
Les données des catalogues de bibliothèques pour créer, sélectionner et évaluer des collections de textes littéraires

Nanette Rißler-Pipka^{*1}, José Calvo Tello^{*2}, and Andreas Lüscho^{*2}

¹Goettingen University – Allemagne

²Göttingen University – Allemagne

Résumé

Face au grand nombre de textes littéraires non étudiés, jamais discutés, survolés puis oubliés par l'histoire littéraire, on a toujours espéré que la numérisation de livres en grande quantité serait une solution qui nous rapproche de l'idéal d'un corpus équilibré. Cependant, les raisons et les mécanismes de sélection de textes littéraires évalués de haute qualité sont assez ostensibles et présents dans la discussion critique : dans de nombreux cas, ces textes n'appartiennent pas au canon littéraire parce qu'ils sont écrits dans des langues minoritaires ou à des endroits périphériques par des personnes marginalisées. Aujourd'hui, nous savons que le système d'exclusion et d'élitisme est tout aussi efficace lorsqu'il s'agit de textes littéraires numériques (Calvo Tello 2016, Robinson 2019).

Cependant, les catalogues de bibliothèque et les bibliographies gardent toutes les informations sur les textes imprimés, ou même sur des manuscrits non publiés. Une révision de l'histoire littéraire est-elle possible à l'aide de catalogues de bibliothèques et de métadonnées bibliographiques ?

Le processus de sélection lors de la constitution de collections de textes est normalement structuré par des critères transparents (Schöch 2017, Gius et al. 2019). Les critères sont plus ou moins équivalents pour les textes imprimés ou numériques. Mais si les chercheurs souhaitent consulter une liste ou un catalogue de tous les romans écrits en français entre 1820 et 1850 pour préparer une étude sur Balzac, ils n'ont pratiquement aucune chance de créer une telle liste. La méthode courante utilisée est de se référer à une liste d'auteurs qui ont influencé Balzac ou qui ont été influencés par ses œuvres (par exemple George Sand); on y ajoutera l'inspiration à partir des études d'autres chercheurs. On finit par comparer deux ou trois romans, plus des données contextuelles comme des incidents biographiques. Ne serait-il pas préférable et nécessaire de dresser une liste des romans publiés pour la période étudiée et pour la langue ou la zone culturelle concernée ?

Afin de créer une telle liste, nous proposons d'utiliser les catalogues des bibliothèques. Il faut un catalogue qui permet aux chercheurs de trouver les données bibliographiques qui correspondent aux questions de recherche et aux catégories littéraires (genre, époque, œuvre, auteur). C'est valable aussi pour les collections de textes imprimés que numériques. Une liste des critères et des catégories garantira la transparence de la sélection des textes d'une

*Intervenant

collection. Les défis concernant certains aspects de la sélection des données et les données FAIR (Wilkinson et al. 2016) peuvent être testés pour une période, un genre et un domaine culturel spécifiques (par exemple, les œuvres littéraires en langues romanes publiées à une période spécifique), appliqués aux données du K10plus (<https://wiki.k10plus.de/> catalogue des bibliothèques allemandes).

Facile à (re)trouver : les catalogues des bibliothèques se concentrent sur des éditions uniques d'une œuvre ou même sur des copies individuelles ; cependant, les corpus littéraires numériques enregistrent principalement des métadonnées sur l'œuvre littéraire elle-même.

Accessible : même si le catalogue de la bibliothèque peut fournir des informations sur la disponibilité et l'accessibilité d'un texte imprimé, ce n'est pas nécessairement le cas pour la copie numérique.

Interopérable : pour dresser une liste complète de toutes les œuvres publiées au cours d'une période, d'un genre et dans une langue donnée, il faut disposer de plusieurs sources de données. Bien que les catalogues de bibliothèques utilisent des normes de métadonnées, la mise en correspondance et la fusion sont des défis communs.

Réutilisable : les catalogues ne contiennent pas toujours les catégories utilisées comme critères pour la constitution de collections de textes (par exemple, les informations sur le genre) ou nécessaires pour l'application de méthodes numériques (par exemple, l'indication du format et de la qualité des données)

Si nous savons comment utiliser les données des catalogues de bibliothèque pour la recherche on réussit aussi à offrir un catalogue avec une fonction de chercher-trouver flexible en offrant aussi une vue globale à propos la relation quantitative entre textes disponible en version numérique ou imprimé.

Références :

Calvo Tello, José. 2016. "Estado de la digitalización de la Edad de Plata: un análisis cuantitativo." *Revista de Humanidades Digitales*, no. 1.

Gius, Evelyn, Katharina Krüger, and Carla Sökefeld. 2019. 'Korpuserstellung Als Literaturwissenschaftliche Aufgabe'. In *DHd 2019 Digital Humanities: Multimedial & Multimodal. Konferenzabstracts*. Frankfurt am Main. <https://doi.org/10.5281/zenodo.2600812>.

Gantert, Klaus. 2016. *Bibliothekarisches Grundwissen*. De Gruyter Saur. <https://www.degruyter.com/view/title/3>

Robinson, Peter. 2019. 'Gender, Feminism, Textual Scholarship, and Digital Humanities'. In *Intersectionality in Digital Humanities*, edited by Barbara Bordalejo and Roopika Risam, 89–108. Collection Development, Cultural Heritage, and Digital Humanities. Leeds: Arc Humanities Press.

Schöch, Christof. 2017. 'Aufbau von Datensammlungen'. In *Digital Humanities*, edited by Fotis Jannidis, Hubertus Kohle, and Malte Rehbein, 223–33. Stuttgart: J.B. Metzler. https://doi.org/10.1007/978-3-476-05446-3_16.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. 'The FAIR Guiding Principles for Scientific Data Management and Stewardship'. *Scientific Data* 3 (March): 160018. <https://doi.org/10.1038/sdata.2016.18>.

Mots-Clés: Bibliothèques, Littérature