

Supplementary material: Methodology of elaboration of the database of international RSS flows

Instead of using the news published by newspapers, we use a collection of item news collected on the “World” or “International” RSS flows of daily newspapers. These solution was developed in the framework of the project Geomedia (2012-2016) funded by the French national research agency. As discussed by M. Severo & al. (Severo, Beauguitte, & Pecout, 2015) it is possible to analyze these specific source is a good candidate for the analysis of so-called ‘news values’ (Galtung & Ruge, 1965), international media agenda setting (M. McCombs, 2005; M. E. McCombs & Shaw, 1972) and more generally detection of international media event (Dayan & Katz, 1994). But we have to take into account two potential limitations concerning the use of newspaper in general and the choice of RSS items in particular.

Studies on media coverage of events are usually based on the analysis of broadcast traditional media such as press and television. In this article we focus only on the daily newspaper which is coherent with initial research of Galtung & Ruge but certainly not exempt from critics (Earl, Martin, McCarthy, & Soule, 2004) and subject to bias related to articles’ content (McCarthy, McPhail, & Smith, 1996). A second issue is the fact that RSS items are very limited pieces of news (a title of 50-100 characters and a summary of 150-300 characters) where it is possible to identify the countries at stake but not much more qualitative information on contender valence. These material can therefore support analysis of agenda setting of first type (“about what to think”) but not from the second type (“how to think”) according to a revised paradigm (M. McCombs, 2005)

Despite these limits, it is possible to underline three great advantages of RSS flows: ‘they are freely accessible, so they may be archived and tagged without limits; they have a quite homogenous structure, so they are easily comparable; they are generally provided as the news is ready and they can therefore be suitable for a real-time analysis’ (Severo et al., 2015). RSS flows has been analyzed from technical point of view by specialist of the web (Hammersley, 2005; Hammond, Hannay, & Lund, 2004). But the application to media studies

is limited to some recent publications concerning plurality of information (Marty, Rebillard, Smyrnaio, & Touboul, 2010) or circulation of international news (Beauguitte, Severo, & Pecout, 2016; Grasland, Lamarche-Perrin, Loveluck, & Pecout, 2016)

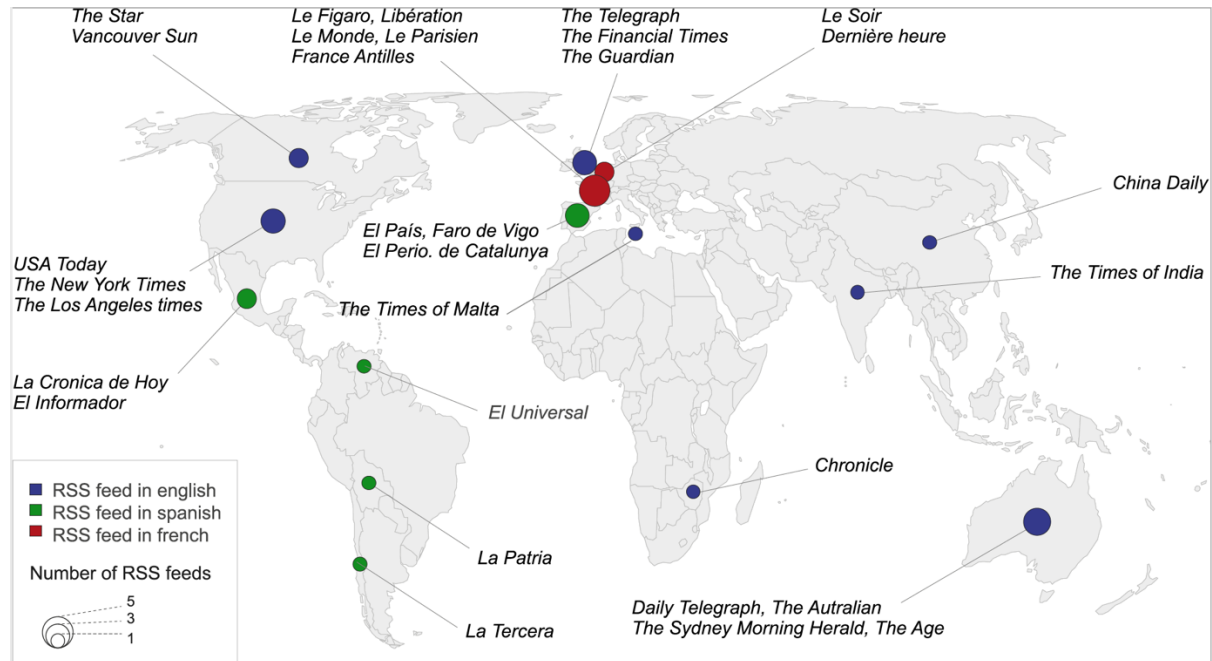
The aim of the data collection is to build a 3-D cube describing the circulation of international information between a set of media (m) toward a set of places (p) during several periods of time (t).

The media dimension: 31 daily newspapers in English, Spanish and French language

We use in this paper a sample of 321269 news published in 2015 by daily newspaper through the channel of rss flows explicitly entitled “international” or “world”. The selection of the newspapers has been realized according to several criteria described precisely elsewhere (Grasland et al., 2016) but it is worse to mention here two crucial criteria. Firstly, the international rss flows should have a universal coverage i.e. would be likely to speak about all countries of the world without exception. We has been therefore obliged to eliminate many newspaper that separated there international news between neighboring countries and remote countries. Secondly, we have chosen rss flows producing a regular number of foreign news per week over the year 2015, in order to support the hypothesis of a selection between competing news. Many newspapers with low production of news did not fulfilled these crucial criteria. Finally, we tried to obtain a regular equilibrium between languages with no more than half of English speaking media. We obtained a final list of 31 newspaper covering foreign news during the 52 weeks of the period of observation which starts the Monday 5th January 2015 and ends the Sunday 3rd January 2016. We have therefore the possibility to explore 1643 experimental situations (31 medias x 52 weeks) where we can analyze the international agenda of a media and measure how much news was allocated to each foreign country.

These sample of media does not intend to be representative of the total volume of news circulating during the period of observations. It has been based on the choice to maximize the diversity of geographical location, under the constraint of availability of newspaper filling the conditions of comparability (*Figure 1*).

Figure 1 : Localization and language of the corpus of 31 daily newspapers



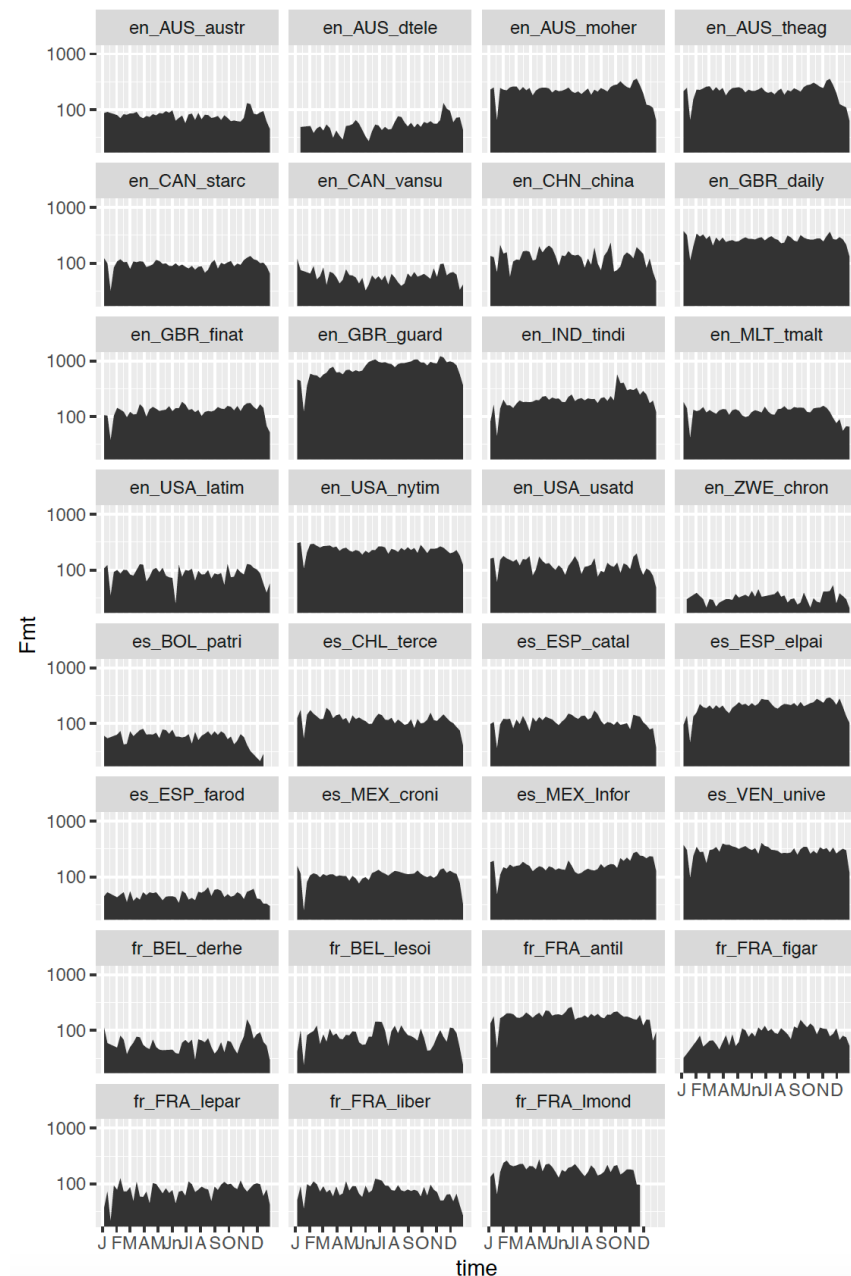
The international dimension: 192 guest countries or autonomous territories

We have extracted from each of the news a list of countries or autonomous territories mentioned in the title or the summary of the RSS item. To do this, we used a dictionary of key-words making possible the identification of 192 countries of the world on the basis of their country names, capital city, political representative (president, king, prime minister, ...) and other clues like gentiles (names of inhabitants), adjectives, major regions or cities. This dictionary currently available for English, French and Spanish language use more than 4000 keywords and insure a pretty good recognition of national entities. Manual check by human operators indicates an accuracy of 90-95% (5-10% of false negative and 2-5% of false positive). For each of these 192 territorial units, we have collected from World bank data describing their size for four different criteria in 2015 (Area, Population, GDP, military expense). We have also extract from CEPII Distance Database for criteria of proximity between places (regional belonging, common language, historical relations, geographical distance). The missing values has been estimated by interpolation, extrapolation and retropolation.

The time dimension: 52 weeks of observation

We have collected the news during 53 weeks but used only 52 weeks in the analysis because we have created a time-lag parameter indicating the news sent by each media for each country during the previous week. We have verified that the collection of news was regular, with the exception of the week starting the 19th January where a technical problem produced a partial break of the harvest tools (*Figure 2*)

Figure 2 : Check of the stationarity of news flow at week level



Concerning the relative decline of news at the end of the year, it is not related to technical problem but to a reduction of activity in the newsroom during the Christmas and New Year period

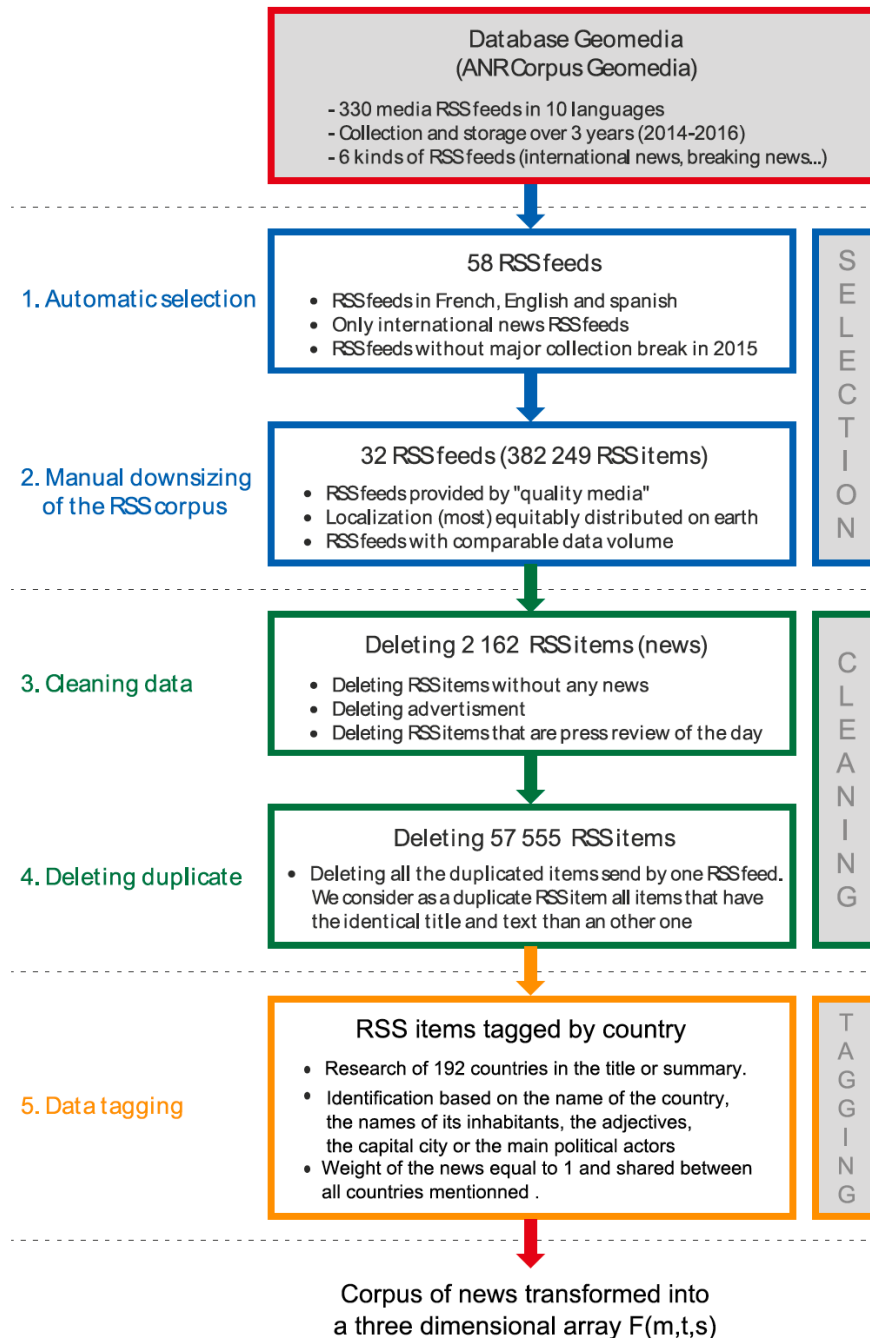
Aggregation of news in 3-D cube and weighting options

We build the cube $F(m,t,p)$ through an aggregation procedure where the total number of news is kept constant. When news are related to only one country, the procedure is straightforward. But when several countries are present in the same news (e.g. "Putin discuss with Obama about Syrian Crisis") we divide the weight of the news between each of the countries mentioned at less one time (e.g. $1/3$ Russia, $1/3$ USA, $1/3$ Syria). We don't take into account the number of time a country is mentioned in the same news. The news where no countries are mentioned (or recognized by the dictionary) are excluded from the analysis. They represent generally 5 to 10% of total news and are related to transnational topics or actors (e.g. Google, UN Assembly, European Union). From the first cube, we have derived a weighted cube $F_w(m,t,p)$, where the number of news is inversely weighted by media and time in order to insure an equal number of news sent by each media during each week.

This weighting scheme avoid the under-representation of "small" newspaper and give the same weight in the analysis to the *Guardian* (55573 news) and the *Chronicle of Zimbabwe* (2095 news). It introduce also a correction of the time sample with equal weight for all weeks. The computation of the number of news received by each country during the year 2015 demonstrate that the hierarchy of countries is not strongly modified by the weighting scheme except for specific cases like Australia that represented 2.0% of news with raw data (9th rank) but only 0.9% of news (32th rank) after weighting procedure. One have to precise immediately that, weighted or not, the sample of news is not representative of world news and is clearly focused on Western media. Its main interest is to offer a relative diversity in terms of language, geographical location and level of quality of newspaper.

The different steps of data elaboration can be finally summarized in Figure 3.

Figure 3: The steps of data elaboration of the information cube



References

- Beauguitte, L., Severo, M., & Pecout, H. (2016). A geomedia analysis of international news: Hierarchies and neighbourhood effects. *Espace Géographique (English Edition)*, 45(2), 1–18.
- Dayan, D., & Katz, E. (1994). *Media events*. harvard university press.
- Earl, J., Martin, A., McCarthy, J. D., & Soule, S. A. (2004). The use of newspaper data in the study of collective action. *Annu. Rev. Sociol.*, 30, 65–80.
- Galtung, J., & Ruge, M. H. (1965). The structure of foreign news: The presentation of the Congo, Cuba and Cyprus crises in four Norwegian newspapers. *Journal of Peace Research*, 2(1), 64–90.
- Grasland, C., Lamarche-Perrin, R., Loveluck, B., & Pecout, H. (2016). International Agenda-setting, the Media and Geography: A multi-dimensional Analysis of News Flows. *L'Espace Géographique*, 45(1), 25–43.
- Hammersley, B. (2005). *Developing Feeds with RSS and Atom: Developers Guide to Syndicating News & Blogs*. O'Reilly Media, Inc.
- Hammond, T., Hannay, T., & Lund, B. (2004). The role of RSS in science publishing. *D-Lib Magazine*, 10(12), 1082–9873.
- Marty, E., Rebillard, F., Smyrniaios, N., & 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*, (2), 107–126.
- McCarthy, J. D., McPhail, C., & Smith, J. (1996). Images of protest: Dimensions of selection bias in media coverage of Washington demonstrations, 1982 and 1991. *American Sociological Review*, 478–499.
- McCombs, M. (2005). A look at agenda-setting: Past, present and future. *Journalism Studies*, 6(4), 543–557.
- McCombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. *Public Opinion Quarterly*, 36(2), 176–187.
- Severo, M., Beauguitte, L., & Pecout, H. (2015). Archiving news on the Web through RSS flows. A new tool for studying international events. In *RESAW 2015. Web Archives as Scholarly Sources: Issues, Practices and Perspectives*.