

# Traces numériques et territoires

Marta Severo, Alberto Romele



## Partie 3

# Les pratiques de la trace numérique

# Dimensions spatiales de l'actualité internationale

## La représentation des individus dans le New York Times

Laurent Beauguitte, Marta Severo

### INTRODUCTION

Les actualités produites par les médias sont de plus en plus utilisées dans la recherche universitaire comme des données permettant l'étude de phénomènes sociaux divers. Leur étude permet ainsi de détecter et de modéliser des phénomènes aussi variés que les mouvements sociaux, les «images-pays» (la façon dont un espace donné est représenté et les thématiques qui lui sont associées) ou les catastrophes naturelles. Ces dernières décennies, plusieurs chercheurs ont également travaillé sur la définition et l'identification des événements médiatiques [Dayan et Katz, 1992; McCombs et Shaw, 1972]. La prise en compte de la dimension spatiale concernant la fabrication de l'actualité internationale a été au cœur de plusieurs travaux théoriques et empiriques majeurs [Galtung et Ruge, 1965; Koopmaas et Vliegthart, 2011], la question essentielle concernant la recherche des mécanismes permettant d'expliquer le niveau d'attention médiatique à des événements donnés. La transformation d'un fait en actualité, c'est-à-dire le passage d'une situation se déroulant dans le monde à sa mention dans un ou plusieurs médias d'information, dépend de facteurs liés tant à la distance qu'aux conditions de production et de consommation de l'actualité, et ceci est encore plus net pour l'actualité internationale [AFP, 2010].

L'un des principaux verrous à lever pour systématiser de telles recherches concerne l'accès aux données, les bases de données médiatiques existantes étant souvent privées et d'accès onéreux (DowJones Factiva). Outre leur coût financier, l'utilisation de ces bases pose plusieurs problèmes méthodologiques importants. Ces dernières années, grâce au développement rapide du Web 2.0, la consommation d'information en ligne a changé radicalement et les principaux médias d'information ont mis en place des dispositifs visant à attirer et fidéliser le lecteur internaute (flux RSS notamment) en fournissant une information gratuite et actualisée en temps réel. Notre hypothèse est que l'étude des actualités contenues dans les flux RSS des journaux en ligne permet simultanément d'étudier la dimension spatiale de l'événement médiatique tout en levant les problèmes méthodologiques liés à l'utilisation des bases de données privées.

Cet article est organisé en deux parties. La première partie détaille les problèmes soulevés par les bases de données médiatiques traditionnelles et les avantages potentiels de l'utilisation des flux RSS, hypothèse actuellement testée dans le cadre

du projet ANR Corpus Géomédia (2013-2016). Au sein de ce projet associant géographes, spécialistes des médias et informaticiens, nous avons construit une base de données stockant les flux RSS concernant l'actualité internationale d'une centaine de quotidiens d'information à travers le monde<sup>1</sup>. Deux types de recherche sont actuellement menés : l'analyse de la structure géographique de l'actualité internationale et la recherche d'événements médiatiques internationaux – seul le premier aspect sera développé ici. Nous présentons brièvement la méthodologie utilisée pour bâtir ce corpus ainsi que ses principales caractéristiques.

Dans la deuxième partie de cet article, un cas d'étude nous sert à valider l'intérêt d'utiliser dans des études spatiales les flux RSS émis par les sites d'informations. Centrée sur la place et l'espace – les espaces – des individus, notre étude montre qu'une analyse qualitative et quantitative de ces flux permet de saisir en partie les logiques sous-jacentes relatives à la production et à la diffusion de l'information. Les logiques de personnalisation de l'information destinées à favoriser l'identification du lectorat sont mises en évidence à travers le cas de l'actualité internationale du *New York Times*. Cette analyse s'intéresse en particulier à deux types d'individus identifiables dans les flux RSS : les stars mondiales et les anonymes.

## ENJEUX DU PROJET ANR CORPUS GÉOMÉDIA

### Les bases de données médiatiques commerciales

Sans prétendre à l'exhaustivité, il est nécessaire de rappeler en premier lieu que les études médiatiques ont une longue histoire et que de nombreuses pistes de recherche ont été explorées ces dernières décennies. La recherche sur la valeur d'actualité d'un fait – *news value* – [Galtung et Ruge, 1965], les travaux relatifs à la mise à l'agenda – *agenda setting* – [McCombs et Shaw, 1972] ou encore les recherches sur l'événement médiatique [Dayan et Katz, 1992] ont ainsi de manière complémentaire permis d'éclairer les logiques médiatiques relatives à l'actualité internationale. Des travaux empiriques se sont notamment intéressés aux mouvements sociaux [Herkenrath et Knoll, 2011 ; Giraud et Severo, 2013], aux catastrophes naturelles [Koopmaas et Vliegthart, 2011] ou à l'image d'un espace donné [Robinet, 2013 ; Chaudhary, 2001], montrant une attention médiatique sensible à la localisation des différents producteurs et diffuseurs d'information [Brooker-Gross, 1983]. Interroger les interactions entre les aspects spatiaux, temporels et thématiques des faits devenus événements médiatiques apparaît comme l'une des pistes prometteuses ouvertes par des travaux récents [Koopmaas et Vliegthart, 2011].

---

1 Nous considérons des journaux en français, anglais, espagnol et portugais.

Ces études sont généralement basées sur l'analyse des médias traditionnels tels que la presse et la télévision<sup>2</sup>. Comme on le sait [Earl *et al.*, 2004], l'utilisation de la presse pour l'étude de différents objets de recherche (par exemple le territoire) peut soulever des critiques relatives à la collecte des données, à leur sélection et à la description du contenu des articles [McCarthy et McPhail, 1996]. Cependant, en accord avec Earl *et al.* [2004], nous estimons que les données de la presse, en dépit de leurs limites, peuvent permettre d'étudier certains aspects que les chercheurs ne pourraient guère approcher avec d'autres méthodes. Il est possible de distinguer deux grands types d'études : des études de nature qualitative qui ciblent le traitement réalisé par un média précis sur un événement donné ; des études de nature quantitative qui cherchent à repérer des structures récurrentes en prenant un corpus large dans le temps et/ou l'espace [Wu, 2000]. Chacune de ces approches présente ses avantages et ses inconvénients.

Les études qualitatives approfondies supposent généralement l'examen manuel d'un corpus médiatique donné (le plus souvent, un ensemble d'articles de presse). Ces analyses sont longues et coûteuses à mettre en place et sont donc généralement appliquées à des corpus d'ampleur limitée dans le temps et l'espace (un État, un événement et une période limitée). De plus, le recours au codage manuel, méthode la plus courante dans ce type d'études, implique nécessairement une part de subjectivité de la part du chercheur : déterminer par exemple si tel ou tel article donne une image positive ou négative d'un espace donné peut être délicat dans la mesure où il est rare qu'un article ne souligne que des aspects positifs ou négatifs. Inversement, les analyses quantitatives portant sur un large échantillon et cherchant à prendre en compte une plus grande étendue chronologique et géographique s'appuient le plus souvent sur des données obtenues *via* les bases de données commerciales telles que Dow Jones Factiva, LexisNexis ou Europresse.

L'utilisation de ces bases de données est non seulement coûteuse, mais elle soulève aussi plusieurs problèmes méthodologiques et techniques. L'extraction des données est généralement limitée (le plus souvent, impossibilité d'extraire plus de 100 items simultanément), que ce soit pour des raisons commerciales ou de droits d'auteur, et la création de bases de données conséquentes est donc chronophage. Au niveau méthodologique, ces bases de données sont peu transparentes en ce qui concerne la variabilité de leur contenu ; des titres apparaissent, d'autres disparaissent et la stabilité du corpus constitué n'est que trop rarement assurée. Si elles permettent en théorie d'étudier de manière longitudinale de vastes corpus, cette possibilité est plus apparente que réelle. Enfin, le problème du partage des données est essentiel : il n'est pas possible avec ce type de fournisseurs de données de s'inscrire dans des démarches d'*open data* et de recherche reproductible. Toutes ces limites nous ont conduit à explorer une piste alternative, celle des flux RSS.

---

2 Nous nous limitons ici aux études portant sur la presse écrite.

## Les flux RSS comme piste alternative ?

Nous proposons d'utiliser un autre type de données qui semble plus adapté à des études de l'actualité internationale. Nous testons l'intérêt d'utiliser les flux RSS fournis par les quotidiens d'information. Les fils RSS sont des fichiers XML régulièrement mis à jour sur un site Internet et donnant une information synthétique sur la mise en ligne d'un nouveau contenu ou l'actualisation d'un contenu déjà présent. Un flux RSS est constitué de plusieurs items, chacun étant construit selon certains standards techniques (RSS 1.0, RSS 2.0 et Atom) et comprenant une série de champs obligatoires (date, titre, description, hyperlien). En ce qui concerne les quotidiens d'information, chaque item correspond à un article et contient un titre, qui coïncide généralement avec le titre de l'article, une description, qui correspond soit à un résumé du contenu soit aux premières lignes de l'article, et un lien vers l'article en question. La description est généralement très courte et ne dépasse pas deux ou trois lignes. De façon plus rare, des métadonnées autres peuvent être présentes (mots-clés, nom de l'auteur de l'article par exemple).

Les flux RSS présentent trois avantages majeurs : ils sont librement accessibles et peuvent donc être archivés et documentés ; leur structure est d'une grande homogénéité, ce qui les rend comparables et facilitent donc la constitution de grands corpus ; ils sont produits en temps réel et permettent des analyses fines de la diffusion d'une information dans l'espace médiatique international. Pourtant, les flux RSS restent simultanément peu étudiés dans leur dimension technique et très sous exploités dans leur dimension thématique [Thewall et Prabowo, 2007]. Les rares études produites à ce jour s'intéressent quasi exclusivement aux dimensions techniques du flux RSS [Hmedeh *et al.*, 2011] et concernent notamment l'utilisation de différentes spécifications [Hammond *et al.*, 2004 ; Hammersley 2005]. Une recherche intensive nous a permis de repérer une poignée d'exceptions, ainsi des études portant sur l'utilisation des flux RSS dans le domaine de l'éducation [Duffy et Bruns, 2006] et de l'édition scientifique [Hammond *et al.*, 2004]. Plus récemment, la valeur informationnelle des flux RSS a été étudiée, Marty *et al.* [2009, 2010, 2012] se sont notamment intéressés à la pluralité de l'information disponible en ligne en analysant le contenu de plusieurs centaines de flux RSS. Si les études sur la presse en ligne font preuve d'un dynamisme général [Dagiral et Parasic, 2012], les flux RSS comme dispositif informationnel original ont à ce jour peu mobilisé les chercheurs.

Si la littérature sur le sujet reste réduite, nous pouvons cependant noter que plusieurs chercheurs ont récemment basé tout ou partie de leurs protocoles de recherche sur des flux RSS, sans d'ailleurs questionner ou justifier leurs choix méthodologiques. En France, l'Agence nationale de la recherche (ANR) a financé plusieurs projets dans ce sens : l'*Observatoire Trans-média* (<http://www.otmedia.fr>) dont l'ambition est de détecter les événements médiatiques dans l'ensemble de la sphère médiatique (Web, presse, radio et télévision) ; le projet *ChronoLines* (<http://chronolines.fr>) dont l'objectif est de modéliser des «chronologies basés sur l'événement» en analysant

l'actualité; et le projet *Webfluence* (<http://webfluence.csregistry.org>) qui étudie et modélise les flux d'information en différenciant les flux produits par les médias reconnus (médias traditionnels et pure-players du web) et ceux produits par des communautés de la blogosphère française. De son côté, l'Union européenne a financé les projets *Glocal Project* (<http://www.glocal-project.eu>) qui vise à identifier et à modéliser les événements locaux et globaux à partir de différentes bases de données médiatiques, et la plate-forme *EMM NewsBrief* (<http://emm.newsbrief.eu>) qui collecte en temps réel et en 43 langues de multiples sources afin d'identifier les événements médiatiques les plus couverts (*top stories*).

Un dernier point mérite d'être rappelé. En 2013, Google Reader, le plus gros agrégateur de flux RSS, a fermé et certains acteurs des médias et de l'Internet ont alors annoncé la mort du RSS. Ces derniers restent pourtant produits et diffusés. Même si des moyens alternatifs d'accès à l'information se développent, notamment *via* les pages Facebook des différents médias ou des plates-formes dédiées à la diffusion de contenu vers les smartphones et assimilés, ce format reste encore largement utilisé aujourd'hui par de nombreux quotidiens d'information. Il apparaît donc utile de continuer à explorer les possibilités offertes par ce dispositif technique.

### **Le projet Géomédia : méthodes et axes de recherche**

Dans le cadre du projet ANR Géomédia Corpus, nous construisons une base de données stockant les flux RSS associés à des articles publiés dans une centaine de journaux dans différentes parties du monde afin d'en extraire plusieurs types d'informations (pour la description détaillée du corpus, voir Beauguitte et Pecout [2014]). Le contenu d'information d'un flux RSS (titre et résumé de quelques lignes) est évidemment plus limité que celui d'un article de presse et les tags thématiques ou géographiques ne sont qu'exceptionnellement fournis par les journaux. Il faut donc, en aval de la collecte des flux RSS, mettre en place des procédures de codage du contenu et vérifier si l'ensemble constitué par le titre et le résumé conduit à des codages géographiques et thématiques conformes à ceux qu'aurait produit une analyse intégrale du contenu de l'article. Pour ce faire, nous menons des études de cas pour tester la validité des RSS comme source d'information pour des études territoriales internationales.

Les objectifs majeurs du projet Corpus Géomédia sont d'une part d'identifier les flux d'information entre Etats (quels sont les lieux d'intérêt en fonction de la localisation du média émetteur? Quels sont les lieux souvent associés?) et d'autre part de définir la catégorie d'événement international (peut-on notamment distinguer au sein de notre corpus entre des événements internationaux, régionaux ou globaux?). En raison de la taille du corpus collecté, des procédures automatiques sont nécessaires pour caractériser les items récoltés.

Une première étape fut d'identifier les éléments de localisation susceptibles d'être repérés dans les items. Avec une faible marge d'erreur (5 à 10 % selon les journaux), identifier les noms d'Etats était possible. A l'inverse, excepté pour des leaders politiques mondialement connus (Obama, le pape ou Poutine par exemple), identifier de manière automatique des personnes n'était pas envisageable. Notre corpus étant constitué d'une centaine de flux issus de 40 pays différents et publiés en quatre langues (français, anglais, espagnol et portugais), chaque journal a un niveau de précision variable dans sa façon de nommer les personnes. Créer un dictionnaire des principaux leaders politiques mondiaux (ex- chef d'Etat et premier ministre de l'ensemble des pays du globe) serait une tâche chronophage et d'utilité relative dans la mesure où ce dictionnaire devrait être actualisé très fréquemment. Identifier les Etats était plus aisé et thématiquement pertinent: cela permet d'étudier les hiérarchies de lieux selon les journaux émetteurs. Le seul problème était de créer un dictionnaire exhaustif tout en évitant les faux positifs, notre objectif étant d'avoir une marge d'erreur inférieure à 5 %. La difficulté principale concerne les Etats-Unis pour lesquels de très nombreux termes peuvent être utilisés, qu'il s'agisse de personnes (Obama), d'institutions (FBI), d'Etats nord-américains (Californie), de villes (Washington), voire de lieux précis (Broadway).

En ce qui concerne les événements, il est nécessaire que leur spectre lexical soit aussi réduit que possible afin de pouvoir être identifié dans notre corpus. Ainsi, l'étude de l'épidémie d'Ebola n'a pas soulevé de problèmes méthodologiques majeurs dans la mesure où, quels que soient le journal et la langue, il n'y a pas d'ambiguïté relative aux termes utilisés. De manière similaire, l'identification d'un nom propre d'une personne comme Edward Snowden ou Dominique Strauss-Kahn [Grasland et Severo, 2014] s'est démontrée très efficace. Inversement, extraire tous les items relatifs au terrorisme par exemple serait beaucoup plus délicat donc la mesure où les termes utilisés pour décrire le phénomène sont nombreux et polysémiques (une attaque peut être cardiaque et toucher une personnalité ou désigner une attaque terroriste); de plus, de nouveaux acteurs et de nouveaux termes émergent régulièrement et de façon imprévisible (*Charlie-Hebdo* est par exemple devenu à l'échelle mondiale synonyme d'attaque terroriste en janvier 2015).

Les traitements quantitatifs en cours à ce jour explorent trois pistes principales. La première piste s'intéresse à la persistance temporelle d'événements donnés tels que la pandémie Ebola, la guerre en Syrie [Giraud *et al.*, 2013], les émeutes de Ferguson ou du Wukan [Severo *et al.*, 2012]. Nous cherchons à modéliser les flux d'information dans une perspective temporelle et spatiale afin de répondre aux questions suivantes: quand et comment un événement donné se diffuse-t-il parmi les journaux à une échelle mondiale? Est-il possible d'identifier des motifs et des tendances de diffusion? Quelles sont les barrières (culturelles, linguistiques, politiques) à ces processus de diffusion? Quand des phénomènes de saturation médiatique apparaissent-ils? La seconde piste concerne la hiérarchie des lieux



selon les journaux et la période étudiés [Severo *et al.*, 2014]. La hiérarchie des lieux dans l'actualité internationale est un sujet de recherche ancien et nous examinons si, malgré les changements majeurs de la production et de la diffusion de l'information liés à l'essor de l'Internet, des règles proposés il y a près de 50 ans restent valides aujourd'hui – les premières études semblent montrer que c'est le cas et qu'Internet ne change guère les configurations d'ensemble [Paterson, 2005]. Enfin, une troisième piste consiste à s'intéresser aux cooccurrences de lieux dans les items RSS [Giraud et Severo, 2013]. Ce sujet de recherche, peu abordé à ce jour, permet de mettre en évidence les hiérarchies et les structures politiques globales [Beauguitte et Severo, 2014].

Nous cherchons par ailleurs à valider l'utilisation des flux RSS pour des approches mixtes, mêlant méthodes quantitatives et qualitatives [Giraud et Severo, 2013], et visant à interroger la dimension spatiale de l'actualité internationale. La deuxième partie de cet article s'inscrit dans cette démarche mixte et s'intéresse à la place de l'individu dans le flux RSS international du *New York Times* et aux espaces qui lui sont associés.

### L'INDIVIDU ET SES ESPACES DANS LES ACTUALITÉS INTERNATIONALES<sup>3</sup>

#### Hypothèses et méthodes

Pourquoi s'intéresser à l'individu dans l'actualité internationale? Notre hypothèse principale, fondée sur les conditions de production de l'information, est liée à la nécessité pour les journalistes de personnaliser l'information en général et l'information internationale en particulier [Gans, 1979]. Intéresser un lecteur d'un pays  $x$  à ce qui se passe dans un pays  $y$  suppose de mettre en place des stratégies permettant de rendre l'information compréhensible et attractive: jouer sur l'identification est l'une des stratégies communément employées dans la sphère médiatique [Ostgaard, 1965; Mercier, 2006].

Suivant des études antérieures, nous nous attendons à trouver trois grands types d'individus: les stars mondiales, les élites et les gens ordinaires («*common people*», voir Galtung et Ruge, 1965). Les caractéristiques attendues sont les suivantes:

- la star mondiale est toujours nommée, elle apparaît à plusieurs reprises dans l'actualité et son action se déploie sur une large gamme d'échelles (du local au global);
- la personne ordinaire n'apparaît qu'une fois et n'est pas nommée. Elle est ancrée localement mais, pour justifier sa place dans l'information internationale,

<sup>3</sup> Ce paragraphe constitue une reprise de Beauguitte et Severo [2014]

son cas est considéré comme le symbole d'une situation plus vaste. Elle peut également jouer le rôle de source ou de témoin d'une actualité ;

- les élites apparaissent à plusieurs reprises, sont généralement nommées, leur action se déploie sur une gamme restreinte d'échelle (Etat ou relation bilatérale entre Etats).

En raison de la faible épaisseur temporelle du corpus analysé pour cette première approche exploratoire, seules les deux premières hypothèses ont été testées ici.

L'ensemble des 1 304 éléments du flux RSS de la rubrique internationale émis par le *New York Times* entre le 1er janvier et le 31 mars 2013 a été codé par les deux auteurs afin de répondre aux questions suivantes : y a-t-il présence d'un individu dans l'item du flux ? L'individu est-il nommé ou anonyme ? Est-il possible de déterminer l'échelle de l'actualité (locale, nationale, mondiale etc.) ? Ce dernier item a été le plus délicat à coder : de rares flux sont a-spatiaux (« *Oscar Pistorius, the double amputee Olympic runner accused of murder, will be allowed to travel abroad under strict conditions* »<sup>4</sup>, item du 28 mars 2013), d'autres mêlent plusieurs espaces de référence (« *The United States government has petitioned the Irish High Court to extradite an Algerian man living in Ireland on terrorism charges relating to a plot to kill a Swedish artist* »<sup>5</sup>, item du 1<sup>er</sup> mars 2013 – cet item n'est absolument pas représentatif dans la mesure où sa complexité spatiale est très largement supérieure à l'immense majorité des items). Les catégories utilisées sont les suivantes : 0 (pas d'espace explicite), local (infra-national, exemple une ville, un Etat américain, etc.), national, inter-national (relation entre deux ou trois Etats), régional (supra-national comme Europe, Asie, etc.), global.

L'étude du corpus montre une information internationale fortement individualisée : plus de la moitié des articles implique au moins un individu (698 sur 1 304) et ces individus sont le plus souvent des personnes clairement identifiées (75 % des individus ont un nom).

Le tableau ci-dessous donne la répartition obtenue lorsque l'on croise identité/anonymat et échelle.

	0	local	national	international	régional	global	total
Anonyme	3	42	57	52	4	13	171
Nommé	26	49	236	148	17	51	527
Total	29	91	293	200	21	64	698

Tableau 1 : Répartition des items par identité/anonymat et échelle

4 Oscar Pistorius, le coureur olympique double amputé accusé d'assassinat, sera autorisé à voyager à l'étranger dans des conditions strictes.

5 Le gouvernement des Etats-Unis a demandé à la Haute Cour irlandaise d'extrader un Algérien vivant en Irlande sur des accusations de terrorisme liées à un complot visant à tuer un artiste suédois.

La forte domination des échelles nationales et internationales (entendue au sens de relations entre deux ou trois Etats) doit être soulignée et ce, que les individus soient anonymes ou non. Et, contrairement à notre hypothèse, l'individu anonyme peut être l'objet d'une information locale, nationale ou globale. Le nombre élevé d'anonymes au niveau global s'explique par nos choix de codage : tout observateur lié aux Nations Unies et rapportant une situation locale a été considéré comme lié à une actualité globale (« *A United Nations investigator, whose report on the C.I.A.'s use of secret detentions and interrogation techniques is to be released, called on the United States to publicly* »<sup>6</sup>, item du 5 mars 2013).

### Stars globales et anonymes : fréquence et espaces d'apparition

Si le lien entre individu et espace est moins flagrant qu'attendu, la fréquence des apparitions conforte nos hypothèses et rejoint d'ailleurs les conclusions des études relatives à la distribution des sujets d'actualités [Smyrnaioi *et al.*, 2010; Marty *et al.*, 2010] : une poignée d'individus est citée régulièrement, la grande majorité apparaît une ou deux fois avant de disparaître de l'actualité. Si 334 noms différents peuvent être identifiés dans les titres et les résumés des items, 249 apparaissent une seule fois (nous sommes donc très proches des 80/20 des distributions de Pareto où 20 % d'une population concentre 80 % d'une propriété donnée). A l'autre extrémité, les personnalités du monde politique dominent très largement (32 occurrences pour Obama et Chavez, suivis de Benoît XVI, François I<sup>er</sup>, Netanyahu, Bashar el Assad, etc.). Sur les 20 personnes évoquées au moins 5 fois durant ces trois mois, seul l'athlète Oscar Pistorius est extérieur au monde politique. La capacité à se maintenir comme sujet ou acteur de l'actualité est très inégalement distribuée : les 9/10 des individus évoqués (anonymes ou non) le sont deux fois maximum durant la période de trois mois considérée.

Lors de la période étudiée, la démission de Benoît XVI et l'élection du pape François I<sup>er</sup> donnent lieu à un grand nombre d'items RSS (19 liés à la démission du premier et 20 liés à l'élection du second) et la multiplicité des échelles mobilisées est impressionnante : les actualités varient du monde (*the Catholic Church*) au local (*St. Peters Square*) en mobilisant toutes les échelles intermédiaires (le Vatican comme Etat, l'Amérique latine comme région).

La figure ci-dessous (Figure 1) montre comment se distribuent les apparitions papales dans les flux RSS en fonction des échelles mobilisées – les articles a-spatiaux ne sont pas représentés ici. Dès son apparition dans l'actualité, le futur François I<sup>er</sup> est sujet ou acteur d'une actualité multiscalaire. La démission de son prédécesseur est, elle, évoquée tantôt au niveau local, tantôt au niveau global.

6 Un enquêteur des Nations Unies, dont un rapport sur l'utilisation par la CIA de détentions secrètes et de techniques d'interrogatoire doit être publié, a accusé les Etats-Unis publiquement.

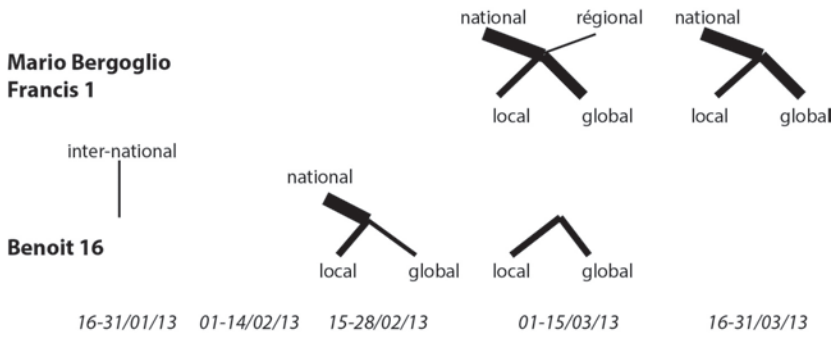


Figure 1 : Benoît XVI et François I<sup>er</sup> dans le flux RSS international du NYT

A titre de comparaison, la figure 2 montre les échelles de référence des deux personnes les plus présentes dans l'actualité internationale du *New York Times* de janvier à mars 2013, Barak Obama et Hugo Chavez.



Figure 2 : Obama et Chavez dans le flux RSS international du NYT

Obama est durant la période considérée la seule personnalité qui apparaisse à une telle fréquence et avec une telle pluralité d'échelles. Il faut signaler qu'il n'y a pas de biais lié à la nationalité du journal : Barak Obama est la personnalité politique la plus citée dans la grande majorité des journaux de notre corpus. Hugo Chavez est présent selon des modalités différentes : son hospitalisation à Cuba crée une actualité internationale (Vénézuéla, Cuba), les conséquences de son décès sont évoquées à toutes les échelles possibles (réactions locales au décès, deuil national, hommage dans des pays tiers, conséquences au niveau régional, etc.), et enfin sa succession devient un problème national avant que son nom ne disparaisse de l'actualité.

Enfin, le cas Pistorius est intéressant car il permet de saisir en partie la temporalité d'une actualité. Si du point de vue de l'échelle, l'actualité est toujours à une échelle locale – explicite ou non (cf. *supra*) –, du point de vue temporel l'articulation est plus complexe. Deux temps apparaissent : un premier temps, très resserré, où 5 items le citent (20 au 22 février) et qui correspond à l'apparition du fait divers dans l'actualité ; un second temps, moins riche (2 items seulement), et plus lâche (un

item le 8 mars et un le 28 mars) et correspondant à des rebondissements mineurs de l'affaire. Prendre une période plus longue permettrait de mettre en évidence la construction narrative en feuilleton des actualités liées aux faits divers, construction expliquant que certains éléments (le lieu notamment) ne soit pas systématiquement rappelé – le lieu est supposé connu par le lectorat «suivant» l'actualité.

Inversement, on peut prendre en considération seulement les articles qui concernent des personnes ordinaires et appliquer la classification de Gans [1979] des personnes anonymes pour étudier leur relation avec l'espace. On peut distinguer notamment : les manifestants, les «émeutiers», les grévistes, mentionnés dans 14 items ; les fonctionnaires tels les policiers, soldats, juges et bureaucrates fédéraux, mentionnés dans 39 items ; les victimes, présentes dans 30 items ; les contrevenants présumés et réels des lois (souvent dans des contextes de guerre), la catégorie la plus représentée avec 46 items. Il y a également un groupe d'individus qui apparaît difficilement catégorisable, comme les cardinaux liés à l'élection du pape (nous avons dû créer une catégorie «autre» pour 15 items).

La relation la plus évidente qui émerge est entre l'échelle locale, d'un côté, et activistes et victimes, de l'autre. Une actualité liée à une protestation ainsi qu'à un fait violent qui touche une victime est souvent localisée de manière précise. Une exception est liée au cas du garçon russe trouvé mort au Texas, ce qui rend l'actualité internationale («*Adopted Boys Death in U.S. Stirs Outrage in Russia. The authorities in Texas are careful about commenting on a 3-year-olds death, with investigations continuing*»<sup>7</sup>). Pour les autres catégories, il n'est pas toujours possible d'identifier des connections aussi claires. Par exemple, les fonctionnaires peuvent être liés à toutes les échelles et dans de contextes très différents. A noter qu'ils sont souvent cités comme source de l'information («... an official in Dar es Salaam said», «... a top official with the World Health Organization said» ou «A judge said that...») et, dans ce cas précis, il faut prendre en compte le phénomène de l'anonymat des sources («said a military official who declined to be named») [Lambert, 2001].

Un sujet d'études prometteur concernant les liens entre espaces et médias est le fait de nommer ou non les personnes, notamment lorsque celles-ci appartiennent à l'élite politique d'un Etat. Ainsi il semble, mais des études approfondies seraient nécessaires pour confirmer ce fait, que si les responsables politiques américains sont systématiquement nommés, ce n'est pas le cas que pour les chefs d'Etat et les ministres étrangers de certaines nationalités (Chine et Royaume-Uni notamment). Les responsables ministériels d'autres Etats ne sont nommés qu'exceptionnellement. Dans notre échantillon, 21 articles concernent des autorités étrangères non nommées («*The foreign ministers of Japan and the Philippines*»,

7 La mort d'un garçon adopté aux Etats-Unis provoque un scandale en Russie. Les autorités du Texas font preuve de prudence en commentant la mort d'un enfant de trois ans, tandis que les enquêtes se poursuivent.

«*Greek Prime Minister*»). Une étude comparative avec les flux émis par d'autres journaux d'origine différente donnerait sans doute des résultats intéressants.

## CONCLUSION

L'objectif de ce chapitre était double : souligner l'intérêt des flux RSS dans les études médiatiques ; s'intéresser à la place de l'individu dans l'actualité internationale. Un troisième objectif était de montrer l'intérêt d'une analyse qualitative d'une information simultanément abondante et pauvre, les flux RSS, dans des études territoriales.

Si le flux RSS est un dispositif largement utilisé par les médias d'information, peu d'études ont à ce jour cherché à les exploiter thématiquement. Le corpus construit au sein de l'ANR Géomédia et destiné à être rendu accessible à la communauté à l'issue du projet (démarche d'open data) fournira une base documentée susceptible d'alimenter des travaux de recherche liés à l'actualité internationale. Le flux RSS est certes une donnée qualitativement pauvre (quelques lignes de texte seulement), il permet pourtant de saisir diverses logiques relatives à la construction et à la diffusion de l'actualité. La facilité de sa collecte pour des médias divers est un atout pour l'étude comparative de médias de différentes origines.

L'analyse exploratoire menée ici montre que l'individu, anonyme ou non, est mobilisé très fréquemment par les journalistes pour mettre en scène l'actualité internationale. Une poignée d'individus est susceptible de revenir fréquemment dans l'actualité et à des échelles variées tandis que la très grande majorité fait une apparition rapide, souvent anonyme et généralement mono-scalaire. Dans le cas de faits divers impliquant des célébrités, une construction feuilletonesque, en épisodes, apparaît. La courte étude de cas présentée dans ce texte ne permet pas de généraliser ces évidences à toute l'actualité internationale et des démarches comparatives (types de journaux différents, nationalités de journaux variées) sur une période plus longue apporterait sans nul doute d'intéressants résultats.

## RÉFÉRENCES BIBLIOGRAPHIQUES

Tous les liens indiqués étaient valides lors de l'écriture de ce chapitre (avril 2015).

[AFP, 2010] AFP, 2010, *Abécédaire de l'AFP. Comment l'Agence informe sur le monde*, Paris : Victoire Editions.

[Beauguitte et Pecout, 2014] Beauguitte, L. et Pecout, H., 2014, « Courte présentation du corpus Géomédia », *Blog Corpus-Géomédia*, <http://geomedia.hypotheses.org/209>.

[Beauguitte et Severo, 2014] Beauguitte, L. et Severo, M., 2014, « Les flux RSS pour les études territoriales : une analyse de l'individu et ses espaces dans les actualités ».

- internationales du *New York Times*», In *Proceedings du 2e colloque international du CIST*, pp 32-37.
- [Brooker-Gross, 1983] Brooker-Gross, S.R., 1983, «Spatial Aspects of Newsworthiness», *Geografiska Annaler, Series B, Human Geography*, 65(1), 1-9.
- [Chaudhary, 2001] Chaudhary, A.G., 2001, «A Comparative Content Analytical Study of Negative News in Western and Third World Newspapers», *Asian Journal of Communication*, 11(1), 25-50.
- [Dagiral et Parasio, 2012] Dagiral, E. et Parasio, S., 2012, «Presse en ligne : où en est la recherche?», *Réseaux*, 2, 13-42.
- [Duffy et Bruns, 2006] Duffy, P.D. et Bruns, A., 2006, «The use of blogs, wikis and RSS in education: A conversation of possibilities», *Proceeding Online Learning and Teaching Conference 2006*, pp 31-38.
- [Earl *et al.*, 2004] Earl, J., Martin, A., McCarthy, J. D. et Soule, S., 2004, «The Use of Newspaper Data in the Study of Collective Action», *Annual Review of Sociology*, 30(1), 65-80.
- [Galtung et Ruge, 1965] Galtung, J. et Ruge, M. H., 1965, «The Structure of Foreign News», *Journal of Peace Research*, 2(1), 64-91.
- [Gans, 1979] Gans, H., 1979, *Deciding what's news: A study of CBS evening news, NBC nightly news, Newsweek, and Time*, New York: Pantheon Books.
- [Gill, 2005] Gill, K., 2005, *Blogging, RSS and the Information Landscape: A Look At Online News*, In WWW2005, [http://faculty.washington.edu/kegill/pub/gill\\_www2005\\_rss.pdf](http://faculty.washington.edu/kegill/pub/gill_www2005_rss.pdf).
- [Giraud et Severo, 2013] Giraud, T. et Severo, M., 2013, «Le périple d'Edward Snowden. Analyse quali-quantitative d'un événement médiatique international», *Netcom. Réseaux, communication et territoires*, 27(3-4), 385-410.
- [Grasland et Severo, 2014] Grasland, C. et Severo, M., 2014, «Qu'est-ce qu'un événement mondial?», In *L'Atlas global*, Paris: Les Arènes, pp 130-133.
- [Herkenrath et Knol, 2011] Herkenrath, M. et Knol, A., 2011, «Protest Events in International Press Coverage: An Empirical Critique of Cross-National Conflict Databases», *International Journal of Comparative Sociology*, 53(3), 163-80.
- [Hmedeh *et al.*, 2011] Hmedeh, Z., Travers, N. et Vouzoukidou, N., 2011, «Everything you would like to know about RSS feeds and you afraid to ask», *Base de Données Avancées Morocco*, [http://hal.archives-ouvertes.fr/docs/00/73/72/43/PDF/art\\_2158.pdf](http://hal.archives-ouvertes.fr/docs/00/73/72/43/PDF/art_2158.pdf).
- [Lambert, 2001] Lambert, F., 2001, *Figures de l'anonymat: médias et société*, Paris: L'Harmattan.
- [Marty *et al.*, 2009] Marty, E., Rebillard, F. et Smyrnaio, N., 2009, «Online News between Redundancy and Diversity: A Quantitative Study of French Speaking News Websites», communication présentée au *Colloque new media and communication*, Athènes.
- [Marty *et al.*, 2010] Marty, E., Rebillard, F., Smyrnaio, N. et Touboul, A., 2010, «Variété et distribution des sujets d'actualité sur Internet. Une analyse quantitative de l'information en ligne», *Mots. Les langages du politique*, 93, 107-126.
- [Marty *et al.*, 2012] Marty, E., Rebillard, F., Pouchot, S. et Lafouge, T., 2012, «Diversité et concentration de l'information sur le web», *Réseaux*, 176(6), 27.

- [McCarthy et McPhail, 1996] McCarthy, J. et McPhail, C., 1996, «Images of Protest: Dimensions of Selection Bias in Media Coverage of Washington, 1982 and 1991», *American sociological review*, 61(3), 478-499, <http://www.jstor.org/stable/10.2307/2096360>.
- [Mercier, 2006] Mercier, A., 2006, «Logiques journalistiques et lecture événementielle des faits d'actualité», *Hermès*, 46, 23-35.
- [Ostgaard, 1965] Ostgaard, E., 1965, «Factors Influencing the Flow of News», *Journal of Peace Research*, 2(1), 39-63.
- [Paterson, 2005] Paterson, C., 2005, «News Agency Dominance in International News on the Internet», In Skinner, D., Compton, J.R. et Gasher, M. (dir.), *Converging Media, Diverging Politics: A Political Economy of News Media in the United States and Canada*, Lanham: Lexington Books, pp 145-64.
- [Robinet, 2013] Robinet, F., 2013, «La fabrique médiatique des événements internationaux : Afrique, un continent en marge (1994–2008) ?», *Le Temps des médias*, 20(1), 152-67.
- [Severo *et al.*, 2014] Severo, M., Douay, N., Giraud, T., Grasland, C., Nielsen, M., Pecout, H. et Texier, M., 2014, *Tools (2011-2014). Feasibility Study on Analytical Tools based on Big Data*, final report, ESPON project.
- [Severo *et al.*, 2012] Severo, M., Giraud, T. et Douay N., 2012, «The Wukan's protests: just-in-time identification of international media event», In *Proceeding of Workshop Just-in-Time Sociology, SocInfo international conference*.
- [Smyrnaioi *et al.*, 2010] Smyrnaioi, N., Marty, E. et Rebillard, F., 2010, «Does the “Long Tail” apply to online news? A quantitative analysis of French-speaking websites», *New Media and Society*, 12(8), 1244-1261.
- [Thelwall et Prabowo, 2007] Thelwall, M. et Prabowo, R., 2007, «Identifying and characterizing public science-related fears from RSS feeds», *Journal of the American Society for Information Science and Technology*, 58(3), 379-390.
- [Wu, 2000] Wu, D.H., 2000, «Systemic Determinants of International News Coverage: A Comparison of 38 Countries», *Journal of Communication*, 50(2), 110-30.



# Table des matières

INTRODUCTION .....	7
<i>Marta Severo, Alberto Romele</i>	
<b>PARTIE 1 - LA TRACE, LES MÉTHODES ET LES DONNÉES.....</b>	<b>11</b>
AU-DELÀ DE LA CRITIQUE <i>BIG DATA</i> .....	13
<i>Richard Rogers</i>	
LES MÉTHODES D'INTERFACE .....	33
<i>Noortje Marres, Carolin Gerlitz</i>	
SOFT DATA .....	61
<i>Marta Severo, Alberto Romele</i>	
L'IDENTITÉ COMME BASE DE DONNÉES .....	87
<i>Jos de Mul</i>	
<b>PARTIE 2 - RENCONTRE ENTRE TRACES NUMÉRIQUES ET TERRITOIRES .....</b>	<b>109</b>
L'ÉCUME NUMÉRIQUE DES TERRITOIRES .....	111
<i>Dominique Boullier</i>	
SPATIALITÉS ALGORITHMIQUES .....	133
<i>Boris Beaude</i>	
L'HYPERVILLE .....	161
<i>Franck Cormerais</i>	
DÉSIRS DE DATA .....	177
<i>Maryse Carmes, Jean-Max Noyer</i>	

---

PARTIE 3 - LES PRATIQUES DE LA TRACE NUMÉRIQUE.....	211
DE LA TRACE À LA CARTE ET DE LA CARTE À LA TRACE .....	213
<i>Matthieu Noucher</i>	
DIMENSIONS SPATIALES DE L'ACTUALITÉ INTERNATIONALE.....	225
<i>Laurent Beauguitte, Marta Severo</i>	
RECONFIGURATION DES PRATIQUES PARTICIPATIVES .....	239
<i>Nicolas Douay, Maryvonne Prévot</i>	
REMERCIEMENTS .....	259
LES AUTEURS.....	261