



**HAL**  
open science

## Qualité et différenciation des biens informationnels. Une étude exploratoire sur l'information d'actualité

Inna Lyubareva, Fabrice Rochelandet, Yannis Haralambous

### ► To cite this version:

Inna Lyubareva, Fabrice Rochelandet, Yannis Haralambous. Qualité et différenciation des biens informationnels. Une étude exploratoire sur l'information d'actualité. *Revue d'économie industrielle*, 2020, 172 (4), pp.133-177. 10.4000/rei.9593 . hal-03373544

**HAL Id: hal-03373544**

**<https://imt-atlantique.hal.science/hal-03373544>**

Submitted on 23 Aug 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Qualité et différenciation des biens informationnels.

### Une étude exploratoire sur l'information d'actualité

Inna Lyubareva\*, Fabrice Rochelandet\*\*, Yannis Haralambous\*\*\*

\*IMT Atlantique, LEGO Brest, France,

\*\* Université Sorbonne Nouvelle, IRCAV, Labex ICCA Paris, France

\*\*\*IMT Atlantique, Lab-STICC Brest, France

Cette étude s'inscrit dans le programme de recherche Pluralisme de l'information en ligne (PIL) soutenu par l'Agence Nationale de la Recherche (ANR-17-CE27-0010).

#### RESUME

Le numérique, que ce soit à travers les innovations de service ou l'usage massif des médias sociaux, influence la qualité des biens informationnels. C'est le cas notamment des biens culturels et de l'information journalistique. Or, la notion et la mesure de la qualité s'avèrent non consensuelles et renvoient à des conceptions et des attentes sociales très différentes. Dans cet article, nous proposons d'analyser la qualité de l'information journalistique à partir d'un modèle de différenciation verticale et horizontale. Nous nous basons sur les méthodes d'analyse de discours en linguistique pour caractériser les choix éditoriaux des producteurs d'informations. En comparant 93K articles de presse portant sur 31 événements entre 2015 et 2019 et publiés par 55 médias de presse représentatifs en France, nous proposons une cartographie des médias français selon leurs choix de différenciation et évaluons, sur cette base, la qualité de l'information journalistique offerte aux consommateurs.

**MOTS CLES** : qualité - différenciation - biens informationnels - information journalistique – analyse de discours

#### ABSTRACT

Digital technology, through service innovations and widespread use of social media, influences the quality of information goods. This is the case of cultural goods in general and journalistic information in particular. There is no consensual definition of quality and its measurement. In the literature, this notion often refers to different concepts and social wishes. This article proposes to use the economic model of vertical and horizontal differentiation to analyze the quality of journalistic information. Our study relies on the linguistic methods of discourse analysis to characterize the editorial choices of news producers. By comparing 93K press articles relating to 31 events between 2015 and 2019 and published by 55 representative news producers in France, we provide a mapping of French media according to their choice of differentiation and assess, on this basis, the quality of journalistic information available on the market.

**KEY WORDS**: quality - differentiation – information goods – journalistic information – discourse analysis

**CODES JEL**: L1 Market Structure, Firm Strategy, and Market Performance, O33 Technological Change: Choices and Consequences; Diffusion Processes

## INTRODUCTION

Le contexte contemporain fondé sur une numérisation des médias et l'explosion de la publication et de la consommation de contenus informationnels en ligne favorise le pluralisme des contenus et des sources d'information. Il s'avère ainsi *a priori* favorable à l'expression et à la formation d'opinions plurielles. Pour autant, ces évolutions ne sont pas synonymes en soi d'une qualité supérieure de l'information (Lyubareva & Rochelandet, 2017). Les exemples sont nombreux, à commencer par les *fake news* qui ont marqué la communication autour du Brexit, la montée des mouvements populistes de la campagne électorale de Donald Trump, les fausses rumeurs sur les trafiquants d'enfants qui ont provoqué une vague d'agression en Inde, les théories du complot qui entourent l'affaire Mélenchon, etc. Sans oublier non plus les conséquences des transformations des conditions de production (*speed-driven journalisme*<sup>1</sup>), d'usage (*snack contents* ou *fast-news*<sup>2</sup>) et de diffusion des informations, processus induits par et soumis aux données d'audiences produites par les plateformes numériques et de plus en plus filtrées par leurs algorithmes.

Dans ce contexte où les médias professionnels écrits et audiovisuels constituent un maillon essentiel du pluralisme, les acteurs politiques et industriels, ainsi que la société civile, cherchent à définir et mettre en place des instruments de régulation des contenus via l'encadrement des pratiques journalistiques, l'indépendance éditoriale, le respect des règles déontologiques et la mise en place d'outils de vérification des faits. L'un des enjeux de cette régulation est de contrôler la qualité des informations en circulation, mais aussi d'informer, d'alerter, voire « d'équiper » les différents acteurs<sup>3</sup>. Malgré l'intérêt potentiel de ces initiatives, aucune d'entre elles ne fait l'objet d'un quelconque consensus et n'arrive à s'imposer en tant que standard dans le domaine de la qualité de l'information journalistique. Faut-il évaluer les caractéristiques de l'information produite ou celle qui a le plus de succès auprès du public ? Les attentes et les objectifs de quels types d'acteurs – producteurs, consommateurs, régulateurs... - faut-il prendre comme base pour évaluer la qualité de l'information ? Comment parvenir, en l'absence de toute évidence empirique, à associer les formes de propriété des titres de presse, les pratiques journalistiques des médias et les modèles économiques à un certain niveau de la qualité de l'information produite (Karppinen, 2018) ?

Plutôt que de proposer un standard de certification de la qualité, l'objectif de cet article est d'élaborer une nouvelle grille d'analyse afin d'évaluer la qualité de l'information journalistique et la situer au sein d'une cartographie des stratégies concurrentielles des producteurs de ces biens informationnels. Dans le prolongement des travaux critiques existants (Karppinen, 2018, McQuail, 2005 ; Napoli, 2011), nous suggérons que la qualité ne réside pas dans les seules caractéristiques de l'information produite. Comme tout bien économique estimé par sa valeur, la qualité de l'information journalistique fait référence à la rencontre entre d'une part, des modèles économiques viables qui sont associés à des choix éditoriaux<sup>4</sup> et d'autre part, les besoins des utilisateurs de cette information, qu'ils soient consommateurs finals, régulateurs, politiciens, acteurs de la société civile, etc. et dont les attentes peuvent être différentes à l'égard des caractéristiques de l'information.

Dans cette perspective, notre étude se base sur le cadre théorique de la différenciation horizontale et verticale. Les caractéristiques de l'information d'actualité s'avèrent non consensuelles et renvoient à des préférences et des attentes sociales très différentes. Gabszewicz & Resende (2012) se fondent notamment sur un tel cadre d'analyse pour développer un modèle de concurrence par les prix entre les éditeurs de presse. Notre article reprend également cette distinction théorique pour cartographier

<sup>1</sup> C'est-à-dire, le journalisme axé sur la vitesse de la production de l'information.

<sup>2</sup> Les informations courtes et fragmentées qui va droit à l'essentiel, « snack content », souvent adaptées aux supports mobiles et au fonctionnement des réseaux sociaux.

<sup>3</sup> Ainsi, les réseaux sociaux (Facebook, Twitter...) et les moteurs de recherche pourraient mettre en avant les sources d'information « sûres ». Les annonceurs pourraient choisir les médias « fiables » pour placer leurs publicités. Les consommateurs seraient outillés pour sélectionner et faire circuler les informations vérifiées.

<sup>4</sup> Le choix éditorial d'un média de presse, appréhendé à partir des caractéristiques de l'information produite (par exemple, format, sources utilisées, sujets traités, opinion(s) politique(s), etc.), consiste à se positionner sur le marché afin d'être clairement identifié par son lectorat, d'entretenir sa notoriété (en générant des reprises par d'autres titres par exemple), d'attirer de nouveaux lecteurs et plus généralement, de garantir sa viabilité.

les stratégies concurrentielles des médias et évaluer la qualité de l'offre disponible pour les consommateurs d'actualités<sup>5</sup>. Dans cette perspective, nous évaluons les choix éditoriaux selon deux critères : (1) l'originalité des articles d'actualité par rapport aux dépêches de l'AFP et (2) l'analyse et l'argumentation développées dans chaque article.

Nous nous fixons comme objectif de savoir si les stratégies de différenciation existantes des médias français sont suffisamment variées pour satisfaire potentiellement toutes les demandes des consommateurs d'informations ayant des dispositions à payer et à supporter des coûts non monétaires différents. En d'autres termes, il s'agit d'évaluer la capacité du système de production de la presse professionnelle à fournir des informations d'actualité aux caractéristiques variées et appariées avec les différents besoins des consommateurs.

Cette analyse s'appuie sur les données inédites de 55 médias en France et un échantillon de 93K articles publiés par ces médias et relatifs à 31 évènements entre 2015 et 2019. Elle montre que les stratégies des médias, qui ont évolué depuis près de vingt ans dans le contexte du numérique, permettent aujourd'hui de produire des informations journalistiques suffisamment différenciées pour répondre à des attentes distinctes des consommateurs. Néanmoins, notre analyse montre également clairement que la variété et la viabilité des choix éditoriaux des titres de presse sont étroitement liées à leurs modèles économiques (et la rentabilité des titres de presse, cf. Lyubareva & Rochelandet, 2016). Cette observation nous permet de proposer en conclusion des pistes de réflexion sur les politiques publiques en matière de régulation des médias.

## CADRE THEORIQUE

### Qualité de l'information journalistique : définitions et questions ouvertes

La notion de qualité de l'information d'actualité (news) ne fait l'objet d'aucun consensus dans la littérature. Elle peut être liée à la notion de diversité, qui constitue souvent une mesure de la mise en œuvre de la qualité : la diversité des opinions exprimées, la possibilité donnée à des points de vue minoritaires de s'exprimer, un bon appariement entre l'offre de news et l'hétérogénéité des opinions individuelles etc. Napoli (1999) opère par exemple une distinction entre la diversité des sources (diversité structurelle), des contenus – « disparité » (qui renvoie elle-même à une diversité externe, la diversité entre les entreprises de médias et une diversité interne, la variété des points de vue au sein d'un même média) et de l'exposition (l'étendue des moyens par lesquels/la variété des contenus auxquels les individus peuvent accéder et qu'ils consomment effectivement (voir également Gibbons, 2015; Helberger, 2012).

Parmi d'autres travaux fondateurs dans le domaine, McQuail (2007) propose une définition de la qualité des médias à travers cinq dimensions dont deux sont particulièrement importantes selon l'auteur : la diversité des points de vue, styles, genres, formats et sujets ; et l'offre de l'information fiable, complète, informative et claire. La réflexivité (le journalisme doit refléter proportionnellement les différences d'opinions ou de représentations politiques, culturelles, sociales dans la société) ; l'égalité (le journalisme doit s'efforcer de proposer un accès égal à tous les points de vue présents dans la société, quelle que soit leur popularité) ; le choix ; et l'ouverture (le journalisme doit mettre l'accent sur les idées nouvelles, les points de vue originaux etc.) constituent les principes normatifs mis en avant par cet auteur.

Un auteur comme Bogart (1989) dont l'approche a été reprise dans d'autres travaux (par exemple Meyer & Kim, 2003) définit comme critères de qualité de l'information produite au sein de la rédaction, la vigueur, la quantité de la production « *in-house* », l'indépendance de la publicité, l'originalité des contenus et la diversité des points de vue. Ces travaux citent également la diversité

---

<sup>5</sup> Dans le cadre de cette étude, l'analyse porte sur la production des titres de presse. Les hypothèses concernant la demande et la consommation de l'information sont basées sur les éléments développés dans la littérature.

des formats, des styles, des sujets et des sources (titres de presse sur le marché mais aussi les sources utilisées par les journalistes) au sein et entre les rédactions.

Les travaux sur le journalisme définissent et mesurent ainsi la qualité de l'information d'actualité de différentes manières aussi bien à un niveau méso (la structure de marché, la propriété des médias etc.) que micro (choix éditoriaux...). Ces définitions concernent tant les modèles économiques, l'organisation de production, que les caractéristiques des contenus tels que son format, sa fiabilité ou sa clarté. Pour autant, les définitions proposées dans la littérature théorique soulèvent quelques problèmes.

Premièrement, la diversité n'est pas synonyme d'égale qualité. D'une part, parce que l'hypothèse selon laquelle la diversité, synonyme de qualité donc, est socialement souhaitable doit être nuancée : elle dépend précisément des définitions que l'on en donne (Karppinen, 2018). D'autre part, parce que dans le contexte numérique, on passe d'une situation de rareté à une abondance de l'information d'actualité rendant plus problématiques les enjeux de diversité qui, en première analyse, ne se posent plus dans les mêmes termes (Lyubareva & Rochelandet, 2017). Les modalités de sélection des informations ont évolué : traditionnellement et jusqu'à une période récente, les journalistes, les salles de rédaction et les organes de la presse ont joué le rôle de *gatekeepers* en filtrant les actualités, notamment selon des normes professionnelles. S'ils continuent à jouer ce rôle avec le développement de l'internet, ils ne sont plus les seuls. De nouveaux acteurs comme les moteurs de recherche, les médias sociaux, les systèmes de recommandation basés sur des algorithmes contribuent à cette sélection selon des critères bien souvent différents (Napoli, 2011, Vos & Heinderyckx, 2015). De surcroît, ces nouveaux acteurs et outils numériques peuvent orienter les choix des journalistes eux-mêmes sur les sujets d'actualité aussi bien à traiter qu'à mettre en avant. La diversité pourrait ainsi être affectée dans ses différentes dimensions par la sur- ou la sous-exposition de points de vue, des thématiques, des minorités, les conditions effectives de recherche et d'accès aux actualités par les citoyens etc. en fonction des stratégies des plateformes numériques et de la redistribution du pouvoir de communication (Helberger, 2011).

Deuxièmement, il n'y a rien d'évident dans le lien entre les conditions de production (formes de propriété, pratiques journalistiques, modèles de revenu, etc.) et les caractéristiques de l'information réellement produite. L'adoption massive des outils et des pratiques numériques a fait émerger autant de nouvelles pratiques de production (blogs, audience et montée des revenus publicitaires, plateformes et recherche d'effets-réseaux...) et de consommation (*snack contents*, partage (compulsif) en ligne...). Ce nouvel environnement contribue à une inflation et une diversification massives des news (par l'entrée de nouveaux producteurs allant des *pure players*<sup>6</sup> aux individus eux-mêmes) tout en favorisant l'entrée en scène de nouvelles voix (médias alternatifs, journalistes indépendants ou activistes...) favorisant la contestation des pouvoirs en place au niveau politique, économique, social et culturel.

Dans ce contexte, les nouveaux acteurs alternatifs peuvent contribuer à accroître la diversité des points de vue tout en produisant des actualités sur la base de "on-dit" et les titres de presse traditionnels de bonne réputation traiter un nombre très large de sujets tout en multipliant les reprises à l'identique de dépêches des agences centrales. Les formes de financements alternatifs (crowdfunding, don) peuvent autant contribuer à la production d'une information riche et originale qu'avoir un impact négatif sur la disparité des sujets traités par les médias (Cariou *et al.*, 2017)<sup>7</sup>. Dans la même lignée, le fait qu'un contenu soit produit en interne par les journalistes "maison" ne garantit en rien sa vigueur. Ainsi, l'analyse du lien entre les modèles économiques et les caractéristiques de l'information produite doit faire partie de l'étude empirique de la qualité. Les travaux sur ce sujet restent très rares (Karppinen, 2018).

---

<sup>6</sup> Un pure player désigne une entreprise d'information utilisant exclusivement le support numérique pour sa diffusion, sans édition papier.

<sup>7</sup> En effet, un engagement trop important des lecteurs-contributeurs peut créer une nouvelle forme de dépendance pour les titres de presse, aussi présente que la dépendance des revenus publicitaires.

Enfin, certains travaux (Lacy, 2000; Rosenstiel *et al.*, 2015; Lacy & Rosenstiel, 2015) suggèrent qu'au lieu de se focaliser sur les caractéristiques de l'information produite, il serait pertinent de mesurer la qualité de l'information via les indicateurs de la satisfaction des utilisateurs de cette information<sup>8</sup>. Dans cette perspective, les différentes caractéristiques de l'information produite par un média – son choix éditorial - constituent un package de services destinés à satisfaire les besoins et les attentes des consommateurs (Lacy & Rosenstiel, 2015). En effet, tandis que pour certaines caractéristiques de l'information, les offreurs et les demandeurs peuvent converger dans leur appréciation (par exemple, l'honnêteté de l'information), les autres caractéristiques (formats, vigueur, types de sources, etc.) vont varier en fonction des besoins et des préférences de chaque utilisateur de l'information.

### **L'information journalistique comme bien économique : différenciation horizontale et verticale**

La grille d'analyse développée dans le présent article propose des éléments de réponse aux problèmes discutés ci-dessus. Côté utilisateurs de l'information, nous prenons comme exemples deux caractéristiques de l'information d'actualité souvent considérées comme étant très affectées dans le contexte numérique (Cagé *et al.*, 2017 ; Vehkoo, 2010 ; Pew, 2019<sup>9</sup>) : (1) l'originalité<sup>10</sup> des informations disponibles dans la presse par rapport aux contenus de l'agence centrale (AFP)<sup>11</sup> et (2) l'analyse ou l'argumentation (désignée aussi par « vigueur » dans la littérature). Les deux caractéristiques de l'information font référence à la valeur ajoutée proposée par les journalistes.

Nous faisons l'hypothèse que la première dimension (l'originalité) correspond à l'une des préférences (quasi) unanimes des acteurs. Par exemple, lorsque les consommateurs décident de payer (par l'abonnement à un titre de presse ou l'achat d'un exemplaire) pour s'informer de l'actualité, ils espèrent y trouver des informations originales par rapport aux contenus AFP, souvent disponibles en libre accès sur les agrégateurs de l'information ou sur le site même de cette agence. Au contraire, la deuxième dimension (l'argumentation) est associée à une hétérogénéité des préférences des consommateurs. Nous formulons ensuite l'hypothèse selon laquelle les différentes combinaisons issues du croisement de ces deux dimensions au sein des choix éditoriaux sont intrinsèquement liées aux modèles économiques des médias : les caractéristiques de l'offre d'informations renvoient à une catégorie de lectorat plus ou moins ciblé et à un modèle de revenu (vente à l'unité, abonnement, publicité, mixte...), le modèle de financement permettant de produire des news et d'atteindre le lectorat correspondant.

A l'instar des biens culturels, les biens médiatiques constituent une catégorie de biens informationnels dont les conditions de production et de consommation présentent des spécificités pouvant justifier des régulations analogues pour préserver un certain niveau de "qualité" et de "diversité" (Guibert *et al.*, 2016). Un bien médiatique combine un bien intangible (l'information d'actualité par exemple) à un support de communication (le média) afin d'informer, mais également de délasser, amuser, éduquer, convaincre, influencer, créer du lien social ou communautaire... (Lacy & Rosenstiel, 2015). Un média se définit lui-même comme un "ensemble homogène de

---

<sup>8</sup> Ces propos sont aussi soutenus par la Federal Communications Commission. 2010 quadrennial regulatory review – review of the Commission's broadcast ownership rules and other rules adopted pursuant to section 202 of the Telecommunications Act of 1996. [http://hraunfoss.fcc.gov/edocs\\_public/attachmatch/FCC-10-92A1.pdf](http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-10-92A1.pdf)

<sup>9</sup> Rapport Pew pour les consommateurs: [https://reutersinstitute.politics.ox.ac.uk/sites/default/files/inline-files/DNR\\_2019\\_FINAL.pdf](https://reutersinstitute.politics.ox.ac.uk/sites/default/files/inline-files/DNR_2019_FINAL.pdf)

<sup>10</sup> L'originalité a fait par ailleurs l'objet d'un certain nombre de travaux sur les marchés de l'art pour évaluer la qualité des peintures et son impact sur les prix, en se référant notamment aux copies et aux *fakes* (Lazaro, 2006 ; Benhamou et Ginsburgh, 2002 ; De Marchi et Van Miegroet, 1996).

<sup>11</sup> Selon les travaux précédents, le numérique favorise, en plus de la prolifération des fake news et des informations dont les sources ne sont pas authentifiées, des nouvelles pratiques de « copier-coller » et de reprise massive de l'information, en particulier des agences centrales (Cagé *et al.*, 2017). Cela met en question la disparité des informations disponibles et le rôle essentiel du journalisme en analyse des actualités (Vehkoo, 2010). Cette pratique de la reprise de l'information, qui se base souvent sur les dépêches des agences centrales comme l'AFP en France, s'explique par un moindre coût contrairement à une production originale.



supports qui relèvent tous d'un même mode de communication [et qui a pour] objectif de faciliter la communication à partir d'une même source vers de nombreux destinataires" (Gabszewicz & Sonnac, 2010). Un quotidien de presse écrite, une chaîne de télévision, une station de radiodiffusion... sont autant d'exemples de médias. Certains s'adressent à un public de masse, d'autres à des niches, des communautés de pratiques, politiques, culturelles, etc.

Les biens journalistiques sont les plus caractéristiques des biens médiatiques. Ils rendent compte du monde qui nous entoure avec un certain rapport à l'actualité (ce qui le distingue d'un écrit scientifique) et au réel (ce qui le distingue d'un écrit fictionnel). L'activité journalistique peut être considérée, dans un sens restreint, dans sa conception défendue par les journalistes professionnels, ou plus largement, lorsqu'elle est exercée par des amateurs ou des militants. Dans cet article, considérant les modèles économiques des médias de presse, nous retenons une conception étroite.

Outre ce premier niveau d'analyse, la différenciation des médias peut porter sur un ensemble d'éléments constitutif du modèle d'affaires d'un titre de presse, à commencer par le modèle en question. Le titre et le logo du média, une grille éditoriale, le mode de présentation et de consultation des articles, la publication de suppléments, le modèle de revenu, un réseau spécifique de partenariats et contributeurs... constituent autant de facteurs permettant à chaque média de se différencier de ses concurrents, c'est-à-dire d'être perçu différemment et durablement par les lecteurs. Une stratégie de différenciation confère à un média un avantage durable si elle est rentable (la disposition à payer des lecteurs pour le média lui confère une prime dont les recettes couvrent normalement les coûts liés à la stratégie de différenciation) et s'avère difficilement reproductible par les concurrents (monopole local en raison de ressources spécifiques comme un réseau de correspondants, réputation et effet marque, crédibilité etc.).

La différenciation d'un média est traditionnellement liée au choix qu'il opère en matière d'informations d'actualité publiées. Les biens journalistiques renvoient alors à la stratégie éditoriale du média et se différencient à partir des caractéristiques des informations journalistiques véhiculées par les médias de presse écrite ou audiovisuelle et leur relation aux préférences des lecteurs, en d'autres termes, leur "qualité" au sens de Lancaster (1966, 1979).

Le présent article propose une grille d'analyse de la qualité de l'information fondée sur la distinction classique entre différenciations verticale et horizontale (Gabszewicz & Thisse, 1979). Comme tout autre bien économique, les caractéristiques de l'information journalistique sont censées répondre aux différents besoins des consommateurs. Dans l'approche à la Lancaster, ces caractéristiques "produisent" des services différents selon les préférences et les attentes des lecteurs (par exemple, être au courant de l'actualité ; se distraire ; se faire une opinion éclairée ; discuter et interagir avec ses proches ; se distinguer socialement par la connaissance d'actualités rares ou insolites etc.). Telle ou telle combinaison de caractéristiques sera plus ou moins apte à satisfaire ces préférences et attentes, donc offrira une "qualité" différente.

Dans cette perspective, un média fournira des news dont les caractéristiques sont différenciables horizontalement ou verticalement : d'un côté, des caractéristiques correspondant aux critères objectivables, pouvant faire l'unanimité entre les lecteurs, les journalistes, les chercheurs... (à l'instar d'un système de freinage dans l'industrie automobile) ; de l'autre, des caractéristiques renvoyant aux critères variables selon les individus en fonction de leurs représentations, préférences et pratiques (comme la couleur d'une voiture).

Dans le premier cas, certains attributs du produit sont classés et hiérarchisés unanimement par l'ensemble des lecteurs. Cela peut être le cas des informations d'actualité selon qu'elles soient fausses (inventées de toutes pièces comme les *fake news*) ou exactes (reflétant fidèlement un événement passé) : on peut supposer que les consommateurs préfèrent *en règle générale* une information "vraie et vérifiée" et donc entre deux news, préfèrent unanimement recevoir celle qui a cette caractéristique (à l'instar d'un faux en art/plagiat ou d'une œuvre authentique/original sur les marchés de l'art contemporain). Cette hypothèse suppose un consensus ou une convention sociale à ce niveau. Toutefois, si les consommateurs apprécient unanimement ce type d'information, sa

production est associée à un coût monétaire ou non-monétaire plus élevé, par exemple en termes de temps de production. Par conséquent, tous les consommateurs ne pourront pas supporter le prix nécessaire (qui peut être aussi de nature non monétaire<sup>12</sup>) pour y accéder. En fonction de cette disposition à payer (et du niveau de recettes commerciales qu'elle permet de générer) et de son modèle économique, un média décidera de fournir ce segment ou non.

Dans le second cas, les produits sont différents et non classables : appliquée à l'information d'actualité, la « qualité » des news (thématiques, formats, points de vue...) correspond à un moment donné aux différentes opinions en place (chaque lecteur choisissant une qualité spécifique).

Le croisement de ces deux types de différenciation fournit une grille de lecture des stratégies concurrentielles visant un positionnement et des tarifications différenciés sur le marché<sup>13</sup>. Dans notre étude, nous faisons l'hypothèse que pour les attributs X et Y de l'information un tel croisement permet d'envisager les stratégies des médias selon quatre possibilités non exclusives (Tableau 1) :

Tableau 1 Stratégie de choix de qualité des informations d'actualité

Différenciation		Horizontale : attribut X	
		-	+
Verticale : attribut Y	+	C	A
	-	D	B

(A) des médias disposant des moyens conséquents pour produire des news coûteuses car combinant des informations différenciées à la fois verticalement (par ex., par la production d'informations exclusives) et horizontalement (par ex., des informations argumentées et approfondies avec envoi de reporters sur le terrain, recours à des spécialistes, multiplication des sources) ; ils s'adressent à un lectorat exigeant et disposé à supporter un prix plus élevé pour la consommation de cette information ;

(B) des médias offrant des news différenciées horizontalement mais non verticalement ; les lecteurs de type B préféreront par exemple des informations peu ou moyennement originales, mais complètes, fraîches (le délai le plus court entre le développement de l'article et la survenance de l'évènement), avec un style particulier du journaliste présentant une analyse neutre et générale du sujet ;

(C) symétriquement au cas B, des médias différenciés verticalement mais offrant une information répondant à des attentes différentes des lecteurs en réunissant par exemple des communautés autour de journalistes "gonzo", des informations ciblées et ultra-subjectives. Dans ce cas, les caractéristiques de ce type d'informations ne font pas l'unanimité tant chez les journalistes que les lecteurs : les critères mis en avant par des journalistes "légitimes" (un style soutenu, une objectivité du ton, des, une multiplicité des sources...) ne seront pas jugés importants, voire évalués négativement, pour certaines catégories de lecteurs ou de journalistes (gonzo) du type C. En revanche, ce lectorat sera prêt à supporter un coût élevé pour accéder à l'information conforme à ses points de vue.

(D) à l'opposé du cas A, des médias proposant des news faiblement différenciées verticalement et offrant des informations pour un lectorat peu exigeant (préférant par exemple les informations courtes et synthétiques) et à faible disposition à payer. Le numérique peut contribuer à la prolifération de ce type d'informations à travers des pratiques comme le *speed-driven journalism* du côté offre et les *snack contents* du côté demande.

Dans la partie suivante, nous testons empiriquement cette grille d'analyse théorique en élaborant des indicateurs pour explorer les quatre types de stratégies des médias de presse en France.

<sup>12</sup> Par exemple, si la production d'une telle information ralentit le temps de publication du média, les consommateurs pourront être amenés à combiner ce média avec d'autres sources d'information pour rester informés en temps réel. Cette consommation peut donc entraîner un niveau d'effort ou un coût cognitif plus importants.

<sup>13</sup> Cf. *inter alia* Gabszewicz & Resende, 2012; Gabszewicz & Wauthy, 2012.



## DONNEES ET METHODOLOGIE

### Médias et leurs modèles économiques

Notre objectif est d'analyser la qualité de l'information journalistique à partir des stratégies de différenciation verticale et horizontale des médias. Afin de tester la pertinence d'une telle approche, nous avons élaboré un échantillon de 55 producteurs d'informations français (Tableau 2<sup>14</sup>, Annexe 1). Nous analysons leur offre en matière d'informations d'actualité pour la période 2015-2019. Ces acteurs ont été choisis en s'appuyant sur les typologies mobilisées habituellement afin de garantir la représentativité en matière de catégories institutionnelles, de fréquence de publication, de localisation et de taille des titres français. Cet échantillon a été construit sur la base des travaux empiriques précédents (Marty *et al.*, 2012 ; Lyubareva & Rochelandet, 2016 ; Cardon *et al.*, 2019), recensant des médias de presse traditionnels (issus de la base des données de l'Alliance pour les Chiffres de la Presse et des Médias<sup>15</sup>) ; les pure players (par exemple, *Reporterre*), classés parmi les médias du "cœur" de l'espace médiatique français (Cardon *et al.*, 2019) ; et les médias « alternatifs » ou « partisans » de gauche et de droite marqués par un positionnement politique plus extrême, qui ont une version papier ou existe uniquement en version web. Pour certains médias nous avons pu collecter séparément les articles publiés dans les versions papier et web des médias, ainsi que leurs éditions spéciales (magazines et blogs) distinctes du contenu principal et venant les compléter (Charon, 2015).

Tableau 2: Les médias analysés

<b>Presse quotidienne : nationale et régionale</b> <b>24 médias</b>
<u>Modèle de revenu : contenu gratuit limité</u> Abonnements web, web+papier, crédit, vente à l'unité, etc. Publicité 68% des médias sont parmi les 10 premiers de diffusion payante (Source : ACPM)
<b>Presse magazine/hebdomadaire/mensuelle</b> <b>11 médias</b>
<u>Modèle de revenu : contenu gratuit limité</u> Abonnements web, web+papier Publicité
<b>Pure players</b> <b>5 médias</b>
<u>Modèles de revenu : variés</u> Contenus gratuits, publicité Contenus gratuits, don, sans publicité Abonnement, contenu gratuit limité, avec ou sans publicité
<b>Médias partisans</b> <b>13 médias</b>
<u>Modèle de revenu : variés</u> Contenus gratuits, don, sans ou avec publicité Abonnement, publicité Contenus gratuits, donations, abonnement, sans ou avec publicité
<b>Editions spéciales/Blogs des médias nationaux et régionaux</b>
<b>Agences : 2 médias</b>
<b>Total médias 55</b>
Total sources 85

<sup>14</sup> Les rédactions de certains médias cités dans le Tableau 2, en particulier L'Expansion ou Rue89, ont intégré respectivement L'Express et L'Obs entre 2015 et 2019. Nous analysons donc leurs données uniquement pour les périodes de leur existence en tant qu'acteurs indépendants.

<sup>15</sup> <https://www.acpm.fr>

Le premier groupe de médias dans notre échantillon est représenté par les titres, leaders en matière de diffusion (et de subventions publiques), appartenant aux grands groupes et publiant les informations d'actualité en continu : *Le Monde*, *La Croix*, *L'Humanité*, *La Libération*, *Le Figaro*; *Le Télégramme*, *La Provence*, *Ouest-France*, *La Voix du Nord* et autres. Afin de protéger leur position des leaders et étant dépendants des modes traditionnels de rémunération qu'ils doivent préserver, ces médias explorent systématiquement des possibilités du numérique pour générer de nouveaux revenus. Ces acteurs utilisent souvent le modèle *Paywall*<sup>16</sup>, complété par les ventes à l'unité (des articles, des numéros) et les modèles de tarification de type « crédit-temps » « crédit-quantité »<sup>17</sup>.

La même approche à la monétisation des informations journalistiques est retrouvée dans le deuxième groupe de la presse magazine, hebdomadaire et (bi)mensuelle. Au cours du développement du marché en ligne, ces acteurs, qui réussissaient majoritairement à préserver leur diffusion grâce au ciblage des publics mais qui souffraient principalement d'un manque de réactivité par rapport aux autres contenus en ligne, proposent aujourd'hui, de plus en plus, des publications quotidiennes ou même en continu sur l'internet (par exemple, *L'Obs*, *Le Point*, *L'Express* ou *Famille chrétienne*). Pour les deux groupes des médias, les revenus publicitaires représentent une source de financement significative.

Les deux autres blocs des médias – les pure players et les médias partisans – se distinguent par leurs modèles économiques. Premièrement, la publicité ne constitue pas pour eux un mode de financement incontournable. Certains acteurs (par exemple, *Mediapart*, *Acrimed* ou *Bastamag*) mettent explicitement en avant le non-recours à la publicité et/ou aux subventions comme critères de leur indépendance. Les autres acteurs dans ces catégories, s'appuyant sur les revenus publicitaires (par exemple *Causeur*), déclarent la séparation de leurs activités commerciale et éditoriale : « ...les personnes en charge de la commercialisation des espaces publicitaires n'ont pas, dans le même temps, des responsabilités éditoriales ou un lien hiérarchique direct avec des membres de la rédaction, sous quelque forme que ce soit » (*Charte de Causeur.fr*).

Le manque des revenus publicitaires est souvent compensé par les donations, très répandues dans les deux groupes, et les abonnements (web et/ou papier), et la commercialisation de produits et de services (par vente de livres, BD, soutien scolaire, dictionnaires, formations, etc.). Les pure players et les médias partisans se distinguent également par la richesse des contenus offerts gratuitement (à l'exception de quelques-uns comme *Atlantico* ou *Mediapart*). On constate aussi une forte implication des lecteurs dans la construction de l'offre éditoriale, à travers les rubriques « ajouter votre analyse, vérification ou commentaires » ou « envoyer un article » (par exemple, *AgoraVox*, *Mediapart*, *Fdesouche*, *Acrimed*, *Lundi matin* et autres). Pour ces acteurs de la presse « alternative » ou « engagée » (Seoane, 2017), créer et fédérer sa communauté de lecteurs et de contributeurs s'avère vital.

Ils sont majoritairement indépendants des grands groupes (sauf quelques exceptions comme *HuffPost* du groupe Verizon ou *Slate* « sauvé » en 2017 par le groupe familial Cattlyea Finance et Lampsane Investissement). Sous forme juridique de sociétés coopératives, d'associations et de fondations, ils disposent de rédactions très restreintes dépassant rarement une dizaine de journalistes (sauf *Mediapart* qui compte plus de 50 journalistes sans compter ses collaborateurs réguliers)<sup>18</sup>.

---

<sup>16</sup> Le site web propose de manière gratuite une partie seulement de son offre ou une fourniture restreinte des contenus (extraits d'articles...), afin de drainer les consommateurs vers des produits payants, la gratuité servant de produit d'appel.

<sup>17</sup> Il s'agit de l'offre gratuite pendant une période limitée (par exemple, le numéro est accessible en ligne pendant 24h selon une logique de « crédit-temps ») ou pour un nombre d'articles limités (« crédit-quantité »)

<sup>18</sup> Une observation intéressante concerne le fait que certains médias partisans ne publient aucune information sur leurs rédactions, dont les membres restent anonymes (par exemple, *Contre info*, *Fdsouche* ou *Lundi*), la plupart des articles ne sont pas signés, et les sites ne fournissent pas d'informations biographiques.

## Les caractéristiques de l'information : originalité et argumentation

Pour introduire la notion de stratégies horizontale et verticale de ces médias et procéder à une mesure de la qualité de l'information journalistique disponible sur le marché, nous distinguons deux dimensions (deux caractéristiques) de l'information d'actualité dans les choix éditoriaux : l'originalité par rapport à l'AFP et l'argumentation dans les articles. Pour mesurer ces dimensions nous mobilisons les méthodes d'analyse de discours, issues de la linguistique informatique (Annexe 2).

L'*originalité des informations* disponibles dans la presse a servi à évaluer la qualité différenciée verticalement. Pour mesurer cette dimension, nous avons calculé la distance sémantique entre les articles de presse et toutes les dépêches AFP précédentes, sur le même sujet d'actualité. Dans le prolongement des travaux précédents (Cagé *et al.*, 2017), notre démarche va au-delà des formes de plagiat « pur et simple » et permet de tenir compte de formes de reprise plus sophistiquées comme la paraphrase. Pour cela, la méthode doc2vec (Le & Mikolov, 2014) a été appliquée à notre échantillon afin d'évaluer la distance entre les documents.

L'*argumentation* ou l'*analyse*, en tant que valeur ajoutée proposée par les journalistes dans les articles, a été utilisée pour évaluer la différenciation horizontale. Pour cette mesure, nous avons procédé à une analyse rhétorique des documents (Roze, 2013) c'est-à-dire une analyse des relations entre phrases et la présence d'indicateurs morphosyntaxiques référencés dans la littérature. La présence de ces relations est quantifiée : le nombre d'indicateurs explicites sur 100 mots correspond à la valeur d'« indice argumentatif ». Calculé de cette manière<sup>19</sup>, à partir des indicateurs explicites, l'indice argumentatif permet de distinguer entre les articles contenant une analyse, un contenu consensuel ou une discussion d'un sujet d'actualité et les articles autour de faits ou des positions précis laissant moins de place à l'analyse.

Pour avoir la même base de comparaison entre les articles des différents titres de presse, 31 événements, considérés comme marquants par l'AFP<sup>20</sup>, ont été sélectionnés pour la période choisie. Pour des raisons techniques propres à l'analyse retenue, en plus de leur importance, les événements choisis ont respecté les critères supplémentaires, à savoir être sans ambiguïté détectable avec les mots clés associés et bien délimités dans le temps. L'extraction des articles pour les 55 médias sur 31 événements a donné lieu à un corpus de 93 648 articles. Le Tableau 3 présente les événements analysés.

Tableau 3 Liste des événements analysés dans l'article

Évènement et ses mots clés, utilisés pour l'extraction des contenus	N articles	N moyen de mots par article	Date de début	Date de fin
Attentat Charlie Hebdo	12748	489	03.01.15	31.12.15
Nicolas Bonnemaïson	656	574	09.01.15	31.12.15
Attentats 13 novembre	20278	580	11.01.15	31.12.15
Accord de Paris climat	795	825	01.02.15	31.12.15
Corruption FIFA Platini	981	696	20.01.15	29.12.15
Airbus A320 accident	983	656	01.03.15	31.12.15
Boris Nemtsov opposition	459	697	01.03.15	26.10.15
Affaire Jacqueline Sauvage grâce partielle	213	617	01.02.16	30.12.16
Procès Luxleaks	518	688	09.02.16	19.12.16
Nuit debout	7906	528	01.04.16	29.12.16

<sup>20</sup> La liste des événements marquant nous a été fournie, dans le cadre d'une convention, par l'AFP.

Fichier TES	179	715	01.11.16	27.05.17
Neutralité du Net	746	765	05.01.17	19.12.18
Affaire Fillon Penelope	1286	744	01.02.17	29.12.17
Referendum Indépendance Catalogne	5363	669	06.02.17	31.12.17
C-star Navire	172	540	11.07.17	25.09.17
Ouragan Irma	8351	517	01.09.17	20.12.17
balance ton porc	393	980	01.11.17	01.12.18
Coupe du monde Russie	14146	614	10.01.18	22.12.18
M5S Italie populiste Salvini	903	780	04.02.18	29.12.18
Avicii	213	396	01.05.18	22.11.18
Benalla Contrescarpe	1497	780	01.08.18	27.12.18
Gilets Jaunes impôt fortune	2502	825	05.10.18	31.10.19
Khashoggi assassine	992	687	07.10.18	18.09.19
Perquisitions siège France insoumise	858	635	01.11.18	28.12.18
Gilets Jaunes prix du carburant hausse	1762	748	02.11.18	22.10.19
Elections européennes Macron LREM	2928	858	01.01.19	14.11.19
Christophe Dettinger gilets jaunes	1543	741	05.01.19	05.11.19
Droits voisins	382	841	12.02.19	31.10.19
Notre dame de paris rénovation	952	708	02.04.19	28.08.19
Balkany fraude blanchiment	1397	829	02.05.19	31.10.19
Rugy Mediapart	1546	678	11.05.19	18.09.19

L'extraction des archives d'articles de presse a été réalisée via différentes sources. Premièrement, suite à un accord avec la plateforme payante *Europresse* qui offre une base de presse très riche, surtout pour la presse nationale et régionale quotidienne ainsi que pour certains magazines, des contenus réservés aux abonnés et des articles en libre accès ont été collectés pour une partie de notre échantillon en utilisant les mots clés associés à chaque événement. Ce corpus a été complété, via l'achat des abonnements en ligne sur les sites web des médias, pour les titres de presse partisans et les pure players qui ne sont pas disponibles via *Europresse*. Ces abonnements nous ont permis l'accès aux moteurs de recherche des sites et d'effectuer la collecte des archives d'articles par mots clés. Cette démarche d'extraction nous a permis d'éviter la collecte des contenus via des grands moteurs de recherche (par exemple, Google) et donc un accès « neutre » à l'information, évitant de la sorte le biais de personnalisation des résultats des recherches par les algorithmes.

## RESULTATS

### Analyse comparative des médias

En premier lieu, on note une forte disparité entre les médias s'agissant de leur productivité en termes d'information d'actualité (Tableau 4). Tandis que la presse nationale et régionale ainsi que les pure players traitent en moyenne quasiment la totalité des sujets d'actualité de la période 2015-2019, la presse magazine/hebdomadaire/mensuelle et les médias partisans se trouvent loin derrière en traitant respectivement en moyenne 17,7 et 13,08 événements (sur 31). Dans la même lignée, les éditions spéciales et les blogs n'abordent aussi qu'un nombre très limité des sujets d'actualité et avec un nombre d'articles très restreint.

Tableau 4: Couverture des actualités par les médias

Médias	Somme d'articles	Moyenne de nombre d'évènements	Moyenne de mots
--------	------------------	--------------------------------	-----------------

Agence	20670	31	708,67
Medias partisans	1335	13,08	1222,29
Presse magazine/hebdomadaire...	5410	17,65	858,76
Presse nationale	20934	25,38	746,38
Presse régionale	42164	25,55	460,94
Pure players	3135	25,80	1229,30
Editions spéciales/Blogs	343	8,25	608,62

En revanche, pour chaque évènement d'actualité abordé, les médias partisans, ainsi que les pure players, produisent des articles significativement plus longs en nombre de mots (le Tableau 4 montre la moyenne des médianes des mots par évènement pour chaque catégorie, les informations par médias sont disponibles en Annexe 3). De manière attendue, ces caractéristiques de la production de l'information se reflètent dans les statistiques d'audience en ligne des médias analysés (Tableau 5). Pour chaque site d'information, nous avons collecté les données Alexa pour 2019 et créé un indicateur agrégé d'audience qui permet de comparer les catégories des médias entre eux<sup>21</sup>.

Tableau 5 : Audience en ligne des médias

	Moyenne PVPU*RPM
Medias partisans	30,73
Presse magazine/hebdomadaire...	123,43
Presse nationale	520,11
Presse régionale	137,75
Pure players	118,96
Total général	168,79

On remarque que les médias sélectifs dans les choix des sujets d'actualité traités et qui occupent des niches d'information politique, géographique ou thématique ont en général une audience web plus faible que les médias plus généralistes qui couvrent en continu tout le spectre des actualités. Par exemple, les médias partisans ont une audience web nettement inférieure aux autres catégories (malgré l'abondance des contenus gratuits). Parmi les médias dans notre échantillon, *Fdesouche* (audience = 128,13) et *AgoraVox* (audience = 78,02) sont deux médias partisans ayant le plus de succès auprès de l'audience en ligne. Les faibles résultats des pure players dans notre étude s'expliquent par le nombre d'acteurs dans cette catégorie et une forte dispersion de leurs audiences<sup>22</sup> entre des sites comme *Le HuffPost* ou *Mediapart* (297,23 et 152 respectivement) et, par exemple, *Reporterre* (audience = 26,4).

En second lieu et conformément à la littérature, le graphe de similarité<sup>23</sup> des articles entre tous les médias montre une faible distance entre les informations disponibles (Figure 1). Les nœuds dans ce

<sup>21</sup> Plus précisément, la variable construite dans l'objectif de cette étude est le produit de deux indicateurs : "Reach" (RPM) qui mesure le pourcentage des internautes qui consultent un site web donné et "Page Views per User" (PVPU) qui mesure un nombre moyen de pages consultées par jour par utilisateur d'un site donné. Cette variable nous fournit l'information complète sur le nombre total de consultations des pages du même site.

<sup>22</sup> L'écart-type est de 112 pour les pure players contre, par exemple, 36,6 pour les médias partisans.

<sup>23</sup> Pour simplifier la visualisation, ce graphe n'inclut pas certains médias de l'échantillon représentant des indices extrêmes. Parmi eux on trouve les éditions spéciales et les blogs (Sud Ouest Blogues, Le Monde diplomatique Carnets, Libération Blogues, Aujourd'hui en France Economie, Le Figaro Magazine, L'Indépendant Blog, Ouest-France L'édition







multidimensionnel (*multidimensional scaling* ou MDS)<sup>24</sup> : les attentats contre Charlie Hebdo, l'ouragan Irma, les perquisitions au siège de la France Insoumise et enfin la neutralité du web (Annexe 4). Pour les deux premiers événements, qui relèvent plus du (triste) fait divers, l'effet de concentration est particulièrement accentué et l'on peut supposer un style rédactionnel plus consensuel, autour de faits précis, chiffrés et factuels, sauf à considérer des titres aux positions politiques plus tranchées (notons sur la figure les positions de *Fdesouche* ou *Minute*, pour les attentats de Charlie Hebdo) ou d'autres à la fréquence de parution plus faible, favorisant la prise de recul (*Courrier International* et *Le Point*). Au contraire, des événements plus politiques comme les perquisitions de La France Insoumise ou la neutralité du web entraînent plus de dispersion dans le nuage de points, ce qui peut s'expliquer par la plus grande subjectivité des articles sur ces thèmes et la prégnance des ancrages politiques (voir Annexe 4).

L'argumentation dans les articles varie aussi en fonction de la nature de l'événement. Par exemple, la Figure 2, qui représente la dispersion de cet indice autour de la médiane pour les médias dans l'échantillon, montre clairement que les événements politiques (par exemple, droits voisins, procès Luxleaks), ceux liés aux mouvements sociaux encourageant la prise de parole des femmes (par exemple, Affaire Jacqueline Sauvage, balance ton porc), ou encore les sujets qui portent sur la protection des données personnelles (par exemple, le Fichier TES) sont à l'origine d'une plus forte dispersion d'argumentation que les faits divers (par exemple, referendum en Catalogne, Nuit debout, attentat Charlie Hebdo et ouragan Irma). On remarque également que les faits divers donnent lieu au nombre d'articles le plus élevé, mais moins longs en moyenne en nombre de mots (cf. Tableau 3).

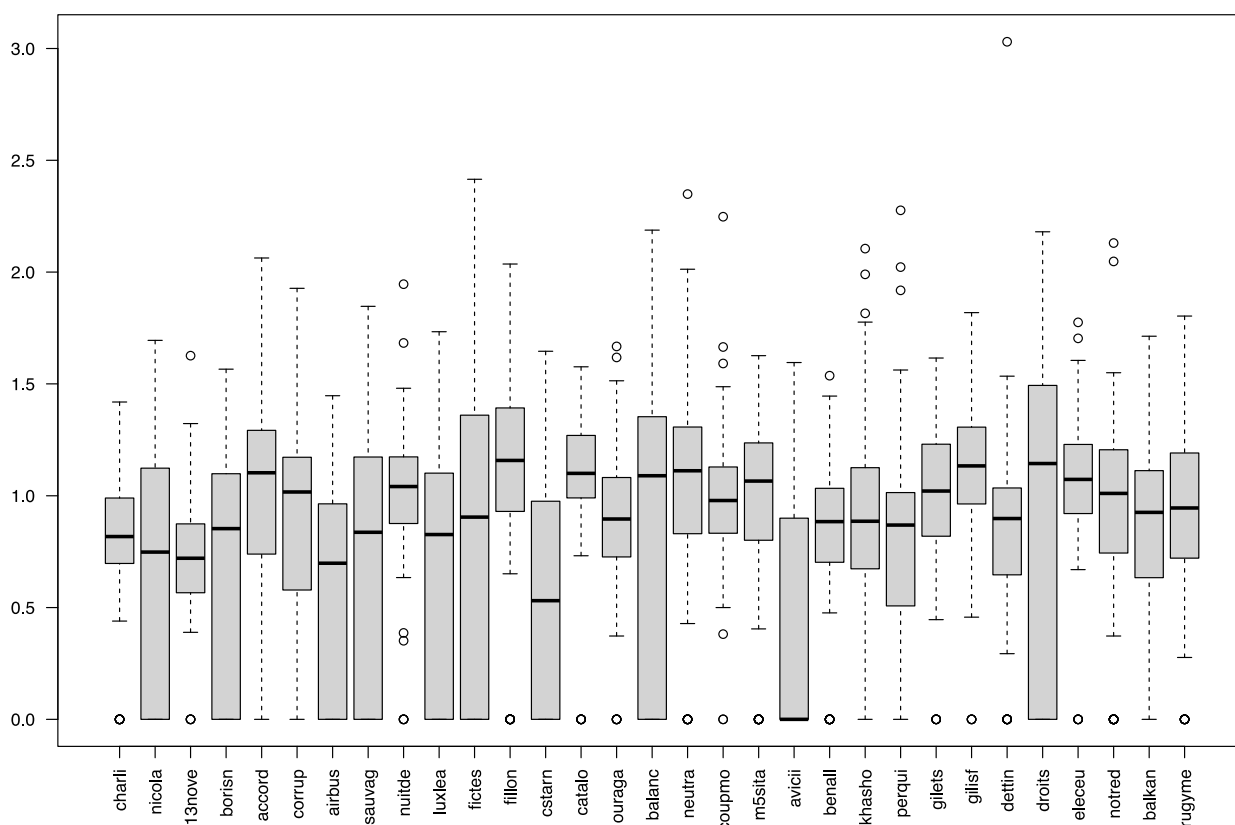


Figure 2: Dispersion de l'indice argumentatif selon les événements

<sup>24</sup> Cette technique de réduction de dimension permet, en partant de la matrice de similarité entre les différents médias, de projeter chacun d'entre eux dans un espace en deux dimensions, plan dans lequel les distances entre chaque point seront les mieux représentées.

Ces premiers résultats permettent de nuancer l'idée selon laquelle le passage des médias en ligne s'accompagne nécessairement d'une convergence des contenus publiés suite à la prolifération des pratiques du copier-coller ou de nouvelles pratiques journalistiques liées aux contraintes du temps de publication. En fonction du type d'évènements, les médias cherchent en fait à produire des informations d'actualité plus ou moins originales par rapport à leurs concurrents.

### **L'argumentation et l'originalité des articles par rapport aux dépêches AFP**

L'*originalité* des contenus par rapport aux dépêches AFP est associée, selon notre hypothèse, à la stratégie de la différenciation verticale. En croisant ces résultats avec la dimension horizontale – *l'analyse* - dépendante des goûts et des préférences hétérogènes des consommateurs d'informations d'actualité, nous pouvons visualiser la variété des stratégies éditoriales des titres de presse selon ces deux critères (Figures 3 et 4). En prenant en compte le fait que les médias traitent de manière différenciée les différents évènements en fonction de leur nature et pour ne pas perdre les informations en agrégeant pour chaque média les données de toutes les années et tous les sujets traités, nous visualisons les résultats à travers deux mesures : a) l'argumentation moyenne dans les articles produits par un média ; b) l'argumentation maximale dans les articles produits par un média.

Les graphiques sont construits de la manière suivante. La ligne horizontale coupe l'axe vertical du graphique au niveau de la valeur médiane de la distance moyenne entre les titres de presse et les dépêches AFP pour les mêmes sujets. La ligne verticale coupe l'axe horizontal au niveau de la valeur médiane de la moyenne (ou du maximum) de l'indice argumentatif. Pour améliorer la qualité de la visualisation des résultats, certains médias ayant un faible nombre d'articles dans notre corpus et des coordonnées extrêmes, en originalité et en analyse, n'apparaissent pas sur les graphiques. Parmi ces acteurs, se trouvent toutes les éditions spéciales et les blogs, ainsi que *L'Expansion* (web+papier) et *Rue89* dont le corpus n'inclut pas toutes les années d'observation en raison de leur restructuration.

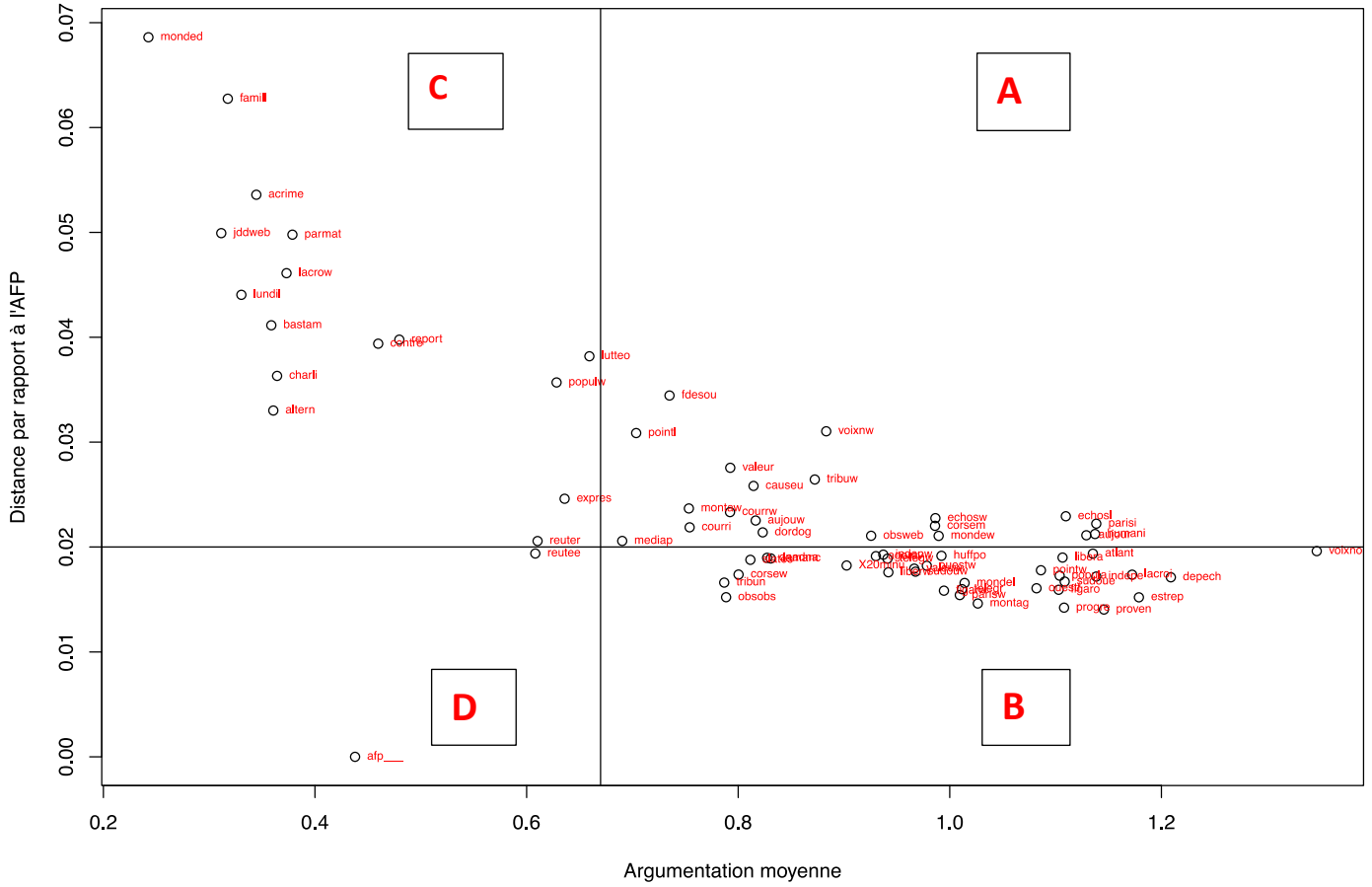


Figure 3: Les stratégies de différenciation horizontale et verticale (argumentation moyenne)

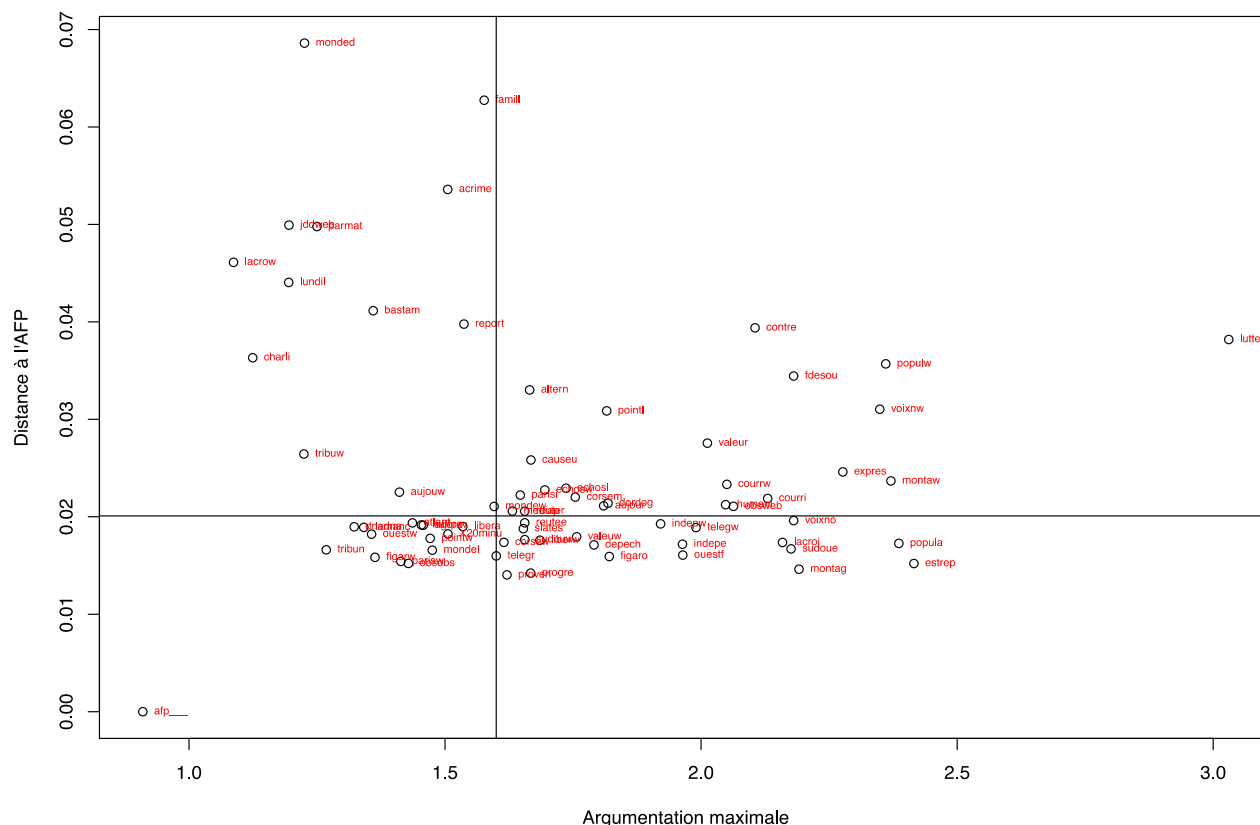


Figure 4: Les stratégies de différenciation horizontale et verticale (argumentation maximale)

En plaçant les titres de presse de notre échantillon sur les deux axes, quatre types de stratégies éditoriales A, B, C et D apparaissent. L’AFP se trouve logiquement en bas du carré D. Le Tableau 6 présente la répartition des titres pour l’argumentation moyenne et le Tableau 7 montre les pourcentages des catégories des médias pour l’argumentation maximale présente dans leurs articles.

Tableau 6: Les stratégies de différenciation des médias (argumentation moyenne)

Argumentation moyenne	A	B	C
<b>Partisans</b>	23.1% Causeur, Fdesouche, Rue89	7.7% AgoraVox	<b>69.2%</b> <sup>25</sup> Acrimed, Alternatives économiques, Bastamag, Contre-info, Là-bas, Lundi, Lutte-ouvrière, Minute, Monde diplomatique
<b>Presse magazine</b>	<b>40%</b> Courrier international (web+papier), Expansion (web+papier), Obs (web), Point, Valeurs actuelle, Tribune (web)	26.7% Obs, Point (web), Valeurs actuelles (web), Tribune	<b>33.3%</b> Charlie Hebdo, Express, Famille chrétienne, JDD, Paris Match (web)
<b>Presse nationale</b>	<b>43%</b> <sup>26</sup> Aujourd’hui en	<b>50%</b> 20minutes, Figaro	7.14% Croix (web)

<sup>25</sup> A ce pourcentage il faut ajouter Le Monde diplomatique carnets.

<sup>26</sup> A ce pourcentage il faut ajouter les blogues et les éditions spéciales : Libération blogues, Aujourd'hui en France (économie), Figaro Magazine, L'Indépendant blogs.

	France (web+papier), Echos (web+papier), Humanité, Monde (web)	(web+papier), Croix, Libé (web+papier), Monde	
<b>Presse régionale</b>	20% <sup>27</sup> Corse-Matin, Dordogne libre, Montagne (web), Parisien, Voix du nord (web)	<b>76%</b> Corse-Matin (web), Dépêche, DNA, Est républicain, Indépendant (web + papier), La manche, Montagne, Ouest-France (web+papier), Parisien (web), Populaire du Centre, Progrès, Provence, Sud Ouest (web+papier), Télégramme (web+papier), Voix du nord	4% Populaire du Centre (web)
<b>Pure Player</b>	20.0% Mediapart	<b>60.0%</b> Atlantico, Huffpost, Slate	20.0% Reporterre
<b>Total</b>	28%	45.3%	24%

Tableau 7: Stratégies de différenciation des médias (argumentation maximale).

Argumentation Max	A	B	C	D
<b>Partisans</b>	<b>46.2%</b> +Monde diplomatique carnets	0.0%	<b>46.2%</b>	7.7%
<b>Presse magazine</b>	<b>40%</b>	6.7%	33.3%	20%
<b>Presse nationale</b>	28.6% + Libération blogues, Aujourd'hui en France (économie), Figaro Magazine, L'Indépendant blogs	21.4%	21.4%	28.6%
<b>Presse régionale</b>	24.0% +Sud Ouest blogues, Ouest-France l'édition du soir, Parisien Week-end	<b>60.0%</b>	0.0%	16.0%
<b>Pure Player</b>	20.0%	20.0%	20.0%	<b>40.0%</b>
<b>Total</b>	32%	22%	20%	20%

Au niveau des catégories de médias, on obtient donc les résultats suivants :

- La **presse magazine** hebdomadaire et mensuelle payante occupe les positions dominantes dans la zone A. Cette stratégie implique logiquement un coût de production élevé, ce qui pourrait expliquer le nombre restreint de sujets traités et d'articles produits. Les exemples du mensuel économique *L'Expansion* et du pure player gratuit *Rue89*, que l'on trouve dans la zone A avant leur disparition, prouvent le fait que la stratégie A nécessite des ressources importantes et peut être risquée sur le long terme. Par exemple, le rachat de *L'Expansion* par SFR Presse associé à une politique de réduction des coûts et de transformation de la ligne éditoriale a conduit à l'effondrement de ses ventes et à sa disparition en 2017.

- La **presse nationale** se présente de manière assez équilibrée dans les quatre stratégies de différenciation verticale et horizontale. Ces choix stratégiques ne sont pas toujours discriminants : le même média peut occuper différentes zones pour, par exemple, la version papier ou web du journal. Les éditions spéciales et les blogs produits par ces acteurs pour compléter leur diffusion quotidienne

<sup>27</sup> A ce pourcentage il faut ajouter les blogues et les éditions spéciales : Sud Ouest blogues, Ouest-France l'édition du soir, Parisien Week-end.

et en continu se trouvent tous dans la zone A. Leur positionnement stratégique dans les différentes zones de différenciation permet de proposer aux consommateurs une gamme suffisamment large : information proche des articles AFP mais renouvelée en continue pour une meilleure couverture des sujets d'actualité et une garantie de fiabilité des sources ; information enrichie par des analyses et explications pour donner une présentation plus exhaustive de la situation ou pour donner une couleur politique aux informations et se différencier des concurrents ; etc.

- La **presse régionale** correspond à une stratégie de différenciation assez similaire à la presse nationale avec néanmoins comme différence principale une domination supérieure de la zone B dans leurs choix. En d'autres termes, en proposant des informations assez proches de l'AFP, ces médias se différencient plutôt horizontalement en matière d'argumentation, d'analyse et d'explication des événements. Ce résultat peut également refléter une répartition spécifique des ressources de production des médias régionaux : en accordant plus d'importance aux événements à l'échelle locale, ces acteurs peuvent allouer des ressources plus importantes à la production originale dans ces domaines qu'aux événements d'actualité nationale ou internationale, analysés dans cet article.

- Sauf exception (par exemple, *Causeur*), les **médias partisans** sont majoritairement présents dans le carré C. Ils se focalisent principalement sur l'originalité de l'information produite. L'explication et la présentation des différents points de vue sur les sujets traités ne constituent pas forcément leur priorité (à l'exception de certains événements pour lesquels ces médias montrent une argumentation élevée et passent dans le carré A pour l'indice de l'argumentation maximale)<sup>28</sup>. Ces titres de presse s'adressent à une communauté de lecteurs bien consolidée, qui partagent des points de vue similaires (Seoane, 2017). Ainsi, au-delà du problème des ressources souvent limitées, les médias partisans ne cherchent pas forcément à convaincre leur lectorat ou à développer une analyse générale d'un sujet reflétant par exemple les enjeux et les positions des différents acteurs concernés. Dans la même veine, on pourrait expliquer la présence dans le carré C du pure player de l'écologie *Reporterre*, à forte orientation thématique.

- Finalement, la zone D, qui est représentée uniquement par les **agences** ou par d'autres médias en fonction de la grille d'analyse choisie (Figures 3 et 4) correspond aux contenus peu originaux et avec de faibles valeurs d'explication et d'argumentation. La totalité des acteurs qui apparaissent dans cette zone sur la Figure 4 viennent de la zone B de la Figure 3. En plus de certains titres nationaux et régionaux, principalement dans leur version web (par exemple, *Le Figaro*, *Ouest-France*, *Le Parisien*, *20Minutes* ou *Le Point*), on y trouve les sites web comme *Atlantico* ou le gratuit *Huffpost*. La mise en place de telles stratégies de différenciation peut aider à drainer les lecteurs vers les offres payantes à haute valeur ajoutée et/ou garantir les revenus publicitaires<sup>29</sup>. Dans ce sens, les informations en ligne jouent le rôle de produit d'appel : la reprise des dépêches AFP (différenciation faible) permet à ces médias de maintenir gratuite une partie de leur offre en ligne (par exemple, la rubrique « Dépêches » chez *Atlantico*) et d'attirer une audience large dont une partie peut être "transformée" au cours du temps en lectorat abonné ou qui peut augmenter l'attractivité du média pour les annonceurs. Cette stratégie éditoriale peut également servir à entretenir une présence en ligne de la "marque" aussi large que possible auprès de types variés de lecteurs.

La cartographie des stratégies des médias en matière de production de l'information présentée ci-dessus se base pour chaque titre sur les valeurs agrégées de l'originalité et l'argumentation pour plusieurs événements d'actualité traités pendant la période 2016-2019. En fonction de la nature de l'événement ainsi que de l'orientation politique, thématique et géographique ou des ressources disponibles, les médias peuvent traiter de manière différenciée les sujets d'actualité. Une telle

---

<sup>28</sup> Par exemple, les articles de *Alternatives économique* pour l'événement « Coupe du monde Russie », *Contre-info* dans le traitement du sujet de l'assassinat du journaliste Jamal Khashoggi, ou *Lutte ouvrière* pour l'analyse de la mobilisation des Gilets jaunes au sujet de l'impôt sur la fortune.

<sup>29</sup> Pour certains acteurs dans la zone B, nous ne disposons pas des archives séparées pour la version papier et numérique. Nous pensons que la liste des acteurs dans le carré D serait plus longue si ces informations étaient disponibles.



dispersion des résultats de la production peut impacter à son tour les choix des consommateurs d'informations et leur disposition à payer : par exemple, lorsqu'il s'agit du choix entre deux médias payants par abonnement, certains consommateurs peuvent avoir une préférence plus forte pour le titre dont la production est plus « stable » pour l'ensemble de sujets d'actualité.

Les Figures 5 et 6 présentent la dispersion, par rapport à la médiane, des caractéristiques de l'information pour les médias en fonction des différents événements traités. En utilisant l'écart interquartile pour mesurer la dispersion de l'originalité et de l'argumentation chez les médias, nous identifions les titres les plus stables (la dispersion est inférieure au 1<sup>er</sup> quartile, 25% de l'échantillon) dans leur style journalistique (Tableau 8). S'agissant de la stabilité du style argumentatif, on trouve principalement les acteurs de la presse nationale et régionale et certains pure players, par exemple *Mediapart*. Concernant l'originalité, les médias partisans comme *Causeur* ou *Bastamag* complètent la liste. En revanche, la majorité des médias partisans, des pure players et de la presse magazine montrent de fortes dispersions dans le traitement des sujets d'actualité, ce qui prouve leur sélectivité dans le choix des sujets à traiter.

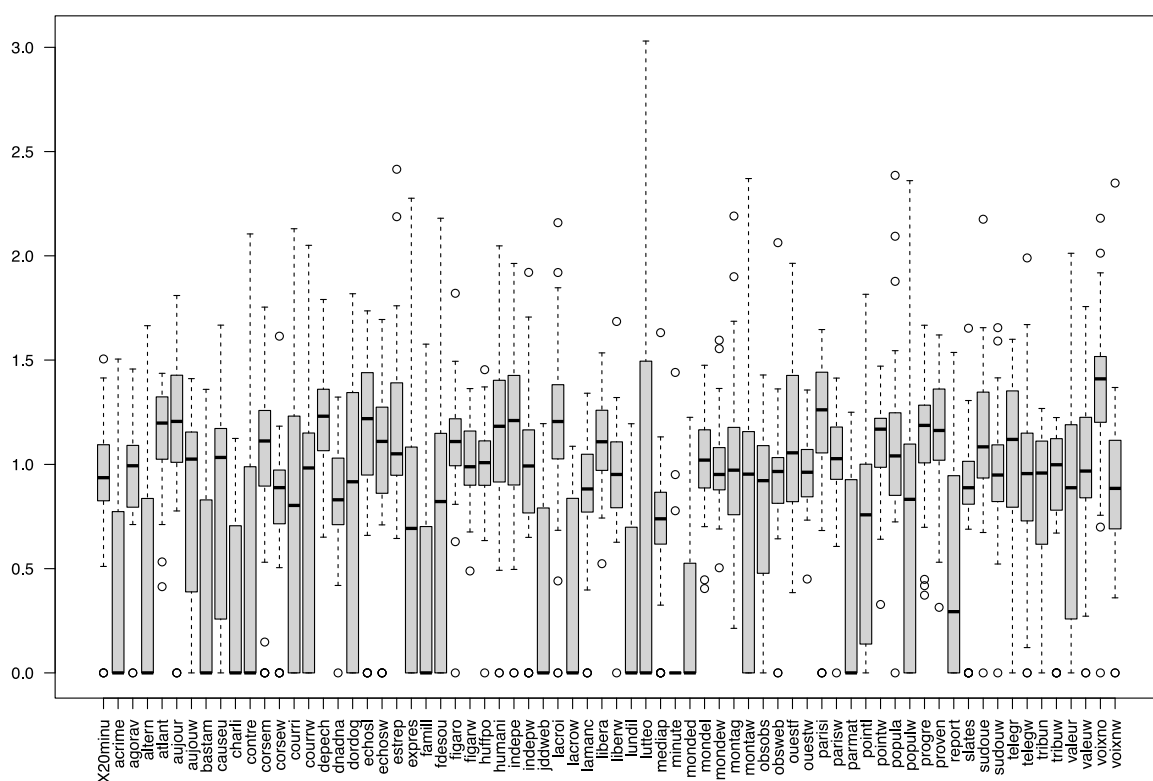


Figure 5: Dispersion de l'indice argumentatif par média

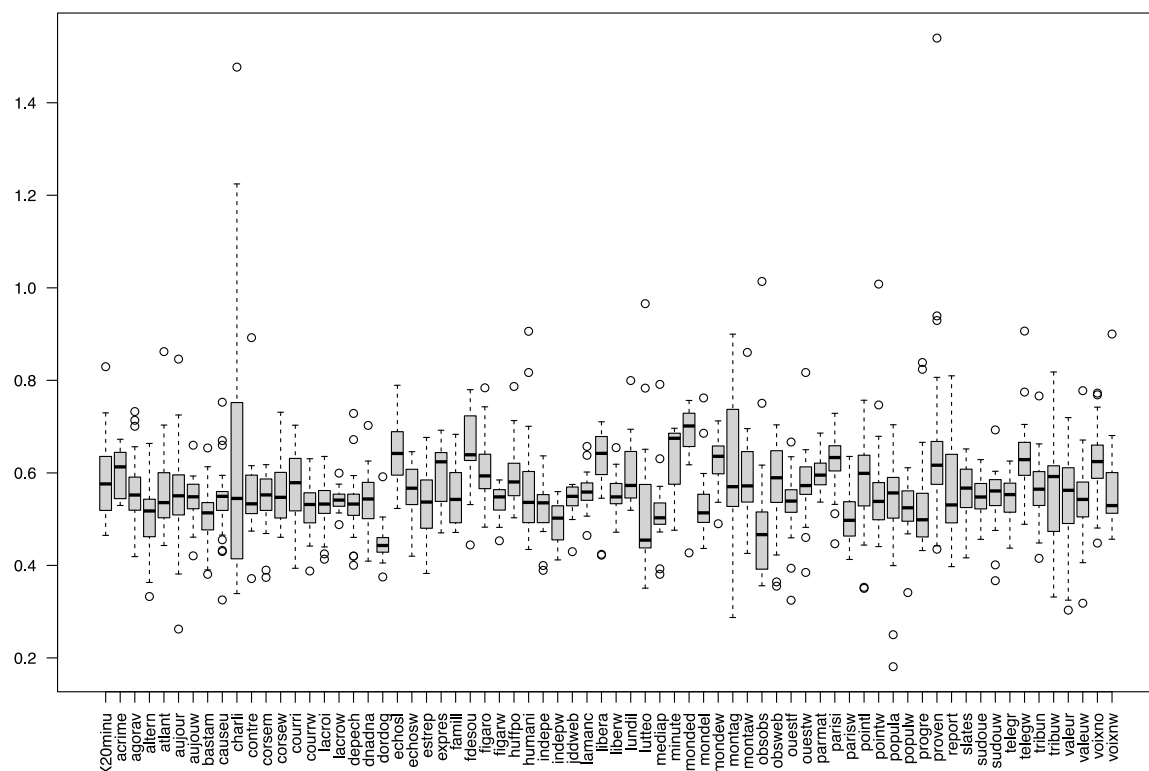


Figure 6: Dispersion de l'originalité par média

Tableau 8: 25% des médias produisant les informations les moins dispersées en originalité et en argumentation

<b>Titre : argumentation stable</b>	<b>Catégorie</b>	<b>IQR : argument</b>	<b>Titre : originalité stable</b>	<b>Catégorie</b>	<b>IQR : originalité</b>
Le Monde (web)	Presse nationale	0.202	La Croix (web)	Presse nationale	0.026
Slate	Pure Player	0.206	Dordogne libre	Presse régionale	0.032
Huffpost	Pure Player	0.213	La Manche libre	Presse régionale	0.036
L'Obs_1 (web)	Presse magazine	0.220	Paris Match (web)	Presse magazine	0.038
Le Figaro	Presse nationale	0.225	Causeur	Médias partisans	0.041
Ouest-France (web)	Presse régionale	0.227	Mediapart	Pure Player	0.043
Le Point (web)	Presse magazine	0.235	Libération (web)	Presse nationale	0.044
Mediapart	Pure Player	0.249	Le Figaro (web)	Presse nationale	0.044
Le Parisien (web)	Presse régionale	0.251	La Dépêche	Presse régionale	0.046
Corse-Matin	Presse régionale	0.258	La Croix	Presse nationale	0.049
Le Figaro (web)	Presse nationale	0.260	Ouest-France	Presse régionale	0.049
20Minutes (web)	Presse nationale	0.269	Sud Ouest (web)	Presse régionale	0.053
SudOuest (web)	Presse régionale	0.272	Aujourd'hui en France (web)	Presse nationale	0.053
Le Progrès	Presse régionale	0.277	Le Parisien	Presse régionale	0.055
La Manche libre	Presse régionale	0.277	Sud Ouest	Presse régionale	0.055
Le Monde	Presse nationale	0.279	Bastamag	Médias partisans	0.059

Libération	Presse nationale	0.289	Ouest-France (web)	Presse régionale	0.059
------------	------------------	-------	--------------------	------------------	-------

## DISCUSSION

La grille d'analyse proposée dans cet article permet de cartographier les médias selon leurs stratégies de différenciation horizontale et verticale. Notre approche permet de visualiser de manière claire les informations d'actualité disponibles sur le marché et d'évaluer, sur cette base, la qualité générale de l'offre journalistique. Nos résultats aboutissent à une comparaison approfondie entre les différentes catégories de médias de presse professionnels. Les médias nationaux et régionaux – leaders de la diffusion payante – couvrent un spectre très large d'évènements. Leurs modèles économiques leur fournissent des moyens pour expérimenter en proposant des services, des modes de tarifications et des informations différenciées pour satisfaire les différents segments de la demande pour l'information d'actualité. La presse magazine propose des informations à haute valeur ajoutée mais priorise certains sujets d'actualité. Les médias partisans, payants et gratuits, fournissent des informations détaillées et originales, mais uniquement sur certains sujets et sous un certain angle. Les pure players, enfin, se divisent entre ceux qui sont payants, comme *Mediapart*, dont la stratégie s'avère proche de la presse magazine, et ceux qui visent le marché grand public, comme *Huffpost*, proposant des informations gratuites en ligne, lesquelles ne sont pas forcément très originales, mais rendent le média attractif pour les annonceurs. La présence de tels médias permet de contrebalancer la prolifération et l'attrait des contenus non librement accessibles et nécessitant un abonnement.

Sur la base de ces observations, on constate un certain niveau de diversification de l'information produite par les médias français sur un plan qualitatif, par rapport aux deux caractéristiques de l'information retenues dans notre analyse : l'originalité par rapport à l'AFP et l'argumentation. Cette variété de l'offre est susceptible de satisfaire les attentes pour les différents types d'information et dispositions à payer (y compris un coût non monétaire) des lecteurs. Par exemple, pour s'informer de manière neutre des actualités, ces derniers ont le choix entre s'abonner à l'un des médias nationaux produisant en continu des informations plus ou moins originales et argumentées selon leur version (en ligne, papier ou édition spéciale), et payer ou soutenir plusieurs médias partisans se spécialisant sur les différents points de vue et sujets. Le panier de consommation peut aussi combiner un média national ou régional et un magazine, ou bien un pure player payant pour approfondir certains sujets d'actualité.

Nous constatons également un lien clair entre la qualité de l'information et les modèles économiques qui se sont établis sur le marché au fur et à mesure des transformations numériques depuis près de deux décennies. Ainsi les médias nationaux et régionaux, afin de protéger leur position de leaders, explorent systématiquement des stratégies éditoriales diversifiées pour générer de nouveaux revenus. Les stratégies de ces acteurs, malgré leurs moyens conséquents, se concentrent dans les zones à plus forte demande (et donc rentabilité) et pas forcément à plus haute valeur ajoutée sur un plan qualitatif (comme tend à le montrer la forte concentration de ces acteurs dans la zone B). Pour les médias alternatifs, guidés par la logique de contenu, le seul moyen de survie face à cette concurrence consiste à se positionner sur des marchés de niche très ciblés. D'une part, si ces acteurs n'arrivent pas à fédérer des communautés fidèles et stables de lecteurs pour financer leur production, ils risquent de disparaître ou de se faire racheter par des médias et des groupes plus puissants. Ces rachats sont souvent accompagnés d'une politique de réduction des coûts, de changement de la ligne éditoriale avec le risque, à terme, d'un retrait du marché. D'autre part, la dépendance trop forte à une communauté de lecteurs peut forcer ces acteurs à produire uniquement les informations conformes aux points de vue et centres d'intérêt du lectorat visé. Pour les pure players plus généralistes, qui étaient majoritairement gratuits à l'époque de leur apparition dans le paysage médiatique, un choix s'opère actuellement : soit adopter un modèle payant à l'instar de *Mediapart*, soit rester gratuits à la condition de produire une information à faible valeur ajoutée

dans la couverture des actualités, dès lors que les revenus publicitaires s'avèrent insuffisants pour financer de manière rentable la production dans le secteur de la presse.

Le problème intrinsèque de ces choix est que le succès financier de la stratégie payante, constaté dans le cas de *Mediapart*, n'est pas garanti si le nombre d'acteurs l'adoptant augmente radicalement (d'autant plus si l'on prend en compte la concurrence assez forte de la part de la presse magazine/hebdomadaire et autre sur ce segment). Nos résultats montrent d'abord que, conformément à la littérature traditionnelle en économie industrielle, les choix stratégiques de différenciation s'expliquent par la recherche d'une atténuation de la concurrence visant à obtenir de meilleures performances économiques. En particulier, les stratégies des acteurs alternatifs face à la concurrence des médias nationaux et régionaux confirment cette hypothèse. Néanmoins, on constate que les titres appartenant à une même catégorie institutionnelle se trouvent souvent dans les mêmes zones de différenciation, laissant supposer davantage des logiques d'imitation guidant les choix en matière de qualité.

Même si une analyse qualitative approfondie est nécessaire pour comprendre finement ces choix stratégiques de différenciation, un élément ressort clairement : les facteurs déterminant les choix de différenciation vont au-delà des seules ressources dont disposent les titres de presse (les médias les plus riches ne se positionnent pas forcément dans les zones à plus forte valeur ajoutée). Il semble que la logique des choix éditoriaux et stratégies de différenciation s'articule avec la construction des modèles économiques sous-jacents. Par exemple, on peut distinguer certains modèles économiques principalement et dès l'origine basés sur le contenu et le marché ciblé qui s'avèrent ainsi des composantes centrales du modèle d'affaires. Les autres composantes - sources du financement, réseau de la production, fréquence de la publication, dispositifs d'interaction avec le lectorat - s'alignent alors sur le choix éditorial du producteur de l'information. D'autres modèles reposeraient sur une configuration dans laquelle, au contraire, le choix éditorial est lui-même le résultat d'optimisation des coûts et de la rentabilité. Des études de cas permettraient là encore d'approfondir cette hypothèse.

Enfin, l'un des apports principaux de l'approche proposée dans cet article tient à son caractère généralisable à d'autres critères d'appréciation de l'information par les consommateurs. Au-delà de l'originalité des articles de presse et de leur style argumentatif, l'analyse peut s'étendre à une variété des caractéristiques de l'information : les formats synthétiques, les sources, les rubriques d'information proposées par le média (économie, culture, politique, écologie, etc.), la véracité, etc. En fonction des attentes et des préférences des utilisateurs de l'information, ces critères retenus pour l'analyse pourraient être classés selon les axes stratégiques vertical et horizontal. Sur cette base, les différentes cartographies des médias ainsi obtenues peuvent permettre d'identifier les zones éditoriales à forte concentration des médias et celles sous-développées et notamment de faire le lien entre les choix éditoriaux des producteurs d'informations d'actualité et leurs modèles économiques, en évaluant leurs viabilité et pérennité face aux besoins du marché.

L'ensemble de ces observations pourrait également servir à nourrir les réflexions des politiques publiques en matière de qualité de l'information. Ainsi, en se basant uniquement sur les critères de l'originalité et de l'argumentation, nos résultats confirment certaines conclusions de la littérature (Cagé, 2015). D'une part, en matière de régulation, la réflexion pourrait être axée sur la mise en œuvre d'un ou des modèle(s) économique(s) alternatif(s), intermédiaire(s) entre les grands groupes et les petites sociétés coopératives, pour que les médias soient économiquement viables et indépendants (y compris de leur lectorat, dont l'implication trop élevée peut aussi être négative pour la neutralité de l'analyse et de la couverture des actualités). D'autre part, les mécanismes de contrôle et de régulation pourraient stimuler davantage la logique de l'information payante (stimulation de la diffusion, régulation des plateformes, défiscalisation des dons, etc.) sans laquelle plusieurs médias seront voués à l'échec favorisant alors une uniformisation de l'information au cours du temps.

## REFERENCES BIBLIOGRAPHIQUES

- BENHAMOU, F., GINSBURGH, V. (2002), "Is There a Market for Copies?", *Journal of Art, Management, Law and Society*, 32(1), pp.37-56.
- BOGART, L. (1989), *Press and Public: Who Reads What, When, Where and Why in American Newspapers*, Hillsdale: Lawrence Erlbaum and Associates Inc. (2nd ed.).
- CAGÉ, J. (2015), *Sauver les médias. Capitalisme, financement participatif et démocratie*, Seuil
- CAGÉ, J., HERVÉ, N., VIAUD, M.-L. (2017). *L'information à tout prix*. Bry-sur-Marne, Ina, coll. Médias et humanités.
- CARDON, D., COINTET, JP, OOGHE, B, PLIQUE, G. (2019), *Unfolding the multi-layered structure of the French mediascape*. Rapport de l'Institut Montaigne
- CARIOU, C., LYUBAREVA, I., ROCHELANDET, F. (2017), "Crowdfunding et qualité de l'information", *Réseaux*, 5(205), pp.23-56.
- CHARON, J.-M. (2015), *Presse et numérique. L'invention d'un nouvel écosystème*. Rapport pour le Ministère de la culture et de la communication. <https://www.culture.gouv.fr/Espace-documentation/Rapports/Rapport-Charon-Presses-et-numerique-L-invention-d-un-nouvel-ecosysteme>
- DE MARCHI, N., VAN MIEGROET, H.J. (1996), "Pricing Invention: 'Originals', 'Copies', and their Relative Value in Seventeenth Century Netherlandish Art Markets.", In V.A. GINSBURGH, P.-M. MENGER (coord.), *Economics of the Arts, Selected Essays*, North Holland, Amsterdam, pp.27-70.
- GABSZEWICZ, J., RESENDE, J. (2012), "Differentiated credence goods and price competition", *Information Economics and Policy*, 24(3-4), pp. 277-287.
- GABSZEWICZ, J., WAUTHY X.Y. (2012) "Nesting horizontal and vertical differentiation", *Regional Science and Urban Economics*, 42(6), pp. 998-1002.
- GABSZEWICZ, J., SONNAC, N. (2010), *L'industrie des médias à l'ère numérique*, La Découverte/Repères.
- GABSZEWICZ, J., THISSE, J.F. (1979), "Price Competition, Quality and Income Disparities", *Journal of Economic Theory*, 20(3), pp.340-359.
- GIBBONS, T. (2015), "Active Pluralism: Dialogue and Engagement as Basic Media Policy Principles", *International Journal of Communication*, n°9, pp. 1382–1399
- GUIBERT, G., REBILLARD, F., ROCHELANDET, F. (2016) *Médias, culture et numérique. Approches socioéconomiques*, Armand Colin.
- HELBERGER, N. (2011), "Diversity by design", *Journal of Information Policy*, 1(1), pp. 441-469
- HELBERGER, N. (2012), "Exposure diversity as a policy goal", *Journal of Media Law*, 4(1), pp. 65-92.
- KARPPINEN, K.E. (2018), "Journalism, Pluralism, and Diversity", in T.P. Vos (coord.) , *Journalism*, Handbooks of Communication Science, 19 , pp. 493-510.
- LACY, S. (2000), "Commitment of financial resources as a measure of quality", in Robert G. Picard (ed.), *Measuring media content, quality and diversity: Approaches and issues in content research*. Turku, Finland: The Media Group, Business and Research Development Centre, Turku School of Economics and Business Administration, pp.25-50.
- LACY, S., ROSENSTIEL, T. (2015), *Defining and measuring quality journalism*. Rutgers School of Communication and Information.
- LAZZARO, E. (2006), "Assessing quality in cultural goods: The hedonic value of originality in Rembrandt's prints", *Journal of Cultural Economics*, 30(1), pp.15-40.
- LANCASTER, K.J. (1979), *Variety, Equity, and Efficiency*, Columbia University Press
- LANCASTER, K.J. (1966), "A New Approach to Consumer Theory", *Journal of Political Economy*, 74 (2), pp. 132-157.
- LE, Q., MIKOLOV, T. (2014), "Distributed Representations of Sentences and Documents", *International conference on machine learning*, pp.1188-1196

- LYUBAREVA, I., ROCHELANDET, F. (2016), "L'évolution des modèles d'affaires dans les industries créatives : l'exemple de la presse en ligne en France (2004-2014)", *Revue d'Economie Industrielle*, 156, pp.123-157.
- LYUBAREVA, I., ROCHELANDET, F. (2017), "Modèles économiques, usages et pluralisme de l'information en ligne. Les nouveaux enjeux du pluralisme de l'information à l'ère des plateformes numériques", *Réseaux*, 5(205), pp. 9-19.
- MARTY, E., REBILLARD, F., POUCHOT, S., LAFOUGE, T. (2012), "Diversité et concentration de l'information sur le web", *Réseaux*, (6), 27-72.
- MCQUAIL, D. (2005), *McQuail's mass communication theory*, Sage publications.
- MCQUAIL, D. (2007), "Revisiting Diversity as a Media Policy Goal", In Werner A. Meier & Josef Trappel (coord.), *Power, Performance and Politics: Media Policy in Europe* Nomos: Baden-Baden, pp. 41-58.
- MEYER, P., KIM, K. H. (2003), Quantifying newspaper quality: I know it when I see it. In *meeting of the Association for Education in Journalism and Mass Communication, Kansas City*.
- NAPOLI, P. (1999), "Deconstructing the diversity principle", *Journal of Communication*, 49(4), pp.7-34.
- NAPOLI, P.M. (2011), "Exposure diversity reconsidered", *Journal of Information Policy*, n°1, pp.246-259.
- ROSENSTIEL, T., SONDERMAN, J, THOMPSON, T. BENZ, J., SONDERMAN, J, SWANSON, E. (2015), "*How Millennials Get News: Inside the Habits of America's First Digital Generation*", American Press Institute and Associated Press-NORC Center for Public Affairs Research.
- ROZE C. (2013), *Vers une algèbre des relations de discours*, thèse de doctorat, Université Paris-Diderot.
- SEOANE, A. (2017), "Effets d'oralité dans la presse engagée: dialogisation et idéologisation du discours", *Mots. Les langages du politique*, 2, pp.167-189
- VEHKOO, J. (2010), What is quality journalism? and how it can be saved. *Reuters Institute Fellowship Paper*.
- VOS, T.P., HEINDERYCKX, F. (coord.) (2015), *Gatekeeping in Transition*, Routledge.



## Annexe 1 : L'échantillon de l'étude

<p><b>Presse quotidienne : nationale et régionale</b></p> <p><u>Modèle de revenu : contenu gratuit limité</u> Abonnements web, web+papier, crédit, vente à l'unité, etc. Publicité</p> <p>* (68%) Parmi les 10 premiers de diffusion payante (Source : ACPM)</p>	<p>20 Minutes web Aujourd'hui en France (web+papier)* Corse-Matin (web+papier) Dordogne libre La Croix (web+papier)* La Dépêche du Midi * La Manche libre La Montagne (web+papier)* La Provence La Voix du Nord (web+papier)* Le Figaro (web+papier)* Le Monde (web+papier)* Le Parisien (web+papier)* Le Télégramme (web+papier)* Le Populaire du Centre (web+papier) Le Progrès* Les Échos (web+papier)* Les Dernières Nouvelles d'Alsace* L'Est républicain L'Humanité* Libération (web+papier)* L'Indépendant (web+papier) Ouest-France (web+papier)* Sud Ouest (web+papier)*</p>
<p><b>Presse magazine/hebdomadaire/mensuelle</b></p> <p><u>Modèle de revenu : Contenu gratuit limité</u> Abonnements web, web+papier Publicité</p>	<p>Charlie Hebdo Courrier international (web+papier) Famille chrétienne JDD L'Obs (web+papier) La Tribune (web+papier) Le Point (web+papier) L'Expansion (web+papier) L'Express Paris Match Valeurs actuelles (web+papier)</p>
<p><b>Pure Players</b></p> <p><u>Modèles de revenu : variés</u> *contenus gratuits, publicité ** Contenus gratuits, don, sans publicité *** Abonnement, contenu gratuit limité, avec ou sans publicité</p>	<p>Atlantico *** Le HuffPost* Mediapart*** Reporterre** Slate*</p>
<p><b>Médias partisans</b></p> <p><u>Modèle de revenu :</u> Contenus gratuits, don, sans ou avec publicité * Abonnement, publicité ** Contenus gratuits, donations, abonnement, sans ou avec publicité</p>	<p>Acrimed AgoraVox Alternatives économiques** Bastamag Causeur* Contre-info Fdesouche** Là-bas si j'y suis** Le Monde diplomatique Lundi matin Lutte ouvrière Minute** Rue89</p>

<b>Editions spéciales/Blogs</b>	Sud Ouest (blogues), Le Monde diplomatique (carnets), Libération (blogues), Aujourd'hui en France (économie), Le Figaro Magazine, L'Indépendant (blogs), Ouest-France (l'édition du soir), Le Parisien Week-end
<b>Agences</b>	AFP - Journal Internet AFP (français), AFP Doc, AFP Infos Economiques, AFP Infos Françaises, Reuters économique, Reuters général
Total médias	55
Total sources	85

## Annexe 2: Les méthodes de la linguistique informatique, utilisées pour mesurer l'originalité et l'argumentation dans les articles des médias.

Pour évaluer l'originalité des contenus des médias, entre eux et avec les dépêches AFP, nous avons décomposé les documents en paragraphes, puis calculé la similarité de toutes les paires de paragraphes ainsi que des moyennes et des maxima. La méthode doc2vec consiste à former un plongement de mots (*word embedding*) dans un espace de grande dimension, avec la particularité qu'aux mots sont également associés des vecteurs de paragraphe de document. Ainsi on tient compte de deux types de relation entre les mots : leurs positions respectives dans le texte (puisque le plongement est calculé à l'aide d'une fenêtre coulissante) et leurs appartenances aux paragraphes (et donc aux documents les contenant). La méthode doc2vec permet donc de mesurer la distance entre documents d'un même corpus. Nous avons utilisé l'implémentation *gensim* de la méthode doc2vec, le document qui a servi à l'apprentissage était d'un volume de 360 Mo (1 280 776 paragraphes, 5 607 818 mots), le modèle créé utilise un espace vectoriel à 300 dimensions (modèle *Distributed-Memory Paragraph Vector*), une fenêtre glissante de 10 et un taux d'apprentissage de 0.025. Une fois le modèle créé, pour une paire de documents donnés nous avons calculé les moyennes des similarités entre leurs paragraphes : chaque paragraphe a été représenté par un vecteur de l'espace vectoriel, la similarité est simplement la valeur absolue du cosinus des deux vecteurs. Comme cela est toujours le cas quand on utilise des méthodes de *deep learning*, c'est-à-dire basées sur les réseaux de neurones récurrents, la similarité ainsi mesurée est diffuse et statistique ; elle peut se baser tout aussi bien sur la synonymie que sur la collocation de mots dans des contextes similaires.

Pour mesurer l'argumentation dans les textes d'articles nous nous sommes basés sur la présence des indicateurs morphosyntaxiques suivants:

### Relations discursives : exemples des indicateurs morphosyntaxiques

**Violation** : néanmoins, malgré tout, hormis que, cependant, ceci étant dit, au lieu, à part ça, à la place...

**Résultat** : ainsi, preuve que, subséquentement, si bien que, pour résumer, par suite, par conséquent, en conséquence....

**Objectif** : pour que, le temps que, jusqu'à ce que, histoire de, en vue de, de manière que, de manière à....

**Explication** : parce que, par le fait que, par exemple, maintenant que, lorsque, faute de, étant donné que...

**Evidence, élaboration, contraste, conséquence, condition, concession, alternation.**

## Annexe 3 : Productivité des médias pour les événements d'actualité

Titre	Acronyme	N d'évènements sur 31	Somme Articles	Médiane de la moyenne des mots	Somme d'articles/ N d'évènements
20 Minutes web	20minu	28	2406	540	85,9
Acrimed	acrimed	12	59	1976	4,9
AFP (AFP - Journal Internet AFP (français), AFP Doc,	afp_	31	16460	678	531,0

AFP Infos Economiques, AFP Infos Françaises)					
AgoraVox	agorav	29	336	1604	11,6
Alternatives économiques	altern	11	115	1074	10,5
Atlantico	atlant	31	812	1051	26,2
Aujourd'hui en France (économie)	aujoue	2	2	368,5	1,0
Aujourd'hui en France	aujourd	29	1130	478	39,0
Aujourd'hui en France (web)	aujouw	23	1978	606	86,0
Bastamag	bastam	13	109	1611	8,4
Causeur (web)	causeu	23	206	1228	9,0
Charlie Hebdo	charli	15	39	1329	2,6
Contre-info	contre	13	34	329	2,6
Corse-Matin	corsem	28	683	473,5	24,4
Corse-Matin (web)	corsew	27	1359	701	50,3
Courrier international	courri	18	146	713	8,1
Courrier international (web)	courrw	21	379	628	18,0
Dépêche du Midi	depech	31	2281	434	73,6
DNA	dnadna	30	539	736,5	18,0
Dordogne libre	dordog	21	227	409	10,8
Les Échos	echosl	28	608	485	21,7
Les Échos (web)	echosw	27	1052	613	39,0
Est républicain	estrep	31	1716	396	55,4
L'Expansion	expans	3	5	667	1,7
L'Expansion (web)	expansw	2	5	551,5	2,5
L'Express	expres	20	145	1282,5	7,3
Famille chrétienne	famill	10	90	761	9,0
Fdesouche	fdesou	22	207	338,5	9,4
Le Figaro Magazine	figarm	18	97	781	5,4
Le Figaro	figaro	31	1369	689	44,2
Le Figaro (web)	figarw	31	4538	609	146,4
HuffPost	huffpo	30	1652	647,5	55,1
L'Humanité	humani	29	572	749	19,7
L'Indépendant (blogs)	indepb	4	18	251,5	4,5
L'Indépendant	indepe	30	1037	355	34,6
L'Indépendant (web)	indepw	27	634	504	23,5
JDD (web)	jddweb	10	529	735,5	52,9
La Croix	lacroi	30	559	457	18,6
La Croix (web)	lacrow	13	1001	687	77,0
La Manche libre	lamanc	28	679	705	24,3
Libération	libera	31	944	829	30,5
Libération (blogues)	liberb	9	19	958	2,1
Libération (web)	liberw	31	2311	662	74,5
Lundi matin	lundil	13	62	2562	4,8
Lutte ouvrière	lutteo	12	42	503,5	3,5
Mediapart	mediap	26	251	1412	9,7

Minute-hebdo	minute	3	6	452	2,0
Le Monde diplomatique (carnets)	mondec	4	9	1127	2,3
Le Monde diplomatique	monded	10	25	1970,5	2,5
Le Monde	mondel	31	1775	880	57,3
Le Monde (web)	mondew	31	2735	912	88,2
La Montagne	montag	31	1356	456	43,7
La Montagne (web)	montaw	20	344	504	17,2
L'Obs	obsobs	26	172	1377,5	6,6
L'Obs (web)	obsweb	29	1241	955	42,8
Ouest-France	ouestf	31	4057	330	130,9
Ouest-France (l'édition du soir)	ouests	12	117	103,5	9,8
Ouest-France (web)	ouestw	31	4516	452	145,7
Le Parisien	parisi	29	1733	441	59,8
Le Parisien Week-end	parism	9	27	750	3,0
Le Parisien (web)	parisw	30	2794	621,5	93,1
Paris Match (web)	parmat	12	402	588	33,5
Le Point	pointl	24	180	1100	7,5
Le Point (web)	pointw	31	1300	820	41,9
Populaire du Centre	popula	31	821	420	26,5
Populaire du Centre (web)	populw	17	161	494	9,5
Progrès	progre	31	2107	404	68,0
La Provence	proven	31	1344	378	43,4
Reporterre	report	16	129	1257,5	8,1
Reuters économique	reutee	27	2121	723	78,6
Reuters général	reuter	27	2089	725	77,4
Rue89	rue89w	7	132	1019	18,9
Slate	slates	26	291	1778,5	11,2
Sud Ouest (blogues)	sudoub	8	54	529,5	6,8
Sud Ouest	sudoue	31	1942	491	62,6
Sud Ouest (web)	sudouw	30	2159	607	72,0
Le Télégramme	telegr	29	1803	370	62,2
Le Télégramme (web)	telegw	30	1731	481	57,7
La Tribune	tribun	25	463	917	18,5
La Tribune (web)	tribuw	27	573	828	21,2
Valeurs actuelles	valeur	23	192	1036	8,3
Valeurs actuelles (web)	valeuw	29	461	524	15,9
La Voix du Nord	voixno	30	980	329,5	32,7
La Voix du Nord (web)	voixnw	29	1862	459	64,2

### Annexe 4: MDS, comparaison de similarité sémantique des médias pour quatre évènements différents.

