HAL open science

Voguer au fil de la transtextualité

Motasem Alrahabi, Marine Riguet

To cite this version:

Motasem Alrahabi, Marine Riguet. Voguer au fil de la transtextualité. Digital Humanities Conference, Aug 2017, Montréal, Canada. hal-03199836

HAL Id: hal-03199836

https://hal.science/hal-03199836

Submitted on 16 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.

Voguer au fil de la transtextualité

Motasem Alrahabi[1] et Marine Riguet[2]

1 Paris-Sorbonne Abu Dhabi
2 Labex OBVIL (Paris IV-Sorbonne)

Introduction

À l'ère du numérique, la littérature ne se pense plus à partir de l'unité du texte, mais se conçoit dans sa dimension réticulaire et dynamique. À l'heure d'un colloque consacré à la notion d'Accès, nous souhaiterions mettre en avant cette exploration de la littérature au prisme du dialogue qu'elle génère et qui l'alimente, c'est-à-dire aborder la valeur singulière d'un texte à partir du tout qui le façonne.

Dans ce dessein, nous proposons une méthode de fouille inédite et adaptable à tout type de questionnement, qui s'attache au traitement des citations au sein d'un vaste corpus. À l'aide d'une interface web de navigation textuelle guidée par les annotations sémantiques et discursives, il s'agira pour nous de traiter la question suivante : sur quels jugements de valeur la littérature se structure-t-elle ? De quelle façon forme-t-elle, à partir des citations, un discours canonique ?

L'étude se concentre sur la critique littéraire française de la seconde moitié du XIXe siècle, à un moment où celle-ci se constitue à proprement parler comme discours ; de Sainte-Beuve à Lanson, les critiques multiplient en effet les initiatives afin de délimiter leur champ et de conquérir leur autonomie vis-à-vis des sciences et de la philosophie. Cet effort de légitimation du discours nous semble d'autant plus intéressant qu'il accompagne la rupture avec un modèle classique et l'autonomisation de la littérature telle qu'elle sera proclamée au début du XXe siècle. Dans ce contexte, le travail sur la citation fait d'autant plus sens que les critiques y recourent abondamment de sorte d'asseoir leur propos ; c'est pourquoi la transtextualité nous paraît au cœur de cette entreprise d'unification et de légitimation du discours critique. L'intérêt est moins pour nous le contenu même de la citation que sa prise en charge, soit la façon dont elle est rendue opérante pour servir la formation de ce nouveau discours sur la littérature.

I. Méthode et choix techniques

Afin de mener l'exploration du corpus, nous nous sommes tournés vers E-Quotes[1], une interface web de navigation textuelle guidée par les annotations sémantiques et discursives [Alrahabi, 2015]. L'application prend en entrée, selon la langue, des corpus segmentés et étiquetés par Excom2, un outil d'annotation à base de règles [Alrahabi, 2010].

L'interface utilisateur d'E-Quotes (*Figure 1*) permet de combiner la technologie classique de recherche d'information (*requêtes de mots clés sur un sac de mots*) et une recherche autour des citations sémantiquement catégorisées. Ainsi, l'utilisateur peut rechercher un terme dans les citations annotées selon une catégorie bien particulière, mais il a également la possibilité d'affiner cette recherche en choisissant de localiser les mots clés à l'intérieur ou bien à

[1] http://www.e-quotes.net

l'extérieur des citations. De cette façon, le corpus est exploré à partir du positionnement de l'énonciateur ou de l'attitude du locuteur vis-à-vis d'un propos rapporté.

En outre, les termes recherchés dans E-Quotes peuvent être enrichis par leurs synonymes ou des mots du même champ lexical (ex. *darwinisme, évolution, classification...*), ou encore par leurs équivalents (ex. *Gustave Flaubert, M. Flaubert…*). L'application permet également d'effectuer des requêtes dans les champs "Titre", "Auteur" ou "Date" des articles du corpus et de combiner plusieurs requêtes ensemble à l'aide des opérateurs ET, OU et NON. Afin de faciliter la navigation entre les résultats, un dispositif dans chaque document permet d'afficher toutes les annotations identifiées, reliées à leur contexte d'origine dans le document. Cette configuration visuelle où annotations sémantiques et mots clés sont mis en relief offre dans les longs documents une lecture optimale et une interprétation pertinente au niveau du sens et du discours. On peut ainsi parcourir les annotations dans le texte, d'une Définition à une Comparaison, d'une Accusation à une Indignation, d'une Opinion positive à une autre négative, etc.

Figure 1 : Interface d'E-Quotes

II. Corpus et évaluation

Cette présente étude s'appuie sur un corpus de 300 ouvrages de critique littéraire française, publiés entre 1830 et 1920[2]. Ceux-ci ont été préalablement numérisés dans le cadre du Labex OBVIL de Paris IV-Sorbonne et sont librement accessibles en ligne[3].

[2] Environ 23 millions de mots.
[3] http://obvil.paris-sorbonne.fr/corpus/critique

Nous avons créé 25 catégories regroupant plus de 600 marqueurs linguistiques manuellement collectés à partir de l'observation du corpus. Il s'agit principalement de marqueurs verbaux introducteurs de citations et de modalisateurs de discours comme les adjectifs et les adverbes.

Nous avons effectué une évaluation métrique pour mesurer la Précision de l'annotation automatique effectuée. À partir d'un échantillon du corpus (*environ 7 %*), nous avons comparé les citations automatiquement catégorisées par Excom2 à celles manuellement annotées par 3 juges. Ainsi, nous avons obtenu 87.6% de Précision pour l'ensemble des catégories évaluées.

Il est vrai que ces résultats dépendent des ressources employées et ne permettent pas encore de découvrir toutes les citations dans le corpus, mais ils nous donnent un très bon indice de la qualité d'annotation automatique.

Soulignons également que l'originalité de notre approche réside dans son aspect opérationnel. En effet, plusieurs travaux ont abordé la question des citations (Pouliquen et al., 2007), mais très peu de recherches ont abouti à des applications réelles qui offres à un utilisateur final la possibilité de fouiller des citations catégorisées selon des critères sémantiques ou discursifs.

Par la suite, nous proposons une interprétation littéraire de certaines catégories qui prédominent de façon remarquable dans notre corpus.

III. Une littérature consensuelle ?

L'exploration de notre corpus de critique littéraire du XIXe siècle via E-Quotes est en ceci significative qu'elle rend rapidement identifiables les auteurs, œuvres ou idées qui font l'objet d'un consensus. Une requête de « Shakespeare » (aidée des variantes orthographiques) autour du discours rapporté renvoie par exemple les résultats suivants :

Auteur	Citation	Annotation automatique
Barbey d'Aurevilly	« …cette ravissante fantaisie de Shakespeare, qui tire, comme on tire un instrument merveilleux d'un étui tombé dans la fange,… »	Appréciation
Barbey d'Aurevilly	« Shakespeare, avec son ironie charmante, appelle quelque part les laquais .... »	Appréciation
Deschanel	« *Macbeth* est, entre toutes les pièces de Shakespeare, celle que Gœthe admirait le plus, ... »	Appréciation
Du Bos	« Le rapprochement ne vaut pas moins pour l'expression dont Lytton Strachey a observé avec justesse "que Shakespeare la porte toujours jusqu'au point d'éclatement". »	Accord
Taine	« …qui véritablement est "de la même substance que ses songes ;" et voilà l'homme tel que Shakspeare l'a conçu. »	Accord

Grâce aux catégories sémantiques annotées automatiquement, on observe très rapidement l'adhésion globale d'une critique qui porte soit directement sur l'œuvre de Shakespeare, soit sur les commentaires exprimés en amont par d'autres écrivains à propos de Shakespeare. À ce titre, on voit de quelle façon la transtextualité, et plus précisément les références à ce qui s'est dit *ailleurs* et *avant*, vient servir en critique littéraire la formation d'une opinion collective. On pourrait multiplier les exemples ; ainsi pour Buffon obtient-on :

Auteur	Citation	Annotation automatique
Bourget	« La doctrine de Flaubert sur le style est renfermée dans cette formule de Buffon qu'il cite quelque part avec admiration : “Toutes les beautés intellectuelles qui se trouvent dans un beau style…” »	Appréciation
Bourget	« Buffon définissait le style : “l'ordre et le mouvement que l'on met dans ses pensées”. » C'est pour obéir à cette règle que Balzac a subordonné, de parti pris, les effets de détail aux effets d'ensemble.	Définition
Deschanel	« Les idées, les faits, les opinions, les connaissances, comme le dit admirablement Buffon, “sont hors de l'homme ; le style est l'homme même...” »	Éloge
Faguet	« C'est de ce ton que Buffon parle toujours du “maître de la terre”, et je ne cite pas, comme trop connu, le passage fameux :…»	Éloge

Ces résultats sont doublement intéressants. Premièrement, ils permettent de constater que seule une partie des critiques littéraires citent Buffon, rappelant la ligne de force qui oppose à l’époque critiques positivistes d’un côté, symbolistes de l’autre. Deuxièmement, ces critiques s’appuient sur la même définition que Buffon donne du style, si ce n’est sur la même citation ; plus encore, cette référence à Buffon vient servir l’analyse critique d’autres écrivains, tel Flaubert ou Balzac. Parce qu’elle provient d’un discours exogène, qui plus est de facture scientifique, la citation fait ici figure d’argument d’autorité admis d’emblée par un ensemble de critiques. Elle participe en ce sens de la construction d’une identité collective.

IV. Façonnements d’un discours canonique

Rappelons que la critique littéraire du XIXe siècle cherche à se détacher du modèle classique des belles-lettres pour se constituer comme discours autonome et légitimé. Dans cette mesure, l’émergence d’un consensus autour de la citation semble prendre part à une stratégie de validation de la critique. Une exploration du discours rapporté via E-Quotes met au jour l’abondance de citations entre critiques, voire intra-corpus.

Auteur	Citation	Annotation automatique
Albalat	« Le style de Taine, confirme M. Faguet, est un miracle de volonté ... »	Affirmation
Faguet	Scherer n'a pas eu tort de dire : « Il y a du Rousseau dans cet homme-là. »	Accord
Deschanel	M. Taine en fait bien la remarque : « Par-dessous l'amateur du Moyen Age... »	Observation

Quand Brunetière se place dans les sillons de Sainte-Beuve ou que Deschanel rapporte l’observation de Taine, l’un et l’autre consolident leur position personnelle par son inclusion au sein d’une communauté. Cette construction discursive accompagne la formation d’un métadiscours au travers duquel la critique conquiert son autonomie et s’émancipe des discours environnants. Se dessinent ainsi les grandes lignes d’un canon d’autorité à partir duquel pourra se déterminer, dès le début du XXe siècle, la nouvelle valeur de la littérature.

Sources bibliographiques

Alrahabi M. (2010). EXCOM-2 : plateforme d'annotation automatique de catégories sémantiques. Applications à la catégorisation des citations en français et en arabe. Thèse de doctorat, sous la direction du Prof. Jean-Pierre Desclés, Université Paris-Sorbonne.

Alrahabi M. (2015). E-Quotes: Enunciative Modalities Analysis Tool for Direct Reported Speech in Arabic. Actes de *The 16th International Conference on Intelligent Text Processing and Computational Linguistics (CICLING)*, Cairo, Egypt.

Compagnon, A. (1979). La Seconde Main ou le travail de la citation. Seuil, Paris.

Desclés, J.-P. (2009). « Prise en charge, engagement et désengagement », Langue française, 162 : 29-53.

Genette, G. (1982). Palimpsestes, ou La Littérature au second degré. Seuil, Paris.

Pouliquen B., Steinberger R., Best C. (2007). Automatic detection of quotations in multilingual news. Actes de *Recent Advances in Natural Language Processing (RANLP)*, Borovets, Bulgaria

Riguet, M. (2016). « L'intertextualité à l'épreuve du numérique », Workshop Data Bahia, UFBA, Salvador de Bahia, Brésil.