

Analyser les commentaires d'actualité sur le web : difficultés méthodologiques et intérêt sociopolitique

La grande majorité des sites d'information en ligne offrent désormais la possibilité aux internautes de réagir à l'actualité par le biais de commentaires en marge des articles. La littérature académique concernant les contenus produits par les utilisateurs sur le web est relativement importante aujourd'hui, mais dans la diversité des genres discursifs en ligne (blogs, forums, réseaux sociaux...), le commentaire, et plus encore le « commentaire à chaud de l'actualité » (Dupret & al., 2010) reste peu étudié. Plus précisément, les travaux qui l'abordent mettent l'accent sur des dimensions pragmatiques¹ ou formelles², mais rares sont ceux qui s'intéressent au contenu propre de ces discours tels qu'ils sont suscités par les dispositifs interactifs des médias en ligne.

Dans un premier temps nous dresserons donc ici un bilan des raisons qui peuvent rendre complexe voire repoussante l'approche des corpus de commentaires. Dans un second temps nous chercherons à dépasser ces contraintes en montrant qu'un ensemble d'éléments tant théoriques que méthodologiques sont à même de légitimer l'approche de leurs contenus. Compte tenu du format court de ce texte, nous ne pourrons présenter d'analyse de corpus à l'appui de nos propositions. Les questionnements épistémologiques afférents à l'utilisation de ces corpus originaux étant très importants, nous avons choisi de nous concentrer sur cette dimension théorique.

Nous concluons en affirmant que les commentaires d'actualité révèlent, par-delà une cacophonie de surface, à la fois une richesse des modes d'interprétation du monde et une structuration de ceux-ci autour d'un nombre restreint de pôles attracteurs. De tels résultats plaident ainsi pour un intérêt sociopolitique de l'analyse qualitative des tableaux de commentaires : il est possible d'établir une typologie des modes d'appréhension d'une question sociale ainsi qu'une liste des topoï et des arguments les plus récurrents au sein de l'espace public autour d'un sujet particulier. Ces données invitent alors à reconsidérer la manière dont les sphères médiatiques ou politiques peuvent problématiser certains enjeux.

I. Analyser les comment boards : des contraintes éthiques, techniques et théoriques

Comme nous l'évoquons en introduction, très peu de travaux se sont consacrés à l'analyse qualitative du contenu des commentaires sur les sites d'information en ligne. Ce silence peut sans doute s'expliquer simplement : le caractère disparate et vociférant du commentaire à chaud de l'actualité décourage l'analyse autant qu'il masque son intérêt. On peut en effet très sincèrement se demander l'apport scientifique de l'étude de corpus de textes dont on ne connaît à peu près rien des conditions de production ni de réception et qui, de surcroît, peuvent apparaître futiles tant ils sont hétérogènes, éphémères et au contenu souvent très négatif. Essayons toutefois dans cette partie de dresser un tableau des principales caractéristiques qui rendent ce matériau particulièrement résistant et repoussant avant de proposer un positionnement à même de dépasser ces contraintes.

II.1. Contraintes éthiques

L'analyse académique des contenus produits par les internautes (*User Generated Content*, UGC) pose avant toute chose un ensemble de problèmes d'ordre éthique (Rakotonelina, 2011 ; Proulx &

-
- 1 Comme par exemple son intrication avec les modèles économiques de la presse en ligne (Trygg, 2012), ses incidences sur l'activité journalistique dans le cas du journalisme participatif (Dagiral & Parisie, 2010), la nouvelle relation entre journaliste et consommateur-producteur, (Calabrese, 2014), etc.
 - 2 Dans le cas des approches ethnographiques de ce genre discursif somme toute récent (Boure, 2013 ; Jouët & Le Caroff, 2013) ou encore de la confrontation des analyses conversationnelles à ce nouveau mode interactif (Maccoccia, 2004 ; 2011)

Latzko-Toth, 2013 ; et cf. le volume 14, n°2 de la revue *éthique publique* consacré à la question en 2012). En effet, l'investissement massif et non réflexif du web par un grand nombre d'individus fait émerger une tension entre *visibilité* et *publicité* de la parole ordinaire en ligne (Cardon, 2009). Cette conception fait écho à la notion de « respect de l'intégrité contextuelle » des productions proposée par Helen Nissenbaum (2004) : si tout individu s'exprimant sur internet est censé savoir que ses propos deviennent immédiatement accessibles, le contexte particulier (forum nécessitant une inscription préalable, intimité du sujet, habitude de n'interagir qu'avec une communauté réduite...) dans lequel il les produit peut participer à créer un *sentiment de privacité*.

Au vu de la difficulté évidente d'obtenir l'accord éclairé des individus dont on corpusse la parole, la question éthique de savoir si l'on peut se permettre de sur-visibiliser des contenus qui ne s'y destinaient pas doit se poser. Elle est par ailleurs renforcée par ce que Serge Proulx et Guillaume Latzko-Toth (2013:43) appellent très joliment le « droit à l'ombre et à la lumière ». D'un côté, les UGC relèvent de la propriété intellectuelle, ce qui invite le chercheur qui les reproduit à la citation nominative et à donner la possibilité de retrouver la page source (en indiquant son URL par exemple). Mais d'un autre, la teneur des propos peut entraîner des conséquences fâcheuses pour leurs émetteurs qui requièrent donc une forme de protection, au moins de leur vie privée.

Il va sans dire que le fonctionnement du web rend cette protection particulièrement difficile à mettre en place. En effet, Internet se caractérise par une forte recherchabilité de l'information. L'indexation en plein texte des pages permet d'accéder à une grande partie des contenus du net par la simple reproduction d'une partie du texte ciblé dans un moteur de recherche. Dans ces conditions l'anonymisation d'un corpus de commentaires relève rapidement de l'illusion : si la traçabilité de la source (ou du moins de son pseudonyme) est légèrement différée, elle n'est en aucun cas empêchée.

Bref, il est clair que chaque situation particulière demande une prise de décision éthique de la part du chercheur : quel est le degré de sensibilité des contenus que l'on utilise ? Nécessitent-ils un consentement explicite pour être utilisés ? Où doit-on placer le curseur entre respect de la propriété intellectuelle, de la vie privée et de la falsifiabilité scientifique ?

II.2. Contraintes techniques

Un second type de problème qui se pose au chercheur qui aborde les corpus UGC issus du web est d'ordre plutôt technique : il ouvre vers des questions relevant tout à la fois d'un versant « pragmatique » de la recherche et de dimensions plus proprement scientifiques.

Sur le premier versant se trouvent les compétences informatiques nécessaires à l'extraction et à l'archivage de documents numériques. Le web étant dynamique, il est également particulièrement instable. Ainsi, des pages peuvent disparaître de manière plus ou moins aléatoire (suppression de contenu par le propriétaire du site, changement d'hébergeur, problème de serveur...) ou se modifier au fil du temps (changement d'interface d'un journal en ligne par exemple, édition de contenu...). Travailler sur un corpus web implique donc un archivage protecteur hors ligne qui demande des habiletés techniques (connaître et savoir utiliser les logiciels d'aspiration de site web ou les archives online comme *wayback machine*...) et qui a éventuellement un coût financier.

Sur le second versant, plus épistémologique, cette instabilité définitoire du web pose à la fois le problème des conditions de production et de la clôture du corpus (Marras, 2004 ; Pierozak, 2011). L'interface d'un site web pouvant changer avec le temps et la production discursive pouvant commencer avant et se terminer après le recueil des données par le chercheur, celui-ci doit se résoudre à n'avoir entre les mains qu'un instantané dont la construction évolutive lui est en partie inconnue. Du point de vue philologique, les problèmes sont bien évidemment très importants : connaissant l'impact d'un dispositif de parole sur la parole elle-même, des textes produits sous deux

dispositifs différents, quand bien même ils relèvent d'un même ensemble, ne peuvent être dotés du même statut.

II.3. Contraintes théoriques

Commençons par tordre le cou à un canard. On brandit souvent comme intérêt majeur des corpus web le fait que la parole des utilisateurs, non suscitée par le chercheur et produite sans conscience de sa présence, serait écologique. C'est vrai. Mais l'écologie du web est différente de l'écologie hors-ligne et croire évacué le fameux paradoxe de l'observateur par la simple intermédiation d'un écran est bien naïf : sur le net l'observateur est partout et tout le temps (Pierozak, 2011:25). La moindre parole, même tenue ailleurs et jadis, peut-être retrouvée. L'internaute même s'il n'en est pas pleinement conscient – nous parlions plus haut de sentiment de privacité – sait bien que ses écrits se fixent et qu'il ne maîtrise pas l'identité de tous ses récepteurs. Les données du web relèvent donc d'une écologie bien spécifique : elle sont une parole émise sans la pression du projecteur du chercheur, mais avec une advertance sans doute accrue pour le regard de l'Autre.

Ce préalable étant posé, essayons de lister les grands problèmes posés par les UGC. Le premier est sans doute la difficulté à disposer de données sociodémographiques individualisées. Une solution peut-être d'utiliser les traces implicites (Ertzscheid, Gallezot & Simonnot, 2013) que laissent les internautes en surfant sur le web (cache, cookies, adresse ip, etc.). Mais outre le fait que l'utilisation de ces métriques pose de nouveau d'évidents problèmes éthiques et techniques, il faut également affirmer qu'elles ne doivent pas être vues comme l'accès à la vérité des pratiques des internautes : si elles indiquent ce que l'internaute a fait avec telle ou telle application, elle ne disent rien de ses intentions, de son parcours, ou du sens qu'il donne à son activité. Bref, l'anonymat sur le web, même partiel, reste un des problèmes les plus prégnants pour le chercheur. Ne pas savoir qui écrit empêche de conférer aisément un statut aux textes étudiés.

Par ailleurs, le contenu des commentaires peut paraître extrêmement hétérogène : de la réponse à l'article, à l'interpellation des autres internautes en passant par la publicité personnelle, l'utilisation des commentaires est assez variable selon les internautes (Dupret & al, 2010). De surcroît, le niveau de langue et le degré d'agressivité rencontrés sur les lieux de discussion les plus grand public peuvent aisément décourager l'analyse.

Autre problème, les modèles économiques de la presse en ligne sont fortement en rapport avec les questions de modération des commentaires (Le Floch & Sonnac, 2013:41-68), ce qui amène les médias à mettre en place un ensemble de stratégies de censure plus ou moins brutales et fluctuantes qui rendent encore une fois les corpus de commentaires problématiques : à quel point sont-ils formatés par une instance éditoriale supérieure, et comment ?

Dans le même ordre d'idée, il est clair que des modes de commentaires différents (anonymes ou non, possibilité de répondre directement dans le corps des commentaires ou non, nombre de caractères limité ou non, etc.) informent les pratiques de contribution. Plus généralement c'est toute la problématique de l'énonciation éditoriale des écrits d'écran (Jeanneret & Souchier, 2005) qui se pose ici : toute parole sur le web, loin d'être parfaitement libre, est soumise à l'ensemble des médiations techniques qui la rendent possible. En plus d'avoir une connaissance aiguë d'interfaces aux tendances mutatoires, il faut donc se demander si il est possible de comparer des écrits produits dans des dispositifs différents.

Ces divers points problématiques sont néanmoins plus des obstacles que des impasses et ne doivent pas mener à conclure à l'impossibilité de traiter ce matériau. Nous présentons ci-après certains éléments qui ouvrent la voie à une approche particulière des commentaires d'actualité : leur sondage comme une part de discours social (Angenot, 2006).

II. Analyser les comment boards : quelques voies d'accès

Commençons par proposer une porte de sortie à l'épineux problème de l'insaisissable population du web et plus particulièrement de la presse en ligne. En effet, si la sociologie fine de ses lecteurs est peu connue, certaines données (métriques et enquêtes) attestent du fait qu'elle est relativement hétérogène (Dagiral & Parasie, 2010). Certes, face au fantasme de l'internet pour tous, il faut sans cesse rappeler que la « fracture numérique » est bien réelle (Charon & Le Floch, 2013:103-116) et l'étude sur les pratiques culturelles des Français à l'ère numérique (Donnat, 2009) montre bien que le lectorat de la presse en ligne s'approche de celui de la presse traditionnelle (en majorité composé d'hommes de classe moyenne à supérieure). S'approcher n'est cependant pas égal et on ne peut nier qu'au moins l'exposition à l'actualité (sinon la *lecture* effective de la presse) est grandement augmentée sur le net (omniprésence sur les portails web, diffusion par les réseaux sociaux...). Aussi, la fonction « bridging » d'internet semble-t-elle vérifiée : l'affaiblissement des barrières physiques et la facilité de se rendre d'un point à un autre du web y augmente l'hétérogénéité sociale (Norris, 2002). Par ailleurs si les données sociodémographiques manquent, les motivations des internautes à commenter, elles, sont mieux connues.

Renee Barnes (2013) montre notamment à partir d'un questionnaire proposé aux visiteurs d'un site australien de « journalisme alternatif » que si finalement assez peu de gens commentent régulièrement, l'immense majorité lit les commentaires des autres et les valorisent souvent au même niveau que l'article journalistique déclencheur (*ibid*, 2013:8-9). De plus, lorsqu'ils participent, les internautes disent le faire avant tout pour exprimer leur réaction émotionnelle vis à vis du contenu de l'article ou des autres contributions (*ibid*, 2013:12). Ces résultats dans un contexte particulier semblent par ailleurs confirmés par des études plus larges sur la large prédominance de la charge affective des commentaires d'actualité sur les sites de presse les plus visités aux États-unis (Paskin, 2010 ; Milioni & al., 2011).

Ainsi, si l'on ne sait pas *qui* parle, on sait *pourquoi* et on sait que les commentaires sont lus³ et valorisés. De plus, de nombreux tableaux de commentaires présentent des signes d'interactivité inter-commentateurs (Paskin, 2010). Ces observations confirment la part d'interlecture importante et soulignent que les commentaires ne semblent pas être seulement des expressions purement individualisées mais entrent au contraire souvent dans une dynamique intersubjective.

Enfin, et c'est ici un point capital, le commentaire d'actualité présente une dimension politique quasi-systématique (Shah et al., 2005) : « que ce soit en ligne ou hors ligne, l'actualité et la discussion politique vont souvent de concert, l'une entraînant l'autre » (Brundidge, 2010:688, nous traduisons). Jenifer Brundidge (*ibid.*) propose un ensemble d'explications à ce phénomène et montre que les internautes engagés dans le commentaire en ligne sont particulièrement à même de constituer des leaders d'opinion dans la société hors ligne, augmentant le caractère représentatif/prescriptif des contenus idéologiques présents dans ces textes. Ces observations sont néanmoins à nuancer par les travaux sur l'engagement politique en ligne. À la suite de la notion de *slacktivism* (Morozov, 2009), de nombreuses études s'accordent à dire que bien souvent, l'engagement en ligne masque paradoxalement un désengagement hors ligne (Blanchard, Gadras & Wojcik, 2013:172). Quand bien même, que les propos soient suivis d'actes concrets ou non, il n'en reste pas moins que les idées deviennent disponibles en circulant dans l'espace public.

Ainsi donc, les commentaires sur les sites d'information en ligne sont moins cacophoniques qu'ils en

3 Un récent sondage Opinion Way (2011) réalisé en France suggère que 92% des internautes déclarant consulter régulièrement (soit près de 75% du panel) la presse en ligne déclarent également lire au moins occasionnellement les commentaires au bas des articles. URL : <http://socialmediacub.fr/2011/10/mais-qui-donc-lit-les-commentaires-darticles-de-presse/>

ont l'air. Si l'on ne sait pas vraiment qui écrit, on sait que les internautes, socialement hétérogènes, ont tendance à s'intéresser au contenu des commentaires. L'interactivité y est bien présente, et pas seulement par le biais de l'interlecture. Que ce soit par simple allusion ou interpellation directe, les commentateurs conversent, à tel point que certains sites voient des chartes internes, plus ou moins implicites réguler le déroulement des discussions. De plus, le contenu des commentaires, même extrêmement violent, n'est pas dénué de logique : il est la plupart du temps l'expression émotionnelle d'une opinion à dimension identitaire et idéologique. Il présente enfin une *disponibilité* forte dans l'espace public, c'est à dire que son contenu est souvent rencontré par le visiteur de l'actualité en ligne, voire diffusé hors ligne par celui-ci en raison de la tendance des lecteurs de presse en ligne à participer au célèbre *two/multiple step flow of communication*⁴ en tant que leaders d'opinion ou, à tout le moins, de discussion.

III. Des commentaires d'actualité au discours social

Nous revendiquons donc le fait que les corpus de commentaires peuvent être homologués à une part de discours social au sens de Marc Angenot (2006). Selon lui, le discours social correspond à « la totalité de ce qui s'écrit, s'imprime et se diffuse à un moment donné dans un état de société » (*ibid.*), du livre de philosophie à la discussion de comptoir en passant par le flyer publicitaire ou la notice administrative. L'ambition étant de saisir derrière cette masse interdiscursive, cette archive au sens de Michel Foucault, la manière dont une société se connaît en se disant. Nous ne tombons évidemment pas de facto dans l'illusion de l'universalisme des corpus internet (Pierozak, 2011) : les développements qui précèdent nuancent clairement cette position. Ces discours sur la toile ne sont donc qu'un miroir partiel de la société dans son ensemble mais il n'en reste pas moins qu'ils en offrent une image intéressante dès lors que l'on s'intéresse à la circulation des idées dans l'espace public.

La théorie du discours social établit par ailleurs qu'au-delà de l'hétérogénéité apparente de toutes les productions discursives d'une époque, se dégagent certaines constances qui se regroupent sous des tendances hégémoniques. L'hégémonie n'est pas entendue ici comme ce qui se dit le plus souvent ou le plus fort mais retient autant les discours des groupes dominants que les contre-doxas subversives de milieux moins audibles. Elle s'oppose en cela non pas au contre-dit mais au pas-encore-dit. Derrière le tohu-bohu apparent et l'intuition naïve que tout peut se dire et, mieux encore, que tout se dit toujours quelque part, il est en fait possible de délimiter dans les modes de dire le monde - et donc de le connaître - des dominances et des récurrences. Ce point est d'une importance capitale dans notre approche puisqu'il autorise un grand mouvement épistémologique : celui de transférer l'habituel crédit de représentativité de la diversité accordé aux très grands corpus à un échantillon plus modeste. En ce qui concerne les modes d'appropriation d'un événement ou les idées circulantes, il apparaît en effet que le modèle sature rapidement⁵ : « la simple observation du fonctionnement des discours sociaux conduit [Bernard Delforce et Jacques Noyer] à une conclusion du même ordre. Sur une thématique socialement admise et stabilisée - au moins provisoirement - (banlieue, immigration, dopage...) ou sur un type d'occurrence-événement (crime, manifestation...), on n'observe pas une infinité de discours mais un nombre limité de discours sociaux qui, correspondant à des façons différentielles de percevoir et d'interpréter, donnent sens, en des versions différentes, à ce qu'on rapporte » (Delforce & Noyer, 1999).

Ainsi, les masses de commentaires, issues de sites variés, ne présentent pas une diversité incompressible mais donnent au contraire à voir des sédimentations autour de pôles attracteurs que l'analyse permet de cartographier. Les masses de commentaires, issues de sites variés, ne présentent pas une diversité incompressible mais donnent au contraire à voir des sédimentations autour de

4 Ou plutôt devrions-nous dire *flow of social construction*.

5 Cette saturation se doit bien évidemment d'être vérifiée par le biais de corpus de contrôle comprenant d'autres genres discursifs.

pôles attracteurs que l'analyse permet de cartographier.

Par ailleurs, travailler sur un corpus à taille humaine permet de proposer une étude fine, au ras du texte, qui n'oublie pas que toute analyse du discours « est avant tout une activité de fréquentation des énoncés, bref un travail qui est une pratique de *lecture* » (Krieg-Planque, 2009, elle souligne). À l'heure où l'*opinion mining*, par définition quantitatif, est en plein essor (Eensoo-Ramdani et al., 2011), les approches plus manuelles d'analyse du discours permettent ainsi d'observer avec une plus grande finesse des sédimentations idéologiques sur le web. Notons néanmoins que cette proposition ne tombe pas non plus dans un manichéisme improductif. Nous revendiquons au contraire – et comme beaucoup – l'intérêt des méthodes mixtes, conformément aux précautions réglant l'usage de la lexicométrie dans une perspective textuelle et non seulement lexicale⁶ (Rastier, 2008, p.48-51). L'outil peut ainsi trouver sa place dans une dialectique permanente entre regards micro- et macro-, tantôt confirmant ou infirmant des hypothèses issues de la lecture fine, tantôt en proposant à la vérification de celle-ci.

IV. Conclusions et perspectives.

Nous entendons ainsi défendre trois conclusions :

1) Les tableaux de commentaires d'actualité ne sont pas les masses ineptes qu'ils peuvent paraître : ils sont au contraire porteurs de contenus idéologiques qui renseignent sur les formes d'appréhension du monde qui peuvent circuler dans l'espace public.

2) Si la fouille d'opinion est un domaine à explorer, une approche qualitative, ou au moins mixte, présente des intérêts non-négligeables en termes de précision et de diversité des données extraites tout en permettant, dans le cadre théorique du discours social, de traiter à moindres frais des corpus de texte de grande taille.

3) Enfin, les résultats de ce type d'étude peuvent apporter des contributions pratiques sur le plan sociopolitique. Outre l'intérêt anthropologique d'établir un cliché du discours social, cartographier les valeurs et les argumentations en confrontation dont le tumulte est renforcé pour les nouveaux outils du numérique peut permettre de fournir un matériau original pour nourrir ou simplement éclairer les débats publics.

Références :

Angenot M. (2006), « Théorie du discours social », ConTEXTES, n°1, [en ligne].

Barnes R. (2013), « Understanding the affective investment produced through commenting on Australian alternative journalism website New Matilda », *New Media and Society*, [en ligne].

Blanchard G., Gadras S., Wojcik S., 2013, « Analyser la participation politique en ligne : des traces numériques aux pratiques sociales », in Barats, C. (dir.), *Manuel d'analyse du web en sciences*

6 Les travaux lexicométriques les moins vigilants prennent en effet le risque d'écraser la complexité des textes d'au moins quatre manières : 1) en étant lexicocentrés alors que l'information linguistique repose sur bien d'autres composantes (syntaxique, narrative, argumentatives, etc.), 2) en imposant la prédominance d'un supposé type sur l'occurrence lexicale (alors qu'au delà des simples phénomènes d'homonymie, un « patron » n'est pas la même chose dans un discours communiste et un discours de droite libérale : les assimiler sous une même forme est de fait questionnable), 3) en reproduisant une doxa quantitativiste, car bien souvent seules les fréquences hautes importent alors qu'un seul mot peut resémantiser totalement un texte (on pensera simplement à un titre ou à la mention d'une filiation intellectuelle ou politique) ; on notera par ailleurs que cette approche est contraire à la théorie de l'information qui stipule à l'inverse que c'est justement ce qui est le plus récurrent qui est le moins porteur d'information !, et 4) en invitant à considérer les sorties machines comme des résultats alors qu'elles ne sont bien sûr que des données de second ordre qui demandent elles-mêmes une théorie de l'interprétation au risque de ne faire que redoubler du sens commun.

- humaines et sociales, Paris, Armand Colin, 166-178.
- Boure R. (2013), « Parler rugby en ligne entre soi », *Réseaux*, 180 (4), 157-187.
- Brundidge J. (2010), « Encountering “Difference” in the Contemporary Public Sphere: The Contribution of the Internet to the Heterogeneity of Political Discussion », *Networks Journal of Communication*, 60, 680–700.
- Calabrese L. (2014), « Rectifier le discours d’information médiatique. Quelle légitimité pour le discours profane dans la presse d’information en ligne ? », *Les Carnets du Cediscor*, 12, [en ligne].
- Cardon D. (2008), « Le design de la visibilité : un essai de typologie du web 2.0 », *Réseaux*, 152, 93-137.
- Charon J.-M., Le Floch P. (2011), *La presse en ligne* Paris, La Découverte, 126 p.
- Dagiral É., Parasie S. (2010), « Presse en ligne : où en est la recherche ? », *Réseaux*, 160-161 (2), 13-42.
- Delforce B., Noyer J. (1999), « Pour une approche interdisciplinaire des phénomènes de médiatisation : constructivisme et discoursivité sociale », *Études de communication*, 22, [en ligne].
- Dupret B, Klauss E, Ghazzal Z. (2010), « Commenter l’actualité sur internet », *Réseaux*, 160-161 (2), 285-317.
- Donnat O. (2009), *Les pratiques culturelles des Français à l’ère numérique. Enquête 2008*, Paris, La Découverte, 288p.
- Eensoo-Ramdani E., Bourion E., Slodzian, M., Valette M. (2011), « De la fouille de données à la fabrique de l’opinion », *Les Cahiers du numérique*, 2 (7), 15-40.
- Ertzscheid O., Gallezot G., Simonnot B. (2013), « À la recherche de la « mémoire » du web : sédiments, traces et temporalités des documents en ligne » in Barats, C. (dir.), *Manuel d’analyse du web en sciences humaines et sociales*, Paris, Armand Colin, 53-68.
- Jeanneret Y., Souchier E. (2005), « L’énonciation éditoriale dans les écrits d’écran », *Communication et langages*, 145, 3-15.
- Jouët J., Le Caroff C. (2013), « L’observation ethnographique en ligne », in Barats, C. (dir.), *Manuel d’analyse du web en sciences humaines et sociales*, Paris, Armand Colin, 147-161.
- Krieg-Planque A. (2006), « « Formules » et « lieux discursifs » : propositions pour l’analyse du discours politique », *Semen*, 21, [en ligne].
- Le Floch P., Sonnac N. (2013), *Économie de la presse à l’ère numérique*, Paris, La Découverte, 128p.
- Marcoccia M. (2004), « L’analyse conversationnelle des forums de discussion : questionnements méthodologiques », *Les Carnets du Cediscor*, 8, [en ligne].
- Milioni D., Vadratsikas K., Papa V. (2011), « Their two cents worth: A content analysis of online readers' comments in mainstream news outlets », Article présenté à la World Association for Public Opinion Research, 64th Annual Conference, Amsterdam, [En ligne].
- Morozov E. (2009), « The Brave New World of Slacktivism », [en ligne].
- Nissenbaum H. (2004), « Privacy as Contextual Integrity », *Washington Law Review*, 1, vol 79, 119-157.
- Norris P. (2002), « The Bridging and Bonding Role of Online Communities », *Harvard International Journal of Press/Politics*, 7 (3), 3-13.
- Paskin D. (2010), « Say what? », *Journal of International Communication*, 16 (2), 67-83.
- Pierozak I. (2011), « Les corpus électroniques en sciences du langage : un eldorado ? », *Le discours et la langue*, 2.1, 15-32.
- Proulx S., Latzko-Toth G. (2013), *Enjeux éthiques de la recherche sur le Web*, in Barats, C. (dir.), *Manuel d’analyse du web en sciences humaines et sociales*, Paris, Armand Colin, 32-45
- Rakotonoelina F. (2011), « Sur l’internet, personne ne sait que tu es un linguiste », *Le discours et la langue*, 2.1, 55-72.
- Rastier F. (2011), *La Mesure et le Grain. Sémantique de corpus*, Paris, Honoré Champion, 280p.
- Shah D., Cho J., Eveland W. P., Kwak N. (2005), « Information and expression in a digital age », *Communication Research*, 32, 531–565.
- Trygg S. (2012), « Is comment free ? Ethical, Editorial and political problems of moderating online

news comment », Nordicom, 1, [en ligne].