
Modéliser le feuilleton. Créer et analyser une nouvelle archive patrimoniale avec les méthodes computationnelles.

Pierre-Carl Langlais*^{1,2}

¹GRIPIC (Groupe de recherches interdisciplinaires sur les processus d'information et de communication)
– Université Paris IV - Paris Sorbonne : EA1498 – 77 rue de Villiers, 92200 Neuilly-sur-Seine, France

²Représenter, Inventer la Réalité du Romantisme à l'Aube du XXIe siècle – Université Paul-Valéry -
Montpellier 3 : EA4209, Université Paul-Valéry - Montpellier 3 : EA4209 – France

Résumé

Cette proposition de communication interroge " l'effet des méthodes computationnelles sur la recherche en SHS " et plus particulièrement en histoire culturelle. Elle présente un nouveau cadre méthodologique d'étude du roman-feuilleton de 1830 à 1950.

Le roman-feuilleton est une production emblématique des premières industries culturelles et médiatiques (Dumasy-Queffelec, Queffelec, et Queffelec 1989; Delafield 2016; Stein et Wiele 2019). Des décennies avant l'apparition des serials cinématographiques ou télévisés, le feuilleton introduit des dispositifs de narration sériels suivis par des centaines de milliers de lecteurs et de lectrices (Letourneux 2017). Les études du feuilletons ont été profondément renouvelé par le " tournant culturel " des études de presse ancienne depuis 2000 (M.-E. Thérenty et Vaillant 2001; M. Thérenty 2007; Kalifa et al. 2011; M.-È. Thérenty, Langlais, et Schuh 2020). Longtemps occultées au profit des structures économiques, sociales et politiques, les formes et pratiques d'écritures deviennent un terrain de recherche privilégié. Malgré ces efforts de recherche soutenus, l'étude du roman-feuilleton est pénalisée par la masse énorme des corpus : au tournant des XIXe et XXe siècle, la plupart des quotidiens nationaux et internationaux publient un voire deux feuilletons chaque jour. Les outils de consultations existants de la presse numérisée ne suffisent pas à saisir ce phénomène culturel global au-delà de quelques grands romans patrimonialisés.

De nouveaux projets français et européens (Europeana Newspaper, Newseye, Numapresse, Impresso) élaborent aujourd'hui des infrastructures de recherche " augmentée " de la presse ancienne dans la continuité de plusieurs travaux précurseurs dans le monde anglo-saxon (Cordell 2015; Mussell 2012; Underwood 2019). Le projet Numapresse s'est spécialisé dans la détection et l'analyse des genres journalistiques, avec le développement de modèles de classification " générationnels " (1840-1860, 1860-1880...). Ces modèles ont été appliqués à l'ensemble des archives de presse quotidienne nationale numérisées par Gallica. Les résultats de la classification permettent de construire rapidement des sous-corpus tels que l'ensemble des pages de cinéma de l'entre-deux-guerre. Ils font également émerger de nombreuses tendances pertinentes pour l'histoire culturelle longue de la presse comme, la spécialisation thématique de certains quotidiens, l'hebdomadairisation des formes journalistiques au début du XXe siècle ou le développement précoce de certains genres, avant leur transformation en

*Intervenant

rubrique (Langlais 2021).

Le roman-feuilleton ne faisait pas initialement partie des trois genres journalistiques privilégiés par Numapresse (reportage, cinéma et sport) mais s'est finalement imposé comme un terrain privilégié pour des collaborations inédites avec les bibliothèques. Numapresse est depuis 2020, partenaire du futur laboratoire en humanités numériques de la BNF : la détection des romans-feuilletons est l'un des principaux enjeux de cette collaboration avec l'accès aux archives sous droits.

Ce projet mobilise plusieurs techniques en intelligence artificielle appliquée aux archives patrimoniales numérisées :

Reconnaissance spatiale : L'objet éditorial du feuilleton est reconnu à partir d'une segmentation automatisée de la page du journal. Pour l'instant, notre approche privilégie un algorithme très rapide de détection de zones géométriques (la marge d'espace entre le feuilleton et le haut du journal) (voir l'illustration n°1). Nous explorons également l'utilisation de modèles d'apprentissage profond (deep learning) entraînés à partir de l'annotation des zones du journal correspondant au feuilleton.

Classification des textes : Les feuilletons ne publient cependant pas que des romans en série mais aussi de nombreux autres textes : critiques théâtrales, chroniques scientifiques, billets d'humeur L'identification du roman-feuilleton repose sur les modèles de détection des genres journalistiques développé par Numapresse (voir l'illustration n°2).

Identification du paratexte : Dans un dernier temps, le projet procède à l'identification des titres, des signatures et d'autres éléments paratextuels à partir d'une analyse croisée du vocabulaire et de la forme visuelle du texte (soit à partir d'une analyse géométrique ou d'une nouvelle classification par deep learning). L'identification des titres est notamment déterminante pour reconstituer la série des feuilletons dans leur continuité.

Cette approche croisée a été pour l'instant appliquée avec succès à un titre de presse parisien, Le Pays. D'ici la conférence elle sera probablement généralisée à d'autres quotidiens.

En revenant sur cette collaboration, nous développerons une réflexion méthodologique sur les transformations de l'archive de presse induite par le déploiement des outils d'analyse computationnelle et/ou par la création de corpus de recherche (Bode 2018). La détection du roman-feuilleton fait en effet émerger des éditions voire des œuvres qui n'existaient pas. Plusieurs projets de recensions ont déjà mis en évidence qu'une large partie des textes publiés en feuilletons (en particulier les plus courts) ne font pas l'objet d'une publication en monographie. Les méthodes computationnelles produisent des œuvres qui n'avaient jamais été cataloguées. Dans le cadre de Numapresse nous avons commencé à réfléchir à de nouvelles interfaces de lecture qui, tout en brisant la structure d'origine de l'archive de presse, restitue la continuité textuelle des feuilletons.

Dans la dernière partie de notre présentation, nous présenterons plusieurs nouveaux champs de recherche ouverts par l'accès sans précédent à un grand corpus de feuilletons :

Recontextualisation de l'émergence du roman-feuilleton : la détection des reprises d'un corpus de 16 000 romans issus de Gallica montre que la presse commence à publier de plus en plus fréquemment des textes littéraires et des bonnes feuilles avant la date de naissance officielle du roman-feuilleton en 1836 avec la publication de la Vieille fille de Balzac (illustration n°3). Nous avons ainsi pu identifier un court feuilleton d'Alphonse Karr publié en deux livraisons dans le Figaro en 1835.

Organisation et structuration du feuilleton dans l'écosystème journalistique : la classification automatisée de l'ensemble du journal permet de situer le feuilleton par rapport à un espace éditorial très normé (illustration n°4). Nous explorerons notamment l'hypothèse que les proximités spatiales (par exemple entre le feuilleton romanesque et

le feuilleton boursier qui prend sous sa suite en dernière page) se traduisent par des interactions intertextuelles.

Feuilletonisation/défeuilletonisation : Nous avons commencé à procéder à une comparaison systématique entre les versions éditées des feuilletons et leur version publiées dans la presse en détournant des outils de détection automatisés des reprises de textes. Lorsque la publication intervient elle introduit des changements substantiels : les comparaisons semi-automatiques du projet Numapresse mettent en évidence des agencements significatifs avec des coupes, des rajouts, des changements de titres, ce qui suggère que les attentes de lecture d'une monographie ne sont pas les mêmes que celles d'une publication en feuilleton.

Recyclage des feuilletons : l'analyse préliminaire des reprises de feuilleton entre les journaux met en évidence une circulation fréquente des feuilletons à succès d'un journal à l'autre. En particulier, plusieurs quotidiens des années 1880 et 1890 reproduisent de nombreux feuilletons de la période classique des années 1840 et 1850. Dans une étude en cours d'écriture nous décrivons ce phénomène de patrimonialisation par analogie avec la daytime TV américaine des années 1980 et 1990 qui remet en en circulation des séries alors vieillissantes.

Évolution des thèmes et des formes d'écriture du feuilleton. Nous souhaitons mobiliser les méthodes appliquées par les cultural analytics anglo-saxonnes à de grands corpus de romans : détection de genres romanesques, de personnages, de lieux et de sujets. Cet axe encore préliminaire s'inscrit dans la perspective d'un enrichissement des nouvelles archives de feuilleton.

Bibliographie

Bode, Katherine. 2018. *A World of Fiction: Digital Collections and the Future of Literary History*. University of Michigan Press.

Cordell, Ryan. 2015. "Reprinting, Circulation, and the Network Author in Antebellum Newspapers". *American Literary History* 27 (3): 417-45. <https://doi.org/10.1093/alh/ajv028>.

Delafield, Catherine. 2016. *Serialization and the Novel in Mid-Victorian Magazines*. Routledge. <https://doi.org/10.4324/9781315608440>.

Dumasy-Queffelec, Lise, Lise Queffelec, et Lise Queffelec. 1989. *Le roman-feuilleton français au XIXe siècle*. Presses universitaires de France.

Kalifa, Dominique, Philippe Régnier, Marie-Ève Thérenty, et Alain Vaillant, éd. 2011. *La civilisation du journal: Histoire culturelle et littéraire de la presse au XIXe siècle*. Paris: Nouveau Monde Editions.

Langlais, Pierre-Carl. 2021. "Classified News, Redefining the history of newspaper genre with supervised models". In *Digital Newspaper: a new Eldorado for the historians*. De Gruyter.

Letourneux, Mathieu. 2017. *Fictions à la chaîne - Littératures sérielles et culture médiatique*. Paris: Le Seuil.

Mussell, James. 2012. *The Nineteenth-Century Press in the Digital Age*. Springer.

Stein, Daniel, et Lisanna Wiele, éd. 2019. *Nineteenth-Century Serial Narrative in Transnational Perspective, 1830s–1860s: Popular Culture–Serial Culture*. Palgrave Studies in Nineteenth-Century Writing and Culture. Palgrave Macmillan. <https://doi.org/10.1007/978-3-030-15895-8>.

Thérenty, Marie-eve. 2007. *La Littérature au quotidien. Poétiques journalistiques au XIXe siècle*. Paris: Le Seuil.

Thérenty, Marie-Ève, Pierre-Carl Langlais, et Julien Schuh. 2020. *Fake News et viralité avant Internet. Les lapins du Père-Lachaise et autres légendes médiatiques*. Paris: Éditions du CNRS.

Thérenty, Marie-Eve, et Alain Vaillant, éd. 2001. *1836: l'an 1 de l'ère médiatique. Etude littéraire et historique du journal La Presse d'Emile de Girardin*. Paris: Nouveau Monde Editions.

Underwood, Ted. 2019. *Distant Horizons: Digital Evidence and Literary Change*. University of Chicago Press.