *La diffusion numérique des données en SHS - Guide des bonnes pratiques éthiques et juridiques*, Presses Universitaires de Provence, 283-295, Aix-en-Provence, 2018.
- Granjon, F., Denouël, J., « Exposition de soi et reconnaissance de singularités subjectives sur les sites de réseaux sociaux », *Sociologie*, n°1, 25-43, [en ligne], 2010.
[<https://www.cairn.info/revue-sociologie-2010-1-page-25.htm>]
- Huyghe, M., Cailly, L. et Oppenheim, N., « Ouverture de données qualitatives à caractère personnel. Approche éthique, juridique et déontologique », In Ginouvès, V. et Gras, I., *La diffusion numérique des données en SHS - Guide des bonnes pratiques éthiques et juridiques*, Presses Universitaires de Provence, 159-168, Aix-en-Provence, 2018.
- Lambert-Chan, M., « Qu'est-ce que l'éthique de la recherche ? », In *Petit guide de survie des étudiants*, Presses de l'Université de Montréal, [en ligne], Montréal, 2012.
[<http://books.openedition.org/pum/7690>]. ISBN : 9791036501043. DOI : <https://doi.org/10.4000/books.pum.7690>]
- Lehti, L., « Blogging politics in various ways: A typology of French politicians' blogs », *Journal of pragmatics*, n°43, 1610-1627, [en ligne], 2011.
[<https://www.sciencedirect.com/science/article/pii/S0378216610004054?via%3Dihub>]
- Maingueneau, D., *Analyser les textes de communication*, 4e édition, Armand Colin, Paris, [2013] 2021.
- Maingueneau, D., *Discours et analyse du discours*, Armand Colin, Paris, 2014.
- Maingueneau, D., « Que cherchent les analystes du discours ? », *Argumentation et Analyse du Discours*, n°9, 1-17, [En ligne], 2012.
[<http://journals.openedition.org/aad/1354>]
- Marcoccia, M., *Analyser la communication numérique écrite*, Armand Colin, Paris, 2016.
- Merzeau, L., « Du signe à la trace : l'information sur mesure », *Hermès La Revue*, n° 53, 21-29, [en ligne], 2009. [<https://www.cairn.info/revue-hermes-la-revue-2009-1-page-21.htm>]
- Nissenbaum, H., « A Contextual Approach to Privacy Online », *Daedalus*, vol. 140, n°4, 32-48, The MIT Press, 2011.
- Nissenbaum, H., *Privacy in context: Technology, policy and the integrity of social life*, Stanford Law, Stanford, 2010.
- Paveau, M.-A., *L'analyse du discours numérique – dictionnaire des formes et des pratiques*, Hermann Éditeurs, Paris, 2017.
- Pereira Couttolenc, E., *La présentation de soi des auteur.e.s de blogs voyage : une identité discursive du voyageur amateur ?*, thèse en sciences du langage dirigée par von Münchow, P., soutenue à l'Université Paris Cité devant un jury composé de Florence Mourlhon-Dallies (Présidente), Sophie Moirand (Rapporteur), Fabienne Baider (Rapporteur) et Michel Marcoccia (Examinateur) le 2 décembre 2022.

Urry, J., « The Tourist Gaze “revisited” », *American Behavioral Scientist*, vol. 36, n°2, 172-186, Sage Publications, [en ligne], 1992. [<https://journals-sagepub-com.ezproxy.u-paris.fr/doi/abs/10.1177/0002764292036002005>]

Urry, J., *The Tourist Gaze. Leisure and Travel in Contemporary Societies*, Sage, London, 1990.

Sitographie

CNIL, « Le règlement général sur la protection des données. RGPD », texte daté du 23 mai 2018, [en ligne], consulté le 15 juin 2023. [<https://www.cnil.fr/fr/reglement-europeen-protection-donnees>]