



HAL
open science

Analyse automatique pour une étude du genre : quels jugements des écrivaines au XIXe siècle ?

Marine Riguet, Motasem Alrahabi

► To cite this version:

Marine Riguet, Motasem Alrahabi. Analyse automatique pour une étude du genre : quels jugements des écrivaines au XIXe siècle?. Digital Humanities Conference, Jul 2020, Ottawa virtual), Canada. hal-03199809

HAL Id: hal-03199809

<https://hal.science/hal-03199809>

Submitted on 15 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Analyse automatique pour une étude du genre : quels jugements des écrivaines au XIXe siècle ?

Marine Riguet et Motasem Alrahabi
Labex OBVIL – Sorbonne Université

L'étude du discours critique français de la seconde moitié du XIX^e siècle, formulé dans les articles de presse et les essais par des auteurs exclusivement masculins, laisse voir une disparité de proportion évidente : on parle en majorité d'hommes écrivains. Néanmoins, certaines écrivaines font bien l'objet d'un jugement critique : George Sand, Germaine de Staël, Madame de Lafayette, Marceline Desbordes-Valmore, George Eliot... Dans une perspective d'étude du genre, il semble alors intéressant de se pencher sur le traitement particulier de ces femmes dans le discours de critique littéraire. Les jugements des critiques du XIX^e siècle sont-ils formulés de la même façon pour les écrivaines que pour les écrivains ? Quels registres sont convoqués, et surtout quels types de valeurs littéraires¹ (esthétique, éthique, sociale, psychoaffective...) leurs sont particulièrement attachés ?

Si de nombreux travaux d'analyse sémantique ont ouvert la voie à de telles études², aucune méthode actuelle en *sentiment analysis et opinion mining* ne permet à notre connaissance de réaliser une analyse automatique suffisamment fine pour se prêter à la spécificité de ces questionnements sur des textes littéraires de langue française. À partir d'un corpus de 134 ouvrages de critique littéraire française publiés entre 1850 et 1914, et numérisés récemment dans le cadre du Labex OBVIL³, nous souhaitons aujourd'hui proposer une méthode d'annotation sémantique adaptée, qui s'appuie sur une analyse du discours de critique littéraire.

Pour ce faire, notre méthode suit plusieurs étapes. Dans un premier temps, des ressources linguistiques ont été établies et catégorisées à partir d'un travail sur notre corpus d'étude :

- Un repérage semi-automatique des entités nommées, réalisé via l'extracteur automatique *UNERD*⁴, puis corrigé manuellement, nous a permis d'identifier le nom de toutes les écrivaines citées dans le corpus.

¹ Nous nous appuyons sur la typologie des valeurs littéraires établie avant nous. Cf. Christine Chollier (dir.), *Qu'est-ce qui fait la valeur des textes ?*, Reims, EPURE, 2011 ; et de Patrick Voisin (dir.), *La Valeur de l'œuvre littéraire, entre pôle artistique et pôle esthétique*, Paris, Classiques Garnier, 2012.

² Voir notamment Catherine Kerbrat-Orecchioni, *L'Énonciation. De la subjectivité dans le langage*, Paris, Armand Colin, 1980 ; et Jean-Pierre Desclés, « Prise en charge, engagement et désengagement », *Langue française* 2, 2009, n° 162, p. 29-53. URL : www.cairn.info/revue-langue-francaise-2009-2-page-29.htm.

³ Le corpus est un échantillon du corpus de critique littéraire de l'OBVIL, consultable en ligne : <https://obvil.sorbonne-universite.fr/corpus/critique/>. Pour cette expérience, seuls les ouvrages mentionnant au moins une femme écrivain ont été retenus.

⁴ Yusra Mosallem, Alaa Abi-Haidar et Jean-Gabriel Ganascia, « Unsupervised Named Entity Recognition and Disambiguation: An Application to Old French Journals », *Lecture Notes in Computer Science series*, vol. 8557, *Proceedings of ICDM 2014*, St. Petersburg, Russia, 2014.

- Un relevé des marqueurs adjectivaux de modalités a été réalisé à l'aide du logiciel de textométrie *TXM*, et enrichi de ressources linguistiques (verbes et locutions) constituées lors de nos travaux antérieurs⁵. Ces marqueurs ont ensuite été classés par catégories sémantiques (*éloge, blâme, dénonciation, accord, injure, dégoût, parodie...*), dans le but de cerner les systèmes de valeurs mis en place par la critique littéraire, mais aussi les différentes orientations de son énonciation (normative, prescriptive, argumentative...).

Cette analyse préliminaire du discours de critique littéraire a conduit dans un second temps à un travail d'annotation automatique du corpus de critique littéraire de la seconde moitié du XIX^e siècle, à l'aide d'un système à base de règles, afin de catégoriser les jugements émis sur les écrivaines. Dans notre corpus, 6887 phrases ont été annotées sur 32 349, parfois plusieurs fois. Par ailleurs, pour mieux interpréter ces résultats, une annotation similaire a été effectuée, sur ce même corpus, autour des noms d'hommes écrivains. Deux constats généraux peuvent être faits de cette comparaison : d'une part, les proportions entre jugements positifs et négatifs semblent très proches, qu'ils portent sur les hommes ou les femmes (avec 71% (hommes) et 65% (femmes) de catégories globalement connotées positives ; et 29% (hommes) et 34% (femmes) de catégories globalement connotées négatives) ; d'autre part, si les grands types de jugement se répartissent de façon similaire, c'est davantage la nature des valeurs associées qui se distinguent (avec une majorité de jugements esthétiques et psycho-affectifs pour les femmes). Ces résultats feront l'objet d'une présentation détaillée.

En d'autres termes, notre approche sémantique a l'ambition de rendre apparents, dans le discours, la part que les écrivaines occupent et les types de traitements dont elles font l'objet ; elle nous permet qui plus est de comparer les modalités et les valeurs potentiellement distinctes par le biais desquels les critiques évaluent les œuvres selon le genre de l'auteur. Nous espérons ainsi circonscrire un discours sur la littérature féminine et questionner une pensée littéraire façonnée au prisme du genre. Mais nous entendons également proposer une méthode d'analyse sémantique exportable à d'autres discours, et adaptable aux besoins spécifiques d'autres recherches littéraires.

⁵ Motasem Alrahabi, « E-Quotes : un outil de navigation textuelle guidée par les annotations sémantiques », *TALN*, Paris, France, 2016 ; Marine Riguet et Motasem Alrahabi, « Pour une analyse automatique du jugement critique : les citations modalisées dans le discours littéraire du XIX^e siècle », *DHQ : Digital Humanities Quarterly*, n°1, vol. 12, 2018.