

Y a-t-il une spécificité thématique de la pastorale ?

Sujet

Le genre de la pastorale a eu une réception plus chaotique (Mortgat-Longuet, 2007) que les tragédies et comédies de l'époque – au point qu'aujourd'hui, le genre est peu connu, et mal identifié. Dans sa *Pratique du théâtre*, l'Abbé d'Aubignac comptait pourtant la pastorale comme un des trois genres fondamentaux du théâtre de son époque :

Comme il y a trois sortes de Vies, celle des Grands dans la Cour des Rois, celle des Bourgeois dans les Villes, et celle des gens de la Campagne ; le Théâtre aussi a reçu trois genres de Poèmes Dramatiques qui portent en particulier le caractère de chacune de ces trois sortes de Vies, savoir la Tragédie, la Comédie, et la Satyre ou Pastorale.

Qu'est ce qui au-delà du caractère tragique ou comique, fait la spécificité de la pastorale ? Quelles sont les délimitations de ce genre ? Notons que beaucoup de pièces aujourd'hui considérées comme ayant un caractère pastoral n'étaient pas explicitement désignées ainsi (Berregard, 2007). D'après l'analyse de leurs thématiques, quelles seraient les pièces qui, bien que n'étant pas étiquetée comme pastorales, mériteraient une telle appellation ?

Retrouve-t-on, en étudiant les thématiques de ces pièces, les différentes variations évoquées par certains critiques (« pastorales pures », « comédies de couleur pastorale », « comédies de type pastoral », (Morel, 1968 ; Scherrer, 1983) ?

Données

On dispose de 493 œuvres dramatiques françaises du XVII^e siècle.

Chaque pièce est fournie sous la forme d'un fichier XML-TEI annoté.

L'ensemble est accompagné d'un tableau de métadonnées qui fournissent, au-delà des informations nécessaires à l'identification de chaque pièce, des données sur la date, la structure, la taille et, ce qui nous intéresse plus particulièrement, le genre et l'inspiration.

Approches retenues

On essaiera de traiter le sujet en abordant successivement les trois questions suivantes :

- question 1 : peut-on vérifier l’assertion de l’Abbé d’Aubignac à travers l’analyse d’un ensemble de pièces identifiées par la critique comme relevant, en tant que genre, de la tragédie, de la comédie ou de la pastorale ?
- question 2 : l’analyse des métadonnées de genre et inspiration montre la présence de pièces d’inspiration pastorale mais ne relevant pas du genre pastoral. On peut donc supposer que les pièces d’inspiration pastorale relèvent également d’une thématique pastorale. Est-ce que cette thématique est différente des pièces relevant du genre pastoral ?
- question 3 : en se penchant plus spécifiquement sur le théâtre de Jean de Rotrou, parvient-on à retrouver les différentes variations de pastorales évoquées par certains critiques ?

Pour extraire les thématiques, on utilise un modèle de *topic modeling* LDA (*Latent Dirichlet Allocation*). On recourt ici l’implémentation [MALLET](#), via l’API proposée par le module Python [Gensim](#).

Prétraitement des données

L’algorithme LDA requiert en entrée une liste de *documents*, c’est-à-dire de termes appartenant à un même ensemble et en nombre relativement peu élevé (environ 700-800). Pour l’ensemble des pièces, on effectue donc le prétraitement suivant :

- on divise chaque pièce en parties plus restreinte d’environ 800 termes, parties qui constitueront les documents analysés par l’algorithme LDA. Dans la mesure où l’on ne s’intéresse qu’aux termes les plus signifiants, on va extraire, pour chacune des pièces, uniquement les lemmes des noms, des adjectifs qualificatifs et des verbes, afin de diminuer le bruit.
- pour faciliter l’analyse, on associe à chaque *document* les métadonnées qu’on juge les plus significatives (identifiant, auteur, titre, genre, inspiration et nombre de lemmes en particulier).

Pour réaliser ce traitement, on a écrit le script [preprocessing_tei2csv.py](#), qui, à partir des fichiers XML-TEI fournis, produit un [tableau](#), qu’on enregistre sous forme de fichier csv et qu’on utilisera par la suite comme fichier source.

Le script et le tableau final sont donc des variantes des fichiers [generate_tm_samples.bash](#) et [tm_samples.tsv](#) initialement proposés avec les données.

Question 1 : pastorales, comédies et tragédies¹

Si l'Abbé d'Aubignac dit vrai, on devrait réussir à dégager trois thématiques, liées respectivement à la cour, à la ville et à la campagne et correspondant à la tragédie, la comédie et la pastorale.

Délimitation du corpus

On essaie de constituer un corpus comprenant un nombre équilibré de *documents* relevant de pièces identifiées comme tragédie, comédie ou pastorale.

On sélectionne d'abord les sept pastorales du corpus : *Lisimène* (Boyer), *Le Berger extravagant* (Thomas Corneille), *Délie* (Donneau de Vizé), *Les Charmes de Félicie* (Pousset de Montauban), *Les Fêtes de l'Amour et de Bacchus* (Quinault), *Amarillis* (Rotrou), *Cléagénor et Doristée* (Rotrou).

On ajoute à celles-ci les comédies et tragédies écrites par les auteurs de pastorales, auxquelles on ajoute quelques œuvres d'auteurs emblématiques (Pierre Corneille, Racine, Molière) : *Le Feint astrologue* (Thomas Corneille), *L'Embarras de Godard* (Donneau de Vizé), *Le Gentilhomme Guespin* (Donneau de Vizé), *La Comédie sans comédie* (Quinault), *Les Sosies* (Rotrou), *L'Illusion comique* (Pierre Corneille), *Les Fourberies de Scapin* (Molière), *Les Plaideurs* (Racine) pour la comédie ; *Agamemnon* (Boyer), *La Mort d'Achille* (Thomas Corneille), *Les Amours du Soleil* (Donneau de Vizé), *Zénobie* (Pousset de Montauban), *Atys* (Quinault), *Le Véritable Saint Genest* (Rotrou), *Horace* (Pierre Corneille), *Britannicus* (Racine) pour la tragédie.

On veille enfin à obtenir un nombre relativement égal de lemmes entre les trois genres, on supprime donc quelques documents parmi les pièces relevant de la tragédie et de la comédie pour parvenir à un équilibre.

Élimination des lemmes les plus fréquents

On établit la table de fréquence des lemmes du corpus afin d'éliminer les termes les plus fréquents. Sachant que n'ont été retenus que les lemmes des noms, verbes et adjectifs qualificatifs, on retient un seuil de fréquence cumulée relativement bas. Après plusieurs essais, on arrête celui-ci à 30 %.

Entraînement et des modèles LDA

Après plusieurs tests, on décide de retenir deux modèles LDA, l'un à trois sujets (ce qui correspond au nombre de genres), l'autre à six sujets (modèle pour lequel le calcul de la cohérence renvoie un taux assez correct).

1. Les traitements et résultats relatifs à cette question sont détaillés dans le fichier [question1.ipynb](#).

Interprétation du modèle à trois sujets

Le modèle parvient à extraire trois thèmes, qui semblent bien correspondre aux trois genres :

- thème 1 : tragédie (lexique qui évoque la grandeur et la noblesse),
- thème 2 : comédie (lexique qui évoque des situations et des relations propres à la bourgeoisie du XVIIe s.)
- thème 3 : pastorale (lexique qui évoque à la fois la campagne et les relations amoureuses).

L'examen des mesures de contributions des thèmes aux pièces soulève quelques questions. *Les Amours du Soleil* (Donneau de Vizé) semblent plutôt relever de la pastorale que de la tragédie (on notera que c'est d'ailleurs le genre retenu par l'éditeur). De même, on constate que la *Comédie sans comédie* (Quinault) emprunte plutôt la thématique de la pastorale que de la comédie. À l'inverse, *Cléagénor et Doristée* (Rotrou) ont davantage de traits de la tragédie que de la pastorale. Enfin, deux pastorales témoignent d'un mélange assez marqués des thématiques de la pastorale et de la comédie : *L'Embaras de Godard* (Donneau de Vizé) et *Les Charmes de Félicie* (Pousset de Montauban).

Interprétation du modèle à six sujets

S'il est difficile de relier précisément les sujets obtenus à l'un ou l'autre des trois genres, le modèle à six sujets permet en revanche de préciser l'analyse pour les pièces ambivalentes :

- le classement des *Amours du Soleil* (Donneau de Vizé) en tant que pastorale plutôt que tragédie est confirmé,
- le doute concernant *L'Embaras de Godard* (Donneau de Vizé) semble levé (la pièce tient bien de la comédie),
- *Les Charmes de Félicie* (Pousset de Montauban) se caractérisent bien par le mélange des thématiques de la pastorale et de la comédie,
- *Cléagénor et Doristée* (Rotrou) apparaissent bien comme un cas particulier : le modèle renvoie une thématique quasiment propre à cette pièce (thématique d'une relation amoureuse contrariée, dans un cadre difficile à déterminer (ni la ville, ni la cour, ni la campagne) ?).

Le modèle semble toutefois mal fonctionner pour la *Comédie sans comédie* (Quinault), qu'on ne parvient pas à classer.

Question 2 : pastorales et inspiration pastorale²

Les pièces d'inspiration pastorale mais identifiées comme tragédies ou comédies ont-elles une thématique qui leur est propre ou qui est partagée avec les pièces relevant du genre

2. Les traitements et résultats relatifs à cette question sont détaillés dans le fichier [question2.ipynb](#).

de la pastorale ?

Délimitation du corpus

On constitue un nouveau corpus, qui comprend :

- les sept pastorales : *Lisimène* (Boyer), *Le Berger extravagant* (Thomas Corneille), *Délie* (Donneau de Vizé), *Les Charmes de Félicie* (Pousset de Montauban), *Les Fêtes de l'Amour et de Bacchus* (Quinault), *Amarillis* (Rotrou), *Cléagénor et Doristée* (Rotrou),
- des textes relevant de genres très voisins de la pastorale (au moins pour leur thématique) : deux pastorales héroïques (*Acis et Galatée* (Campistron) et *La Silvanire* (Urfé)) et deux égloques (*La Grotte de Versailles* (Quinault) et *Le Temple de la paix* (Quinault)),
- les pièces d'inspiration pastorale classées dans un autre genre : *Les Noces de Vaugirard* (Discret), *Astrée* (La Fontaine), *Mélicerte* (Molière), *La Généreuse ingratitude* (Quinault), *Palinice, Circéine et Florice* (Rayssiguier), *Antigone* (Rotrou), *La Bague de l'oubli* (Rotrou), *La Céliane* (Rotrou).

Élimination des lemmes les plus fréquents

Comme précédemment, on établit la table de fréquence des lemmes du corpus afin d'éliminer les termes les plus fréquents. Après plusieurs essais, on arrête le seuil de fréquence cumulée à 25 %.

Essai de construction d'un modèle pertinent

Après avoir calculé le taux de cohérence pour des modèles dont le nombre de sujets varie de un à vingt, on constate assez peu de variation, ce qui ne permet pas d'arrêter un choix. On lance donc manuellement le modèle sur 3, 5, 7 et 9 sujets. Aucun ne donne de résultat probant et surtout cohérent (deux pièces seront proches pour un modèle, éloignées pour d'autres). À titre d'exemple sont présentés les résultats pour un modèle à cinq sujets.

Interprétation

En dépit de tests sur différents modèles (avec pour variable le nombre de sujets), on ne parvient pas à construire d'ensembles stables permettant de dégager une opposition entre pièces dont le genre est la pastorale et pièces relevant d'une inspiration pastorale mais étant classées dans un genre différent. Cela laisse à penser la thématique pastorale n'est pas spécifique au genre du même nom.

Pour aller plus loin, on pourrait reprendre le corpus précédent (question 1) et essayer de l'étendre de manière systématique aux pièces d'inspiration pastorale.

Question 3 : focus sur Jean de Rotrou³

Les pièces de Rotrou sont particulièrement malaisées à classer de manière univoque au sein des différents genres dramatiques. L’auteur fait fréquemment usage, au sein d’une même œuvre, de caractéristiques propres à des genres multiples. Par ailleurs, il n’hésite pas à changer la dénomination du genre d’une pièce d’une édition à l’autre⁴.

Ce jeu entre les genres est particulièrement marqué pour quelques comédies et tragi-comédies que le critique Jacques Morel rassemble sous l’appellation de « pastorales pures »⁵ :

On désignera sous le nom de *pastorale pure* une œuvre dramatique de structure *fermée*, dont le style s’apparente tour à tour à la manière lyrique et à la manière élégiaque, et dont l’action, qui exclut l’aventure et les surprises non préparées, se déroule selon l’exigence d’une certaine logique [...]. Les bergers n’y sont pas indispensables.

Morel définit donc cette catégorie moins par la matière que par le style et les modalités de l’action. On tentera de vérifier cette absence de spécificité thématique.

Délimitation du corpus

Le jeu de données dont on dispose ne comprend qu’une seule de ces pièces, *La Céliane*. On la mettra en regard des deux pièces de Rotrou classées comme pastorale, *Amarillis* et *Cléagénor et Doristée* (cette dernière se distinguant fortement de l’ensemble des pastorales, comme on l’a vu dans la question 1).

Élimination des lemmes les plus fréquents

Comme précédemment, on établit la table de fréquence des lemmes du corpus afin d’éliminer les termes les plus fréquents. Après plusieurs essais, on arrête le seuil de fréquence cumulée à 25 %.

Essai de construction d’un modèle pertinent

Une analyse à trois sujets dispose d’un bon niveau de cohérence. Vu la taille du corpus, on n’ira pas au-delà.

Interprétation

Des trois pièces, *Amarillis* est celle qui fait preuve de la plus grande unité thématique. À l’inverse, *Cléagénor et Doristée* maintient son statut d’œuvre inclassable.

La Céliane puise dans deux thématiques, l’une proche du registre tragique (thème 2 : "mort", "honneur"), l’autre du registre amoureux (thème 3 : "flamme", "désir", "ardeur").

3. Les traitements et résultats relatifs à cette question sont détaillés dans le fichier [question3.ipynb](#).

4. Voir BERREGARD Sandrine, « La mixité des genres dramatiques dans le théâtre de Rotrou », *Littératures classiques*, 2007, no 2, p. 97-106.

5. MOREL Jacques, *Jean Rotrou, dramaturge de l’ambiguïté*, Paris, Armand Colin, 1968, p. 137.

S'il est difficile de conclure, un point est sans équivoque : les bergers (tout comme l'ensemble du lexique de la campagne) sont bien absents, comme le souligne Morel.

Conclusion

Si l'Abbé d'Aubignac distingue tragédie, comédie et pastorale par des thématiques qui leur seraient spécifiques, et s'il est effectivement possible de vérifier ces spécificités sur un corpus restreint et composé essentiellement de pièces qui pratiquent peu le mélange (pas de tragi-comédie par exemple), l'analyse semble montrer que, plutôt qu'un genre bien circonscrit, la pastorale constitue « un ensemble de thèmes et d'images susceptibles de s'incarner dans différents genres »⁶.

Les méthodes de *topic modeling* permettent donc de saisir une part des modalités d'un genre, mais doivent être complétées par d'autres approches, par exemple d'étude du style ou des modalités de l'action.

6. BERREGARD Sandrine, « La mixité des genres dramatiques dans le théâtre de Rotrou », *Littératures classiques*, 2007, no 2, p. 102.