Les représentations numériques des rivières urbaines : la construction de relations ville-rivière renouvelées ?

Digital representations of urban rivers: a construction of renewed city-river relations?

Liolia Bajemon, Lise Vaudor, Hervé Piégay

UMR 5600 EVS, ENS de Lyon, liolia.bajemon@ens-lyon.fr

RÉSUMÉ

Les villes et leurs cours d'eau ont aujourd'hui une présence et une image numériques. Cette empreinte numérique véhicule des discours sur l'environnement qui se diffusent largement et influencent potentiellement les interactions socio-environnementales. Nous proposons ici de nous intéresser aux représentations numériques des rivières urbaines à l'échelle du globe, car ces milieux apparaissent comme stratégiques dans le cadre d'une urbanisation croissante. En se basant sur les discours du web, nous cherchons à identifier et caractériser les relations villes-rivières qui sont exprimées. Un corpus a été construit grâce à des méthodes de *web scraping*, puis analysé par des approches qualitatives et textométriques. Les résultats montrent les multiples visages, enjeux et imaginaires entourant les rivières urbaines, qui apparaissent à la fois comme dangers, espaces récréatifs ou espaces à mettre en valeur. La mobilisation des données du web permet aussi d'interroger la production des discours et de mettre en évidence les relations de pouvoir qui sont en jeu.

ABSTRACT

Today, cities and their rivers have a digital presence and image. This digital print conveys discourses about the environment, which are diffused largely and potentially influence socio-environmental interactions. Here we propose to study digital representations of urban rivers on a global scale, because these areas appear as strategic in the context of a growing urbanisation. Drawing on web discourses, we seek to identify how city-river relationship are defined. A corpus was built thanks to web scraping and was analysed through qualitative and textometric approaches. Results show the multiple faces, stakes and imaginaries surrounding urban rivers, which appear as dangers, recreational spaces or areas to be developed. The mobilisation of web data also allows to question the discourses production and highlights the power relations at play.

MOTS CLÉS

Discours, numérique, représentations, rivières urbaines, textométrie, web scraping.

KEYWORDS

Digital, discourse, representations, textometry, urban rivers, web scraping.

1   URBAN RIVERS ON THE INTERNET

In the face of a growing and global urbanisation of floodplains and their associated rivers (Wu et al., 2023), these spaces appear as strategic and vulnerable. In particular, urban rivers are at the interface between societies and their environment. More and more space is being given to rivers, materialized for example through the development of riverfronts and waterfronts in many cities. In addition, cities and rivers are no longer only material, but also digital. Digital tools and technologies have an impact on social, economic, politic interactions and thus influence human relations to nature, *via* digital representations (Graham, 2013). Indeed, representations play a major role in behaviours and practices. Discourses are a way to access them – in particular because they participate in the production of space (Gumuchian, 1991). However, digital representations of nature are still an under-exploited subject, although it might help us comprehend how the digital era impacts our relations to the environment.

In this contribution, we propose to focus on digital representations regarding urban rivers around the world. We apply the framework of urban political ecology to examine web discourses as products of social relations. Using web scraping methods, we build a digital corpus on a selection of cities on a global scale. Because urbanisation is a global and systemic process, it requires global urban studies (Robinson, 2011). As such, this scale seems relevant and is an opportunity for examining local manifestations of a global phenomenon. Through this research, we aim at identifying how rivers are represented and debated on the web. It is also a question of knowing how these discourses are produced.

2   METHOD FOR CONSTRUCTING A DIGITAL CORPUS FROM WEB SCRAPING

First of all, 302 cities of more than 300 000 inhabitants (demographic criteria of the United Nations to define an agglomeration) and intersected by a river over a width 30 meters (detection limit by satellite images) have been selected on a global scale. This selection was based on hydrologic, climatic and socioeconomic criteria, in order to represent a diversity of situations and stakes. Starting from the assumption that Google dominates the information landscape and that it is the most used search engine in the world, the search engine pages results (SERP) from this search engine were collected for each city-river combination. A SERP corresponds to the page which is generated in response to keywords (the query) sent to a search engine (here, Google) and which contains links towards other web pages. Up to 99 pages have been collected for each query – for some queries, there were not as many pages available. The queries have been sent in English and assuming no particular location for the user (*e.g.* "rhône river AND lyon") on the one hand, and in the local language(s) and assuming the user sent the request from the concerned country (*e.g.* "fleuve rhône AND lyon" from France) on the other. Several information on each of the web pages have been gathered in this way, such as their URL address, their title and their snippet (which corresponds to a small text bringing forwards the content of the page based on the sent query). Finally, using the identified URLs, we implemented web scraping methods in order to collect the textual content from each web pages. For those which could not be scraped automatically, the text was replaced by the snippet – this is the case for some websites which prevent web scraping. This method enabled us to get a relatively consistent textual corpus (approximately 76 000 pages) about urban rivers on a global scale.

Once the raw corpus is produced, the text must be prepared for the analysis. The particularity of this digital corpus is its heterogeneity: it consists of an assembling of diverse web pages, which do not share the same structure, the same writing style, *etc.* After translating everything into English, the textual content is thus cleaned: removal of stop words (words which are so common that they are not significant), dates, special characters (*e.g.* # and @ which are characteristic of social networks), URLs, *etc.* Lastly, the text is lemmatized, *i.e.* to each word is assigned its base form; for example, "is" becomes "be" and "rivers" becomes "river". The corpus is finally analysed thanks to mixed methods, using simultaneously a qualitative approach (citations extracting) and a quantitative approach (textometry). In particular, this contribution shows results from a descending hierarchical classification (DHC), based on the Reinert method (Reinert, 1983). Through a classification of text segments, lexical universes appear and highlight the main topics covered in our corpus.

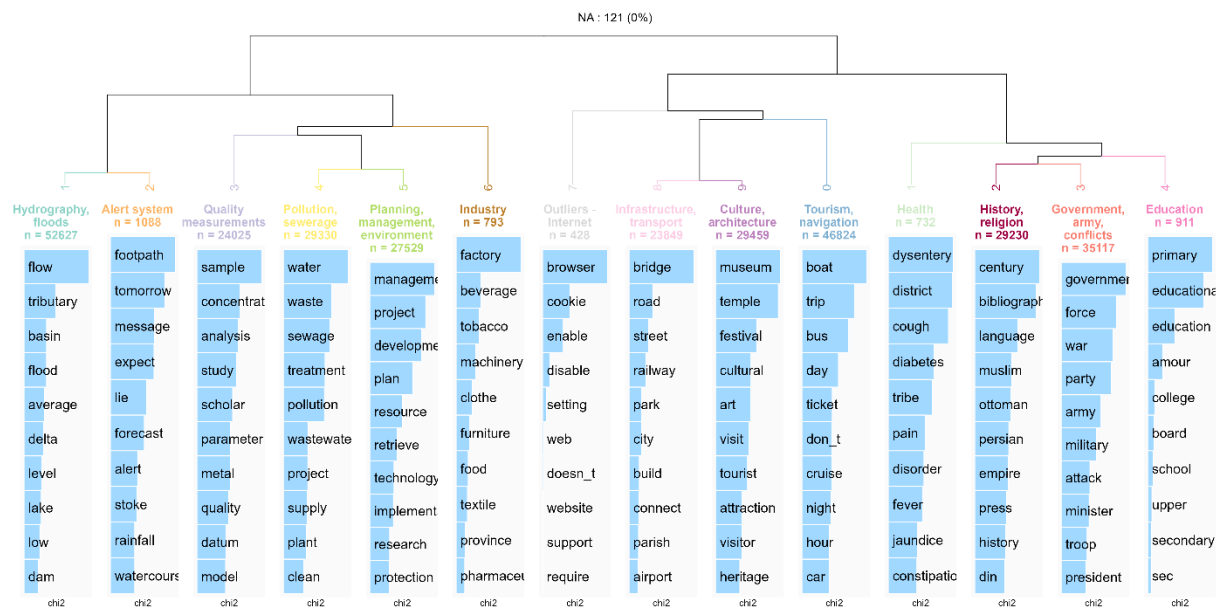# 3 RESULTS AND DISCUSSION: A DIGITAL GEOGRAPHY OF RIVER REPRESENTATIONS

Figure 1: Results of a descending hierarchical classification of web pages on urban rivers

The results show that urban rivers are complex and diverse environments around the globe, but also in a given city. Because rivers have multiple faces and stakes, this research shows the interest of having a holistic approach when studying them. The lexical universes go from floods, pollution, wastewater to leisure, tourism, and history, as well as to conflicts, industry and land planning (figure 1). This shows that a river can be seen as (i) a danger or a threat, (ii) a tourism capital and a heritage, (iii) a strategic area to protect or to develop, or (iv) a resource – but one aspect does not exclude another. Sometimes the river is not the main subject of the discourse, but rather appears as a landmark: it serves as a reference point in the city to describe or locate other elements.

The Rio Paraná is an example of a multifaceted space. In Posadas (Argentina), online discourses speak both of floods and tourism. The river is a space of recreation and leisure, offering multiple aquatic activities, a waterfront, cruises, *etc.* However, several web pages, including press articles, report floods in the city and its surroundings. Moreover, the Yacyretá-Apipé dam is also a major topic. This hydroelectric plant was built upstream of the city of Posadas on the Rio Paraná. It is praised for its economic impact – because it provides electricity to the region – while it also causes floods elsewhere: "[…] *Ayolas, a town in Paraguay, was flooded and dozens of its inhabitants had to leave their homes because of the rising of the Paraná River but also because the Yacyretá dam opened its floodgates without warning*" (translation, *Diario del Norte*, 31/10/2023). Web discourses thus show the river's dual identity: it is a space of economic, urban and environmental stakes, as much as a recreative and touristic environment.

In other cities, rivers have become a showcase of planning, identity and heritage. They are considered as recreational elements, offering leisure spaces such as parks, gardens and banks, as well as activities such as cruises or aquatic sports. For example, in Ahmedabad (India), the main topic on the web is that of the Sabarmati riverfront, a development project which has been under construction for about twenty years, and which has a strong presence on the web, in particular through multiple social networks. This urban river is reclaimed as a way to enhance the city. However, a finer analysis also shows that this project generated a strong debate in terms of socio-environmental impacts, because it caused population displacement and a transformation of the hydrological regime of the river: "[…] *the Sabarmati riverfront development is symbolic of how a natural river, not perennial but dependent on the monsoon, has been turned into an infrastructure project with emphasis on commercial and recreational aspects rather than on ecological conservation.*" (*Questions of cities*, 4/11/2022). The highlight of the riverfront as a positive transformation of the river on social networks benefits investors and project actors before anything else, setting aside social inequalities. In the framework of urban political ecology, this shows the impact of social, political and economic relations on the production of discourses.

If this method allows for outlining topics and by extent the main stakes of our corpus, it also allows for underlining which stakes are not visible. For example, climate change, drought or aquatic biodiversity do not stand out among the different topics, although they would seem to be major stakes when talking about rivers.

Aquatic biodiversity can be found in tourism and leisure subjects through the mention of fishing, while conservation stakes and endangered species are mentioned in some cities, but it does not stand out among the topics. This would mean that compared to floods, leisure or tourism, these subjects receive less attention and thus produce less discourses (although a qualitative analysis shows that they are mentioned). This invites us to reflect upon the results that are shown by search engines: a vicious cycle might be created by the web construct and referral. Receiving less attention and thus less clicks, results and discourses about invisible topics fall out behind and receive even less attention – because people usually click on the first pages of SERP. This might contribute to silence these subjects even more. This also questions the responsibility of search engines in creating collective images and representations.

This contribution shows the relevance of mobilising a web corpus for a geographic study, which should also include a critical analysis of data distribution and quality. Indeed, if digital discourses are available everywhere, they are not necessarily produced locally. Their distribution echoes the digital divide: most data is produced by and on countries of the North. Through this global approach, spatial patterns can be found in thematic results as well as in data availability. It also allows to identify multiple stakes at the same time for a given city, without focusing on a single subject like it is usually the case. Thus, the results (1) show the various aspects of urban rivers, (2) highlight stakes for a given city and river and (3) question the impact of web discourses on our collective representations and imaginaries.

BIBLIOGRAPHY

Graham, M. (2013). The Virtual Dimension. In M. Acuto & W. Steele (Éds.), *Global City Challenges* (p. 117-139). Palgrave Macmillan UK. https://doi.org/10.1057/9781137286871_8

Gumuchian, H. (1991). *Représentations et aménagement du territoire*. Anthropos : Diffusion, Economica.

Reinert, M. (1983). Une méthode de classification descendante hiérarchique : Application à l'analyse lexicale par contexte. *Cahiers de l'Analyse de Données*, *8*(2), 187-198.

Robinson, J. (2011). Cities in a World of Cities: The Comparative Gesture: Cities in a world of cities compared. *International Journal of Urban and Regional Research*, *35*(1), 1-23. https://doi.org/10.1111/j.1468-2427.2010.00982.x

Wu, Q., Ke, L., Wang, J., Pavelsky, T. M., Allen, G. H., Sheng, Y., Duan, X., Zhu, Y., Wu, J., Wang, L., Liu, K., Chen, T., Zhang, W., Fan, C., Yong, B., & Song, C. (2023). Satellites reveal hotspots of global river extent change. *Nature Communications*, *14*(1), 1587. https://doi.org/10.1038/s41467-023-37061-3

ACKNOWLEDGEMENTS AND FUNDING