



HAL
open science

Combattre les silences

Violette Saïag, Marie Puren

► **To cite this version:**

Violette Saïag, Marie Puren. Combattre les silences. *Humanistica* 2024, Association francophone des humanités numériques, May 2024, Meknès, Maroc. hal-04761030

HAL Id: hal-04761030

<https://hal.science/hal-04761030v1>

Submitted on 30 Oct 2024

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.



Distributed under a Creative Commons Attribution 4.0 International License

Combattre les silences. Une analyse des représentations et des (re)configurations de la féminité dans la littérature lesbienne du Paris fin-de-siècle (1880-1970) avec la modélisation de sujets

Violette Saïag¹ et Marie Puren^{2,3}

¹École des hautes études en sciences sociales

violette.saiag@chartes.psl.eu

²ÉPITA, Laboratoire de Recherche et de Développement

marie.puren@epita.fr

³École nationale des Chartes | PSL, Centre Jean Mabillon

Résumé

L'article examine les représentations de la féminité dans la littérature lesbienne du Paris fin-de-siècle, en analysant 45 œuvres écrites entre 1880 et 1970. Nous essayons de déterminer si ces œuvres proposent une nouvelle conception de la féminité, ou bien si elles se contentent de reprendre les normes de l'époque. En utilisant la modélisation de sujets, notre étude révèle que ces écrits contestent en partie les normes de genre, et contribuent à l'expression d'une forme d'émancipation féminine. Les méthodes basées sur LDA, Top2Vec, et BERTopic offrent des perspectives complémentaires, soulignant la richesse et la diversité des représentations de la féminité dans ce corpus.

1 Introduction

La fin du XIX^e siècle et le début du XX^e correspondent à la multiplication dans le paysage parisien des lieux de rencontre lesbiens (Albert, 1998). Du paysage parisien au paysage littéraire français, les autrices lesbiennes se frayent peu à peu un chemin dans la culture. Toutefois, les œuvres écrites par les femmes lesbiennes à la Belle Époque ont été largement ignorées par l'histoire littéraire (Turbiau et al., 2022). Le domaine des études sur le genre ne s'y est pas penché non plus, car ces autrices étaient considérées comme des esthètes oisives qui n'étaient pas féministes (Lazard, 2008). Pourtant, leurs écrits ont jeté les bases des revendications des mouvements féministes des années 1970, en représentant des personnages féminins forts et actifs qui contestent les rôles sexués.

Nous étudierons la conception que ces autrices lesbiennes du Paris fin-de-siècle ont de leur propre féminité au sein d'un corpus de 45 romans, recueils de nouvelles et mémoires, afin de déterminer si elles créent de nouveaux critères de féminité. Nous avons choisi la période large de 1880 à 1970 afin

d'inclure dans le corpus les œuvres tardives d'autrices influencées par les salons de la fin du XIX^e siècle et du début du XX^e siècle, comme Jeanne Galzy (1883-†1977) qui a vécu presque 100 ans et a écrit tout au long de sa vie. Les autrices des œuvres qui constituent ce corpus répondent à deux caractéristiques : elles entretiennent des relations amoureuses, affectives et sexuelles avec d'autres femmes, et elles ont fréquenté les cercles littéraires lesbiens du Paris de la Belle Époque (Albert, 2006). Notre hypothèse principale est que ces écrits lesbiens pourraient témoigner d'une féminité spécifique aux lesbiennes, qui ne pouvait être assimilée à la féminité hétérosexuelle (Castle, 1993).

Au cours de la Belle Époque, la relative liberté de la vie parisienne en fait une ville cosmopolite, un creuset d'émulation intellectuelle et d'échanges littéraires et amoureux sous l'égide de Sappho, dont les autrices lesbiennes revendiquent l'héritage (Albert, 1998). Certaines d'entre elles ont non seulement vécu leur homosexualité de manière relativement ouverte, mais l'ont également évoquée dans leurs œuvres – une nouveauté, puisque le thème du lesbianisme était jusqu'alors cantonné aux romans, à la poésie et aux essais écrits par des hommes hétérosexuels. Ces écrivaines lesbiennes occupent une position ambiguë dans la société parisienne, car elles écrivent dans des cercles littéraires et sociaux hermétiques (Waelti-Walters, 1997). L'existence de ces microcosmes, qui leur sont exclusivement réservés, permet de supposer que leur féminité est plus autonome que la féminité hétérosexuelle car, dans le cadre de leur vie intime et sociale échappant au joug masculin, elles entretiennent une forme de féminité anandryne. Leur féminité ne serait donc pas définie en termes négatifs par rapport à la masculinité, comme le voudraient les normes de l'époque. Les représentations

artistiques des femmes par les hommes de cette période témoignent de cette conception réductrice de la féminité (Nochlin et Bonis, 1993).

Dans la mesure où les autrices lesbiennes de la Belle Époque ont collaboré artistiquement au sein de leurs propres cercles, peut-on y déceler une conception commune de la féminité ? Créent-elles leurs propres critères de féminité, ou se contentent-elles de reprendre les codes traditionnels ? Ou bien constaterons-nous au contraire qu’il n’existe pas de conception commune de la féminité lesbienne ?

Pour tenter de répondre à ces questions, notre approche combine lecture proche et lecture distante (Moretti, 2013). Pour « lire à distance » notre corpus, nous utilisons la modélisation de sujets ou *topic modeling*. Cette méthode, particulièrement bien adaptée à l’étude d’importants corpus textuels, permet de faire « émerger » automatiquement des sujets ou *topics* de grands corpus de textes (Shawn et al., 2016; Lemercier et Zalc, 2008). Particulièrement bien adaptée aux corpus sériels (comme la presse, cf. Marjanen et al. 2020; Mayor et Miani 2023), la modélisation de sujets peut aussi être utilisée de manière tout à fait pertinente pour analyser des collections de textes littéraires (Navarro-Colorado, 2018).

Plusieurs méthodes sont aujourd’hui à notre disposition pour identifier les sujets. Si l’allocation de Dirichlet latente ou LDA (Blei et al., 2003) est certainement la plus connue et la plus utilisée, ces dernières années ont aussi vu l’émergence de représentations distribuées des mots sous forme de vecteurs appelés *word embeddings* (Mikolov et al., 2013b) ou plongements de mots. Pour représenter les mots sous forme de vecteurs, on utilise des modèles de langage pré-entraînés comme Word2Vec (Mikolov et al., 2013a) ou BERT. Introduit par Google en 2018, BERT (*Bidirectional Encoder Representations from Transformers*, cf. Devlin et al. 2019) a marqué un tournant significatif dans la compréhension du langage par les machines. Si des modèles bidirectionnels tels que les Bi-LSTM existaient déjà, BERT repose sur l’architecture des transformeurs, une évolution du mécanisme d’attention proposé par Bahdanau et al. (Bahdanau et al., 2016) qui permet une meilleure gestion des dépendances à longue distance dans les séquences de texte.

Nous avons ainsi choisi de consacrer cet article à la comparaison entre trois méthodes : l’allocation de Dirichlet latente, les plongements de mots ou

word embeddings et le *Transfer learning*. LDA présente l’avantage d’offrir une perspective globale sur le contenu du corpus. En retour, les techniques d’*embedding* de mots offrent une exploration plus détaillée de nos textes, en intégrant notamment le contexte et la structure des phrases, ce qui permet d’obtenir une compréhension plus fine des sujets abordés. Lorsque l’on se penche sur un thème spécifique et son évolution, il s’avère particulièrement bénéfique de combiner ces deux approches (Puren et Pellet, 2023). Il nous semble ainsi judicieux de commencer par une analyse préliminaire avec LDA, puis d’affiner l’exploration de certains aspects en recourant aux plongements de mots.

2 Les trois méthodes utilisées

L’allocation de Dirichlet latente ou LDA (Blei et al., 2003) est certainement la méthode la plus connue et la plus utilisée. LDA part de l’idée que les documents sont générés par un processus spécifique où pour chaque document, l’on choisit une distribution de sujets, puis, pour chaque mot dans le document, l’on sélectionne un sujet selon cette distribution. L’on choisit ensuite un mot basé sur une distribution de mots associée à ce sujet. Pour chaque sujet, LDA calcule la probabilité qu’un mot donné soit généré dans le contexte de ce sujet. Simultanément, LDA détermine pour chaque document la distribution des sujets, c’est-à-dire la probabilité qu’un sujet donné soit sélectionné dans le contexte de ce document. Contrairement à BERTopic ou à Top2Vec, le modèle LDA ne génère pas lui-même un nombre de *topics*. Ainsi, par tâtonnements, nous ajustons le nombre de *topics* et nous fions à la mesure de cohérence de la librairie Gensim pour choisir un nombre adéquat.

En se basant sur des *embeddings* de mots, Top2Vec (Angelov, 2020) vise à découvrir la structure thématique des données textuelles sans supervision, en apprenant conjointement les représentations vectorielles des documents et des mots dans un espace vectoriel commun. Il utilise ces représentations pour regrouper des documents similaires, ce qui permet de découvrir les sujets ou thèmes principaux contenus dans un vaste ensemble de données textuelles. Cela permet de capturer la sémantique des mots et des documents, où les vecteurs de mots et de documents similaires dans leur signification sont positionnés plus près les uns des autres dans cet espace.

Basé sur BERT, BERTopic (Grootendorst, 2022)

que le lesbianisme était rarement évoqué dans des termes propres à l'orientation sexuelle ; la modélisation de sujets révèle quant à elle que l'amour entre deux femmes était généralement évoqué de manière très descriptive et que les autrices s'attachaient à décrire la dimension corporelle de l'acte amoureux.

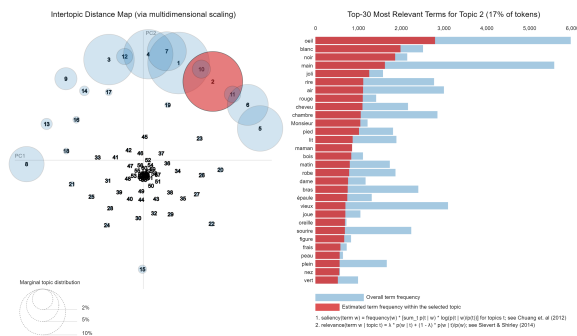


FIGURE 4 – Topic montrant un vocabulaire corporel et sensoriel

3.2.2 Un amour différent

Le *topic* 3 (fig. 1) donne un aperçu des connotations de l'amour dans notre corpus, que nous avons déjà en partie repérées lors de notre lecture proche. Les mots « nature » et « naturel » peuvent évoquer le renversement des valeurs qu'opèrent les autrices lesbiennes en montrant que ce sont leurs amours qui sont naturelles, et les critiques qui leur sont adressées qui ne le sont pas.

Le *topic* 4 (2), que nous avons déjà évoqué, a la particularité de contenir le mot « femme » mais pas le mot « homme ». Cette observation permet de vérifier partiellement l'hypothèse formulée par Nicole G. Albert d'une « écriture séparatiste » (Albert, 1998) porteuse d'une voix poétique spécifiquement lesbienne. Ce thème nous montre que, même dans leurs romans, les autrices lesbiennes utilisent une voix lyrique et poétique qui s'adresse uniquement à d'autres femmes et n'inclut pas de présence masculine.

Si les analyses menées avec BERTopic ont en grande partie confirmé les résultats obtenus avec LDA, elles font aussi apparaître un sujet qui, jusque-là, était passé totalement inaperçu : la liberté dans l'amour. Dans le *topic* 26, on voit apparaître une concordance entre l'amour, la liberté et le respect de l'autre.

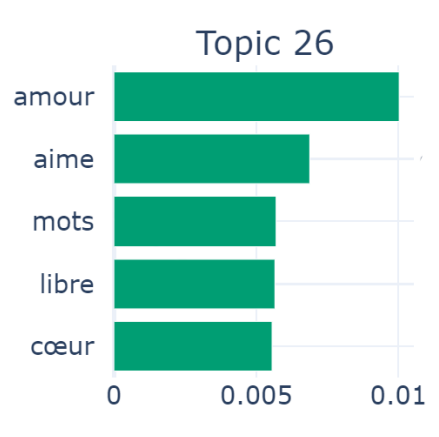


FIGURE 5 – Topic exprimant un amour libre

L'amour peut en effet être une libération ; celle-ci se manifeste également par la présence d'un *topic* 33 parlant de féminisme, thème qui n'avait pas été détecté par LDA, et qui provient de l'ouvrage *La Femme vierge* de Madeleine Pelletier (6). LDA et BERTopic s'avèrent donc tout à fait complémentaires dans le cadre de notre étude.

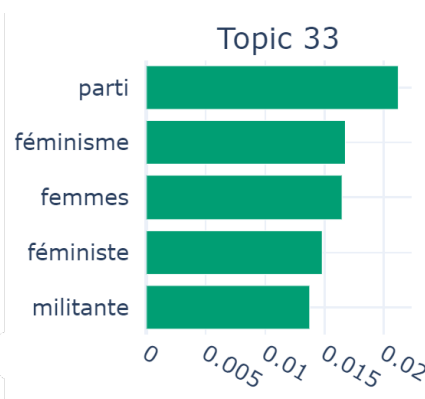


FIGURE 6 – Topic parlant de féminisme

3.2.3 Une remise en cause des rôles sexués ?

Explicitement lié au champ lexical de l'amour, le *topic* 3 (1) généré par LDA comprend les mots « femme » et « homme » (le premier étant plus fréquent que le second). Le fait que l'amour désigne ici aussi bien les hommes que les femmes peut être interprété de différentes manières. D'une part, cela montre que certaines des histoires d'amour décrites dans notre corpus sont des histoires hétérosexuelles – l'amour lesbien étant parfois comparé à l'amour hétérosexuel. D'autre part, cela témoigne du fait que l'amour soit un lieu privilégié de réflexion sur les rôles sexués, notamment dans les œuvres de Natalie Clifford Barney. Pour rédiger ses ouvrages littéraires et ses mémoires, cette autrice puise souvent dans des histoires d'amour. Celles-ci inspirent

ses réflexions sur les différents rôles attendus des femmes et des hommes dans une relation amoureuse.

Avec Top2Vec, il est nécessaire de fournir un mot-clé pour rechercher les sujets qui sont sémantiquement liés à ce mot. Nous avons ainsi effectué plusieurs recherches en utilisant trois mots-clés différents : « femme », « amour » et « Sapho ». Les thèmes retournés par Top2Vec pour chaque mot-clé se chevauchent, ce qui nous montre que ces trois termes sont compris par le modèle comme étant relativement similaires. Nous avons été surprises de constater que le thème ayant la plus grande similarité avec le mot-clé « femme » était lié à l'activité politique, avec des mots comme « politique », « parti » ou encore « réunions » :

Topic 14



FIGURE 7 – Topic le plus associé à « femme »

Trois œuvres ont principalement irrigué ce sujet. Il s'agit de *Les Sévriennes* de Gabriel Reval (1900), *Claudine à l'école* de Colette (1900) et *La Femme vierge* de Madeleine Pelletier (1933). Ce thème montre que, dans notre corpus, une discussion s'est engagée sur la place des femmes dans certains milieux tels que la politique, la culture et l'enseignement. Il ne faut néanmoins pas conclure de ce topic que le rôle des femmes dans notre corpus est fortement politisé, car ce n'est pas le cas : une lecture proche montre que c'est la place des femmes dans les instances politiques qui est débattue, et qu'il n'y est pas question de leur émancipation politique.

4 Conclusion

Ces trois approches nous ont permis de mieux comprendre comment les autrices lesbiennes caractérisaient leur propre féminité. Nous avons remarqué que si LDA était particulièrement efficace pour extraire les grands tendances dans le corpus, l'usage commun de Top2Vec et de BERTopic pouvait permettre de révéler des sujets inattendus et particulièrement intéressants pour notre étude.

Jusqu'à la fin du XIX^e siècle, le lesbianisme n'avait été vu qu'à travers le regard masculin dans la littérature française ; nous constatons alors que les autrices écrivant dans les premiers temps de la

littérature lesbienne n'avaient pas encore développé un vocabulaire qui leur soit propre. Néanmoins, bien qu'elle soit parfois définie dans les mêmes termes, la manière dont la féminité y est présentée témoigne d'une affirmation d'un sujet féminin actif, qui s'exprime librement et parfois en opposition à ses homologues masculins. Ainsi, le développement des milieux lesbiens à Paris se reflète dans la littérature lesbienne, dont les personnages féminins manifestent une certaine conscience de groupe, et constitue une étape primordiale dans l'émergence de la mémoire et de l'histoire d'une communauté et de son émancipation.

Si certaines écrivaines lesbiennes cautionnent les représentations classiques de l'amour, d'autres l'utilisent comme un levier d'émancipation. Les personnages de ces œuvres s'écartent des rôles féminins sexués : les femmes lesbiennes sont présentées ainsi comme des êtres hybrides. La littérature lesbienne diffère des écrits des femmes et des hommes hétérosexuel.le.s qui écrivent sur le lesbianisme, en ce qu'elle remet en question les rôles de genre. Des réflexions sur la misogynie dans l'œuvre de Natalie Clifford Barney (1876-†1972) à la description du plaisir féminin dans le roman de la Marquise de Mannoury d'Ectot (1815-†1899), en passant par la présence de personnages féminins forts et engagés qui s'éloignent de la passivité imposée aux femmes dans les écrits de Madeleine Pelletier (1874-†1939), la littérature lesbienne apparaît comme un moteur de changement social (Turbiu et al., 2022).

Bibliographie

- Nicole G. Albert. 1998. *Saphisme et décadence dans l'art et la littérature en Europe à la fin du XIXe siècle*. Thèse de doctorat, Paris IV-Sorbonne.
- Nicole G. Albert. 2006. De la topographie invisible à l'espace public et littéraire : les lieux de plaisir lesbien dans le paris de la belle époque. *Revue d'histoire moderne contemporaine*, 53(4) :87–105.
- Dimo Angelov. 2020. [Top2vec: Distributed representations of topics](#).
- Dzmitry Bahdanau, Kyunghyun Cho, et Yoshua Bengio. 2016. [Neural machine translation by jointly learning to align and translate](#).
- David Blei, Andrew Ng, et Michael Jordan. 2003. [Latent Dirichlet Allocation](#). *Journal of Machine Learning Research*, 3 :993–1022.
- Terry Castle. 1993. *The apparitional lesbian : Female homosexuality and modern culture*. Columbia University Press, New York.

- Jacob Devlin, Ming-Wei Chang, Kenton Lee, et Kristina Toutanova. 2019. [BERT: Pre-training of deep bidirectional transformers for language understanding](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Maarten Grootendorst. 2022. Bertopic : Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv :2203.05794*.
- Madeleine Lazard. 2008. *Colette*, volume 35. Gallimard, Paris.
- Claire Lemerrier et Claire Zalc. 2008. *Méthodes quantitatives pour l'historien*, Repères, pages 3–7. La Découverte, Paris.
- Jani Marjanen, Elaine Zosa, Simon Hengchen, Lidia Pivovarova, et Mikko Tolonen. 2020. [Topic modelling discourse dynamics in historical newspapers](#).
- Eric Mayor et Alessandro Miani. 2023. [A topic models analysis of the news coverage of the omicron variant in the united kingdom press](#). *BMC Public Health*, 23(1).
- Tomas Mikolov, Kai Chen, Gregory S. Corrado, et Jeffrey Dean. 2013a. [Efficient Estimation of Word Representations in Vector Space](#). In *International Conference on Learning Representations*, Scottsdale, Arizona.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, et Jeff Dean. 2013b. [Distributed representations of words and phrases and their compositionality](#). In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc.
- Franco Moretti. 2013. *Distant reading*. Verso, Londres.
- Borja Navarro-Colorado. 2018. [On poetic topic modeling: Extracting themes and motifs from a corpus of spanish poetry](#). *Frontiers in Digital Humanities*, 5.
- Linda Nochlin et Oristelle Bonis. 1993. *Femmes, art et pouvoir ; et autres essais*. Editions Jacqueline Chambon.
- Marie Puren et Aurélien Pellet. 2023. [Explorer les débats parlementaires français de la Troisième République par leurs sujets](#). In *Humanistica 2023*, Corpus, Genève, Suisse. Association francophone des humanités numériques.
- Graham Shawn, Ian Milligan, et Scott Weingart. 2016. *Exploring big historical data: the historian's macro-scope*. Imperial College Press, Londres.
- Aurore Turbiau, Margot Lachkar, Camille Isler, Manon Berthier, et Alexandre Antolin. 2022. *Écrire à l'encre violette. Littératures lesbiennes en France de 1900 à nos jours*. Le Cavalier bleu, Paris.
- Jennifer Waelti-Walters. 1997. [Les lesbiennes et le roman français 1796-1996](#). *Nouvelles questions féministes*, pages 45–56.