

Sources numériques et flux mondiaux d'information.

Laurent Beauguitte (Chargé de recherche, CNRS, UMR IDEES - laurent.beauguitte@cnrs.fr)
Hugues Pecout (Ingénieur d'études, CNRS, UMS RIATE - hugues.pecout@gis-cist.fr)

Rencontres Réseaux et histoire, Nice, 22-24 septembre 2016

Les données numériques contemporaines permettent *a priori* de constituer des corpus exhaustifs et homogènes de grande taille, ce que ne permettent qu'exceptionnellement les archives utilisées par les historiens ou les traces des archéologues. Cette vision naïve mais répandue oublie que, comme toute donnée mobilisée en science sociale ou « dure », la donnée numérique est toujours construite et travaillée en fonction d'un objectif de recherche donné (Bastard *et al.*, 2013). Les traces numériques ne permettent d'améliorer notre connaissance du monde social qu'une fois comprises leurs logiques de production et leurs limites interprétatives (Ollion et Boelaert, 2016). Même la supposée complétude des données recueillies, données qui nécessitent toujours un important travail de nettoyage et de (re)codage, est sujette à caution : qu'il s'agisse des utilisations de facebook (projet Algopol, <http://algopol.huma-num.fr/>), de données de téléphonie mobile (Prieur *et al.*, 2009) ou des usages de twitter (Raux, 2014), tous les articles publiés à ce jour supposent la construction d'échantillons dont la représentativité n'est que rarement assurée.

L'expérience acquise au sein de l'ANR Corpus-Géomédia (2012-2016), dont l'objectif était de construire un corpus de flux RSS¹ relatifs à l'actualité internationale, nous a confirmé que, tant en termes de taille que d'homogénéité, la source numérique reste difficilement exploitable et les résultats obtenus fragiles. Traiter des *big data* ou des archives expose certes à des problèmes différents mais le regard critique sur les sources et leurs lacunes est une exigence commune. Traditionnellement, l'étude de l'actualité internationale était basée sur la reconstruction de semaine-type, se limitait souvent à un quotidien et s'intéressait à la hiérarchie de l'information ainsi qu'aux images-pays produites (vision négative ou positive). Cette tradition de recherche datée des années 40 dans le monde anglophone a été bouleversée par l'apparition des outils numériques en ligne mais les questions de recherche se sont peu renouvelées (Beauguitte, 2014).

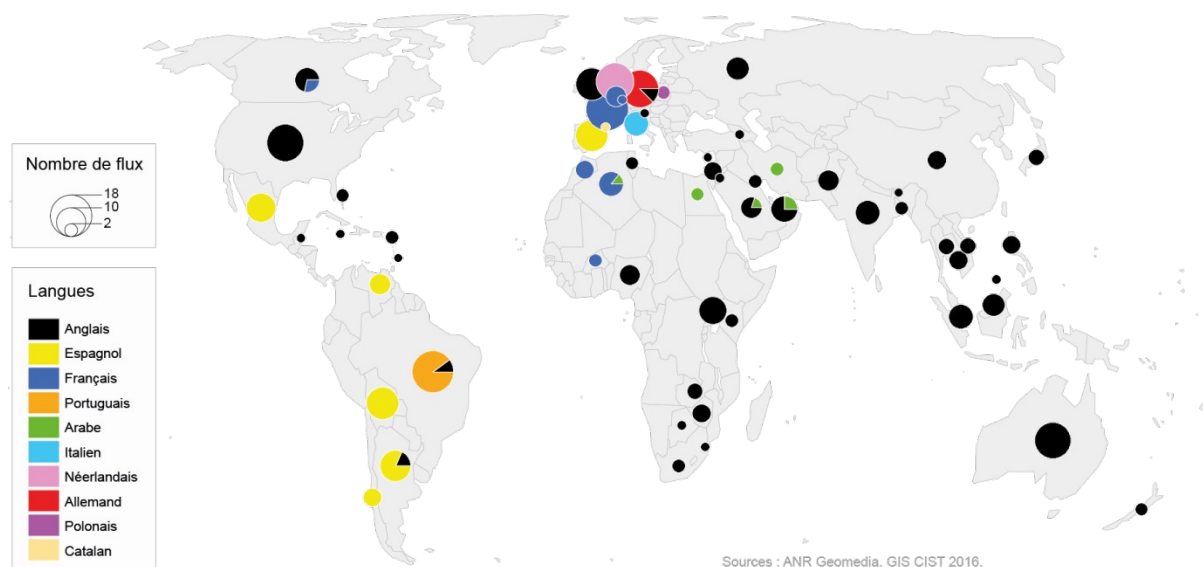
Le billet est organisé en trois parties : une première partie présente la base de données récoltée, une deuxième partie montre comment l'analyse de réseaux est mobilisée pour conceptualiser les questions de recherche, enfin la troisième partie s'intéresse aux cooccurrences de lieux cités, cooccurrences supposées refléter la structure politique du système-Monde contemporain.

1. La collecte des données numériques

Dès fin 2013, un outil développé dans le cadre de l'ANR Corpus-Géomédia a commencé à capter progressivement des données provenant de flux RSS médiatiques. Aujourd'hui (juillet 2016), les données en provenance de 328 flux RSS médiatiques sont collectés en continue. Ces flux RSS sont proposés par 193 différents média, originaires de 62 pays. Ainsi, 25 000 nouveaux items sont collectés chaque jour, et la base de données est actuellement composée de plus de 15 millions d'items RSS médiatiques en 10 langues (80% du corpus est composé de flux RSS anglais, espagnols ou français).

¹ Le RSS est un format d'échange de données utilisé pour la syndication de contenu web. Il s'agit d'un fichier au format XML produit automatiquement et qui synthétise le contenu d'une page web. Il est très largement utilisé par les sites d'actualités pour diffuser en temps réel des informations sur les nouveaux articles publiés en ligne.

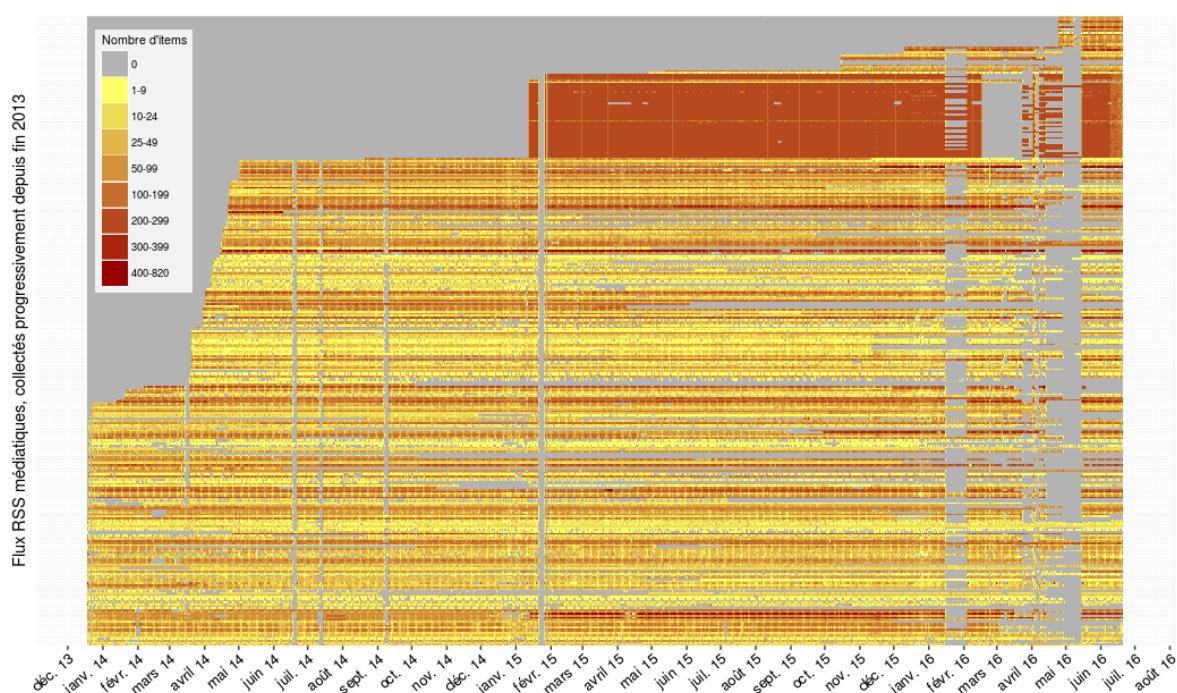
Figure 1 : Répartition géographique et linguistique des flux RSS collectés (juillet 2016)



La masse d'information collectée progressivement depuis 2 ans et demi est donc considérable. Cependant, deux problèmes majeurs altèrent la qualité de cette base : l'aspect volatil des sources et l'homogénéité relative des données.

Tout d'abord, l'alimentation en continue de la base de données est dépendante de la technique (serveur en maintenance, coupure réseau) et des interventions opérées par les média (refonte de site web, nouvelle catégorisation des articles, utilisation d'un nouvel outil de diffusion...). Même un suivi régulier et manuel de la base de données ne permet pas d'éviter les ruptures de collecte. La stabilité et la pérennité de la captation des données RSS médiatiques n'est donc jamais garantie.

Figure 2 : Nombre d'items collectés quotidiennement, pour chaque flux RSS branché



De plus, le maintien d'une telle base de données à long terme est sans doute compromis, car la technologie RSS semble petit à petit être délaissée au profit des réseaux sociaux (Twitter, Facebook...)

L'homogénéité de l'information reçue peut également être remise en question. L'objectif étant de construire un corpus relatif à l'actualité internationale, les flux RSS collectés sont essentiellement des flux RSS nommés «monde» ou encore « international » sur les différents sites web. Mais la classification thématique des actualités médiatiques est propre à chaque média. Elle est influencée par sa localisation géographique et bien sûr par ses choix éditoriaux. Le corpus final est donc avant tout composé d'items catégorisés « international » par leur producteur, plus que d'items relatifs à l'actualité internationale.

Au-delà de l'impact des nomenclatures thématiques établies par les médias, l'homogénéité du corpus est également altérée par la variété de contenu et de masse d'information transmise. Bien que la structure d'un item RSS réponde à des normes établies, leur contenu dépend directement de choix éditoriaux. Par exemple, chaque média choisit (et peut modifier sans aviser les utilisateurs) l'information transmise dans le champ 'description' d'un item RSS. La description peut ainsi contenir soit un résumé d'article, soit un début d'article, soit un article dans sa globalité, ou encore ne comporter aucune information. Cette hétérogénéité s'observe également au niveau de la masse d'item mise à disposition par chaque flux RSS. L'homogénéité du corpus est donc très relative, et plus les sources sont nombreuses, plus l'hétérogénéité et la présence de particularités sont assurées.

Une dernière critique peut être soulignée quant à l'homogénéité spatiale de ce corpus, qui se trouve logiquement liée à la qualité de l'infrastructure internet, à l'offre médiatique et à la langue officielle locale. Certaines régions de monde sont ainsi moins bien couvertes que d'autres (Afrique, Asie centrale, Europe de l'Est). L'exploration de la base de données dans son ensemble est donc problématique, et il est préférable de construire des sous-corpus homogènes plus comparables avant de réaliser une analyse.

Afin de constituer un échantillon de flux RSS le plus représentatif de la presse quotidienne mondiale pour ce cas d'étude, nous avons décidé de retenir des flux provenant de quotidiens d'origine géographique variée, tout en prenant en compte nos limites linguistiques. Il était également nécessaire pour assurer la comparabilité des données que le volume d'items produit ne soit pas trop variable d'un flux à l'autre et que la structure même des items soit comparable. Tous les flux RSS retenus comportent ainsi un titre court et une description de deux ou trois lignes (soit 250 à 300 caractères) correspondant à un résumé de l'article. En ce qui concerne les critères quantitatifs de sélection, les flux retenus ont une moyenne hebdomadaire d'items envoyés supérieure ou égale à 30 et un minimum hebdomadaire supérieur à zéro (ces derniers cas indiquant des ruptures de collecte).

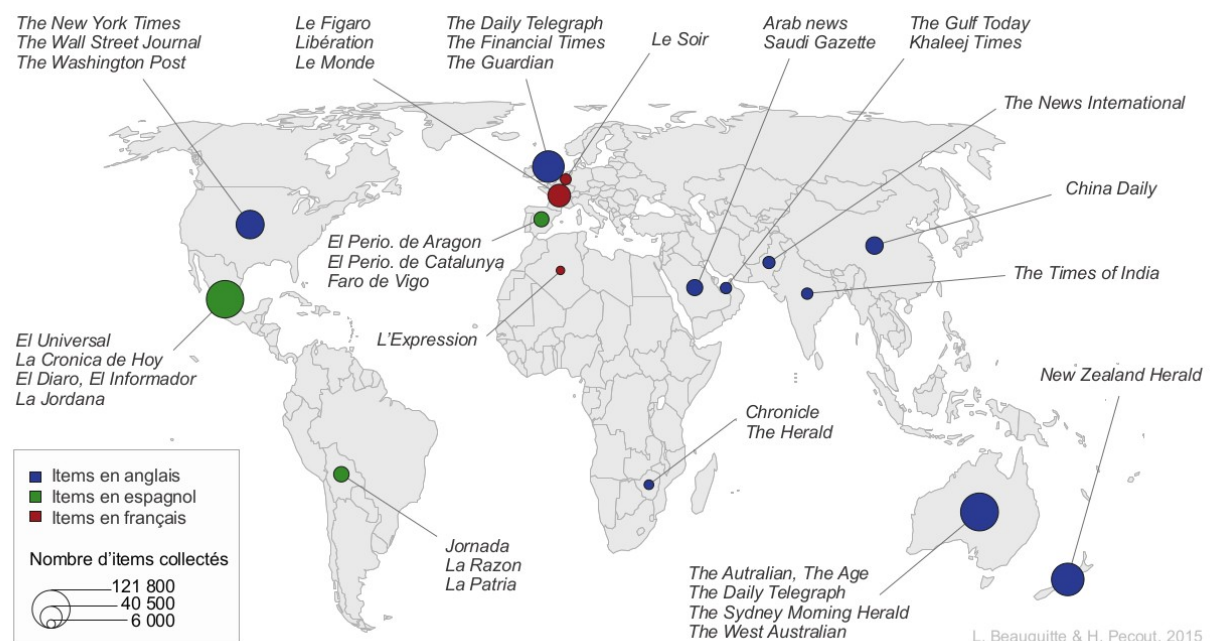
Ces critères qualitatifs et quantitatifs minimisent l'hétérogénéité du corpus en terme de contenu, mais ne garantit pas une parfaite homogénéité. Ils ne permettent pas non plus d'avoir une couverture mondiale équilibrée, d'autant que si pour les pays francophones, anglophones et hispanophones, les journaux étudiés s'adressent avant tout à des audiences nationales, ce n'est pas nécessairement le cas pour les autres journaux du corpus. Étudier le contenu d'un journal chinois ou saoudien diffusé en anglais, c'est en grande partie étudier un contenu pensé et diffusé pour un public étranger.

Ainsi, le sous-corpus sélectionné pour ce cas d'étude se compose de trente-sept flux RSS relatifs à l'actualité internationale et générés par des quotidiens de presse écrite en français,

en anglais ou en espagnol. La première période d'étude s'étend d'avril à décembre 2014, la seconde du 1 février au 31 octobre 2015. Le corpus étudié dans cette seconde période comprend 33 quotidiens, certains journaux ayant cessé d'émettre des flux RSS internationaux (*Khaleej Times* et *El Universal*), d'autres flux ayant connu des ruptures de collecte (*The Washington Post*, *La Razon*).

Chaque item de ces flux RSS correspond à un article paru dans la version en ligne du quotidien étudié. Il peut être soit un article rédactionnel qui apparaîtra également dans la version papier, soit un article rédactionnel qui apparaîtra seulement dans la version en ligne, soit la simple reprise d'un dépêche d'agence (non publié ensuite dans la version papier), soit l'actualisation en temps réel d'un fait donné (précision sur le nombre de victimes dans le cas d'un accident grave par exemple).

Figure 3 : Composition du corpus étudié



2. Des flux RSS aux réseaux

Dès les premiers stades du projet, l'item RSS a été conceptualisé de différentes manières sous forme de graphe uni et biparti. Il ne saurait ici être question de présenter l'ensemble des structures réticulaires envisagées mais de présenter celles ayant été effectivement mobilisées dans divers travaux (Beauguitte *et al.*, 2014 et 2016 ; Queyroi *et al.*, 2015).

Considéré de manière individuelle, l'item RSS, une fois étiqueté au niveau géographique et/ou thématique, peut être envisagé comme support d'un graphe biparti de trois manières complémentaires :

- réseau de liens entre items et entités géographiques ;
- réseau de liens entre entités géographiques citées dans un même item (cooccurrence) ;
- réseau de liens entre items et thématiques.

Si l'on agrège les items de journaux d'un même État, on peut constituer une matrice origine - destination de taille $N \times M$ où N correspond au nombre d'États où des flux sont capturés et M l'ensemble des États considérés, chaque case indiquant combien de fois, durant une période

donnée, un État M_j est cité par les journaux d'un État N_i . Cette matrice incomplète et à diagonale non vide (présence de boucles) ne saurait être assimilée à un réseau unimodal mais permet cependant des analyses relatives à la hiérarchie différentielle des acteurs étatiques. Étudier ces matrices durant plusieurs années permet de différencier les acteurs majeurs de l'actualité internationale, toujours cités souvent quelle que soit l'origine du média émetteur, les acteurs régionaux cités par les seuls quotidiens proches et les acteurs intermittents susceptibles de « faire l'actualité » de manière épisodique.

L'étude des cooccurrences de lieux dans un même item complète l'analyse hiérarchique des acteurs car elle permet de dessiner les paysages géopolitiques relationnels perçus et/ou construits par les quotidiens. Les matrices obtenues, de taille $M \times M$, fournissent un premier aperçu sur les dyades les plus fréquentes au sein de notre corpus, information restant sommaire dans la mesure où elles ne permettent pas de différencier entre relations conflictuelles et relations coopératives.

Les graphes obtenus, combinés aux données sur les hiérarchies des acteurs, peuvent cependant être interprétés dans le cadre du modèle théorique centre-périphérie. En effet, si l'on reprend la typologie d'Alain Reynaud (1981), typologie construite sur la structure inégale des échanges (de biens, de capitaux, de personnes, etc.), on devrait pouvoir distinguer des centres, des périphéries intégrées, des périphéries délaissées et des isolats. Quelle pourrait être la traduction médiatique de ces catégories ? Les centres seraient les États apparaissant systématiquement parmi les États les plus cités, et le plus souvent de façon isolée : en d'autres termes, l'actualité interne à un État du centre est jugée suffisamment importante à l'extérieur pour devenir une actualité internationale de façon durable et hors événement dramatique. La périphérie intégrée pourrait être constituée par les États régulièrement cités mais très souvent associés aux États centraux. La périphérie délaissée serait elle composée d'États qui n'apparaissent que rarement dans l'actualité, exceptionnellement seuls, et le plus souvent pour des événements particulièrement dramatiques. Enfin, les isolats seraient les absents des flux médiatiques.

Ajouter une approche thématique permet de produire plusieurs types de réseaux différents :

- réseau bimodal thème – lieu d'émission ;
- réseau bimodal thème – lieu cité

et d'en dériver des réseaux unimodaux valués permettant de déterminer quels journaux parlent des mêmes lieux en fonction des mêmes thématiques à une période donnée. Si cette approche est potentiellement très riche, elle est cependant coûteuse dans la mesure où elle suppose la construction et la validation d'un dictionnaire multilingue lié au thème. Le risque de faux positifs (item rattaché à un thème par erreur) et de faux négatifs (item non taggué alors que le thème est traité) est plus important ici et seuls deux thématiques ont à ce jour été traités dans le cadre de l'ANR Corpus-Géomédia : les séismes et l'épidémie d'Ebola (Le Texier *et al.*, 2016). Si ce dernier cas était relativement simple, le nom de la maladie permettant d'identifier sans problème les items concernés, l'identification des séismes a été plus difficile en raison du caractère polysémique du terme (les journalistes francophones aiment l'expression « séisme politique ») que du caractère ambigu de certains événements (lorsque Fukushima est cité, est-ce en raison du séisme initial, du tsunami, des conséquences nucléaires, ou de tout ceci à la fois?).

Si construire des réseaux à partir des données récoltées n'a pas soulevé de problème conceptuel ou méthodologique majeur, le choix des mesures et traitements a lui donné lieu à davantage de discussions, le nombre de mesures disponibles n'ayant cessé d'augmenter ces deux dernières décennies (Newman, 2010). Nous avons fait le choix de nous limiter à quelques indicateurs de base, essentiellement le degré (matrices émetteur – État $N \times M$ et matrices de cooccurrences étatiques $M \times M$). Combiner le degré et le taux de cooccurrence

permet de cartographier le résultat et de mettre en évidence d'éventuelles logiques de voisinage. Des tests de détection de communautés, comparant différents algorithmes, ont été utilisés pour produire une régionalisation mondiale sur les matrices $M \times M$. Nous n'avons pas souhaité utiliser des indicateurs autres pour deux raisons principales : nos graphes sont évalués or les outils disponibles permettent rarement de prendre en compte cette valuation des liens ; les données de départ restent, malgré nos précautions, difficilement comparables tant dans le temps (évolution des politiques éditoriales) que dans l'espace. Il nous semble, peut-être à tort, que plus les données sont fragiles, plus les mesures utilisées doivent être robustes et simples à interpréter.

3. Les couples de lieux médiatiques

L'étude des cooccurrences a rarement été abordée dans la littérature sur les déterminants de l'actualité internationale. Elle présente pourtant le double avantage de s'intéresser aux aspects relationnels et de pouvoir s'étudier dans le cadre théorique du modèle centre-périphérie. Ce modèle a été appliqué aux données journalistiques lors du débat sur le Nouvel ordre mondial de l'actualité porté par l'Unesco dans les années 80. Un double constat fondait ce débat : les agences de presse sont (si le débat est ancien, la localisation des agences de presse n'a pas varié) toutes situées au nord et parlent de fait peu des pays du Sud ; l'image des pays du sud est le plus souvent négative et constitue un frein au développement. L'utilisation du modèle centre-périphérie est ici différente dans la mesure où nous lions nombre d'occurrences et fréquence des cooccurrences pour tester les hypothèses suivantes :

- les acteurs majeurs sont souvent cités et souvent cités seuls (l'actualité interne d'un acteur majeur est *de facto* une actualité digne d'intérêt à l'étranger) ;
- les acteurs dominés sont rarement cités et rarement cités seuls (un acteur dominé n'est évoqué que si son actualité est liée à celle des acteurs dominants) ;
- les cooccurrences systématiquement présentes renvoient de préférence à des situations conflictuelles (priorité donnée aux mauvaises nouvelles).

Avant de commenter les principaux résultats, il convient tout d'abord de souligner la distribution parétienne des États dans les flux RSS : quels que soient le quotidien et la période étudiée, 20 % des États assurent plus de 80 % des items RSS de l'actualité internationale, l'État le plus cité étant dans plus de 90 % des cas les États-Unis. Les rares journaux faisant exception sont ceux où l'auto-citation est importante (cas du quotidien chinois par exemple où plus de 25 % des items mentionnent la Chine). D'importances variations régionales existent : les journaux d'un État donné tendent à privilégier les États voisins et culturellement proches.

Cette concentration de l'actualité sur une poignée d'États se retrouve lorsqu'on étudie les cooccurrences. Le tableau 1 donne les principales caractéristiques des graphes de cooccurrences étatiques pour deux périodes (avril à décembre 2014 et février à octobre 2015) et pour les sous-corpus nationaux (ensemble des journaux issus d'un même État) dont le nombre d'items reste stable. On notera tout d'abord l'écart toujours très important entre l'intensité médiane des cooccurrences et l'intensité maximale. Cette dernière ne concerne que très rarement la même dyade étatique, il y a donc bien des visions géopolitiques différentes du monde selon la nationalité des médias émetteurs. On note enfin que les caractéristiques structurales des graphes restent stables d'une période à l'autre : si la densité par construction varie peu, le degré de centralisation des graphes reste également très stable. Enfin, la hiérarchie des degrés (pondérés ou non) est très forte dans tous les cas.

Tableau 1 : Caractéristiques des graphes de cooccurrences par corpus national en 2014 (1) et en 2015 (2)

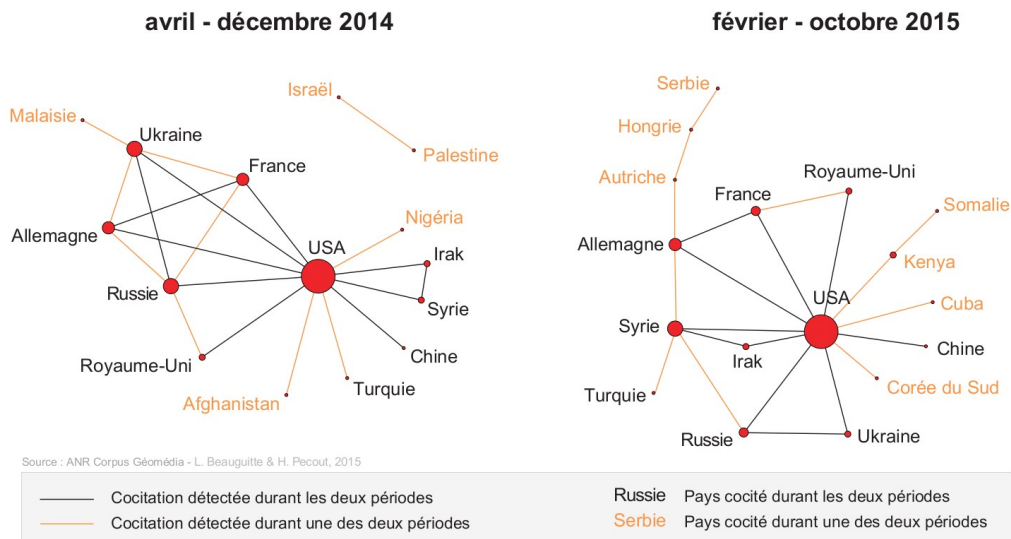
| Corpus | Nb sommets | Nb liens | Densité | Intensité médiane | Intensité max | Centralisation (degré) |
|--------------------|------------|----------|---------|-------------------|----------------|------------------------|
| <i>Australie 1</i> | 198 | 4110 | 0.21 | 2 | 881 (UKR-RUS) | 1.86 |
| <i>Australie 2</i> | 202 | 4740 | 0.23 | 2 | 644 (UKR-RUS) | 1.81 |
| Chine 1 | 182 | 788 | 0.05 | 2 | 701 (USA-CHN) | 0.59 |
| Chine 2 | 180 | 718 | 0.04 | 2 | 1122 (USA-CHN) | 0.53 |
| <i>Algérie 1</i> | 131 | 459 | 0.05 | 1 | 123 (PSE-ISR) | 0.38 |
| <i>Algérie 2</i> | 125 | 419 | 0.05 | 1 | 68 (PSE-ISR) | 0.32 |
| France 1 | 188 | 2122 | 0.12 | 5 | 411 (UKR-RUS) | 0.97 |
| France 2 | 194 | 2229 | 0.12 | 5 | 193 (PSE-ISR) | 0.99 |
| <i>Mexique 1</i> | 205 | 7594 | 0.36 | 5 | 433 (UKR-RUS) | 1.84 |
| <i>Mexique 2</i> | 207 | 8017 | 0.41 | 4.8 | 318 (USA-GBR) | 1.7 |
| N.-Zélande 1 | 204 | 2308 | 0.11 | 2 | 788 (UKR-RUS) | 0.65 |
| N.-Zélande 2 | 205 | 2657 | 0.13 | 2 | 632 (USA-CHN) | 0.62 |
| <i>Pakistan 1</i> | 166 | 709 | 0.05 | 1 | 169 (UKR-RUS) | 0.45 |
| <i>Pakistan 2</i> | 163 | 710 | 0.05 | 1 | 118 (SYR-LBN) | 0.46 |
| Zimbabwe 1 | 167 | 914 | 0.06 | 1 | 60 (PSE-ISR) | 0.7 |
| Zimbabwe 2 | 173 | 969 | 0.06 | 1 | 34 (PSE-ISR) | 0.64 |

Deux dyades reviennent de façon récurrente dans différents corpus nationaux : la dyade Russie-Ukraine, présente essentiellement dans la première partie ; et la dyade Palestine-Israël présente aussi souvent en 2014 qu'en 2015. Dans le premier cas, il s'agit clairement d'une dyade conjoncturelle liée à la partition de l'Ukraine ; dans le second cas, il s'agit d'une dyade géopolitique structurelle, le conflit israëlo-palestinien étant devenu ces 40 dernières années l'un des plus médiatisés et des plus symboliques.

Si l'on étudie les cooccurrences systématiquement présentes dans l'ensemble du corpus, le nombre de liens et de sommets chute de façon drastique comme le montre la figure 1. On remarque dans les deux cas une forte hiérarchie des sommets (la taille est fonction du degré non pondéré, prendre en compte la valuation des liens ne ferait qu'accentuer cette hiérarchie) et une structuration autour des États-Unis. Sans grande surprise, les liens

présents dans l'ensemble du corpus pour les deux périodes concernant soit les relations économiques et géopolitiques majeures (France – Allemagne ; USA – Chine) soit les conflits en cours faisant l'objet d'un traitement médiatique conséquent (Syrie - Irak ; Russie – Ukraine). À l'inverse, les liens partout présents mais pour une seule période concerne des événements à faible portée temporelle car le lien Ukraine – Malaisie liée au drame de la Malaysian Airlines en 2014, ou les liens Serbie – Hongrie – Autriche – Allemagne liés à un épisode migratoire européen en 2015.

Figure 2 : Cooccurrences détectées dans les items RSS



Conclusion

Le corpus étudié ici, issu d'une collecte automatique d'informations numériques, présente des caractéristiques que pourraient envier nombre d'historien.ne.s : forte couverture géographique, nombre élevé d'items, continuité temporelle. Ces différents aspects doivent être fortement nuancés. La couverture géographique est certes non négligeable mais une grande partie des journaux sélectionnés ne peut être considérée comme représentative des quotidiens des États concernés. Ainsi, le seul quotidien chinois présent est un quotidien publié en anglais et donc très clairement un outil de propagande destiné à un public international. Si le nombre d'items est élevé, il s'agit d'une information pauvre et exclusivement textuelle, or l'information est aussi affaire d'images. Les milliers d'articles sur les migrant.e.s cherchant à atteindre l'Europe ont eu moins d'impact semble-t-il que la seule photo d'un enfant syrien mort sur une plage turque. Enfin, la continuité temporelle est toute relative : la base géomédia capture en théorie plus de 200 flux issus de plus de 150 journaux. Or, lorsque nous cherchons à construire un corpus autorisant des comparaisons (à peu près) raisonnées, il nous reste 38 journaux en 2014 et 35 en 2015...

De façon plus générale, la donnée numérique, comme la donnée archivistique, ne parle pas d'elle-même et doit être travaillée. Sa récolte suppose déjà l'établissement de règles – plus ou moins explicitées – et donc des choix à justifier. Une réflexion sur les conditions de production de ces données est également nécessaire sous peine de faire des interprétations totalement erronées des résultats. L'attrait du nombre (le fort volume semble permettre des tests statistiques plus rigoureux) est bien souvent un leurre : il nous a fallu plusieurs mois par exemple pour réaliser que les différents quotidiens australiens (sauf un) du corpus émettaient *exactement* les mêmes flux RSS internationaux... Ajoutons pour terminer sur cet

aspect que les données numériques sont toujours sales, coûteuses à nettoyer et à modifier pour autoriser ensuite une quelconque exploration.

En ce qui concerne l'analyse des cooccurrences présentée rapidement ici, elle fournit des éléments intéressants pour lire et comprendre le paysage géopolitique contemporain. Au delà de certains aspects triviaux – les journalistes parlent d'abord des grandes puissances et des conflits meurtriers dans lesquels ces grandes puissances sont impliquées (la Syrie et l'Irak plutôt que la République démocratique du Congo donc) -, les variations selon les origines géographiques des quotidiens sont dignes d'intérêt. Que la cooccurrence Israël – Palestine soit en 2014 et en 2015 la plus fréquente dans les quotidiens du Zimbabwe et celui d'Algérie n'est pas neutre. La stabilité des structures de graphes sur les deux périodes testées est également intéressante même si elle gagnerait à être testée de façon plus approfondie.

Références

- I. Bastard, D. Cardon, G. Fouetillou, C. Prieur et S. Raux, 2013, Travail et travailleurs de la donnée, InternetActu.net, <http://www.internetactu.net/2013/12/13/travail-et-travailleurs-de-la-donnee/>
- L. Beauguitte, 2014, Les déterminants de l'actualité internationale : synthèse bibliographique, Carnet de recherche de l'ANR Corpus Geomedia, <http://geomedia.hypotheses.org/71>
- L. Beauguitte, M. Severo et H. Pecout, 2014, Do International News Reflect World Structure? A Network Approach, *1st European Conference on Social Networks*, Barcelone, <http://geomedia.hypotheses.org/176>
- L. Beauguitte, M. Severo et H. Pecout, 2016, Une analyse géomédiatique de l'actualité internationale : hiérarchies et effets de voisinage, *L'Espace géographique*, 45(2), p. 124-141
- M. Le Texier, M.H. Devès, C. Grasland et J.-B. De Chabalier, 2016, « La couverture médiatique des séismes à l'ère numérique », *L'Espace géographique*, 45(1), p. 5-24
- M. Newman, 2010), *Networks : An introduction*, Oxford, Oxford University Press.
- É. Ollion et J. Boelaert, 2015, « Au-delà des big data », *Sociologie* , 3(6), mis en ligne le 20 janvier 2016, <http://sociologie.revues.org/2613>
- C. Prieur, A. Stoica et Z. Smoreda, 2009, Extraction de réseaux égo centrés dans un (très grand) réseau social, *Bulletin de Méthodologie Sociologique*, 101, p. 5-27
- F. Queyroi, L. Beauguitte et H. Pecout, 2015, RSS Flows, World Structure & Community Detection, *Actes du Colloque européen de géographie théorique et quantitative*, Bari, <http://geomedia.hypotheses.org/345>
- S. Raux, 2014, *Dynamiques des réseaux sociaux en ligne : recommandations et interactions*, 2014, Thèse informatique, Paris Diderot
- A. Reynaud, 1981, *Société, espace et justice. Inégalités régionales et justice socio-spatiale*. Paris : Presses universitaires de France, coll. « Espace et liberté », 264 p.

