



9e Congrès des Enseignants Documentalistes de l'Éducation Nationale
organisé par la FADBEN

Objets documentaires numériques : nouvel enseignement ?
Paris Ile de France
22, 23 et 24 Mars 201

Pourquoi et comment évaluer le pluralisme du journalisme sur le web ?

*Nikos Smyrnaiois,
maître de conférences, LERASS, Université Toulouse 3*

Évaluer le pluralisme du journalisme en ligne est un défi pour la recherche. La multiplicité des supports et la complexité des modes de consommation de l'information d'actualité font qu'il est particulièrement difficile d'estimer le degré de pluralisme que le web est censé incarner. Seule une combinaison d'analyses à niveaux multiples mettant en œuvre des méthodes de recherche quantitatives et qualitatives est en mesure de donner une réponse satisfaisante. Notre projet de recherche appelé IPRI (Internet, Pluralisme et Redondance de l'Information)¹ vise à mesurer la diversité des nouvelles en ligne en France, à travers une étude transdisciplinaire de plusieurs catégories de sites web d'information (médias en ligne, portails, blogs, *pure-players*²). Notre but est ici de présenter les aspects théoriques et méthodologiques de cette étude ainsi qu'un aperçu des résultats initiaux afin de tirer quelques conclusions quant aux enjeux de cette problématique pour l'éducation aux médias.

Le cadre théorique

Le degré de pluralisme des opinions dans la sphère publique est un enjeu politique et social majeur qui dépend fortement de la diversité du journalisme et des médias : plus les médias sont diversifiés du point de vue de leur ligne éditoriale et des sujets couverts, mieux le public sera informé sur les enjeux sociaux qui en découlent. Cette conception idéal-typique de la sphère publique, en grande partie issue des travaux de Jürgen Habermas (1991), a inspiré une vision de l'internet en tant que parangon de la démocratie (Hindman, 2009). Le raisonnement est simple : depuis l'apparition de l'internet, la publication de l'information à grande échelle est beaucoup plus aisée et moins chère que les formes traditionnelles de communication de masse. Par conséquent l'expression publique médiatisée serait à la portée de chaque citoyen ayant accès à l'internet. Cet argument a été employé par exemple dans plusieurs rapports gouvernementaux (Lancelot, 2005, Tessier, 2007), mais aussi dans

¹ Le programme de recherche IPRI (Internet, pluralisme et redondance de l'information) dirigé par Franck Rebillard (CIM, Université Paris 3) a été soutenu par l'Agence Nationale de la Recherche (ANR-09-JCJC-0125-01b). Plusieurs chercheurs appartenant à différents laboratoires y ont participé : Emmanuel Marty (I3M, Université de Nice), Annelise Touboul et Stéphanie Pouchot (ELICO, Université de Lyon), Béatrice Damian-Gaillard (CRAPE, Université Rennes 1), Elöd Egyed-Zsigmond et Sylvie Calabretto (LIRIS, INSA Lyon), Bernhard Rieder (Université d'Amsterdam), Thierry Lafouge (RECODOC, Université de Lyon) et Dominique Fackler (INA).

² Le terme *pure-players* désigne les sites d'informations nés en ligne et qui ne sont pas les émanations d'un média traditionnel.

les documents produits par des organisations supranationales comme l'ONU et l'Union européenne sur l'avènement d'une supposée « société de l'information » (Bangemann, 1994, ONU, 2005). Plus généralement, le web est souvent considéré comme un canal par le biais duquel les contenus culturels et informationnels marginaux peuvent atteindre facilement un public plus large qu'ils ne le font par le biais des canaux de distribution traditionnels, au point de supplanter les traditionnels *best-sellers*. Cette idée a été popularisée par Chris Anderson et sa théorie de la Longue Traîne (Anderson, 2006).

Cependant, cette vision séduisante mais quelque peu simpliste a été depuis remise en cause par une série de recherches empiriques (Benghozi & Benhamou, 2010, Elberse & Oberholzer-Gee, 2008, Brynjolfsson *et al.*, 2006). Entre autres choses, cette approche ignore le fait que, au cours des dernières années, l'internet est devenu un champ de compétition acharnée entre groupes sociaux, organisations politiques et entreprises multinationales pour la distribution du pouvoir sur les canaux de communication numériques (Mansell, 2004). Désormais, la structuration du secteur de l'information en ligne est le résultat des relations complexes et parfois antagoniques entre médias professionnels, communautés d'amateurs productrices de contenu et intermédiaires puissants, tels que Google ou Facebook (Smyrnaiois & Rebillard, 2011).

En effet, pour acquérir une visibilité importante sur l'internet, il ne suffit pas simplement d'y publier une information. Encore faut-il être en mesure de mettre en œuvre une stratégie de référencement efficace, ce qui implique des moyens techniques, humains et économiques importants. Il faut aussi, pour maîtriser le processus de diffusion virale de l'information (le fameux *buzz* tant convoité par le marketing), disposer d'un savoir-faire et des moyens humains conséquents souvent réservés aux organisations puissantes. Il en résulte ainsi une concentration de l'audience sur un petit nombre de sites d'information (médias traditionnels, portails et agrégateurs) au détriment des acteurs plus marginaux, porteurs de diversité (Granjon & Le Foulgoc, 2010, Cevipof, 2010). De ce point de vue, la sphère publique électronique doit être considérée comme une arène conflictuelle, plutôt que comme une *agora pacifique* (Peters, 2004), où les choix journalistiques et éditoriaux incarnent des stratégies de rivalité politiques et industrielles (Mosco, 2009, Fenton, 2009). La diversité de l'information en ligne dépend au moins autant du résultat de ces stratégies et du rapport de force entre les acteurs qui en découle que des caractéristiques supposées « naturelles » du média.

On constate de fait à ce propos que les discours des décideurs, comme ceux de nombreux observateurs et professionnels (journalistes, consultants, spécialistes du marketing, etc.), sont bien souvent caractérisés par un certain déterminisme technologique. Or, une étude plus approfondie fait apparaître une piste d'investigation inverse : la multiplicité des espaces de publication sur l'internet pourrait favoriser le pluralisme autant qu'une certaine redondance des informations en circulation, appelée aussi *shovelware phenomenon* (Paterson & Domingo, 2008). En effet, nombre d'articles publiés sur des sites de presse en ligne, fondés dans bien des cas sur des dépêches d'agence, sont reproduits à une échelle de masse par un grand nombre de médias en ligne, d'agrégateurs, de portails, de réseaux sociaux ou de blogs (Rebillard, 2006). Il s'agit là d'un phénomène bien étudié par la sociologie des médias, à savoir l'effet d'agenda qui postule que les médias ont tendance à se centrer sur des sujets identiques, au même moment et en des termes comparables (Dearing

& Rogers, 1992). Par conséquent, l'envergure du spectre des sujets abordés par les médias et la diversité de leur traitement peuvent constituer un indicateur de pluralisme politique dans la mesure où « la position d'un problème sur l'agenda des médias détermine surtout le poids de cette question sur l'agenda public » (Dearing & Rogers, *op. cit.*, p. 92). Cependant, si l'hypothèse de l'effet d'agenda a été largement testée dans le domaine des médias traditionnels, il n'en est pas de même pour les médias en ligne. La difficulté tient non seulement à la multiplicité des espaces de production et de publication de l'information de presse sur le web, mais également à la multitude des espaces de discussion et de commentaire qui lui sont associés.

En outre, par rapport aux médias traditionnels, un tel environnement médiatique complexe est caractérisé par deux autres aspects particuliers. Premièrement, l'énorme quantité d'information quotidienne produite et reproduite en ligne par un large éventail d'entités provoque une situation de surproduction — une tendance qui est intrinsèque aux industries culturelles, mais qui a été exacerbée sur le web (Hesmondhalgh, 2007). Deuxièmement, cette information est systématiquement numérisée, stockée et traitée dans des systèmes composés d'ordinateurs et de réseaux. Cette offre de contenu numérique est ensuite fournie à des millions d'utilisateurs dans le monde entier à travers une multiplicité de canaux et d'outils (flux RSS, moteurs de recherche, réseaux sociaux, portails, blogs, etc.) qui permettent une interaction sociale complexe autour de l'actualité — de la simple transmission à son réseau social à la transformation de l'information en profondeur (Im *et al.*, 2011). Ainsi, si l'on vise une étude approfondie et complète d'une question aussi complexe et difficile à saisir que celle de la diversité de l'information en ligne, les méthodes traditionnelles des sciences sociales s'avèrent insuffisantes. Afin de tirer parti des vastes quantités de données disponibles sur l'internet, que ce soit du côté de la production mais aussi de la consommation des *news*, il est nécessaire de combiner une approche ancrée dans les sciences humaines et sociales avec des méthodes de collecte et de traitement automatisées, comme l'analyse de contenu médiatique assistée par ordinateur (Riffe *et al.*, 2005) et les « digital methods » (Rogers, 2009) qui permettent d'exploiter les « traces » laissées par les internautes.

Le projet IPRI

En 2010 et 2011, nous avons participé à un projet de recherche appelé IPRI (Internet, Pluralisme et Redondance de l'Information) visant à mesurer, à travers une étude transdisciplinaire, la diversité des contenus d'actualité en ligne en France. L'aspect principal du projet a porté sur une analyse quantitative : nous avons utilisé le logiciel IPRI News Analyzer (IPRI-NA)³ pour collecter et traiter automatiquement les titres des articles en provenance de plusieurs dizaines de sites d'information publiant informations et/ou commentaires sur l'actualité générale et politique, à caractère essentiellement national. Après collecte et élimination des articles hors critères, le corpus comprend 37 569 articles publiés entre le 7 et le 17 mars 2011 par 199 sources réparties en quatre catégories : médias en ligne (presse et audiovisuel), *pure-players* (uniquement présents sur le web), portails/agrégateurs (télécommunications et services informatiques) et blogs.

³ IPRI-NA est un logiciel libre disponible à l'adresse <http://liris.cnrs.fr/ipri/pmwiki/index.php?n=Public.IpriNA>. Il a été développé par Samuel Gesche et Elöd Egyed-Zsigmond dans le cadre du projet.

Afin de tester la prégnance de l'effet d'agenda sur le journalisme en ligne, nous avons ensuite élaboré une méthode de classification semi-manuelle des articles à l'intérieur des « sujets d'actualité »⁴ à partir de deux mesures : la variété et l'équilibre (Benhamou, Peltier, 2006). La variété correspond au nombre de sujets au sein de l'échantillon considéré, l'équilibre est pour sa part déterminé par la répartition des articles entre les différents sujets. Parallèlement, nous avons exécuté une opération similaire sur Twitter. Notre objectif était de repérer des divergences ou des convergences entre l'agenda dominant des médias en ligne et les contenus les plus partagés et discutés sur ce réseau social par un échantillon d'utilisateurs français. Pour ce faire, nous avons constitué un échantillon de comptes Twitter appartenant à des utilisateurs particulièrement intéressés par l'actualité et nous avons aspiré les messages et les interactions qu'ils ont produits sur une même période de onze jours. Le jeu de données final, constitué du 7 au 17 mars 2011, se compose de 14 627 comptes d'utilisateurs ayant produit 1 145 135 messages, ainsi que de toutes les connexions entre eux.

En troisième lieu, une analyse qualitative des articles de presse en texte intégral a été menée afin d'identifier les différentes stratégies éditoriales et routines linguistiques mobilisées dans le traitement des mêmes faits d'actualité par différentes catégories de médias. Cette analyse s'est concentrée sur deux sujets particuliers largement traités lors de la journée du 8 mars 2011 : la révolution libyenne (178 articles dans le corpus IPRI-NA) et la publication d'un sondage donnant Marine Le Pen en tête des intentions de vote pour l'élection présidentielle (165 articles). Enfin, une étude socio-économique a été réalisée visant à analyser les conditions concrètes de production de l'information en ligne à travers une articulation de l'analyse des modèles d'affaires avec l'observation des pratiques journalistiques mises en œuvre par les sites d'information. Notre hypothèse a été la suivante : le degré de dépendance de la ressource publicitaire est un facteur déterminant de l'inscription ou pas d'un site d'information à l'agenda dominant et, par extension, des pratiques journalistiques qu'il met en œuvre. Au lieu de nous intéresser aux grands médias, nous nous sommes au contraire concentrés sur un échantillon de pure-players indépendants⁵ qui tentent de s'affranchir de l'agenda dominant imposé par la course à l'audience et de proposer une information différenciée.

Un aperçu des résultats initiaux

De manière générale, le projet IPRI a confirmé les résultats des travaux que nous avons pu mener antérieurement (Smyrnaioi *et al.*, 2010). Il apparaît à travers notre observation que si l'information en ligne est diverse de par le grand nombre de sujets d'actualité couverts, elle fait également l'objet d'une forte concentration autour d'un petit nombre de sujets dominants. Ce phénomène est particulièrement visible dans le cas des événements qui ont eu lieu au Japon pendant la période de notre observation (tremblement de terre, tsunami, accident dans la centrale nucléaire de Fukushima). C'est aussi le cas, dans un moindre degré,

⁴ Par *sujet*, nous désignons une expérience factualisée, telle que définie par Jean-Pierre Esquenazi (2002) : la « factualisation de l'expérience » revient « dans le temps et l'espace, à délimiter, circonscrire le fait en le présentant comme une discontinuité ». En ce sens le sujet préexiste à son traitement journalistique qui en fait un article ou un reportage.

⁵ Owni, Arrêt sur images, Electron libre, Rue89 et Rezo.net. Entre mai 2010 et août 2011, nous avons effectué neuf entretiens semi-directifs avec des journalistes ou des responsables de cinq sites ainsi que des observations au sein des rédactions quand cela a été possible.

du conflit en Libye. Ces deux sujets d'actualité représentent à eux seuls presque le tiers de la totalité des articles produits par notre échantillon de sites sur une période de onze jours.

Dans le même temps, l'analyse qualitative montre que s'il existe sur l'internet une grande diversité des registres discursifs dans le traitement d'une information, on trouve malgré tout des récurrences qui correspondent à des catégories de sites spécifiques. On observe en particulier une dichotomie entre, d'une part, les sites à forte audience (médias traditionnels, portails et agrégateurs) et, de l'autre, les blogs et les *pure-players*. Pour décrire cette dichotomie à grands traits, nous pouvons dire que les premiers semblent adopter un traitement de l'actualité très dépendant des sources officielles (agences de presse, et notamment l'AFP, communiqués, etc.) ainsi qu'un langage neutre et descriptif. Cela apparaît dans leur utilisation intense du discours rapporté et des citations. Globalement, il s'agit d'une information normalisée et formatée qui correspond aux standards professionnels des grands médias. Les blogs et les *pure-players* de leur côté sont avant tout caractérisés par une extrême hétérogénéité des propos de fond, à laquelle répond également une hétérogénéité des stratégies énonciatives et des styles.

Ces résultats sont confirmés par les conclusions du volet socio-économique de notre enquête qui montrent qu'il existe bien une corrélation entre le degré de dépendance envers la ressource publicitaire et le traitement formaté de l'agenda médiatique dominant. La principale préoccupation des sites que nous avons étudiés est de réussir à mettre en place un modèle d'affaires rentable tout en s'affranchissant, en partie ou complètement, de la contrainte publicitaire et de l'obligation d'une audience de masse qui lui est associée. Ainsi, Rue 89, Arrêt sur images et Owni misent sur une diversification des sources de revenus en faisant appel aux abonnements ou à la vente de prestations aux entreprises. De même, les sites en question testent des modèles journalistiques innovants (rédaction décentralisée, information participative, curation de la blogosphère, utilisation intensive des réseaux sociaux, journalisme de données) afin de compenser le manque de moyens qui les caractérise. Ils tentent ainsi de se démarquer des sujets qui dominent l'actualité et des formats normalisés qu'on trouve sur les sites des grands médias. Enfin, à la différence de ces derniers, les relations entre journalistes et lecteurs des sites que nous avons étudiés — mais aussi les relations entre lecteurs eux-mêmes — sont caractérisées par une proximité élevée qui va parfois jusqu'à établir des communautés participant à la construction collective de l'information.

Enfin, l'étude des échanges sur Twitter renseigne sur la manière dont une partie du public s'approprie les contenus d'actualité dans le cadre de sa sociabilité en ligne. La première remarque qu'on peut faire à ce sujet est qu'il existe une forte corrélation entre l'intensité de l'agenda médiatique et la productivité des utilisateurs de Twitter. Autrement dit, plus les médias produisent des contenus d'actualité sur un sujet donné, plus les internautes en parlent sur les réseaux sociaux. On observe également un mouvement cyclique : les utilisateurs de Twitter citent les sites d'information les plus populaires, renforçant par là même cette popularité. Finalement, ce qui distingue Twitter par rapport à l'offre des sites d'information est l'importance des sources d'information spécialisées dans le domaine de la technologie ainsi que celle des sources étrangères. Le même phénomène est observé quant à la répartition des sujets d'actualité. En effet, les sujets les plus discutés sur Twitter sont souvent les mêmes que ceux qui dominent l'agenda médiatique en ligne. Cependant, un examen plus détaillé révèle un phénomène intéressant : des articles isolés, correspondant à

des sujets politiques très peu ou pas du tout repris par les principaux médias en ligne, ont un succès conséquent sur Twitter. Cette tendance semble indiquer qu'un nombre non négligeable d'internautes trouve sur ce réseau social des informations de nature politique délaissées ou même ignorées par les grands médias.

Conclusion

Si nous voulons tirer quelques conclusions relatives à la question de l'éducation aux médias à partir des résultats de notre projet de recherche, une chose semble évidente : le paysage médiatique en ligne est beaucoup plus complexe et difficile à saisir dans sa totalité que celui des médias traditionnels. En effet, une très grande diversité des sources et des informations y cohabite avec une forte redondance et une inflation de contenus produits et formatés selon des méthodes industrielles. Nous constatons alors sur l'internet ce que Pierre Bourdieu a appelé la « circulation circulaire des informations » (Bourdieu, 1996), à une échelle quantitativement beaucoup plus massive que celle qu'on trouve sur les médias traditionnels. De plus, les études des usages de l'information en ligne que nous avons citées précédemment (Granjon & Le Foulgoc, 2010, Cevipof, 2010 mais aussi Comby *et al.*, 2011) montrent que, contrairement à une idée reçue, le pluralisme offert réel de l'internet ne correspond pas forcément à un pluralisme consommé. Autrement dit, malgré la grande richesse des contenus disponibles en ligne, la majorité des internautes se concentre sur un petit nombre de sites à forte audience. Cela est également vrai pour les fameux *digital natives* : un grand nombre d'entre eux développe des usages informationnels qui se limitent à un éventail restreint de sources d'informations composé de multinationales comme Google, Facebook, Orange ou Microsoft et des groupes de médias établis. En fait, ce sont les inégalités économiques et d'éducation qui caractérisent le public qui sont déterminantes dans ce contexte d'usage.

Pourtant, le pluralisme de l'information en ligne existe, et il est fort. Des *pure-players* jusqu'à la blogosphère et aux réseaux sociaux, le choix est grand. Néanmoins, profiter de ce potentiel nécessite un investissement cognitif important qui se traduit par du temps et de l'énergie consacrés à la découverte de sources pointues et par la mise en œuvre d'outils sophistiqués. La capacité d'optimiser la recherche d'information, d'automatiser les processus de veille, d'exploiter le potentiel des liens faibles dans les réseaux sociaux est désormais essentielle pour comparer le traitement d'un même événement par plusieurs médias et par la même comprendre les enjeux de notre monde. Les internautes, jeunes et moins jeunes, ne sont pas tous en position de développer ces types d'usage complexes par eux-mêmes, loin de là. Il appartient donc aux éducateurs que nous sommes de les assister, ce qui implique notre propre investissement approfondi dans le monde de l'information numérique.

Bibliographie

Anderson C. (2006). *The Long Tail: Why the Future of Business is Selling Less of More*. New York: Hyperion.

Bangemann, M. (1994). *Europe and the Global Information Society*, Recommendations to the European council. Brussels: EU.

Benghozi J.P., Benhamou F. (2010). « The Long Tail : Myth or reality ? ». *Marketing Management*, volume 12, number 3, p. 43-53.

Benhamou, F., Peltier, S. (2006). « Une méthode multicritère d'évaluation de la diversité culturelle: application à l'édition de livres en France ». In: Greffe X (ed.) *Création et diversité au miroir des industries culturelles*. Actes des journées d'économie de la culture. Paris: La Documentation Française, p. 313-344.

Bourdieu, P., (1996). *Sur la télévision*. Paris : Liber.

Brynjolfsson, E., Hu, Y.J., Smith, M. D. (2006). « From Niches to Riches: Anatomy of the Long Tail ». *Sloan Management Review*, Vol. 47, No. 4, p. 67-71.

CEVIPOF (2010), *Mediapolis - Résultats de l'enquête* [<http://www.cevipof.com/fr/mediapolis/rapport/>]

Comby, J.B., Devillard, V., Dolez, C., Rieffel, R., (2011). « Les appropriations différenciées de l'information en ligne au sein des catégories sociales supérieures », *Réseaux* 29(170), p. 75–102.

Dearing, J.W., Rogers, E.M. (1992). *Communication Concepts 6: Agenda-Setting*. Thousand Oaks, CA: Sage.

Elberse, A., Oberholzer-Gee F. (2008). « Superstars and Underdogs: An Examination of the Long Tail Phenomenon in Video Sales ». *Harvard Business School Working Paper* No. 07-015.

Fenton, N. (2009). *New Media, Old News: Journalism & Democracy in the Digital Age*. London: Sage Publications Ltd.

Esquenazi Jean-Pierre. (2002). *L'écriture de l'actualité. Pour une sociologie du discours médiatique*, Grenoble :PUG.

Habermas, J. (1991). *The Structural Transformation of the Public Sphere: An Inquiry into a category of Bourgeois Society*. Cambridge, MA: MIT Press.

Hesmondhalgh, D. (2007). *The Cultural Industries*. London : Sage Publications.

Hindman M. (2009). *The Myth of Digital Democracy*. Princeton, NJ and Oxford: Princeton University Press.

Granjon F., Le Foulgoc A., (2010). « Les usages sociaux de l'actualité. L'expérience médiatique des publics internautes », *Réseaux*, n°160, p. 225-253.

Im, Y. H., Kim, E. M., Kim, K., Kim, Y. (2011). « The Emerging Mediascape, Same Old Theories? A Case Study of Online News Diffusion in Korea ». *New Media & Society*, available online

ahead of print on March 2011 at
[<http://nms.sagepub.com/content/early/2010/12/16/1461444810377916.abstract>]

Lancelot, A. (2005). *Les problèmes de concentration dans le domaine des médias*, Rapport pour le Premier ministre. Paris: La Documentation française.

Mansell, R. (2004). *Political economy, power and new media*. London: LSE Research Online.

Paterson C.A., Domingo D. (2008). *Making Online News: The Ethnography of New Media Production*. New York :Peter Lang.

Rebillard F., (2006). « Du traitement de l'information à son retraitement. La publication de l'information journalistique sur l'internet », *Réseaux*, n°137, p. 29-68.

Riffe, D., Lacy, S., Fico, F.G. (2005). *Analyzing Media Messages: Using Quantitative Content Analysis in Research*, London: Taylor & Francis, 2nd Edition.

Rogers, R. (2009). *The End of the Virtual: Digital Methods*, Amsterdam: Amsterdam University Press.

Smyrnaiois, N., Marty, E., Rebillard, F. (2010). "Does the Long Tail apply to online news? A quantitative study of French-speaking news websites". *New Media and Society* 12 (8), p. 1244-1261.

Smyrnaiois N., Rebillard F., (2011), « Entre coopération et concurrence : Les relations entre infomédiaires et éditeurs de contenus d'actualité », *Concurrences*, N° 3-2011, n°37281.

Tessier M., Baffert M., (2007). *La presse au défi du numérique*, Rapport pour le Ministre de la culture et de la communication.

UN, (2005). *Tunis Agenda for the Information Society*, World Summit on the Information Society (WSIS), Mars 2011
[http://www.itu.int/wsis/documents/doc_multi.asp?lang=en&id=2267|0]