ONLINE NEWS BETWEEN REDUNDANCY AND DIVERSITY: A QUANTITATIVE STUDY OF FRENCH SPEAKING NEWS WEBSITES

Emmanuel Marty,
University of Toulouse 3, France
e-mail: emmanuel.marty@iut-tlse3.fr
Franck Rebillard,
University of Lyon 2, France,
e-mail: franck.rebillard@univ-lyon2.fr
Nikos Smyrnaios,
University of Toulouse 3,
e-mail: France, smyrnaios@gmail.com

ABSTRACT

On the internet, the pluralism of information is supposed to result from the multiplicity of sources. The web is expected to offer a wider range of contents than offline media (as developed in the Long Tail theory). But such an assumption requires to be proved by empirical evidence. Our research aims to test this hypothesis on the basis of a transdisciplinary quantitative study of French speaking websites. Based on several thousands of articles coming from different categories of websites, the editorial identification of topical issues and the lexicometric analysis of the titles both highlight a more complex situation. The wide diversity of issues dealt with, during a day on the web, simultaneously presents a high concentration of a few major issues, often treated in a redundant way. These results show that the ideal of pluralism that the internet is supposed to embody needs to be put into perspective.

Keywords : online news, journalism, Long Tail, internet, content analysis.

1. INTRODUCTION

Since the late 90's, when the web began to confirm its position as a mass media worldwide, much of the scientific literature as well as public policy reports on the subject have considered the new media as the paragon of cultural and informational diversity. Indeed, the open "nature" of the Internet Protocol and the easiness of online content distribution have progressively established the web as one of the main means of massive public expression. From an economic point of view, this trend is often seen as a way for marginal cultural and informational products to reach a larger public than they do through the bottleneck of traditional offline distribution channels. This aspect was popularized by Chris Anderson and his Long Tail theory (2006). Paradoxically, in spite of its popularity, the Long Tail theory has brought about few scientific studies. Those who actually tried explicitly to test its validity on an empirical basis have produced mixed results (Elberse, Oberholzer-Gee, 2007; Benghozi, Benhamou, 2008).

Yet the question of diversity is a major political and social issue. This is particularly true when it comes to news and journalism, whose role in the public sphere is undoubtedly central. Despite the assumption that the internet offers citizens a much wider spectrum of opinions than traditional media, facts seem to be more complicated. The agenda-setting effect has not disappeared on the web. Rather, it has become more complex by the emergence of numerous intermediaries in the news

distribution circuit. Search engines, aggregators, portals, digg-like platforms, even individual blogs act as infomediaries inside the online media sector by distributing large amounts of journalistic content. Nevertheless, this content is essentially redundant and originates from press agencies and traditional media corporations. The actual process of dissemination of redundant news among professional media themselves seems to have accelerated dramatically online as well.

In this light, the quantitative growth of online information circulation does not necessarily mean that online news is more diverse in journalistic terms, or that the spectrum of social, political and economic issues covered on the internet is much wider than that of its offline counterparts. Such an assumption needs to be proved by empirical evidence. Our research aims to test this hypothesis on the basis of a transdisciplinary quantitative study of French speaking websites[1]. Our approach examines the total production of more than eighty websites that provide online news over a given period of time. It is based on a semi-automatic extraction and treatment process, completed by a "manual" content analysis. The study includes three levels of analysis. Firstly, we will describe our research method and its findings. Then we will discuss the findings in relation with our theoretical frame.

2. A FIRST OVERVIEW OF ONLINE NEWS PRODUCTION

One major difference between the internet and traditional media is the great variety of websites that are involved in the process of producing and broadcasting news. Alongside newspapers, radio stations and TV channels there is a multitude of newcomers that one could qualify as pure-players. For example, Mark Deuze classifies news websites according to their position on a *continuum* which extends from a focus on producing original news to an activity of connecting users to a variety of external content (Deuze, 2003). He also adds a second variable, related to the participatory aspect of online journalism, which is the degree of editing and moderating user generated content. Then the author distinguishes four main categories of journalism sites : mainstream news sites, index and category sites, meta and comment sites, share and discussion sites. This typology helped us build a renewed and exhaustive description of the online news arena (Rebillard, 2006) which revealed a complex set of relations between seven categories of websites : newsmedia and press agencies digital outlets; weblogs; webzines; participatory journalism websites; portals and aggregators.

Therefore, if one aims at an analysis of web journalism as a whole, it is necessary to take into account a representative sample of this variety of players. Academic research on the subject has mainly focused on comparing two categories at a time, such as professional publications and amateur weblogs (Reese et al., 2008) or traditional newsmedia websites and aggregators (Thurman, 2007). In order to address the question in a more comprehensive manner we included in our study a large sample of more than eighty French-speaking news sites which cover the entire spectrum of

---

the seven categories listed above. At least five sites of each category were included in our sample. From now on, we will refer to these websites as *sources*.

We then applied an automated crawling method to the selected sources using their RSS feeds in order to extract two sorts of data: the titles of the articles and their descriptions. The crawling took place in 2008 between November 1 and 20, on a 24 hour basis. The gathered data were then processed in order to extract lemmas, that is the canonical form of a lexeme[2]. This operation helped us obtain a panoramic view of the news agenda over that period.

Table 1

**The most frequent lemmas between November 1 and 20 (2008)**

Lemma	countTitle	Lemma	countDesc
obama	**2825**	avoir	28463
être	2197	être	20923
A	2113	**président**	**3982**
france	1562	**obama**	**3708**
avoir	1545	premier	3676
Ps	1396	source	3265
sncf	1143	**barack**	**3027**
royal	992	mardi	3013
crise	947	faire	2977
Frequency of top ten lemmas in the article titles		Frequency of top ten lemmas in the article descriptions	

As we can see in Table 1, except for two ordinary verbs ('être' = to be ; 'avoir' = to have), the most frequent lemma in both the titles and the descriptions is *obama*. Some other frequent lemmas refer to French politics (*'ps' for Parti Socialiste, 'royal' for Ségolène Royal*), to the French national railroad company (*'sncf'*) as well as to the financial crisis (*'crise'*). This indicates that the U.S. President election dominated the agenda of our sample of French-speaking news websites. The intensity of news production related to Barack Obama is also visible in the graph of the repartition of the lemma *obama* through our sample period.

---

[2] e.g. in English, run, runs, ran and running are forms of the same lexeme, with run as the lemma.

Table 2

Frequency of *obama* between November 1 and 20 (2008)
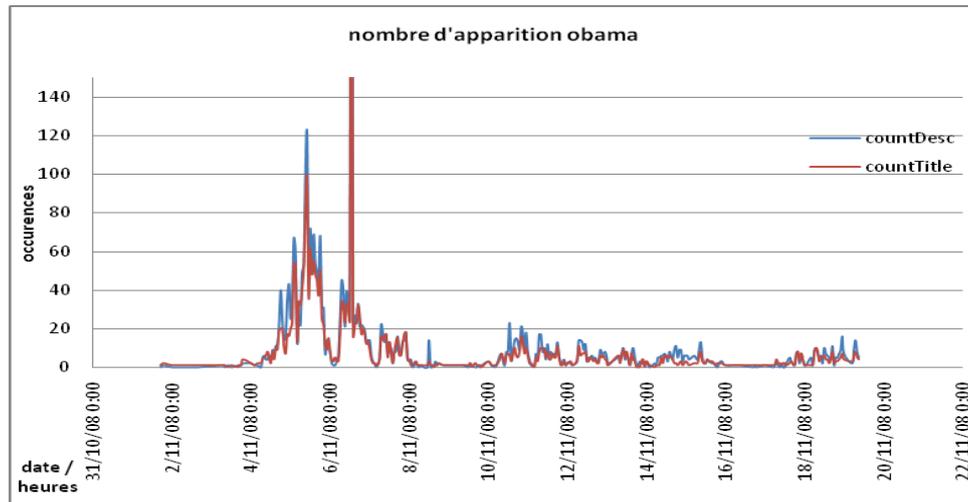
We observe that news production related to Barack Obama starts growing rapidly from election day and on. It presents an enormous peak on November 6 when comments on election results multiplied in European media. Some other smaller peaks coincide with events such as the fist press conference of Obama as President elect (November 7-8), his first visit to the White House (November 10-11) and so forth. This first level of analysis of the collected data gives us some evidence that the agenda effect remains quite strong on the web. Nevertheless, a second level of analysis is obviously necessary in order to access a more detailed view.

3. QUANTIFYING PLURALISM: VARIETY AND BALANCE AMONG TOPICS

Thereafter we focused on the news that were published by our sources sample during two full days, the 6th and the 10th of November 2008. The data that we gathered include 61 sources for 2 617 titles on the 6th and 60 sources for 2 040 titles on the 10th of November[3]. The choice of those two days was motivated by our intention to compare two periods of different informational density: November 6 was dominated by the election of Barack Obama to the U.S. presidency that took place two days earlier. News production of the first day of our study was expected to be dominated by stories related to the Obama election. The production of the second day was expected to be less focused on that subject and therefore more diverse.

In order to examine the degree of diversity of our corpus of articles we classified them according to their topic. Our definition of a "topic" is that of an event that occurred in a specific spatiotemporal context. A "topic" becomes a "story" or an article after it has been recounted as such by journalists (Esquenazi, 2002; Ringoot and Rochard, 2005). A topic is much broader than a story, in the sense that it can be

---

[3] Our method consisted in extracting automatically all the titles of the articles that where published by the selected websites on a period of 24 hours. The reason why there are 61 websites for the first day and only 60 in the second one is that some of our sources (mostly weblogs) did not publish a new entry on both days.

approached through different angles or mental frames, but still refers to the same facts. By proceeding this way, we can embrace the spectrum of issues evoked by French-speaking websources during a given day. At last, this classification gives us the opportunity to measure the number of articles dedicated to each one of them.

We carried out the operation of classification through a semi-automated and inductive method. Firstly, we applied on our corpus of article titles a software of lexical analysis, Lexico, in order to identify repeated phrase segments. This step allowed us to spot the most redundant topics in our data. The counting of single words, mostly names of individuals or countries, was also used to isolate some frequent topics that were not visible in repeated phrase segments because the wording of titles differed from one source to another.

In a second step, after eliminating duplicates, it was necessary to carry out a manual categorization of titles into less frequent topics. Such a method raises the question of arbitrariness. Nevertheless, this solution appeared to be the least biased since the nature of editorial content does not allow a purely automated categorization. This initial treatment led to the creation of a database including three parameters: the name of the web source, the article titles that it published and the topic to which each article referred to.

In order to measure the pluralism of information in our data, we applied to the case of journalistic content formulas that have already been tested in other sectors of the cultural industries (Benhamou and Peltier, 2006). The two main criteria we used are those of variety and balance. Variety in this case depends on the number of topics that we isolated in our sample of titles. Balance depends on the number of titles per topic.

Before exploring the results in details, let us mention that news production was variable for each source. At the top of the scale, aggregators and portals such as Wikio, MSN and Canoë broadcast several hundreds of articles on a daily basis. At the bottom, the production of personal weblogs is quantitatively very limited. Between these two extremities, we found a range of sources (online press, agencies, webzines and participatory sources) whose daily production ranges between 10 and 100 articles. The differences in production rate are important according to the categories of source, and the weight of a few players in the entire sample is extremely important (MSN and Wikio alone weigh about 20 % of the sample).

Analysis of the first sample of titles (November 6)

On that day, 2617 titles were extracted from 61 sources. 385 different topics were identified. There is an average of 7 titles by topic, with vast differences between extremities. The median is 1, and 301 of the 385 topics identified include less than 5 titles each. The five most frequent topics include more than 100 titles each.

Table 3

The most frequent topics of the November 6 sample

Topic	Number of articles	% of sample

Obama's election	435	16,62%
The congress of the French Socialist party	187	7,15%
International institutions facing the financial crisis	178	6,80%
Stock market	121	4,62%
Rail strike in France	120	4,59%

Analysis of the second sample of titles (November 10)

On that day, 2040 titles were extracted from 60 sources. 309 different topics were identified. There is still an average of 7 titles by topic, and the differences between extremities remain huge, albeit they are a little less important than in the sample of November 6. That is because, in this sample, there is no extremely dominant topic like the election of Barack Obama. The median is also close to 1 and 236 topics out of 309 include less than five titles each. On the other extremity, the five most frequent topics include more than 80 titles each. This concentration, although lower than the one of November 6, is still quite high.

Table 4

The most frequent topics of the November 10 sample

| Topic	Nb articles	% of sample
The French Socialist Party after the congress	129	6,32%
Obama and Bush meeting in the White House	114	5,59%
Sabotage in French rail	112	5,49%
Goncourt (litterature prize) awarded toAtiq Rahimi	108	5,29%
The situation in République Démocratique du Congo	83	4,07%

Both samples have very similar characteristics: a wide variety of topics that are very unequally treated by sources. Some topics include substantial amounts of titles (several hundreds), while the majority of topics include a single article. The production of information as we have measured it on the web displays a relatively constant structure made of an extreme concentration of articles on a small number of topics and, at the same time, a dispersion of the remaining articles on a high number of isolated topics.

Despite the extraordinary media coverage of the election of Barack Obama in the November 6 sample (approximately 17% of the articles), there is a very similar structure in both samples. The graph of titles balance for the November 6 sample reveals a very unequal structure, close to the Pareto distribution following the

classical 20/80 rule: 20% of the topics include 82% of the articles and 10% include 71%. The concentration is almost identical in the sample of November 10 2008 : 20% of the topics include 81% of the articles and 10% include 71% of them.

Table 5

Repartition of articles per topic

Repartition of articles per topic
(November 6th, 2008)

Repartition of articles per topic
(November 10th, 2008)
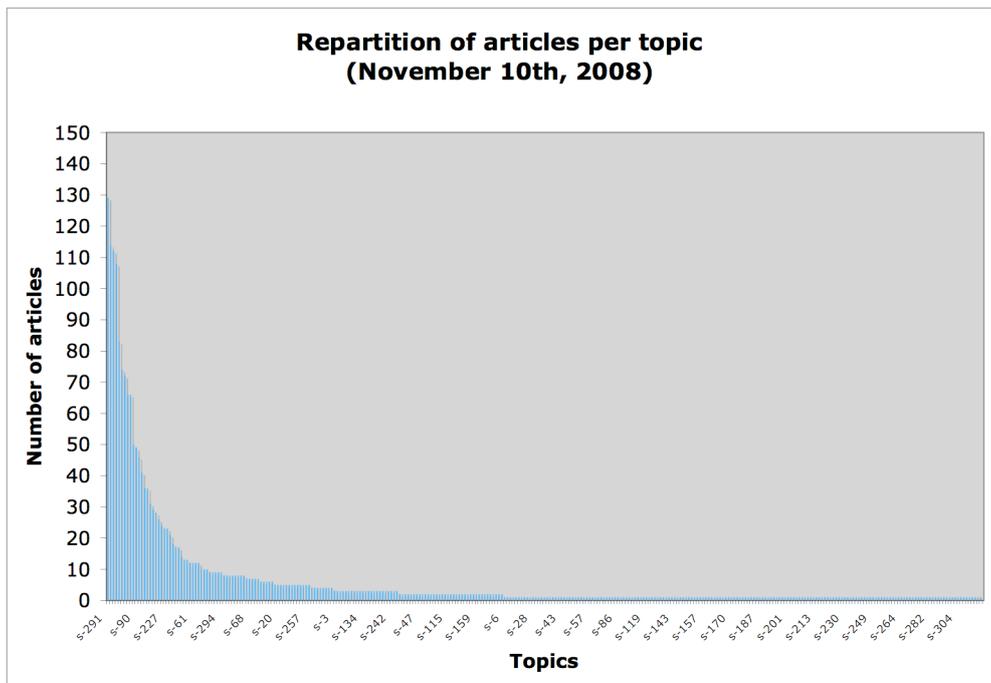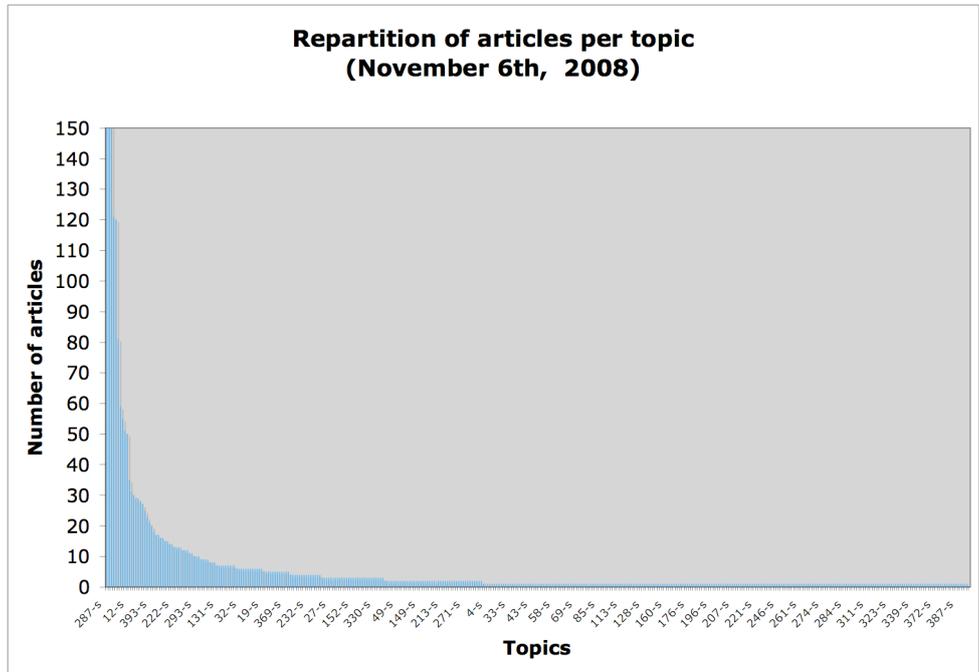
By including French-speaking sources from North America, Africa, Europe and Asia we wanted to take into account the international dimension of the internet in

our analysis. Only this could create an artificial fragmentation of the topics variety because of sources referring to heterogeneous media agendas (e.g. ice hockey competitions which are very present on the Quebec sites).

A reduction of the sample to French sites only was carried out in order to test the potential impact of such a bias. Analysis of the reduced sample confirms the permanence of the structure already observed : a wide variety of subjects divided between extreme concentration and high dispersion. Indeed, on each of our sample days the first five topics are identical between French and French-speaking sources and the magnitudes are very similar : the median remains at 1 and the 20/80 rule is verified.

This first quantitative analysis has provided a numerical representation of the French language news production on the internet based on samples of consistent data (over 2000 articles per day collected on more than sixty sites). News appears to be both varied (more than 300 different topics were covered each day) and very unevenly distributed (classical rule of 20/80). On top of that, many issues emerge through a single article or a handful of articles, so that pluralism (high variety of topics) coexists on the web with some strong redundancy of information (extreme concentration on dominant topics).

4. LEXICAL AND EDITORIAL REDUNDANCY AMONG SOURCES

The main indication of the two initial levels of analysis is that the phenomena of pluralism and redundancy coexist on the web. Based on the concepts of variety (number of topics in the sample) and balance (number of titles per topic), we obtained a comprehensive panorama of the global news production in a given period. However, a third level of analysis is necessary in order to integrate a key factor : the specificities of the different categories of sources. Having identified and quantified the variety and balance - or more precisely the imbalance - of information, the next step is to distinguish the players. The challenge is to be able to locate sources on a *continuum* ranging from actors of redundancy to actors of pluralism.

In order to do so, we used two different measures to examine both the degree of diversity and of redundancy for each source. The first measure was based on repeated segments of text that we isolated in our database of titles using Lexico. The existence of frequent repeated segments in the titles of articles is a mark of redundancy. It results from the repetition of the same stereotypical phrases in the titles that were published by our sources sample. For example, news websites tend to republish press agency material under the same original title. Although title resemblance does not always mean that the content of the articles is identical, it is a strong sign of editorial redundancy. The second measure was based on *hapax*, that is words that appear only once in our database. The existence of *hapax* in titles is a mark of lexical diversity. The lexical diversity indicates some degree of editorial pluralism, although this is not always true given that journalists can use different words to cover the same topic. The combination of those two measures provides additional evidence to the issue of diversity and redundancy from the point of view of the production of original titles (or, on the opposite, of the recurrence of their wording). Thus, the third level of analysis takes into account the degree of diversity of the language used by news websites.

Table 6

Repeated segments of at least ten words for November 6

Segment	Frequency of the repeated segment	Number of titles for corresponding topic
Japon des tissus cérébraux créés à partir de cellules souches	28	28
Obama le premier président noir élu face à des défis colossaux	27	435
perpétuité pour le beau père 30 ans pour la mère	23	51
Les marchés replongent malgré une baisse des taux en Europe	21	121
Enfant mort de coups perpétuité pour le beau père 30 ans	19	51
Obama met en place l' équipe chargée de préparer la transition	18	435
La BCE abaisse ses taux et garde des cartouches pour décembre	18	178
Ligue des champions Lyon en tête de son groupe la Juventus	17	81
Le FMI annonce pour 2009 la 1ere récession dans les pays	15	178
Discrimination la Halde recommande de lutter via les programmes et manuels	15	31

Given these results, we can observe that topics that concentrate large numbers of titles are also those that present higher "lexical banality", that is "a massive recourse to the most common words" (Marchand, 2008). Instead of receiving a more diverse treatment, dominant topics seem to be also those that concentrate many stereotypical phrases, at least as far as the titles are concerned. On the basis of that finding we can deduce that there is a strong correlation between editorial and lexical redundancy. In other words, the more a topic dominates the news agenda on the web the more it is likely to generate recurrent wording. The election of Barack Obama is a good example of that trend. While the topic has generated a huge amount of articles, it is mainly addressed through stereotyped formulas (see Tables 6 and 7). We can assume the existence of a link between dominant topics and dominant sources,

characterized by an intense activity of distribution of content and, at the same time, a high degree of editorial and lexical redundancy.

Table 7

Repeated segments of at least ten words for November 10

Segment	Frequency of the repeated segment	Number of titles for corresponding topic
attentat à Bagdad fait 28 morts et des dizaines de blessés	46	66
Le Goncourt à l Afghan Atiq Rahimi le Renaudot au Guinéen	33	108
SNCF nouvel acte de malveillance le parquet anti terroriste saisi	23	112
Etats Unis Obama doit rencontrer Bush à la Maison Blanche	23	114
l UE lance une opération navale historique contre les pirates	21	36
Gilles Simon bat Roger Federer son 2e succès face au Suisse	19	72
les deux camps se font face dans l est sans affrontement	19	83
Le plan de relance chinois applaudi par des marchés en nette	18	74
Partenariat avec la Russie l UE prête à reprendre les négociations	16	49
PS Ségolène Royal réunit ses représentants lundi après midi au Sénat	15	129
Le plan de relance chinois et le G20 offrent un bref	15	74

Over the two days, the sources that appear to be the most redundant in terms of both choice of topics and use of stereotypical titles are the three portals, MSN News, Yahoo News, Orange News, and the AFP press agency. This trend is due to the policy of the portals that publish newswire content in a continuous flow and focus on reactivity rather than creativity. From a quantitative point of view, these players occupy a central position in the French speaking online news sector because they broadcast high volumes of information and aggregate large audiences. However, from a qualitative point of view their content is redundant and stereotypical. Other sources

that appear to be very close to the model proposed by the AFP are the sites of French television, especially France 2 and France 3 (TF1 and France 24 appear to be less redundant in comparison), as well as RTL radio station.

On the diversity side, the most notable sources are all situated outside France: Africa Online, Canoë (Canada) and Ria Novosti (Russia). This results from the fact that French sources were the majority of our sample and thus the French news agenda was dominant in our data. Among sources in France, four types of media can be distinguished based on their production of original titles: Agoravox a citizen journalism website, Backchich a webzine on politics, The Post another participatory journalism site, and some blogs. Then come electronic versions of print media such as Le Journal du Dimanche, Le Point, Les Echos, and to a lesser extent, Libération, L'Humanité and Le Monde, which also have a fairly original lexical identity. Finally some sources appear at the crossroads of diversity and lexical banality. This position is evident for the digital version of the magazine Nouvel Observateur, but it is also true to a lesser extent for two free dailies 20Minutes, Métro, and for two radio digital outlets, RFI and RMC : these sources combine an overall under-representation of both segments and repeated hapax, which reflects a real productive activity of titles, but also the use of a very common language which is used also by other sources.

5. DOES THE LONG TAIL APPLY TO ONLINE NEWS?

After having presented our study we can now try to see to which extent the Long Tail effect apply to online news. There are two main arguments in the Long Tail theory. The first argument is that, in online markets, distribution costs of informational goods are very low compared to offline markets. According to economic theories, informational goods are costly to produce but inexpensive to reproduce. That means that they have high fixed costs and low marginal costs as far as production is concerned. On the web, this characteristic also applies to distribution (Shapiro, Varian, 1998). If we compare the traditional press industry to its online counterpart this characteristic is obvious. Distribution costs of newspapers are particularly high. This trend can be an obstacle to pluralism because it may push editors and distributors to concentrate on sectors of high profitability. Marginal newspapers and magazines that cost a lot to distribute and have low sales are disadvantaged in such a system. This is the reason why in some countries like France the state has established a cooperative system of distribution which is based on the principle of equalization of distribution costs among editors.

On the other hand, offline distribution circuits are also limited by geographical constraints. This is particularly true in the newspaper sector whose products need to be delivered very quickly. This trend means that readers situated in a given geographical area can mainly access local or national newspapers. From the editors' point of view this aspect of the newspaper industry limits their market potential. On this point, the second argument of the Long Tail theory is that the aggregation of worldwide niche markets on the web can sustain financially marginal cultural and informational products. Indeed, news websites can broadcast information worldwide with very low costs. Therefore, in theory, the Long Train effect should apply to the online news sector. Given that, editors are expected to propose on the web a wider variety of content than they do offline since online distribution characteristics allow them to aggregate niche markets. More precisely, following the Long Tail theory, the total quantity of marginal original news that is distributed (long tail) should be higher

than that of front-page redundant news (short head) given that distribution costs in both cases are similar.

However our findings do not totally confirm this hypothesis. Our quantitative study of French-speaking news websites shows that their production has quite similar characteristics to those of traditional media. News appear to be both varied and very unevenly distributed. Indeed, our study confirmed the classical rule of Pareto distribution: 20% of the most important topics of the news agenda generate 80% of the distributed articles. While the head of the news repartition graph is both thick and short as expected, the thin tail of the graph is not very consistent because many single topics correspond to only one article (seeTable 5). More importantly, there seems to be a strong correlation between editorial and lexical redundancy in our study sample. That means that a limited number of issues dominate online news and that those issues are mostly covered in a stereotyped manner.

Many factors can explain our results. First, even if our sample is very large it is not totally comprehensive, especially when it comes to marginal sources such as blogs and webzines. Consequently, part of the long tail of news production in French language is necessarily absent from our study that considers single day samples while many blogs and webzines do not necessarily publish news on an everyday basis. Other factors that may explain our results are related to the particularities of the online news sector. The Long Tail theory was built on data coming from particular sectors of the cultural industries such as music and film whose products are perennial. In the music market for example the music of The Beatles can be commercially exploited through a long period of time, a trend that reinforces the Long Tail effect. On the contrary, the utility of news is extremely ephemeral. An editorial of The New York Times from the 70's can only be of interest to a very specific public (historians, researchers) and as such is not a profitable commodity for the editor. This is a very important differentiation that makes a comparison between music, film and news sectors quite difficult.

Another factor that one has to take in account is the particular economic conditions in which the online news sector develops. Recent research on the actual conditions of production in online newsrooms shows a strong tendency toward high productivity (Estienne, 2007; Rebillard and al., 2007). Online journalism tends to privilege rewriting and republishing existing material at the expense of original reporting. An element of our study that confirms such a trend is that the most redundant sources of our sample are portals without any journalist stuff, free dailies with small newsrooms and radio and TV channels with a limited production of textual content. We can deduce that the particular conditions of production inside each news website have a major impact on its content diversity. Indeed, even if online distribution costs are lower than those of traditional distribution, the cost of producing original content on a regular basis is still quite high.

Finally, an important factor concerning the question of diversity is its double nature (Benhamou and Peltier, op. cit.). On the one hand there is the offered diversity, that is the spectrum of choices offered to the public. On the other hand there is the consumed diversity, meaning the actual choices that the public operates. Even if the existence of an offered diversity is a *sine qua non* condition for consumed diversity, it is not a sufficient one. In the case of an extreme concentration of online audiences into a small number of sources, even if the spectrum of online information is very

wide, pluralism is not effective. In order to examine this aspect, it is necessary in future research to take into account trends in online news consumption.

REFERENCES

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

Benghozi P.J., Benhamou F. (2008). "The Long Tail: digital leverage for cultural diversity?". *Note de recherché pour le Département des études, de la prospective et des statistiques*. http://www2.culture.gouv.fr/deps/fr/traine.pdf [Accessed the 12th of Mars 2009]

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

Deuze M. (2003). "The web and its journalisms : considering the consequences of different types of newsmedia online", *New Media and Society*, 2 (5): 203-230.

Elberse A., Oberholzer-Gee F. (2007) "Superstars and Underdogs : an Examination of the Long Tail Phenomenon in Video Sales". *Harvard Busisness School Working Papers Series*, n° 07-015.

Esquenazi J.P. (2002). *L'écriture de l'actualité; pour une sociologie du discours médiatique.* Grenoble : Presses Universitaires de Grenoble.

Estienne, Y. (2007). *Le journalisme après Internet*, Paris : L'Harmattan.

Marchand P. (2008). "La déclaration de politique générale a-t-elle vécu ? ", Parlement(s), Revue d'histoire politique, 10 (2) : 152-164.

Rebillard F. (2006). "Du traitement de l'information à son retraitement. La publication de l'information journalistique sur l'internet", *Réseaux*, 137 : 29-68.

Rebillard F., Cabedoche B., Damian B., Smyrnaios N. (2007). "Les mutations de la filière Presse et information " in *Diversité culturelle et mutations des industries de la culture,* Research report for the French Ministry of Culture and Communication, (MSH Paris-Nord).

Reese S. D., Rutigliano L., Hyun K., Jeong J. (2007). "Mapping the blogosphere : Professional and citizen-based media in the global news arena", *Journalism*, 3 (8) : 235-261.

Ringoot R, Rochard Y. (2005). " Proximité éditoriale : normes et usages des genres journalistiques ", *Mots. Les langages du politique*, 77 : 73-90.

Shapiro C., Varian H.R. (1998). *Information Rules: A Strategic Guide to the Network Economy*. Boston, MA : Harvard Business School Press.

Thurman N. (2007). "The globalization of journalism online. A transatlantic study of news websites and their international readers", *Journalism*, 3 (8) : 285-307.

AUTHORS

Emmanuel Marty is a Doctoral fellow and a Sessional Lecturer in Information and Communication Sciences in the University of Toulouse 3, LERASS Laboratory.

Franck Rebillard is a Lecturer in Information and Communication Sciences at the University of Lyon 2, ELICO Laboratory.

Nikos Smyrnaios is a Lecturer in Information and Communication Sciences at the University of Toulouse 3, LERASS Laboratory. www.smyrnaios.com