Les quanta de transition ou le trouble d'Amphitryon

Stephan Vonfelt

Email: stephan.vonfelt@free.fr - Web: graphoscopie.free.fr

ABSTRACT: This article introduces an intertextual distance according with the rhythm, alternating repetitions and variations. While our prior measurement focused on the return of characters, the last one integrates their transitions. Applied to the classical French theatre, the distance appears to be relevant to the genre, the form and the critic. It doesn't encourage to unite Corneille and Molière, the various work of the latter writer suggesting a rhapsody weaved across Europe.

KEYWORDS: stylometry, author attribution, rhythm, transition, character.

RÉSUMÉ: Cet article introduit une distance intertextuelle accordée sur le rythme, alternance de répétitions et de variations. Alors que notre première mesure se focalisait sur les retours des caractères, la dernière intègre leurs transitions. Appliquée au théâtre classique, la distance se montre pertinente à l'égard du genre, de la forme et de la critique. Elle n'incite guère à unir Corneille et Molière, l'œuvre variée du comédien suggérant une rhapsodie tissée à travers l'Europe.

Mots-clés: stylométrie, attribution d'auteur, rythme, transition, caractère.

1. Introduction

Que le lecteur se rassure, il ne trouvera ici ni physique quantique, ni transition atomique, ni inégalité d'Heisenberg. Ce modèle ténu et lointain a cependant inspiré notre modeste recherche : un texte est vu comme une succession d'atomes – les lettres – transitant d'un état à un autre selon des aléas que nous n'expliquerons pas.

La méthode est appliquée à un corpus aussi classique que controversé, le théâtre de Corneille, Molière et Racine. Corneille est-il la plume d'ombre de Molière? À l'aide d'une distance intertextuelle qui généralise celle que nous avons expérimentée jusqu'ici, nous comparerons ces trois auteurs, et tenterons d'attribuer leurs pièces.

2. Distance intertextuelle

Le premier choix est celui des unités linguistiques: vocables lemmatisés, catégories grammaticales, sons élémentaires, sont privilégiés par les études. Mais ces variables de seconde main impliquent un étiquetage préalable, parfois douteux. Fondamentalement, un texte reste une succession de caractères, éléments tangibles et abondants, sources de statistiques fiables et significatives. Dans cette veine, citons Markov et ses processus aléatoires, Shannon et sa théorie de l'information, Brunet et son étude du vocabulaire français, Khmelev & Tweedie puis Jardino pour l'attribution d'auteur¹. Plus récemment, notre thèse a constaté l'efficacité des caractères par rapport aux catégories grammaticales et aux vocables². Notre première étude sur le corpus de Corneille, Molière et Racine suit cette voie étroite³.

De ces unités, les statistiques textuelles retiennent traditionnellement les fréquences d'apparition : le texte est appréhendé comme un sac dont on pèse les billes, les yeux fermés sur l'agencement. Pour pallier cette lacune, une première méthode consiste à segmenter un texte en morceaux, puis à suivre les évolutions des fréquences. Mais le découpage est souvent délicat, a fortiori la comparaison de deux architectures différentes, ainsi que le soulignent Longrée, Luong & Mellet⁴. Prolongeant la marche vers l'infiniment petit, nous adoptons une division élémentaire qui suit le fil ténu de chaque occurrence. Voyant dans le rythme le fondement de l'art, notre première approche se concentrait sur les temps de retour d'une unité. Or le rythme est fait tant de variation que de répétition, d'où la généralisation proposée par cet article, intégrant les temps de transition.

2.1 Caractérisation d'un texte

Pour fixer les idées, soit la séquence de lettres ABBAABBA, dont nous retenons les temps de transition T entre les lettres A et B. Dans la diagonale de la matrice ci-dessous se retrouvent les temps de retour : si n_{α} désigne la taille de l'alphabet, n_{α}^2 cases sont analysées, contre n_{α} antérieurement.

Т	A	В
A	3, 1, 3	1,1
В	1,1	1, 3, 1

Parmi les transitions du caractère i au caractère j, les temps successifs T_{ij} ne présentent généralement aucune corrélation, et se résument à leur distribution $P_{ij}^{\ 5}$. Ces dernières présentant un aspect chaotique, nous recourons aux répartitions F_{ij} qui les lissent par intégration, soit : $F_{ij}(t) = P_{ij}(T_{ij} \le t)$.

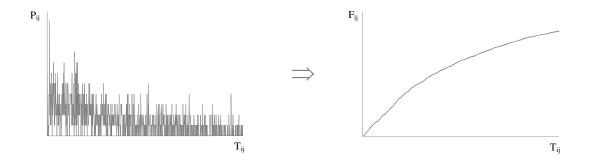
 4 Longrée D, Luong X., Mellet S. (2004).

¹ MARKOV A. (1913), SHANNON C.E. (1951), BRUNET E. (1981), KHMELEV D. & TWEEDIE F.J. (2001), JARDINO M. (2006).

² VONFELT S. (2008), chapitre 8.

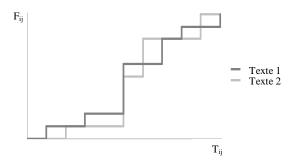
³ VONFELT S. (2009).

⁵ Sur le corpus de notre thèse, les corrélations sont nulles dans le cas des lettres ; légèrement perceptibles pour les espaces et la ponctuation, elles confondent les œuvres et sont donc inadaptées à notre problématique : Vonfelt S. (2008), chapitre7.



2.2 Comparaison de deux textes

La comparaison porte dans un premier temps sur les transitions T_{ij} : suivant les valeurs discrètes et croissantes des temps de transition T_{ijk} , les écarts quadratiques des répartitions entre deux textes sont pondérés par les effectifs n_{ijk} : $d_{12ij}^2 = \Sigma_k (n_{1ijk} + n_{2ijk}) \left(F(T_{1ijk}) - F(T_{2ijk}) \right)^2 / (n_{1ij} + n_{2ij})$.



Ces écarts particuliers sont intégrés sur l'ensemble des transitions pour définir une distance intertextuelle : $d_{12}{}^2 = \Sigma_i \Sigma_j \left(n_{1ij} + n_{2ij} \right) \left(d_{12ij}{}^2 / \left(n_1 + n_2 \right) = \Sigma_i \Sigma_j \Sigma_k \left(n_{1ijk} + n_{2ijk} \right) \left(F(T_{1ijk}) - F(T_{2ijk}) \right)^2 / \left(n_1 + n_2 \right)$.

Moyenne des écarts quadratiques entre les répartitions, d est comprise entre 0 et 1. C'est une pseudo distance⁶ : $d_{11} = 0$, $d_{12} = d_{21}$, et $d_{13} \le d_{12} + d_{23}$ d'après l'inégalité de Cauchy-Schwarz. Formellement, nous noterons : $d_{12} = |F_1 - F_2|$.

La longueur est une caractéristique importante d'un texte : sous le joug du nombre, les écrits des uns et des autres se rendent à la langue commune. La distance décroit naturellement avec les effectifs⁷ : si $n = (n_1^{-1} + n_2^{-1})^{-1}$, d est de l'ordre de $n^{-0.5}$.

_

⁶ La nullité de la distance n'implique pas l'identité des textes : sans la propriété de séparation, cette mesure n'est pas une vraie distance, mais la limitation est ici sans incidence.

⁷ Dans le cas de variables continues, si deux échantillons de tailles n_1 et n_2 sont tirés d'une même population, $(n_1^{-1}+n_2^{-1})^{-1}$ d' tend vers une loi indépendante de cette population : ANDERSON T.W. (1962). Nous admettrons ce gabarit pour des variables discrètes.

3. Application à l'affaire Molière

3.1 Historique

Dans un article de la revue *Le Temps* paru en 1919, le poète Louÿs attribua l'*Amphitryon* de Molière à Corneille : « Il est évident que Pierre Corneille domine toute la vie de Molière, qu'il a collaboré à plusieurs de ses pièces, et que l'une d'elle, Amphitryon, est toute entière de sa plume si l'on néglige quelques "interruptions" très faciles à détacher et quelques rares fragments de scènes ». Quelques semaines plus tard, dans un second article paru dans la revue *Comædia*, il généralisa son postulat à l'ensemble de l'œuvre de Molière⁸.

Si la thèse fut reprise par certains romanciers, elle prit une dimension nouvelle lorsque Labbé l'éprouva par les statistiques textuelles⁹. Basé sur le vocabulaire, son travail rejoignait l'intuition du poète. De nombreuses réactions suivirent, contestant ou modérant les conclusions du chercheur ; face à la polémique, une table ronde permit de conforter les points de vue¹⁰. Quelques années plus tard, notre étude fondée sur les caractères n'engageait guère à rapprocher Molière de Corneille.

3.2 Corpus

Outre Corneille et Molière, le corpus inclut Racine, témoin de la comparaison.

Les 33 pièces de Corneille remontent entre 1630 et 1674; principalement tragiques, elles sont formées de vers. Les 32 pièces de Molière sont datées entre 1654 et 1673; franchement comiques, elles favorisent quelque peu la prose. Enfin les 12 pièces de Racine se placent entre 1664 et 1691; résolument tragiques, elles sont écrites en vers¹¹.

La liste des pièces, selon la chronologie, le genre et la forme, est donnée en annexe 2.

3.3 Traitements linguistiques

Le corpus numérisé a été aimablement fourni par Dominique Labbé. À partir de la version de la base *Frantext*, celui-ci a réalisé quelques corrections et normalisations orthographiques sans incidence statistique.

Les indications de scènes et de personnages, isolées par des balises, sont passées sous silence pour se limiter au texte dramatique. Afin de s'affranchir de la mise en page, les retours à la ligne sont éliminés. Chaque œuvre se voit ramenée à sa plus simple expression, une suite de caractères.

L'intégralité des unités est lue et mémorisée : les espaces, la ponctuation, les lettres en distinguant minuscules, majuscules et caractères accentués, ainsi que les apostrophes et les traits d'union.

⁸ Louÿs P. (1919).

⁹ LABBÉ C. & LABBÉ D. (2001).

¹⁰ Brunet E. (2004), Viprey J.M. (2004), Baudouin V. & Yvon F. (2004), Table-ronde (2004).

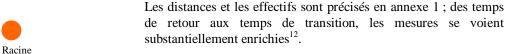
¹¹ 71 % des pièces de Corneille sont tragiques, 100 % sont en vers ; 97 % des pièces de Molière sont comiques, 53 % sont en prose ; 92 % des pièces de Racine sont tragiques, 100 % sont en vers.

3.4 Mesures

3.4.1 Carte des auteurs



Les auteurs sont situés par les distances entre leurs œuvres, soit Corneille-Racine (2.9 %), Corneille-Molière (3.4 %), Molière-Racine (3.7 %). Trois points formant un plan, ils peuvent être représentés sur une carte.



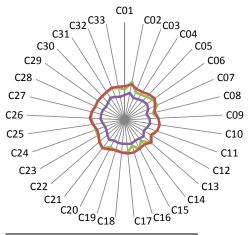
Corneille et Racine sont les auteurs les plus proches, unis par le genre tragique et la forme versifiée. A contrario, Molière prend ses distances avec ce couple primordial et s'oppose notamment à Racine.

Le mariage pressenti par Louÿs entre Corneille et Molière se présente sous de sombres auspices. Quant à la méthode, cette distance apparaît plus réaliste que la précédente, qui renvoyait les auteurs vers un triangle uniforme.

3.4.2 Attributions des pièces

La stylométrie permet-elle de cerner un auteur ? Sans débattre ici, une pièce est attribuée à l'œuvre la plus proche, par la comparaison d'une distance interne et de deux distances externes. Afin que le jeu soit équitable, la pièce en question est boutée hors de l'œuvre qui la contient.

Pour chaque auteur et chaque pièce, les graphes tracent les distances entre cette pièce et l'œuvre de Corneille, de Molière, et de Racine. Les valeurs des distances et les codes des pièces sont précisés en annexe 2.

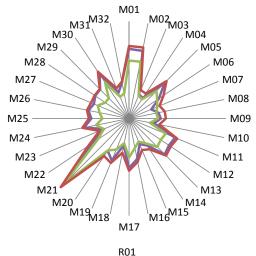


Le tour de Corneille est réglé et serré.

Toutes les pièces lui sont attribuées, mais l'affectation hésite pour *Le Menteur* (C14) et sa *Suite* (C16) : la comédie a manifestement marqué Molière¹³, d'où ce rapprochement déjà opéré par Labbé sur le plan sémantique.

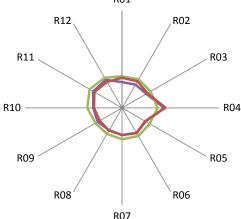
¹² La relation entre le nombre de transitions et de retours est complexe ; le ratio croît depuis l'origine pour se stabiliser vers 12 quand les effectifs tendent vers l'infini.

¹³ « Je dois beaucoup au Menteur » : Molière rapporté par Boileau, dans FRANÇOIS DE NEUFCHÂTEAU N. (1819), p. 149.



La course de Molière est vagabonde : la burlesque et brève *Pastorale comique* (M21) se détache visiblement ; non datées, les farces de *La Jalousie du Barbouillé* (M01) et du *Médecin volant* (M02) ressortent dans une moindre mesure.

Dom Garcie de Navarre (M07), comédie héroïque en vers, avatar du Don Sanche d'Aragon (C20), est logiquement donnée à Corneille; plus troublante est l'attribution d'Amphitryon (M23) au même Corneille, réhabilitant l'intuition de Louÿs.



Le chemin de Racine renoue avec la solidité.

Précoces, La Thébaïde (R01), Alexandre Le Grand (R02), Andromaque (R03), sont attribués à leur modèle Corneille ; l'unique comédie des Plaideurs (R04) est donnée à Molière ; par excellence cornélien, Mithridate (R08) est assigné à Corneille.

91 % des pièces sont attribuées à leur père naturel, et les rares transfuges se comptent d'abord parmi les tragédiens : quatre entre Corneille et Racine, contre deux entre Molière et Corneille, et un entre Molière et Racine, à l'image des distances séparant ces auteurs. Par comparaison, les temps de retour marquaient seulement deux croisements entre les tragédiens, avec les mêmes parités ailleurs. Les différences mineures traduisent le rapprochement déjà observé entre Corneille et Racine. La tripartition apparaît authentique, et la distance réaliste.

4. Conclusion

D'un point de vue théorique, la distance intertextuelle épouse fidèlement le rythme imprimé par les caractères : par rapport aux temps de retour, les temps de transition enrichissent substantiellement la mesure. Dans la pratique, les distances obtenues sont pertinentes à l'égard du genre, de la forme et de la critique théâtrale.

Quant à la question moliéresque, les mesures d'ensemble et de rares accointances n'incitent pas à unir le comédien à Corneille. Si un rapprochement devait être fait, il joindrait ce dernier à Racine, conformément au genre tragique et à la forme versifiée. La variété de l'œuvre de Molière suggère plutôt une rhapsodie tissée à travers l'Europe¹⁴.

_

¹⁴ Critique émise du vivant de Molière par Donneau de Visé; sur les influences françaises, italiennes est espagnoles, voir FORESTIER G. (2003).

Références

- And Annals of Mathematical Statistics, Beachwood, Institute of Mathematical Statistics, vol. 33, n° 3, p. 1148-1159.
- ASSOCIATION CORNÉLIENNE DE FRANCE, L'affaire Corneille-Molière, www.corneille-moliere.org.
- BAUDOUIN V. & YVON F. (2004), « Contribution de la métrique à la stylométrie », Actes des Journées internationales d'Analyse statistique des Données Textuelles, Louvain, Presses Universitaires de Louvain.
- Brunet E. (1981), Le vocabulaire français de 1789 à nos jours, Paris, Champion.
- BRUNET E. (2004), « Où l'on mesure la distance entre deux textes », *Texto!*, www.revue-texto.net, rubrique « Dits et inédits ».
- FORESTIER G. (2003), « D'un vrai canular à une fausse découverte scientifique à propos des travaux de Dominique et Cyril Labbé », *Centre de Recherche sur l'Histoire du Théâtre*, www.crht.org, rubrique « Ressources / Dossiers ».
- François de Neufchâteau N. (1819), L'esprit du grand Corneille, Paris, Didot.
- JARDINO M. (2006), « Identification des auteurs de textes courts avec des n-grammes de caractères », *Actes des Journées internationales d'Analyse statistique des Données Textuelles*, Besançon, Presses Universitaires de Franche-Comté.
- KHMELEV D. & TWEEDIE F.J. (2001), « Using Markov Chains for Identification of Writers », *Literary and Linguistics Computing*, Oxford, Oxford University Press, vol. 16, n° 4, p. 299-307.
- LABBÉ C. & LABBÉ D. (2001), « Intertextual Distance and Authorship Attribution : Corneille and Molière », *Journal of Quantitative Linguistics*, London, Routledge, vol. 8, n° 3, p. 213-231.
- LONGRÉE D., LUONG X., MELLET S. (2004), « Temps verbaux, axe syntagmatique, topologie textuelle : analyse d'un corpus lemmatisé », *Actes des Journées internationales d'Analyse statistique des Données Textuelles*, Louvain, Presses Universitaires de Louvain.
- LOUŸS P. (1919), « L'auteur d'Amphitryon », Le Temps, Paris.
- LOUŸS P. (1919), « Molière est un chef d'œuvre de Corneille », Comædia, Paris.
- MARKOV A. (1913), « Un exemple de recherche statistique sur le texte d'Eugène Onéguine illustrant la liaison des épreuves en chaînes », *Bulletin de l'Académie Impériale des Sciences*, vol. 7, Saint-Pétersbourg, p. 153-162.
- SHANNON C.E. (1951), « Prediction and Entropy of Printed English », *Bell Systems Technical Journal*, Hoboken, Wiley, vol. 30, p. 50-64.
- TABLE-RONDE (2004), « Corneille et Molière », Actes des Journées internationales d'Analyse statistique des Données Textuelles, Louvain, Presses Universitaires de Louvain.
- VIPREY J.M. (2004), « Analyse séquencée de la micro-distribution lexicale », *Actes des Journées internationales d'Analyse statistique des Données Textuelles*, Louvain, Presses Universitaires de Louvain.
- VONFELT S. (2008), La musique des lettres : variations sur Yourcenar, Tournier et Le Clézio, Université de Toulouse, thèse.
- VONFELT S. (2009), « Le graphonaute ou Molière retrouvé », Lexicometrica, 2009.

Annexes

1 Auteurs

	d (%)	n (transitions)	n (retours)
Molière-Racine	3,683	34 597 274	2 759 721
Corneille-Racine	2,867	46 948 486	3 708 407
Corneille-Molière	3,367	58 729 392	4 694 940

2 Pièces

Auteur	Titre	Date	Genre	Forme	Code	d _C (%)	d _M (%)	d _R (%)
	Mélite	1630	comédie	vers	C01	3.953	4.829	5.339
	Clitandre	1631	tragédie	vers	C02	4.341	6.139	5.712
	La Veuve	1632	comédie	vers	C03	3.726	4.689	5.248
	La Galerie du Palais	1633	comédie	vers	C04	3.897	4.770	5.241
	La Suivante	1634	comédie	vers	C05	4.146	5.075	5.592
	La Place Royale	1634	comédie	vers	C06	4.321	5.485	5.853
	Médée	1635	tragédie	vers	C07	4.124	5.688	5.264
	L'illusion comique	1636	comédie	vers	C08	3.565	4.577	4.924
	Le Cid	1636	tragédie	vers	C09	3.978	5.338	5.369
	Horace	1640	tragédie	vers	C10	4.073	5.623	5.606
	Cinna	1641	tragédie	vers	C11	3.758	5.432	5.411
	Polyeucte	1643	tragédie	vers	C12	3.483	4.951	4.794
	Pompée	1644	tragédie	vers	C13	3.979	5.692	5.496
	Le Menteur	1644	comédie	vers	C14	3.759	4.160	4.997
	Rodogune	1644	tragédie	vers	C15	3.684	5.237	5.140
Corneille	La Suite du Menteur	1645	comédie	vers	C16	3.925	4.256	5.363
	Théodore	1646	tragédie	vers	C17	3.832	5.164	5.362
	Héraclius	1647	tragédie	vers	C18	3.805	5.200	5.205
	Andromède	1650	tragédie	vers	C19	3.783	5.133	4.983
	Don Sanche d'Aragon	1650	comédie héroïque	vers	C20	3.681	4.844	5.274
	Nicomède	1651	tragédie	vers	C21	4.019	5.685	5.708
	Pertharite	1652	tragédie	vers	C22	3.622	5.207	5.353
	Œdipe	1659	tragédie	vers	C23	3.340	4.931	4.770
	La Toison d'or	1660	tragédie	vers	C24	3.275	4.816	4.640
	Sertorius	1662	tragédie	vers	C25	3.527	5.038	5.084
	Sophonisbe	1663	tragédie	vers	C26	3.840	5.334	5.537
	Othon	1664	tragédie	vers	C27	3.835	5.317	5.386
	Agésilas	1666	tragédie	vers	C28	3.781	5.043	5.134
	Attila	1667	tragédie	vers	C29	3.608	5.227	5.215
	Tite et Bérénice	1670	comédie héroïque	vers	C30	3.787	5.117	5.180
	Psyché	1671	tragi-comédie	vers	C31	4.181	5.201	5.284
	Pulchérie	1672	tragédie	vers	C32	3.920	5.362	5.503
	Suréna	1674	tragédie	vers	C33	3.731	5.167	5.350
	La Jalousie du Barbouillé	?	comédie	prose	M01	11.118	9.238	11.580
	Le Médecin volant	?	comédie	prose	M02	11.000	9.250	11.549

1	L'Étourdi	1654	comédie	vers	M03	4.161	3.957	4.910
	Le Dépit amoureux	1656	comédie	vers	M04	3.966	3.698	4.798
	Les Précieuses ridicules	1659	comédie	prose	M05	8.053	6.206	8.460
	Sganarelle	1660	comédie	vers	M06	5.744	5.398	6.635
	Dom Garcie de Navarre	1661	comédie héroïque	vers	M07	3.842	4.870	5.275
	L'École des Maris	1661	comédie	vers	M08	4.789	4.445	5.824
	Les Fâcheux	1661	comédie	vers	M09	5.080	4.625	5.870
	L'École des Femmes	1662	comédie	vers	M10	4.042	3.419	4.934
	La Critique de l'École des Femmes	1663	comédie	prose	M11	7.950	6.347	8.349
	L'Impromptu de Versailles	1663	comédie	prose	M12	7.679	6.100	8.012
	Le Mariage forcé	1664	comédie	prose	M13	8.067	6.374	8.369
	La Princesse d'Élide	1664	comédie	vers/prose	M14	5.167	4.737	5.891
	Le Tartuffe	1664	comédie	vers	M15	4.420	3.988	5.417
Molière	Dom Juan	1665	comédie	prose	M16	6.576	4.784	7.047
	L'Amour médecin	1665	comédie	prose	M17	7.865	6.208	8.328
	Le Misanthrope	1666	comédie	vers	M18	4.389	4.028	5.547
	Le Médecin malgré lui	1666	comédie	prose	M19	7.233	5.358	7.676
	Mélicerte	1666	comédie héroïque	vers	M20	6.187	6.098	7.182
	Pastorale comique	1667	comédie	vers	M21	15.134	14.908	15.476
	Le Sicilien	1667	comédie	vers/prose	M22	7.257	6.037	7.883
	Amphitryon	1668	comédie	vers	M23	4.359	4.380	5.128
	George Dandin	1668	comédie	prose	M24	7.317	5.387	7.625
	L'Avare	1668	comédie	prose	M25	6.048	4.288	6.429
	Monsieur de Pourceaugnac	1669	comédie	prose/vers	M26	6.636	4.837	6.873
	Les Amants magnifiques	1670	comédie	vers/prose	M27	6.104	5.415	6.600
	Le Bourgeois gentilhomme	1670	comédie	prose	M28	6.510	4.575	6.641
	Les Fourberies de Scapin	1671	comédie	prose	M29	6.558	4.612	6.650
	La Comtesse d'Escarbagnas	1671	comédie	prose	M30	8.500	6.932	8.985
	Les Femmes savantes	1672	comédie	vers	M31	4.431	3.745	5.245
	Le Malade imaginaire	1673	comédie	prose	M32	5.872	4.093	5.983
Racine	La Thébaïde	1664	tragédie	vers	R01	4.102	5.053	4.801
	Alexandre le Grand	1665	tragédie	vers	R02	4.280	5.263	4.905
	Andromaque	1667	tragédie	vers	R03	4.481	5.240	4.685
	Les Plaideurs	1668	comédie	vers	R04	6.720	5.660	6.933
	Britannicus	1669	tragédie	vers	R05	4.354	5.092	4.258
	Bérénice	1670	tragédie	vers	R06	4.671	5.203	4.618
	Bajazet	1672	tragédie	vers	R07	4.357	5.019	4.355
	Mithridate	1673	tragédie	vers	R08	4.108	4.814	4.194
	Iphigénie	1674	tragédie	vers	R09	4.488	4.912	4.104
	Phèdre	1677	tragédie	vers	R10	4.597	5.550	4.482
	Esther	1689	tragédie	vers	R11	5.310	5.966	5.076
	Athalie	1691	tragédie	vers	R12	5.182	5.538	4.780