



HAL
open science

Détection automatique de l'ironie dans la stratégie argumentative des querelles théâtrales

Motasem Alrahabi, Chiara Mainardi

► **To cite this version:**

Motasem Alrahabi, Chiara Mainardi. Détection automatique de l'ironie dans la stratégie argumentative des querelles théâtrales. 14th International Conference on statistical analysis of textual data, Jun 2018, Rome, Italie. hal-03199810

HAL Id: hal-03199810

<https://hal.science/hal-03199810>

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.

Détection automatique de l'ironie dans la stratégie argumentative des querelles théâtrales

Motasem ALRAHABI, Chiara MAINARDI
Sorbonne Université

RÉSUMÉ

Cet article présente les résultats de notre analyse automatique d'un corpus de débats sur le théâtre (16^{ème} – 19^{ème} siècle). Notre objectif est d'illustrer le rôle important que jouent les expressions de l'ironie dans la polémique autour du théâtre et de mettre en évidence la position des auteurs ou des autorités antiques citées vis-à-vis des spectacles. Les résultats obtenus sont encourageants malgré la difficulté de la tâche et ils nous permettent de comparer d'une façon inédite les textes des auteurs défenseurs avec ceux des auteurs pourfendeurs du théâtre, ce qui offre une meilleure compréhension de certains arguments et stratégies d'auteurs dans le champ de la controverse.

ABSTRACT

This paper presents the results of an automatic analysis on a corpus of French texts about theatre debates (16th –19th centuries). Our purpose is to highlight the important role of irony in the theatre controversy and to reveal the stand point of authors and established authorities towards theatre performances. Despite the difficulty of this task, our research shows encouraging results. This unprecedented comparison of these kind of texts, in which authors condemn the theatre or approve it, which enables to a broader understanding of the authors' positions, arguments and rhetorical strategies relating to theatre controversies.

MOTS-CLÉS : Ironie, théâtre, marqueurs linguistiques, annotation, système à base de règles.

KEYWORDS: Irony, theatre, linguistic markers, annotation, rule-based system.

1 Prémisses

Depuis son origine, le théâtre est conçu comme un vecteur d'expression d'un sentiment collectif qui génère le développement de controverses de la part de ses détracteurs. Les querelles du théâtre qui se sont succédé au fil du temps sont donc lourdes d'enjeux politiques, économiques et religieux aussi importants que rarement explicités (Lecerclé et al., 2016). Par conséquent, les attaques contre le théâtre traduisent souvent des tensions assez extérieures aux questions traitées explicitement.

Avec cet article, nous proposons une analyse automatique d'un corpus de textes théâtrophiles ou théâtrophobes écrits en français entre le milieu du 16^e siècle jusque dans les années 1840. Notre objectif est d'illustrer le poids des expressions ironiques dans la stratégie argumentative de la polémique autour du théâtre.

Après une courte introduction à l'ironie, nous présentons le corpus de travail, le choix technique et les ressources linguistiques créées. Ensuite, nous abordons l'étape de l'évaluation et de l'interprétation des résultats, avant de conclure sur les prolongements et perspectives à venir.

2 Définition de l'ironie

Comme le mot lui-même (en grec εἰρωνεία¹) l'indique, l'ironie, une *dissimulatio*, est un fait de langue utilisé afin de transmettre un message explicitement ou implicitement opposé à ce qui est dit littéralement. Largement étudiée en philosophie, en rhétorique ou en linguistique (Berrendonner, Sperber et Wilson, Kerbrat-Orecchioni, Ducrot, Grice...), l'ironie représente un concept hétérogène extrêmement difficile à définir du fait de ses nombreuses formes et de la complexité des phénomènes qui sont en jeu. L'ironie fonctionne à l'aide d'indices laissés par le locuteur à l'interlocuteur pour lui faire comprendre ses intentions par des jeux de parallélismes, de contradictions, d'exagérations et d'hyperboles plus ou moins marqués. Ces indices – souvent pragmatiques ou extralinguistiques – sont plus ou moins évidents, d'où l'importance de la prise en compte du contexte (référentiel, locuteur, interlocuteur...), des connaissances partagées et des normes sociales et culturelles. Comment alors détecter l'ironie dans les propos des auteurs, quand avec des subtilités stylistiques ils indiquent l'opposé de ce qu'ils affirment ? Dans notre approche, nous nous basons uniquement sur les marqueurs textuels observables qui indiquent des valeurs ironiques. Exemple :

De sorte qu'on ne peut mieux définir la Comédie, qu'une « assemblée de railleurs où personne ne se connaît, et où chacun rit des défauts qui les rendent tous également coupables et ridicules ». (Lelevel, 1694)

L'ironie est présente dans plusieurs genres littéraires, tel que la critique littéraire, la satire, l'article de presse, et le roman épistolaire.

3 Le corpus de la Haine du théâtre (HdT)

Le projet HdT² du Labex OBVIL vise à étudier les enjeux des querelles théâtrales en Europe depuis la moitié du 16^e siècle jusqu'aux résurgences contemporaines. Le corpus que nous avons utilisé a été préalablement numérisé et édité dans le cadre de ce projet, il rassemble 59 textes (*environ un million de mots*) écrits en langue française

¹ Action d'interroger en feignant l'ignorance.

² La « Haine du Théâtre », projet dirigé, au sein du Labex OBVIL, par François Lecerclé et Clotilde Thouret : <http://obvil.paris-sorbonne.fr/projets/lahainedutheatre>.

depuis le milieu du 16^e siècle jusque dans les années 1840³. En particulier, cette présente étude s'appuie sur des textes à argumentation théâtrophile ou théâtrophobe, et sur des textes adoptant une stratégie « mesurée ». Cette dernière consiste à dénoncer des abus de la scène pour, ensuite, convaincre le lecteur à préserver l'utilité intrinsèque du théâtre⁴. Ces trois types de textes possèdent une logique souvent détournée et déconcertante pour le lecteur : phrases redondantes, métaphores douteuses et images incongrues mariant l'abstrait au concret se retrouvent dans le vaste corpus de la HdT. L'usage de l'ironie dans les textes de critique autour du théâtre est l'un des moyens employés pour mettre en évidence les dépréciations et les mérites des différents auteurs.

4 Approche technique et ressources linguistiques

L'énonciation ironique est difficile à identifier (Joshi et al., 2016) : que l'on reprenne pour réitérer ou au contraire pour retourner l'argument contre l'adversaire, la machine ne fait pas toujours la différence. En l'absence de ressources linguistiques et de corpus annotés adaptés à nos besoins particuliers, nous nous sommes tournés vers une approche orientée connaissances en nous basant sur d'autres travaux autour de l'annotation automatique des modalités énonciatives (Riguet et Alrahabi, 2017). Cela nous a d'abord permis de réutiliser des ressources linguistiques existantes et de les enrichir par la suite. Il s'agit principalement de marqueurs verbaux comme *se moquer*, *ironiser*, *parodier*... Ensuite, par l'observation du corpus, nous avons enrichi ces ressources par des *substantifs* (*satire*, *dérision*, *raillerie*...), des *adjectifs* (*ridicule*, *piquant*...) et des adverbes. Nous avons ensuite classé ces marqueurs dans des sous-catégories selon différentes nuances sémantiques liées à l'ironie. Ainsi, selon le contexte et les indices complémentaires, certains marqueurs peuvent indiquer clairement l'ironie (*ironique*, *dérision*, *se moquer*, *sarcastique*, *parodier*...). D'autres marqueurs, tout en indiquant l'ironie, sont porteurs d'autres valeurs sémantiques, voici quelques exemples de groupes de marqueurs :

- *faire rire*, *comique*, *pitre*, *grotesque*, *idiot*...
- *chicaner*, *taquiner*, *narguer*...
- *mordant*, *piquant*, *pinçant*, *aigre*...
- *mépriser*, *dénigrer*, *sous-estimer*, *vilipender*...
- *calomnier*, *hypocrisie*, *ruse*, *malice*...

En tout, nous avons collecté autour de 70 marqueurs linguistiques.

Pour l'annotation du corpus, nous avons utilisé Excom2 (Alrahabi, 2010), un outil d'annotation à base de règles qui nous a permis d'avoir le contrôle sur le processus d'annotation et d'améliorer progressivement la pertinence des ressources linguistiques exploitées. Pour ce système, la présence dans une phrase d'un marqueur saillant déclenche les règles associées qui explorent le contexte et vérifient la présence ou l'absence d'indices complémentaires. Les règles dans Excom2 sont organisées selon un ordre de priorité et utilisent en entrée les résultats d'autres règles. Avant l'étape de l'annotation, l'outil procède à la segmentation des textes afin de les découper en sections, paragraphes et phrases. Pour l'ironie, nous avons créé 8 règles que nous avons associées aux différents marqueurs linguistiques.

5 Tests et évaluation

Une première phase de tests sur un échantillon du corpus a été nécessaire pour stabiliser les règles d'identification et de désambiguïsation. Nous nous sommes heurtés à plusieurs difficultés. Au niveau du lexique, peu de changements ont été apportés à nos marqueurs, comme par exemple le mot *satire* qui se trouve avec les deux orthographes *satire* (88 occurrences) et *satyre* (68 occurrences). En français, le dernier désigne le demi-dieu compagnon de Dionysos ou Bacchus. Cependant, dans certains textes qui n'ont pas encore été modernisés, et sont en langue française du 16^e ou 17^e siècle, ce mot indique plus largement la « satire ». Certains marqueurs sont

³ Nous renvoyons à la liste de la bibliographie française qui constitue le corpus total de la HdT : http://obvil.paris-sorbonne.fr/corpus/haine-theatre/bibliographie_querelle-france/

⁴ Cette stratégie de défense double, utilisée depuis le 16^e siècle, consiste à admettre les excès du théâtre afin de protéger la partie extraordinairement utile de cet art.

polysémiques, comme *ridicule* (437 occurrences, le marqueur le plus fréquent), *plaisanter* (176 occurrences) et *comique* (131 occurrences). Concernant la syntaxe du 17^e et 18^e siècle, nous avons observé une certaine complexité au niveau des phrases qui sont parfois très longues (*cinq lignes ou plus*), et au niveau des signes de ponctuation qui ne sont pas stables. Plusieurs virgules, points virgules, *etc.* peuvent en effet se succéder dans une seule phrase. De plus, les auteurs de notre corpus utilisent des tournures complexes. Dans les stratégies polémiques, le discours adverse se trouve présent pour être réfuté ; il faut ainsi être attentif dans l'évaluation de l'ironie dans les différentes séquences.

Afin d'évaluer la qualité des annotations automatiques, nous nous sommes focalisés dans un premier temps sur le calcul de la précision. Nous avons alors annoté avec Excom2 une partie du corpus (7 articles, 215675 mots) et nous avons obtenu 416 annotations. Ensuite, nous avons demandé à une personne experte de vérifier manuellement les sorties selon un guide d'annotation. Pour chaque annotation, l'évaluatrice devait choisir entre deux options : « Correct » ou « Incorrect ». Après la comparaison, nous avons obtenu 93.9 % de précision.

6 Interprétation des résultats

On constate d'abord que tout au long des siècles couverts par le projet Haine du Théâtre (16^e – 19^e siècles), l'usage de l'ironie se situe entre les valeurs de 0,20 à 0,30 % (1265 annotations en total). Nous avons ensuite analysé les marqueurs d'ironie en étudiant leur présence relative selon les siècles et avons pris en compte les marqueurs avec un pourcentage supérieur à 5% (à l'intérieur d'au moins un siècle)⁵.

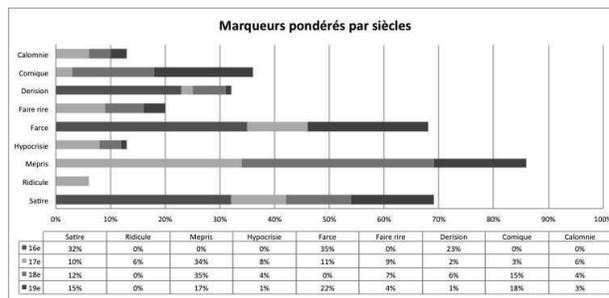


FIGURE 1: Les marqueurs d'ironie dans le corpus HdT pondérés par siècle

De manière générale, l'ironie est utilisée dans le corpus comme procédé éthique et stylistique, ce qui rend les auteurs bien efficaces dans l'élaboration de leur vision de la querelle. Qu'ils soient théâtrophobes ou théâtrophiles, ils peuvent jouer avec les nuances des marqueurs d'ironie, dissimuler un double-sens dans leurs phrases, s'exprimer figurément de manière contraire à ce qu'ils communiquent littéralement.

Par exemple, nous retrouvons une présence considérable du lemme « mépris » au 17^e et 18^e siècles. Il s'agit principalement d'un usage de l'ironie en tant que mécanisme de régulation de la vie sociale. Notamment, Conti et Voisin utilisent un humour inoffensif contre les excès de l'art et mettent en avant la bienséance :

Ceux qui vont aux Spectacles, non par hasard, mais de propos délibéré, et avec tant d'ardeur, qu'ils abandonnent l'Eglise par un mépris insupportable pour y aller, où ils passent tout le jour à regarder ces femmes infâmes, auront-ils l'impudence de dire qu'ils ne les voient pas pour les désirer (Conti 1667, Voisin 1671)

⁵ Signalons que le marqueur « ironie » et toutes ses variantes n'ont que 11 occurrences dans le corpus.

L'« hypocrisie » commence à être utilisée au 17^e et son utilisation se réduit avec le temps (jusqu'à 1% au 19^e). Le lemme en question est essentiellement appliqué à des phrases où l'ironie n'est qu'un « autre nom du malheur » (Martin 2009), une manière de renforcer le point de vue de l'auteur.

L'hypocrisie est un vice privilégié, qui ferme la bouche à tout le monde, et qui jouit en repos d'une impunité souveraine. (Coustel, 1694)

Très répandu dans le corpus est l'usage de l'ironie comme écho satirique. Le lemme « calomnier », présent dans les textes du 17^e au 19^e siècle, en est l'exemple :

[...] cessez de calomnier vos contemporains selon l'usage immémorial de ceux qui profèrent de vaines paroles. (Senancour 1825)

Ces premiers résultats et statistiques nous ont ainsi permis d'effectuer des comparaisons très intéressantes entre les textes des auteurs défenseurs et les textes des auteurs pourfendeurs du théâtre. A partir du nombre d'expressions ironiques correctement identifiées comme telles, nous avons recensé leur nombre et dressé des statistiques pour chaque auteur du petit échantillon.

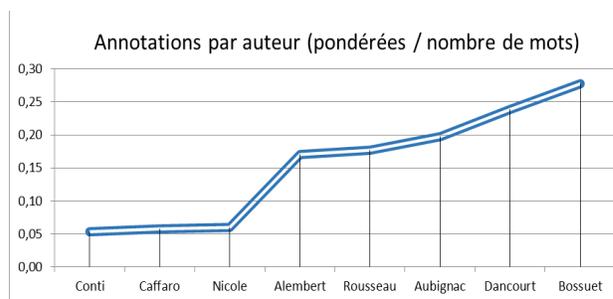


FIGURE 2: Valeurs pondérées de l'annotation de l'ironie dans le corpus

En données relatives, les auteurs qui utilisent le plus les marqueurs d'ironie appartiennent à la « querelle Rousseau » (moitié du 18^e siècle). Cela est à analyser en perspective mais, en l'espèce, dans cet article nous pouvons le mettre en lien avec l'usage de l'ironie au 18^e siècle, comme plusieurs écrits sur Voltaire le témoignent (Loriot, 2015). Les mots de D'Alembert sont très parlants sur ce sujet et éclairent le rôle de l'ironie :

Si la satire et l'injure n'étaient pas aujourd'hui le ton favori de la critique, elle serait plus honorable à ceux qui l'exercent, et plus utile à ceux qui en sont l'objet. (Alembert, 1759)

Les marqueurs linguistiques qui ont été détectés pour cette période appartiennent à la sphère sémantique du *ridicule*, de la *satire*, de la *farce* et du *comique*. D'autres marqueurs verbaux, tels que *se moquer* et *plaisanter* sont présents dans cette querelle et sont communs aux écrits de la précédente controverse datant du milieu du 17^e siècle. Les valeurs ironiques de cette dernière, dont les représentants sont Conti et Nicole, parmi d'autres, sont cependant moins importantes (0,06 vs. 0,17). Outre ces marqueurs verbaux, nous pouvons citer les catégories de substantifs tels que le *ridicule* et le *faire rire*. A la même période, Aubignac, auteur de la stratégie offensive-défensive, part d'une critique du théâtre pour arriver à sa défense. Il s'inspire des marqueurs habituels pour la période du 17^e siècle et reprend dans ses phrases les propos de ses collègues, pour ensuite les réfuter. De plus, il recourt plus spécifiquement à des marqueurs ironiques tel que *railler* et *idiot*. Contemporaine à d'Aubignac, la querelle entre Caffaro et Bossuet nous donne des résultats surprenants : si Caffaro emploie peu de marqueurs relevant de l'ironie (0,05), Bossuet est lui chef de file parmi ses contemporains (valeur de 0,27). Comme les autres auteurs, Bossuet puise dans les marqueurs du *comique* et du *ridicule*, tout comme la forme verbale *plaisanter*. Néanmoins, nous retrouvons dans ses résultats des mots appartenant à la catégorie de marqueurs piquants :

Il ne faut pas s'étonner que l'église ait improuvé en général tout ce genre de plaisirs [les spectacles...] à cause que communément, ainsi que nous l'avons remarqué, par sa bonté et par sa prudence, elle épargne la multitude dans les censures publiques : néanmoins parmi ces défenses, elle jette toujours des traits piquants contre ces sortes de spectacles, pour en détourner tous les fidèles. (Bossuet, 1694)

Nous comprenons ainsi que pour juger le théâtre incompatible avec la morale chrétienne, Bossuet privilégie un style vif et mordant, il appuie l'église tout en dénigrant les défenseurs du théâtre.

La recherche sur les stratégies de la querelle du théâtre, tout en se questionnant sur les modalités argumentatives et les objectifs circonstanciels de chaque auteur, nous dévoile également certaines idées récurrentes autour de la considération du théâtre. Les différents textes partagent un certain nombre de lieux communs, comme par exemple l'idée de perversion, l'inflation temporelle, ou les arguments économiques et politiques.

7 Conclusion et perspectives

Dans cet article, nous avons présenté une approche à base de règles pour la détection automatique de l'ironie dans un corpus de débats sur le théâtre (16^e – 19^e siècle), en réutilisant et enrichissant des ressources linguistiques existantes. La méthode que nous avons adoptée nous a fourni une matière abondante et des données quantitatives pour mieux cerner l'objet d'étude. Vu la particularité du phénomène langagier étudié et la simplicité de notre approche par analyse de surface, nous considérons que ces premiers résultats sont très encourageants (93.9 % de précision). A ce titre, ils méritent d'être approfondis afin d'en tirer le plus grand bénéfice en terme d'exploitation et de précision de résultats. Dans ce sens, nous venons de créer un corpus standard de référence et nous procédons actuellement au calcul du rappel dans l'annotation. Ce corpus nous aidera à améliorer la couverture des marqueurs et à trouver des règles pour identifier les sources des segments annotés (les locuteurs). A long terme, l'un de nos objectifs consiste à annoter les phrases négatives et à analyser leur association avec l'ironie (Mainardi et al., 2015). En effet, la liaison de ces deux types de marqueurs avec d'autres catégories que nous pouvons détecter automatiquement nous permettra de dégager des pistes de recherche inédites.

Références

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, Université Paris-Sorbonne.

Conti A. (1667), *Traité de la comédie et des spectacles*, Paris : L. Billaine.

Costel P. (1694), *Sentiments de l'Eglise et des SS. Pères pour servir de décision sur la comédie et les comédiens*, Paris : Veuve de Charles Coignard.

Joshi A., Bhattacharyya P., Carman M. J., (2016). Automatic Sarcasm Detection: A Survey *ACM Comput. Surv.* V, N, Article A (January 2016).

Lecerclé F., Mainardi C., Thouret C. (2016). Pour une exploration numérique des polémiques sur le théâtre, *RHLF*, n°116 / 4 dir. Didier Alexandre, Littérature et humanités numériques, Paris : PUF.

Lelevel H. (1694), *Réponse à la lettre du théologien, défenseur de la comédie*, Paris : T. Girard.

Loriot C. (2015), *Rire et sourire dans l'opéra-comique en France aux XVIIIe et XIXe siècles*, Lyon : Symétrie.

Mainardi C., Sellami Z., Jolivet V., (2015). A Semantic Exploration Method Based on an Ontology of 17th Century Texts on Theatre: la Haine du Théâtre, *First International Workshop on Semantic Web for Cultural Heritage (SW4CH 2015)*, *New Trends in Databases and Information Systems*, 539, *Communications in Computer and Information Science*, 468-476.

Martin L. (2009), "Le rire est une arme. L'humour et la satire dans la stratégie argumentative du Canard enchaîné", *A contrario* 2009/2 (n° 12), 26-45.

Riguet M., Alrahabi M. (2017), "Pour une analyse automatique du Jugement Critique: les citations modalisées dans le discours littéraire du XIXe siècle", in *DHQ: Digital Humanities Quarterly* 2017

Senancour E de. (1825), De quelques naïves coutumes, in *Le Mercure du dix-neuvième siècle*, t. IX, Paris : Au bureau du Mercure, 262-266.

Voisin J. (1671), *Défense du traité de Mgr le Prince de Conti touchant la comédie et les spectacles ou la réfutation d'un livre intitulé Dissertation sur la condamnation des théâtres*, Paris : Coignard.