

Les Sganarelle de Molière : un nom, des syntaxes ? [en]

Élodie Bénard <elodiebenard27_at_gmail_dot_com>, Université Paris-Sorbonne

Francesca Frontini <francesca_dot_frontini_at_univ-montp3_dot_fr>, Université Paul Valéry Montpellier

Abstract

Les études quantitatives consacrées à la syntaxe dans le théâtre de Molière sont rares. Cet article propose une analyse syntaxique du discours des personnages de Molière qui repose sur l'extraction de séquences de catégories grammaticales (ou parties du discours) et leur tri et filtrage grâce à l'analyse des correspondances. Il s'agit ici d'étudier l'évolution du personnage de Sganarelle, qui apparaît dans plusieurs pièces de Molière, en vers et en prose, et qui représente souvent le « bourgeois de Paris », dont les valeurs vont à l'encontre de celles que partagent les mondains. L'analyse des résultats est menée du point de vue de ce que D. Biber et S. Conrad appellent la *register perspective*, afin d'identifier les séquences syntaxiques qui caractérisent les différents Sganarelle en fonction de la situation de communication dans laquelle s'inscrit leur discours. Les résultats montrent les différentes façons dont Molière exploite le personnage à mesure que s'affirme le nouveau comique qu'il met en œuvre, un comique fondé sur le jeu avec les valeurs de la société mondaine.

Introduction

La bibliographie sur Molière et son œuvre est immense, mais les travaux spécifiquement consacrés à la syntaxe de l'auteur sont peu nombreux. Pierre Larthomas, dans son ouvrage consacré au langage dramatique, cherche à apprécier l'efficacité du style dramaturgique, en particulier celle de Molière, en tenant compte des spécificités du genre, notamment du compromis nécessaire entre la langue écrite et la langue parlée [Larthomas 1972/2001]. Son analyse montre que c'est moins sur le plan du lexique que sur le plan de la syntaxe que s'opère le compromis entre les deux langues : au théâtre, le lexique peut avoir la complexité caractéristique de la langue écrite, s'il est justifié par la situation dramatique et qu'il est intégré dans une construction syntaxique qui soit « suffisamment "parlée" » [Larthomas 1972/2001, 184]. La vraisemblance se situe au niveau de la syntaxe, d'où la place qu'elle se voit accorder dans l'analyse du style. L'étude de [Conesa 1983] envisage le dialogue moliéresque du point de vue de sa macrostructure et de sa microstructure. La syntaxe est l'objet d'une attention particulière en raison de son rôle fondamental dans l'attaque du dialogue, l'enchaînement des répliques et la modification des rapports de force entre les interlocuteurs ; de la syntaxe naissent aussi les effets rythmiques et mélodiques caractéristiques de l'écriture de Molière. Dans ces deux ouvrages, l'analyse porte sur un nombre limité de séquences syntaxiques sélectionnées par l'intuition et l'intelligence du chercheur, parce qu'elles donnent à voir des phénomènes caractéristiques du dialogue chez Molière. Parallèlement à ce type d'études, le théâtre classique a fait l'objet d'une série de travaux relevant de la linguistique computationnelle, initiée par l'étude de statistique lexicale menée par [Muller 1967] sur les pièces de Corneille. Cette approche, qui cherche à évaluer la richesse lexicale ou la fréquence des classes grammaticales, a inspiré les recherches de [Bernet 1983] sur le vocabulaire des tragédies de Racine et de [Kylander 1995] sur les pièces en vers de Molière^[1]. Si, dans ces études, les mots sont analysés sur le plan grammatical, ils sont considérés isolément et non pas par séquences (*n-grams*). Les méthodes statistiques et automatiques développées plus récemment permettent d'analyser quantitativement des séquences syntaxiques, composées soit d'une suite d'étiquettes grammaticales – c'est le cas dans cette étude –, soit d'éléments dont la nature est mixte, des formes fixes, des lemmes et des catégories morpho-syntaxiques [Quiniou, Cellier, Charnois et Legallois 2012]. Elles suscitent notamment des travaux qui revisitent les interprétations et lectures de la tradition critique à l'aune de ces nouvelles approches. Ainsi, dans son étude sur Racine,

Christof Schöch [Schöch 2016] s'est proposé d'examiner le fameux « effet de sourdine » décrit par Leo Spitzer en mesurant automatiquement les *stylistic patterns* associés à cet effet.

La méthode utilisée dans cette étude repose sur l'extraction automatique de séquences syntaxiques appartenant à différents textes et sur leur comparaison grâce à l'analyse des correspondances [Benzécri 1982]. Utile pour l'étude du style d'un auteur, cette méthode a été expérimentée pour l'analyse de romans, afin de comparer les propriétés syntaxiques du style de quatre écrivains français du ^{xix}^e s., Balzac, Hugo, Flaubert et Zola [Frontini, Boukhaled et Ganascia 2015b]. Elle s'est révélée également efficace quant à la caractérisation des personnages et de leur fonction dramatique. Les propriétés syntaxiques des discours de différents personnages créés par Molière ont ainsi été analysées dans une perspective communicationnelle [Frontini, Boukhaled et Ganascia 2015a] [Frontini, Boukhaled et Ganascia 2017]. En effet, comme le rappelle Larthomas, la parole au théâtre, qu'elle soit en vers ou en prose, est avant tout un acte de communication ; aussi la syntaxe est-elle déterminée par la situation^[2]. Dans cette optique – Biber et Conrad parlent, pour leur part, de *register perspective*^[3] –, il s'agit de mettre en relation les traits linguistiques d'un texte donné et ses propriétés socio-pragmatiques, c'est-à-dire son contexte d'énonciation et les fonctions de communication qu'il remplit [Biber et Conrad 2009].

Nous nous sommes intéressées, dans ce travail, aux personnages de Molière qui portent le nom « Sganarelle ». Ils apparaissent dans sept pièces, représentées entre 1660 et 1666 : *Sganarelle ou le Cocu imaginaire* (1660), *L'École des maris* (1661), *Le Mariage forcé* (1664), *Le Médecin volant*^[4] (1664), *Le Festin de Pierre* (1665), *L'Amour médecin* (1665) et *Le Médecin malgré lui* (1666^[5]). Comme les personnages de Jodelet et de Mascarille^[6], Sganarelle représente un type dans le répertoire comique. Après avoir réuni Mascarille et Jodelet dans *Les Précieuses ridicules* (1659), Molière invente, dans *Sganarelle ou le Cocu imaginaire* (1660), un nouveau personnage : « celui du bon bourgeois anti-galant, fort de ses prérogatives de mari et en même temps persuadé de l'infériorité et de la légèreté des femmes, qu'il va soutenir à travers deux Sganarelle (celui du *Cocu* et celui de *L'École des maris*) jusqu'à Arnolphe de *L'École des femmes*^[7]. » Georges Forestier et Claude Bourqui ont mis au jour une continuité entre des pièces qui, par-delà leur diversité (comédies/comédies-ballets, petites comédies/comédies en cinq actes, pièces en vers/en prose), sont quasiment toutes traversées, à des degrés divers, par l'idéologie et la culture mondaines. Dans les deux comédies-ballets, *Le Mariage forcé* et *L'Amour médecin*, le comportement des Sganarelle est dénoncé au nom de l'idéologie mondaine : l'un souhaitant se marier à la condition d'avoir un contrôle absolu sur son épouse et redoutant plus que tout une femme émancipée qui pourrait le tromper ; l'autre faisant primer la conservation de son bien sur le droit individuel au désir, au plaisir et au bonheur. Cependant, ils ne sont pas amenés à exprimer et à justifier leurs idées anti-galantes de façon aussi approfondie que les Sganarelle du *Cocu imaginaire* et de *L'École des maris*. On peut en dire autant du héros du *Médecin malgré lui* : par son autoritarisme et sa brutalité envers sa femme, il incarne le mari anti-galant, mais sa conception du mariage et du statut de la femme est formulée de manière très sommaire. En revanche, du point de vue axiologique, les Sganarelle du *Festin de Pierre* et du *Médecin volant* ne sont pas les héritiers du Sganarelle du *Cocu imaginaire* : ce sont des valets, à la fois naïfs et ingénieux, non pas des bourgeois rétrogrades. On voit alors se dessiner des lignes de partage déterminées par le type de comédie. Les comédies en vers du *Cocu imaginaire* et de *L'École des maris* témoignent des premiers efforts de Molière pour élever la petite comédie, en y intégrant les débats agités dans les salons. Le lien étroit entre le discours de Sganarelle et les valeurs de la cour et de la ville – qui le distingue de Mascarille et de Jodelet – introduit en effet un comique inédit, fondé sur la connivence avec le public, inhérente à la création littéraire mondaine^[8]. Ces tentatives sont rapidement suivies par la création de comédies en cinq actes où la complexité du personnage ridicule va croissant. Dès lors, Sganarelle laisse place à Arnolphe (*L'École des femmes*), Harpagon (*L'Avare*) ou Orgon (*Le Tartuffe*^[9]). Aussi, après 1663 et *L'École des femmes*, le nom « Sganarelle » est-il réservé à des personnages qui s'inscrivent dans la lignée du *Cocu imaginaire*, mais dont le caractère n'est pas développé en raison de la nature de la comédie – comédies-ballets du *Mariage forcé* et de *L'Amour médecin* et petite comédie du faux médecin (*Le Médecin malgré lui*) –, ou à des personnages qui s'en écartent et portent ce nom uniquement parce qu'ils sont joués par Molière (*Le Festin de Pierre* et *Le Médecin volant*^[10]).

Il existe aussi des ressemblances entre les Sganarelle du point de vue de l'intrigue. Plusieurs souhaitent un mariage

inapproprié, soit pour eux (*L'École des maris* et *Le Mariage forcé*), soit pour leurs enfants (*L'Amour médecin*). Deux d'entre eux ont un rôle de « solliciteur » : l'un prend conseil pour savoir s'il doit se marier ou non (*Le Mariage forcé*) et l'autre, pour trouver un remède à la maladie de sa fille (*L'Amour médecin*). *Le Médecin volant* et *Le Médecin malgré lui* ont également en commun l'épisode de l'imposture médicale – le personnage de Sganarelle jouant le rôle d'un médecin pour faire obstacle à un mariage^[11].

L'enjeu de notre étude est de vérifier si, lorsque les Sganarelle sont considérés du point de vue de la syntaxe de leur discours et que l'on se fonde sur des techniques d'extraction automatique, les distinctions entre les différents personnages – qui reposent tantôt sur l'intrigue (bourgeois cherchant à se marier ou à marier leurs enfants vs autres ; faux médecins vs autres), tantôt sur la forme de la comédie (petites comédies en vers centrées sur l'antagonisme des valeurs vs comédies-ballets et petites comédies dont l'efficacité comique repose sur d'autres ressorts que le conflit des valeurs) – se maintiennent ou si d'autres lignes de partage émergent. Il s'agit de voir, par cette analyse exploratoire, si ces méthodes algorithmiques, qui se sont révélées efficaces pour vérifier la caractérisation des personnages ou la fonction dramatique qui leur est assignée, permettent une analyse critique plus approfondie.

5

Résultats quantitatifs

La comparaison des différents Sganarelle est fondée sur les traits syntaxiques caractéristiques associés à chacun d'entre eux. La méthode^[12] comporte cinq étapes :

6

1. Étiquetage des textes ;
2. Extraction des séquences syntaxiques ;
3. Filtrage des séquences syntaxiques ;
4. Analyse des correspondances ;
5. Analyse des occurrences.

Les sept textes, comprenant chacun l'ensemble des répliques attribuées à Sganarelle dans une pièce, sont analysés, dans un premier temps, à l'aide de l'étiqueteur TreeTagger ([Schmid 1995]; [Stein 2003], pour les étiquettes spécifiques au français^[13]). Ils sont découpés en phrases et, dans chacune des phrases, chaque partie discours (*part of speech*, ou *PoS* : mot, ponctuation) se voit attribuer une des étiquettes syntaxiques suivantes (*PoS tag*) :

TAG	DESCRIPTION
ABR	Abréviation
ADJ	Adjectif
ADV	Adverbe
DET	Déterminant
INT	Interjection
KON	Conjonction
NAM	Nom propre
NUM	Chiffre
PRO	Pronom
PRP	Préposition
PUN	Ponctuation
SENT	Ponctuation de fin de phrase
VER	Verbe

Table 1.

La phrase « Le livre est sur la table. » est étiquetée de la façon suivante : <DET><NOM><VER><PRP><DET><NOM><SENT>. Après quoi, des séquences syntaxiques de longueur donnée (*PoS n-grams*) sont extraites de chaque texte^[14] et la fréquence de chaque séquence syntaxique est mesurée. Si l'on choisit une fenêtre de 2 à 4, on peut extraire dans la phrase choisie en exemple des séquences de longueur 2, 3 ou 4, comme :

- (<DET><NOM>) : deux occurrences, « Le livre » et « la table » ;
- (<DET><NOM><VER>) : une occurrence, « Le livre est » ;
- (<DET><NOM><VER><PRP>) : une occurrence, « livre est sur la », etc.

Les séquences syntaxiques les plus utiles pour les besoins de la présente étude sont de longueur 3, 4 ou 5. Les séquences très courtes sont très fréquentes et ne sont pas assez spécifiques, tandis que les séquences longues, trop rares, sont trop spécifiques. Les séquences intermédiaires, de 3 à 5, correspondent à des structures grammaticales récurrentes d'une certaine complexité et suffisamment caractéristiques.

Titre de la pièce et son abréviation	Année de représentation / année de publication	Forme de la pièce	Nombre de <i>tokens</i> dans l'ensemble des répliques de Sganarelle	Nombre de séquences extraites ^[15]
<i>Sganarelle, ou le Cocu imaginaire (CI)</i>	1660/1662	Comédie en 1 acte, en vers	2589	251
<i>L'École des maris (EM)</i>	1661/1661	Comédie en 3 actes, en vers	5224	553
<i>Le Mariage forcé (MF)</i>	1664/1668	Comédie-ballet en 1 acte, en prose	2986	294
<i>Le Médecin volant (MV)</i>	1664/1819	Comédie en 1 acte, en prose	2243	232
<i>Le Festin de Pierre (DJ)</i>	1665/1683	Comédie en 5 actes, en prose	4856	500
<i>L'Amour médecin (AM)</i>	1665/1666	Comédie-ballet en 3 actes, en prose	1942	173
<i>Le Médecin malgré lui (MML)</i>	1666/1667	Comédie en 3 actes, en prose	4607	461

Table 2. Le corpus

Les textes ayant des longueurs variées, les fréquences absolues sont transformées en fréquences relatives.

Pour comparer les données obtenues, nous avons eu recours à un type de méthode d'analyse factorielle très utilisé pour l'analyse textuelle : l'analyse factorielle des correspondances (AFC^[16]). Dans le cadre de l'analyse des correspondances, chaque séquence syntaxique est une *variable* et sa fréquence est sa *valeur*. Les résultats de l'analyse des correspondances permettent d'identifier, pour chaque variable à haute contribution, le ou les texte(s) le(s)

plus associé(s)^[17].

La Figure 1, qui montre le positionnement des sept Sganarelle dans l'espace bidimensionnel, permet de comparer visuellement les distances entre les textes, calculées en tenant compte de toutes les (micro-)variations des fréquences des séquences extraites^[18].

10

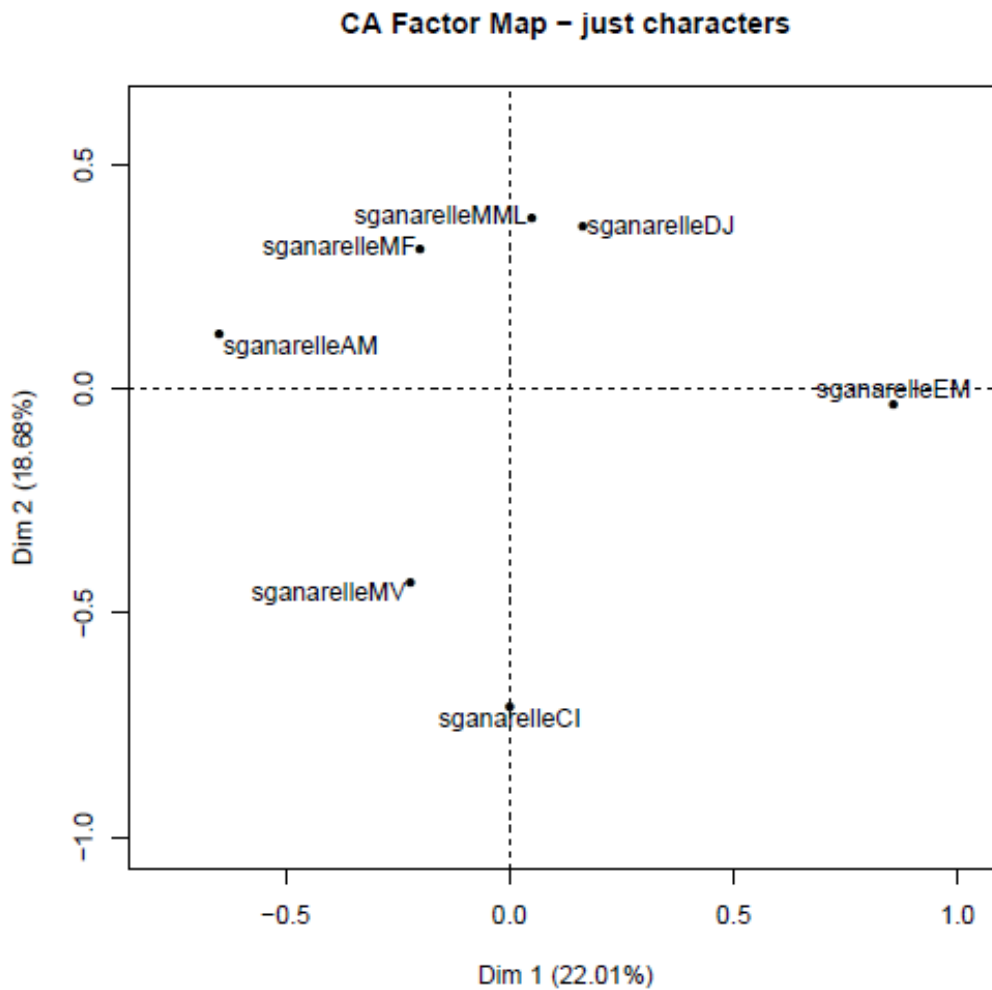


Figure 1. Les sept Sganarelle

Dans le graphique, chaque variable peut également être représentée en fonction de sa valeur dans chaque texte (voir Figure 2). On peut ainsi apprécier l'influence de chaque variable sur la distance entre les différents textes. Une variable dont la valeur est la même dans chaque texte occupe une position plutôt centrale ; une variable surreprésentée dans un texte contribue à déplacer celui-ci, à la manière d'un aimant, vers l'extrémité d'un axe ou vers celles des deux axes. Un texte idiosyncratique est représenté comme un point isolé dans l'espace, avec un nuage de variables qui l'éloigne du centre du repère.

11

Selon le texte et la longueur des séquences choisie, l'extraction des séquences peut produire une grande quantité de données. La syntaxe de la langue limitant fortement la variabilité syntagmatique, la plupart des séquences syntaxiques extraites ont une fréquence relative similaire dans les différents textes. Ces variables qui convergent vers l'intersection des deux axes du graphique ne sont pas intéressantes pour étudier les différences entre les textes. L'AFC permet de les filtrer en fonction de leur *contribution*, donnée qui mesure l'influence exercée par chaque variable sur l'éloignement ou le rapprochement des textes^[19]. Calculée pour chaque variable, la contribution indique la pertinence de chacune dans le cadre de la comparaison des textes. Seules les variables les plus contributives sont conservées.

12

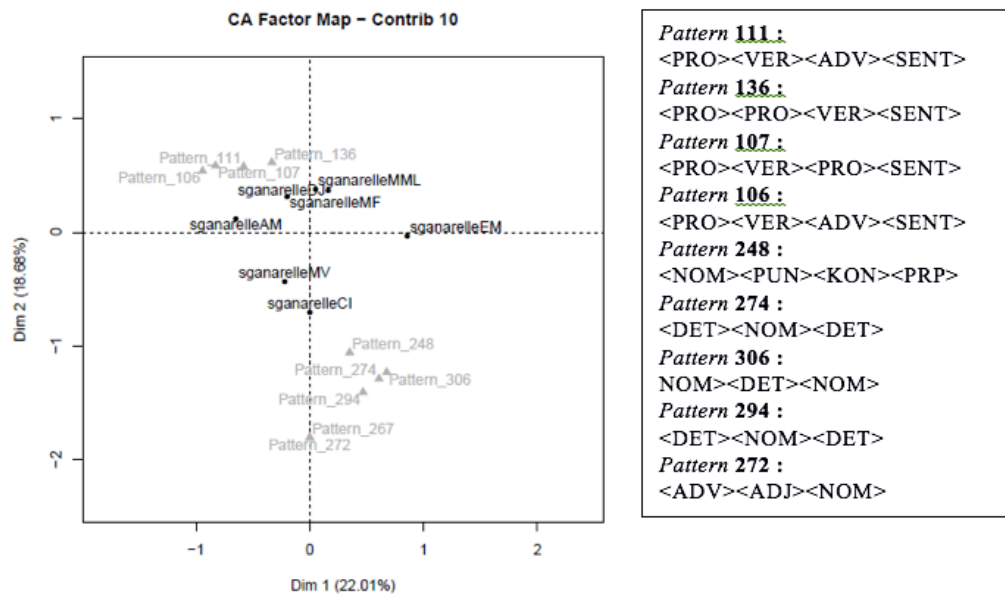


Figure 2. Les sept textes et les dix séquences à plus haute contribution

La Figure 2 représente les dix variables qui contribuent le plus à l'éloignement de chaque texte du centre du repère et à l'éloignement des textes entre eux. On peut également observer, parmi ces variables à plus haute contribution, celles qui sont projetées dans l'espace et qui sont fortement associées à un texte. Ainsi la visualisation bidimensionnelle peut-elle être exploitée, en observant soit la position de chaque texte sur les deux axes, soit celle des textes les uns par rapport aux autres (notamment l'opposition des textes sur l'axe des abscisses ou sur l'axe des ordonnées), soit l'association d'une variable à un texte.

13

Visuellement, les variables les plus surreprésentées dans un texte ne sont pas forcément celles qui lui sont le plus proches, car plus une variable est surreprésentée, plus elle est projetée loin vers les extrémités des axes du repère, selon sa contribution. Cependant, si l'on part d'une variable à forte contribution, il est possible d'identifier le texte qui lui est le plus proche : c'est celui qui se trouve sur la ligne imaginaire reliant la variable au point d'intersection des deux axes. Toutefois, étant donné les déformations causées par la représentation en deux dimensions, il est plus exact de considérer que le texte qui est le plus associé à une variable est celui dans lequel la fréquence relative de cette variable est la plus importante.

14

Grâce à l'AFC, nous produisons une liste des séquences syntaxiques classées par ordre décroissant en fonction de leur contribution et nous identifions, pour chaque séquence, le texte dans lequel elle est la plus fréquente. Parmi ces séquences à plus haute contribution, nous sélectionnons les cinq premières pour chaque texte, c'est-à-dire celles qui sont le plus surreprésentées par rapport aux autres^[20]. Le retour au texte et l'analyse qualitative sont possibles grâce à l'extraction automatique de toutes les occurrences textuelles correspondant à une séquence donnée. Par l'analyse qualitative des cinq séquences les plus distinctives associées à chaque personnage, nous tenterons de formuler une hypothèse concernant la signification des deux axes du repère.

15

Les deux comédies en vers : *Sganarelle ou le Cocu imaginaire* et *L'École des maris*

	Identifiant de la séquence	Séquence
1	Pattern 306	<NOM><DET><NOM>
2	Pattern 294	<DET><NOM><DET>
3	Pattern 267	<PRP><DET><NOM><DET>
4	Pattern 272	<ADV><ADJ><NOM>
5	Pattern 274	<KON><DET><NOM><PRP>

Table 3. Les cinq séquences les plus distinctives associées à Sganarelle *Cl* et classées par ordre d'importance

La fréquence des noms dans les séquences associées à Sganarelle *Cl* est notable. Les noms ont une corrélation positive avec les phrases longues : caractéristiques des registres écrits, celles-ci privilégient en effet les noms par opposition aux verbes [Biber et Conrad 2009, 114–115, 247]. Les noms sont aussi en corrélation négative avec les verbes et les pronoms, ce qui introduit la notion d'interaction.

16

Titre de la pièce	Réplique moyenne (taille moyenne en lignes : 1 l. = 60 signes)
<i>Le Médecin volant</i>	2,7 l.
<i>Sganarelle ou le Cocu imaginaire</i>	2,5 l.
<i>Le Festin de Pierre (Don Juan)</i>	2,3 l.
<i>L'École des maris</i>	2,2 l.
<i>L'Amour médecin</i>	1,9 l.
<i>Le Mariage forcé</i>	1,4 l.
<i>Le Médecin malgré lui</i>	1,4 l.

Table 4. Longueur moyenne des répliques des Sganarelle dans les pièces^[21]

Le Table 4 peut être mis en relation avec les séquences les plus distinctives associées aux Sganarelle. On remarque en effet que les pièces qui contiennent les répliques les plus longues (souvent associées à des phrases longues) sont celles dont les séquences distinctives comprennent le plus de noms.

17

Pour ce qui est des séquences syntaxiques, on note que les occurrences correspondant aux trois premières séquences associées à Sganarelle *Cl* se recoupent. De fait, les segments (<NOM><DET><NOM>) et (<DET><NOM><DET>), présents dans les trois séquences, renvoient aux mêmes phénomènes syntaxiques.

18

Pattern 306^[22] : <NOM><DET><NOM> (total de 14 occurrences)

1. Tu ne m'entends que trop, madame la carogne
2. Nous l'avons, et je puis voir à l'aise la trogne
Du malheureux pendard qui cause ma vergogne.
3. Nous savons **Dieu merci le souci** qui vous tient
4. Je ne sais pas si j'ai dans sa galanterie
L'honneur d'être connu de votre seigneurie ;
5. C'est-à-dire qu'il faut toucher au doigt la chose.
6. Voici **ma foi la chose** en propre original.
7. Je hais **de tout mon cœur les Esprits** colériques,
8. Et mettons **sous nos pieds les soupirs** et les larmes ;
9. Courage mon enfant, soit un peu vigoureux,

10. L'on ne demandait pas carogne ta venue,
 11. Accepte **sans façon le marché** qu'on propose.
Pattern 294 : <DET><NOM><DET> (total de 10 occurrences)

12. La chose est avérée, et je tiens **dans mes mains**
Un bon certificat du mal dont je me plains.
 13. Doucement s'il vous plaît cet homme a bien la mine
 D'avoir le sang bouillant et l'âme un peu mutine,
 14. Et d'attacher l'honneur de l'homme le plus sage
 Aux choses que peut faire une femme volage ;
 15. Vous croyez qu'en ce fait la plus forte apparence
 Peut jeter **dans l'esprit une** fausse créance

Pattern 267 : <PRP><DET><NOM><DET> (total de 7 occurrences)
 Aucune nouvelle occurrence, puisque la séquence 3 est enchâssée dans la séquence 2.

Dans un certain nombre d'occurrences, le verbe est séparé de son complément d'objet par un circonstant ou par une locution-phrase (« Dieu merci », « ma foi ») ou apostrophe (« carogne »), ce qui implique que deux noms se suivent, séparés par un déterminant. La contrainte de l'alexandrin – ou sa souplesse, c'est selon – explique l'antéposition de ces syntagmes [Fournier 1999]. On a alors affaire à trois séquences syntaxiques fortement déterminées par le choix de la forme versifiée. Mais, d'une part, on ne trouve pas les segments (<NOM><DET><NOM>) et (<DET><NOM><DET>) parmi les cinq séquences les plus distinctives associées à la comédie en vers de *L'École des maris*, de l'autre, une seule de ces cinq séquences correspond à un ordre des mots inhabituel et déterminé par l'alexandrin, mais il s'agit d'un phénomène limité.

26

	Identifiant de la séquence	Séquence
1	<i>Pattern 502</i>	<KON><PRP><NOM>
2	<i>Pattern 744</i>	<ADJ><NOM><PUN><PRO>
3	<i>Pattern 747</i>	<PRO><PUN><PRO><VER>
4	<i>Pattern 438</i>	<NOM><PUN><KON><DET>
5	<i>Pattern 269</i>	<PUN><KON><DET><NOM>

Table 5. Les cinq séquences les plus distinctives associées à Sganarelle *EM* et classées par ordre d'importance

La séquence (<KON><PRP><NOM>) peut renvoyer à une antéposition du complément déterminée par l'alexandrin, mais six occurrences seulement correspondent à ce phénomène syntaxique.

27

Pattern 502 : <KON><PRP><NOM> (total de 17 occurrences)

28

- De ces manches qu'à table on voit tâter les sauces
- Et de père, et d'époux donner pleine puissance,
- Et de père, et d'époux donner pleine puissance,
- Qu'aux discours des muguets, elle ferme l'oreille,
- Son cœur qu'avec excès, votre poursuite outrage,
- Que du don de ta foi je ne suis pas jaloux,

Ajoutons que dans l'occurrence (1), l'antéposition du thème-cadre est attendue^[23] et l'on aurait le même ordre des mots en prose.

Dans les trois premières séquences associées à Sganarelle *CI*, c'est donc moins l'ordre des mots qui est significatif – ce serait le même en prose du fait de la relation serrée du complément avec le support (2, 6, 8, 11, 13, 15) et de l'antéposition du thème-cadre (4) – que la présence de circonstants et de locutions-phrases servant à accréditer le discours de Sganarelle. Ce dernier est le personnage « originel », qui est systématiquement trompé par les apparences et qui se croit cocu parce qu'il interprète les faits de façon erronée, actualisant parfaitement le nom « Sganarelle », qui signifie « dupe de soi-même^[24] ». Les circonstants et locutions-phrases « voir à l'aise », « nous savons *Dieu merci* », « toucher *au doigt* », « voici *ma foi* » et « tiens *dans mes mains* » sont autant d'éléments censés garantir la force persuasive du discours.

29

Sganarelle s'évertue à convaincre les autres de la vérité de ce qu'il dit, mais on le voit aussi s'efforcer de se convaincre lui-même de la justesse de son attitude. C'est ce qui apparaît à travers la séquence (<KON><DET><NOM><PRP>).

30

Pattern 274 : <KON><DET><NOM><PRP> (total de 6 occurrences)

31

1. Doux objet de mes vœux j'ai grand tort de crier,
Et mon front de vos dons vous doit remercier.
2. Et songez **que les nœuds du** sacré mariage
3. [...]. Peut-être sans raison,
Me suis-je mis en tête ces visions cornues,
Et les sueurs au front m'en sont trop tôt venues.
4. Mais c'est peu **que l'honneur dans** mon affliction,
L'on me dérobe encor la réputation,
5. Quand j'aurai fait le brave, et **qu'un fer pour** ma peine
M'aura d'un vilain coup transpercé la bedaine,
6. Elles font la sottise, et nous sommes les sots :
C'est un vilain abus, **et les gens de** police
Nous devraient bien régler une telle injustice.

Trois de ces occurrences figurent dans des scènes où Sganarelle est seul : le court monologue de la scène 13 (3) et le monologue de la scène 17 où Sganarelle s'exhorte, sur 166 vers, à combattre son rival tout en essayant de trouver des arguments à sa lâcheté (5, 6). Les conjonctions de coordination (« et ») et de subordination (« que ») mettent en évidence la visée démonstrative du discours que le héros s'adresse à lui-même. Dans le long monologue, Sganarelle mêle aux arguments traditionnels du poltron – les risques qu'une vengeance lui ferait courir (5) – ceux que mobilise l'honnête homme, qui considère qu'il faut prendre avec indifférence ce qui n'entache pas son honneur, mais celui de l'épouse infidèle (6). La séquence (<KON><DET><NOM><PRP>) met donc en valeur un morceau de bravoure de la pièce, où Molière fait de la traditionnelle tirade de poltronnerie le lieu d'une confrontation des valeurs mondaines et des valeurs qui leur sont antagonistes.

39

Deux séquences associées à Sganarelle *EM*, (<PUN><KON><DET><NOM>) et (<NOM><PUN><KON><DET>), contiennent aussi une conjonction de coordination. Leurs occurrences se recoupent en partie.

40

Pattern 438 : <NOM><PUN><KON><DET> (total de 17 occurrences)

1. [à Ariste]
À lui souffrir en cervelle troublée,
De courir tous les **bals, et les** lieux d'assemblée ?
2. [à Valère]
Je vous l'apprends donc, et qu'il est à **propos**,

Que vos feux, s'il vous plaît, la laissent en repos.

3. [à Valère]

N'a que trop de vos yeux entendu le langage ;
Que vos secrets désirs, lui sont assez connus,

4. [à Valère]

Elle vous eût plutôt fait savoir sa pensée,
Si son cœur avait eu dans son émotion,
À qui pouvoir donner cette commission ;

5. [à part]

Appelons Isabelle, elle montre le fruit,
Que l'éducation dans une âme produit,
La vertu fait ses soins, et son cœur s'y consomme,

6. [à Isabelle]

Un plein effet
A suivi tes discours, et ton homme a son fait ;

7. [à Valère]

Qu'elle vous a fait voir assez quel est son choix,
Que son cœur tout à moi d'un tel projet s'offense,

8. [à Isabelle]

Tous ses désirs étaient de t'obtenir pour femme,
Si les destins en moi qui captive ton cœur,

9. [à Ariste]

On gagne les esprits par beaucoup de douceur ;
Et les soins défiants, les verrous et les grilles,
Ne font pas la vertu des femmes, ni des filles,

10. [à Ariste]

Nous les portons au mal par tant d'austérité,
Et leur sexe demande un peu de liberté.

11. [à Ariste]

On voit ce qu'en deux sœurs nos leçons ont produit,
L'une fuit ce galant, et l'autre le poursuit.

Les occurrences correspondent à :

- des répliques adressées à Ariste, dont les valeurs sont radicalement opposées à celles de son frère Sganarelle. La confrontation des deux personnages débouche systématiquement sur une dispute concernant l'éducation des filles, le mariage, la mode, etc. L'opposition entre les noms correspondant à la séquence (<NOM><PUN><KON><DET>), « douceur » et « austérité » (9 et 10), illustre les manières radicalement opposées dont les deux frères élèvent leur pupille.
- des répliques adressées aux amoureux, Valère et Isabelle. Sganarelle tente de dissuader Valère de faire la cour à Isabelle en rapportant les refus de celle-ci et il se félicite ensuite de la réussite de sa démarche auprès d'Isabelle en lui rapportant les propos de Valère. Le discours de Sganarelle « entremetteur malgré lui^[25] » a donc un caractère argumentatif et contient, de surcroît, des propos rapportés, ce qui explique la présence des conjonctions.

La séquence (<NOM><PUN><KON><DET>) renvoie donc à la fois à l'intrigue, fondée sur le schéma de « l'entremetteur malgré lui », et au débat sur les valeurs, que Molière introduit en mettant en scène deux frères que tout

oppose sur le plan axiologique.

La séquence (<ADJ><NOM><PUN><PRO>) attire aussi l'attention, car elle contient un adjectif, classe grammaticale assez rare dans les séquences distinctives associées aux différents personnages.

56

Pattern 744 : <ADJ><NOM><PUN><PRO> (total de 12 occurrences)

- M'obliger à porter de ces **petits chapeaux**,
Qui laissent éventer leurs débiles cerveaux,
- Peste soit du **gros bœuf**, **qui** pour me faire choir,
Se vient devant mes pas planter comme une perche.
- Que vos **secrets désirs**, **lui** sont assez connus,
Et que c'est vous donner des soucis superflus,
- Ne t'afflige point tant, va ma **petite femme**,
Je m'en vais le trouver, et lui chanter sa gamme.
- Voilà comme il faut que les femmes soient faites,
Et non comme j'en sais, de ces **franches coquettes**,
Qui s'en laissent conter [...]
- Hé, hé, mon petit nez, pauvre **petit bouchon**,
Tu ne languiras pas longtemps, je t'en réponds,
- Venez beau directeur, **suranné damoiseau**,
On veut vous faire voir quelque chose de beau.
- Vous l'avez bien stylée ;
Il n'est pas bon de vivre en **sévère censeur**,
On gagne les esprits par beaucoup de douceur ;
- **Pauvre esprit**, je vous dis, et vous redis encor,

L'antéposition de l'adjectif témoigne d'une prise en charge énonciative ou d'une subjectivité plus marquée. Dans le discours de Sganarelle, elle traduit soit le ton ironique et satirique sur lequel il s'adresse à ceux qui prônent d'autres valeurs que les siennes (1, 3, 5, 7, 8, 9), soit le ton attendri et amoureux qu'il adopte avec Isabelle (4, 6). Ces deux attitudes sont tout-à-fait caractéristiques du personnage de bourgeois anti-galant qu'invente Molière et elles éclatent dans un héros comme Arnolphe (*L'École des femmes*).

57

Les comédies-ballets : *L'Amour médecin* et *Le Mariage forcé*

	Identifiant de la séquence	Séquence
1	<i>Pattern 136</i>	<PRO><PRO><VER><SENT>
2	<i>Pattern 107</i>	<PRO><VER><PRO><SENT>
3	<i>Pattern 106</i>	<PRO><VER><ADV><SENT>
4	<i>Pattern 111</i>	<VER><PRO><SENT>
5	<i>Pattern 144</i>	<VER><VER><SENT>

Table 6. Les cinq séquences les plus distinctives associées à Sganarelle AM et classées par ordre d'importance

Sur les cinq séquences, trois correspondent à des occurrences qui ne sont pas réparties de façon homogène dans la pièce, mais circonscrites dans une scène ou une partie d'une scène. Dans notre corpus, ces séquences sont très idiosyncratiques : elles renvoient à un effet comique, fondé sur la répétition et propre à *L'Amour médecin*. La séquence (<PRO><VER><PRO><SENT>) et sa variante (<VER><PRO><SENT>) renvoient au dialogue de sourds entre Sganarelle et Lisette (I, 6) : la servante entre en scène en s'écriant « Ah Malheur ! ah disgrâce ! », sans expliquer la raison de ses cris, laissant Sganarelle imaginer le pire au sujet de sa fille.

58

Pattern 107 : <PRO><VER><PRO><SENT> (total de 8 occurrences)

59

1. Hé bien, qu'est-ce ?
2. Que sera-ce ?
3. Qu'est-ce ?
4. Qu'y a-t-il ?
5. Qu'est-ce ?

La séquence (<PRO><VER><ADV><SENT>) correspond à un autre dialogue de sourds. Cette fois, c'est Sganarelle qui refuse d'entendre ce que lui répète huit fois Lisette – que la maladie de sa fille, un mutisme soudain, requiert un seul remède : un mari.

60

Pattern 106 : <PRO><VER><ADV><SENT> (total de 8 occurrences)

61

1. Non, ne m'en parlez point.
2. Ne m'en parlez point.
3. Ne m'en parlez point.

La relative brièveté du texte (c'est le plus court des sept) explique aussi la surreprésentation de ces séquences très spécifiques.

La seule séquence qui ne renvoie pas à un effet de répétition, mais à l'intrigue, est la séquence (<VER><VER><SENT>).

62

Pattern 144 : <VER><VER><SENT> (total de 14 occurrences)

63

1. Je n'avais qu'une seule femme qui est morte.
2. je vous prie de me conseiller tous ce que je dois faire.
3. Aimerais-tu quelqu'un, et souhaiterais-tu d'être mariée ?
4. Ce n'est pas la récompense de t'avoir élevée comme j'ai fait.
5. Je suis perdu.
6. Elle s'est jetée.
7. Est-ce que les médecins font mourir ?
8. je vous prie de me dire vite ce que vous avez résolu.
9. Et vous suis infiniment obligé de la peine que vous avez prise.
10. Il faut que j'aïlle chercher de l'orviétan, et que je lui fasse prendre.
11. Ma fille est guérie.
12. Voilà qui est fait.

Elle correspond à :

64

– des syntagmes verbaux à la première ou troisième personne et au passé composé (1, 4, 5, 6, 11, 12) : Sganarelle, passif, se contente d'enregistrer ce qui a eu lieu.

65

– des modaux (2), des périphrases (10), des groupes verbaux à la deuxième personne et au passé composé, dont le sémantisme renvoie à la délibération (8 et 9) : ces syntagmes expriment la requête de Sganarelle, soucieux de guérir sa fille pour hâter le mariage qu'il a décidé.

66

Ces occurrences ont en commun la brièveté des phrases et la présence des pronoms des première et deuxième personnes, marques d'une forte interaction. Les séquences associées à Sganarelle *AM* donnent à voir les « deux styles » de Molière : « l'un, très proche des propos quotidiens » et « l'autre, très rythmé et, de ce fait, aussi peu réaliste que possible » [Larthomas 1972/2001, 317].

67

On retrouve ces « deux styles » à travers les séquences relatives à Sganarelle *MF*.

68

	Identifiant de la séquence	Séquence
1	Pattern 149	<PRO><VER><SENT>
2	Pattern 34	<NOM><PUN><KON><PRO><VER>
3	Pattern 337	<PRP><PRO><VER><SENT>
4	Pattern 530	<PRO><PRO><VER><PUN><KON>
5	Pattern 533	<ADV><VER><PRO>

Table 7. Les cinq séquences les plus distinctives associées à Sganarelle *MF* et classées par ordre d'importance

Comme pour Sganarelle *AM*, on repère une séquence dont les occurrences ne sont pas réparties de façon homogène et qui correspondent à une répétition comique.

Pattern 533 : <ADV><VER><PRO> (total de 9 occurrences)

1. N'ai-je pas tous les mouvements de mon corps aussi bons que jamais ?
2. N'ai-je pas encore toutes mes dents les meilleures du monde ?
3. Ne fais-je pas vigoureusement mes quatre repas par jour ?
4. N'ai-je pas raison, d'avoir fait ce choix ?
5. N'êtes-vous pas bien aise de ce mariage, mon aimable pouponne ?

Les occurrences (1) à (3) figurent dans la scène 1 où Sganarelle vient demander conseil à son ami Géronimo à propos de son projet de mariage. La réponse de Géronimo est dubitative et pour dissuader son ami de se lancer dans une telle entreprise, il lui fait remarquer son âge. La série de phrases interro-négatives prononcées par Sganarelle répond à cet argument. Plus globalement, la forme interro-négative illustre l'état d'esprit du personnage qui, du moins au début de la comédie, cherche moins à être bien conseillé qu'à être conforté dans son choix, ce que montrent les occurrences (4) et (5), adressées à Géronimo et à Dorimène, la future épouse.

Au contraire, les occurrences correspondant à la séquence (<PRO><VER><SENT>) et à la séquence (<PRP><PRO><VER><SENT>) sont réparties dans la pièce de façon homogène et renvoient à l'intrigue.

Pattern 337 : <PRP><PRO><VER><SENT> (total de 11 occurrences)

1. C'est que je veux savoir de vous, si je ferai bien de me marier.
2. Peut-il y avoir un homme, qui n'ait, en la voyant, des démangeaisons de se marier ?
3. j'ai quelque chose à vous communiquer.
4. je vous prie de m'écouter.
5. elle me plaît beaucoup, et est ravie de m'épouser.
6. et prenez la peine de m'écouter.
7. Je viens vous dire que j'ai envie de me marier.
8. Ferai-je bien, ou mal, de l'épouser ?

Cette séquence qui est composée d'un verbe, dont le sémantisme renvoie au projet de mariage de Sganarelle (« marier », « épouser ») et à la requête qui en découle (« communiquer », « écouter »), et d'un pronom de la première personne, désignant le locuteur, ou de la deuxième personne, représentant celui dont on sollicite les conseils, met en évidence le sujet de la comédie : comme le personnage de Panurge dans le *Tiers Livre*, Sganarelle cherche désespérément une autorité qui lui dise s'il doit se marier ou pas.

La quasi-absence de noms dans les séquences associées aux Sganarelle *AM* et *MF* et son corollaire, la forte présence des pronoms et des verbes [Kylander 1995, 94], montrent que dans ces deux pièces le débat sur les valeurs – que l'on repère dans *Le Cocu imaginaire* et *L'École des maris* par la coordination/subordination et des noms dont le sémantisme renvoie aux valeurs en jeu – n'est pas au premier plan. Sganarelle *AM* et Sganarelle *MF* font partie de la lignée issue

du *Cocu imaginaire* et de *L'École des maris* et sont porteurs des mêmes valeurs rétrogrades, mais l'affirmation des conceptions qui sous-tendent leur comportement est secondaire. *L'Amour médecin* et *Le Mariage forcé* sont des comédies-ballets dont la composition se rapproche du ballet à entrées : une série de consultations médicales dans la première et une série de « consultations conjugales » dans la seconde. Différents protagonistes défilent devant Sganarelle, presque toujours en scène : ainsi celui-ci permet-il de relier des scènes entre lesquelles le rapport n'est pas évident. D'où un personnage peu caractérisé sur le plan des valeurs et très interactif, puisque sa principale fonction est de répéter sa requête.

De faux médecins : les Sganarelle du *Médecin malgré lui* et du *Médecin volant*

	Identifiant de la séquence	Séquence
1	Pattern 144	<VER><KON><DET>
2	Pattern 405	<DET><NOM><PUN><VER>
3	Pattern 194	<PRP><VER><PUN>
4	Pattern 455	<VER><KON><DET><NOM>
5	Pattern 730	<PRO><VER><PUN><PRP>

Table 8. Les cinq séquences les plus distinctives associées à Sganarelle *MML* et classées par ordre d'importance

Sganarelle *MML* incarne les valeurs anti-mondaines défendues par les personnages du *Cocu imaginaire* et de *L'École des maris*, mais l'imposture médicale est le thème dominant de la comédie. La séquence (<VER><KON><DET><NOM>) et sa variante (<VER><KON><DET>) correspondent ainsi à des propositions subordonnées complétives exprimant les assertions péremptoires, diagnostics et prescriptions du faux médecin.

76

Pattern 455 : <VER><KON><DET><NOM> (total de 11 occurrences)

77

1. je **tiens qu'un homme** bien sain s'en accommoderait assez.
2. je **vous apprend** **que** votre fille est muette.
3. **il se trouve que** le poumon
4. **on voit que** l'inégalité de leurs opinions
5. **vous voyez que** l'ardeur

La séquence (<DET><NOM><PUN><VER>) renvoie aussi à la consultation du faux médecin.

78

Pattern 405 : <DET><NOM><PUN><VER> (total de 13 occurrences)

79

1. je tiens que cet empêchement de l'action de sa langue, est causé par de certaines humeurs qu'entre nous, savants, nous appelons humeurs peccantes, peccantes, c'est-à-dire ... humeurs peccantes
2. on voit que l'inégalité de leurs opinions, dépend du mouvement oblique, du cercle de la lune
3. et comme le soleil qui darde ses rayons sur la concavité de la terre, trouve ...

Dans le théâtre imprimé du XVII^e siècle, la ponctuation est autant orale que grammaticale^[26] et joue donc un rôle dans la prononciation ou la lecture des vers. Dans la consultation médicale fantaisiste de Sganarelle, on peut supposer que la ponctuation, placée entre le sujet et son verbe, sert, d'une part, à ralentir le tempo et à marquer les hésitations d'un discours qui s'invente au fur et à mesure (comme les points de suspension), de l'autre, à créer des effets de cadence majeure ou mineure (2).

80

Le faux discours médical n'est pas aussi développé dans *Le Médecin volant* et les séquences associées au personnage en témoignent.

81

	Identifiant de la séquence	Séquence
1	Pattern 248	<NOM><PUN><KON><PRP>
2	Pattern 212	<NOM><PRO><VER><PRP>
3	Pattern 843	<PUN><NOM><NAM>
4	Pattern 842	<NOM><NAM><PUN>
5	Pattern 280	<PRP><DET><NOM><PRO>

Table 9. Les cinq séquences les plus distinctives associées à Sganarelle *MV* et classées par ordre d'importance

À titre d'exemple, la séquence (<PRP><DET><NOM><PRO>) fait apparaître des circonstants ou propositions relatives à valeur explicative, censés étayer le raisonnement du faux médecin, mais les occurrences relatives à la consultation sont peu nombreuses :

Pattern 280 : <PRP><DET><NOM><PRO> (9 occurrences)

1. Voilà de l'urine **qui** marque grande chaleur.
2. parce qu'avec le goût **je** discerne bien mieux la cause et les suites de la maladie
3. et que la bile qui se répand par le corps **nous** fait devenir jaunes

Par ailleurs, deux séquences (<PUN><NOM><NAM>) et (<NOM><NAM><PUN>) déclinent le même segment séquentiel correspondant à l'apostrophe, « Monsieur Gorgibus », qui figure dans les scènes 14 et 15 où Sganarelle passe d'un lieu à un autre en changeant d'apparence pour éviter d'être démasqué. La répétition de l'apostrophe ponctue la fausse dispute entre les « deux » Sganarelle qui s'adressent chacun à leur tour à Gorgibus. La séquence d'ubiquité est fondée sur un effet de répétition, que souligne la récurrence de l'apostrophe, à la différence du comique médical qui ne repose pas sur des procédés de répétition, mais sur le jargon médical, le latin macaronique, les propositions de remèdes absurdes ou nocifs, la querelle sur la nature de la maladie, etc.

Le valet d'un impie : le Sganarelle du Festin de Pierre

	Identifiant de la séquence	Séquence
1	Pattern 543	<KON><DET><NOM><VER>
2	Pattern 548	<PUN><KON><ADV>
3	Pattern 552	<VER><ADV><PRP><NOM>
4	Pattern 542	<ADV><PRP><NOM><PUN>
5	Pattern 519	<PRP><PRO><PRO>

Table 10. Les cinq séquences les plus distinctives associées à Sganarelle *DJ* et classées par ordre d'importance

Les séquences distinctives associées à Sganarelle *DJ* sont également caractérisées par la présence de conjonctions qui soulignent les efforts du valet pour faire revenir son maître de son égarement.

Pattern 548 : <PUN><KON><ADV> (total de 14 occurrences)

1. non seulement il réjouit et purge les cerveaux humains, mais encore il instruit les âmes à la vertu, et leur apprend avec lui à demeurer honnête homme.
2. Je pourrais peut-être me tromper, mais enfin sur de tels sujets l'expérience m'a donné quelque lumière.
3. Vous tournez les choses d'une manière qu'il semble que vous ayez raison, et cependant il est vrai que vous ne l'avez pas ;
4. Osez-vous bien ainsi vous jouer du Ciel, et ne tremblez-vous point de vous moquer comme vous faites des

choses les plus saintes ;

5. et je leur disais que si quelqu'un leur venait dire du mal de vous, elles se gardassent bien de le croire, **et ne** manquaient pas de lui dire qu'il en avait menti.
6. car il n'y a rien de plus vrai que le Moine bourru ; et je me ferais pendre pour celui-là ; **mais encore** faut-il croire quelque chose dans le monde
7. Est-ce que vous vous êtes fait tout seul, **et n'a-t-il** pas fallu que votre père ait engrossé votre mère pour vous faire ?
8. Après cela, **si ne** vous rendez, tant pis pour vous.

Ces occurrences figurent dans des phrases relativement longues, dans lesquelles le raisonnement de Sganarelle se déploie, ou plutôt s'embrouille, avec force conjonctions de coordination. 87

Une autre caractéristique du discours de Sganarelle apparaît à travers la séquence (<VER><ADV><PRP><NOM>). 88

Pattern 552 : <VER><ADV><PRP><NOM> (17 occurrences) 89

1. vous **ne croyez pas au séné** [...] ?
2. est-il possible que vous **ne croyez point du tout** au Ciel ?
3. vous **ne croyez rien du tout**
4. qui **n'a point de loi** vit en bête brute

Le recours à la forme négative est déterminé par l'impiété de Don Juan. 90

Conclusion

L'étude présentée repose sur une méthode exploratoire de comparaison textuelle de type inductif et mêle les approches quantitative et qualitative. Les textes sont en effet analysés de manière automatique pour mettre au jour d'éventuelles différences dans l'utilisation de la syntaxe, sans qu'un nombre de traits grammaticaux soit ciblé a priori, et l'analyse des résultats passe par le retour au texte et l'approche herméneutique, propre aux études littéraires. Elle permet d'illustrer la manière dont les méthodes numériques peuvent s'intégrer à la pratique de l'analyse littéraire, sans changer ses pratiques scientifiques ni imposer des méthodes empruntées aux sciences dures, mais en respectant les paradigmes disciplinaires dans le sens d'une « herméneutique numérique », telle que la conçoit S. Ramsay [Ramsay 2008], entre autres^[27]. 91

L'examen des cinq séquences les plus distinctives associées à chaque Sganarelle de notre corpus nous amène à formuler les hypothèses suivantes concernant la visualisation des résultats quantitatifs. 92

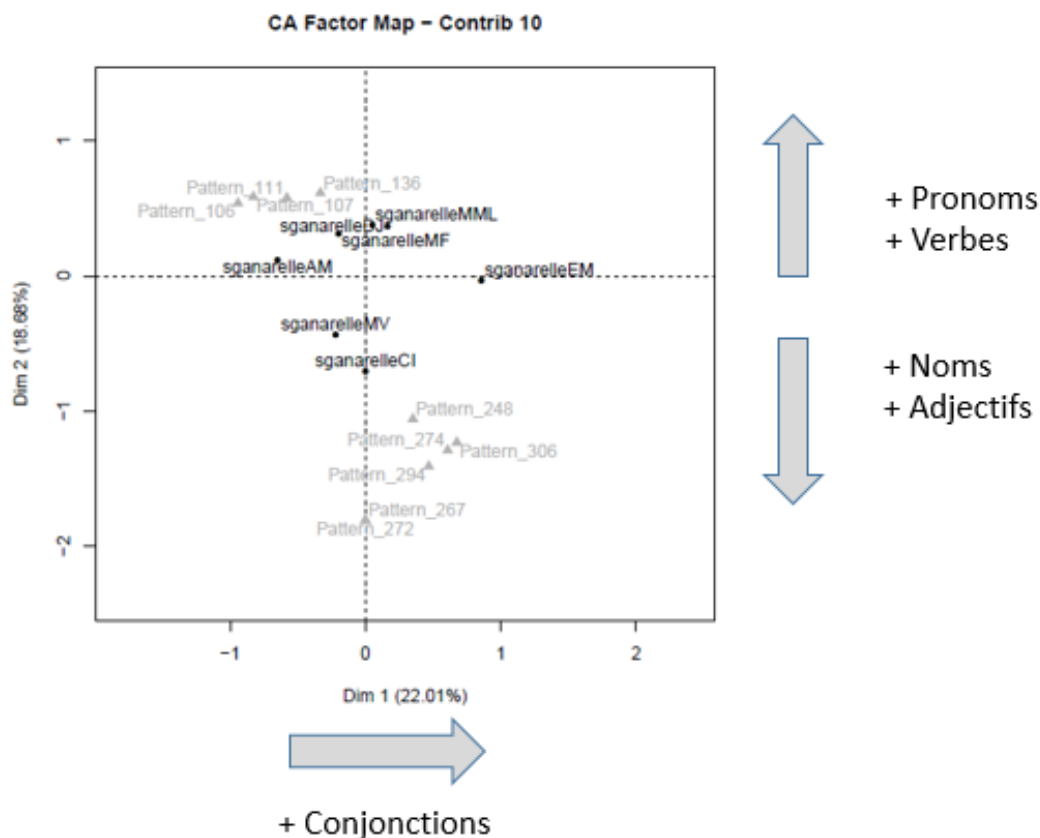


Figure 3. Hypothèses concernant les deux axes de la représentation bidimensionnelle

L'écart entre Sganarelle *CI* et Sganarelle *MF* sur l'axe des ordonnées invite à mettre en relation cet axe et la dimension interpersonnelle du dialogue. Plus l'ordonnée diminue, plus la fréquence des pronoms des première et deuxième personnes et des verbes – signe d'un degré d'interaction élevé – est importante. En effet, aucune des cinq séquences distinctives associées à Sganarelle *CI* ne comporte un pronom ou un verbe, tandis que dans celles qui se rapportent à Sganarelle *DJ*, ainsi qu'aux Sganarelle *MML* et *MF* situés à proximité, les verbes sont nombreux et parfois accompagnés de pronoms. Les pronoms et les verbes étant en corrélation négative avec les noms, plus l'ordonnée augmente, plus la fréquence des noms – signe que le discours est porteur d'informations – est importante. Il ne s'agit pas de dire que certains personnages seraient du côté de l'interaction et d'autres du côté de l'information. Les deux modes de fonctionnement sont présents et coexistent tout au long de la pièce. Cependant, la comparaison des personnages montre que l'une ou l'autre dimension peut être plus importante chez certains. L'axe horizontal est plus complexe à analyser. D'après la distance séparant *L'Amour médecin*, dont les cinq séquences distinctives sont dépourvues de conjonctions, et *L'École des maris*, dont les cinq séquences distinctives présentent le plus de conjonctions, il semble que le degré de complexité du discours soit en jeu.

93

Il est difficile, à partir de la visualisation et de l'interprétation qu'on peut en faire, de caractériser les différentes comédies qui ont comme héros Sganarelle. Tout au plus peut-on distinguer le fonctionnement des comédies-ballets (*L'Amour médecin* et *Le Mariage forcé*), dont la structure sérielle explique que Sganarelle, confronté à de nombreux personnages, ait une forte dimension interactive, et celui des deux comédies en vers où les répliques plus longues et la place accordée à l'exposé des idées et des valeurs diminuent le degré d'interaction (*Le Cocu imaginaire* et *L'École des maris*). En revanche, la méthode apparaît pertinente en ce qui concerne la caractérisation des personnages et leur fonction dramaturgique, particulièrement sensibles dans la syntaxe. Le rôle du « solliciteur », assumé par Sganarelle dans *Le Mariage forcé*, son rôle d'« entremetteur » dans *L'École des maris* et de valet « raisonneur » dans *Le Festin de Pierre* apparaissent clairement. Du point de vue de la caractérisation, l'antéposition de l'adjectif met en évidence des traits saillants du personnage de *L'École des maris* : tempérament satirique, mais aussi perméable à la passion amoureuse. En outre, les effets comiques reposant sur la répétition sont bien repérés dans ce type d'analyse.

94

Concernant la caractérisation, l'extraction automatique fournit des données plus nombreuses susceptibles d'étayer des constats déjà formulés par la tradition critique. Cependant, c'est pour l'étude du rythme, du tempo et du nombre que la méthode nous semble la plus prometteuse^[28]. En effet, au cours de l'étude, nous avons repéré un certain nombre de phénomènes : l'antéposition des groupes prépositionnels, la séparation du sujet et de son verbe par une virgule, l'insertion d'apostrophes, des répétitions. Or, des caractéristiques du dialogue moliéresque, comme la « prose cadencée^[29] » ou l'imitation de la conversation mondaine^[30], pourraient être mieux appréhendées si l'on pouvait mesurer ces phénomènes de manière automatique et statistique.

Notes

[1] On peut aussi mentionner les études de statistique lexicale, fondées sur la distance intertextuelle, qui nourrissent la polémique autour de l'attribution des œuvres de Molière, connue sous le nom de « L'affaire Corneille-Molière » ([Labbé et Labbé 2001] ; [Brunet 2004] ; [Viprey et Ledoux 2006]).

[2] [Larthomas 1972/2001, 48]

[3] Biber et Conrad distinguent trois approches : *register perspective*, *genre perspective*, *style perspective*.

[4] Il s'agit d'un manuscrit retrouvé au XVIII^e siècle, dont l'attribution est incertaine. Sur l'attribution de la pièce à Molière, voir Forestier et Bourqui, 2010, II, 1719-1720.

[5] Il s'agit des dates de création. Concernant *Le Médecin volant*, la date de création est très incertaine. Georges Forestier et Claude Bourqui avancent l'hypothèse selon laquelle le valet, d'abord prénommé « Mascarille », aurait été rebaptisé « Sganarelle » lorsque la pièce a été jouée en 1664. Si l'on suit cette hypothèse, c'est cette date qui nous intéresse puisque le personnage a alors été rattaché à la lignée des Sganarelle [Forestier et Bourqui 2010, II, 1720].

[6] Mascarille est le type du valet rusé à l'italienne : on le retrouve dans deux pièces de Molière, *Les Précieuses ridicules* et *L'Étourdi ou les Contretemps*. Jodelet est un personnage qui a été conçu pour l'acteur Julien Bedeau et dont la principale caractéristique est de faire montre d'assurance tout en étant un grand poltron. Scarron est le principal auteur de pièces « à Jodelet », mais Molière se sert aussi de Jodelet dans *Les Précieuses ridicules*.

[7] [Forestier et Bourqui 2010, I, 1227].

[8] *Ibid.*, « Introduction », I, xxvi-xliii.

[9] Voir [Forestier et Bourqui 2010, I, 1249].

[10] D'après Forestier et Bourqui, en 1664, lorsqu'est représentée la comédie du *Médecin volant*, « en dehors de ses grandes comédies, les personnages qu'il [Molière] s'attribuait portaient tous le nom de Sganarelle ». *Le Festin de Pierre* est créée en 1665 [Forestier et Bourqui 2010, II, 1720].

[11] L'imposture médicale est aussi au centre de *L'Amour médecin*, mais dans cette pièce c'est l'amoureux Clitandre qui se déguise en médecin, et non Sganarelle, qui est le père.

[12] Voir [Frontini, Boukhaled et Ganascia 2015a] [Frontini, Boukhaled et Ganascia 2015b] [Frontini, Boukhaled et Ganascia 2017].

[13] Cet outil d'étiquetage est entraîné sur un corpus de français contemporain en prose. Le texte de Molière que nous avons utilisé est une édition normalisée du point de vue orthographique (<http://obvil.paris-sorbonne.fr/corpus/moliere/moliere>). Un prétraitement a été effectué afin de supprimer les majuscules à l'initiale des vers. La performance de *Tree Tagger* sur ce corpus ne diminue pas de façon significative : 700 mots du corpus ont été annotés manuellement et le taux d'erreur est de 3 %. Ce résultat satisfaisant ne nous a pas paru justifier le recours à des modèles spécifiques, tels que le modèle développé dans le cadre du projet PRESTO (http://presto.ens-lyon.fr/?page_id=197).

[14] L'extraction des *PoS n-grams* est effectuée grâce à EReMoS, un outil développé au sein du LIP6 (Laboratoire d'informatique de Paris VI) et dans le cadre de la thèse de M. A. Boukhaled, 2016. Il en existe une version en ligne : <http://eremos.lip6.fr/>. EReMoS est un outil qui permet d'extraire des séquences lexico-grammaticales de complexité variable (*PoS n-grams*, mais aussi *n-grams* composés d'éléments de natures différentes). Dans cette étude, nous utilisons la fonction la plus basique, qui permet l'extraction de *PoS n-grams* de longueur déterminée.

[15] Dans cette phase, nous appliquons un pré-filtrage minimal qui exclut les séquences à très basses fréquences, autrement dit celles que l'on trouve dans moins de 5 % des phrases du texte et qui correspondent à moins de cinq occurrences.

[16] Nous avons utilisé la librairie FactoMineR sous R pour l'analyse des correspondances (Lê et al., 2008). Le code R avec lequel nous analysons les données d'EReMoS à l'aide de l'AFC est disponible à l'adresse suivante: <https://github.com/francescafrontini/CAforEREMOS>.

[17] L'AFC est souvent utilisée avec des fréquences absolues (tableau de contingence). Toutefois, certaines études ont recours aux fréquences relatives [Cichocki 2006]. Nous avons constaté de manière empirique qu'étant donné les différences importantes entre les longueurs des textes, le recours aux fréquences absolues tendait à donner plus de poids aux séquences surreprésentées dans les textes les plus longs. Du point de vue de la caractérisation, cela revient à dire que les personnages qui ont un volume de parole plus important sont les plus marqués. Nous avons donc privilégié une approche qui analyse les textes prononcés par les Sganarelle comme s'ils avaient tous le même volume de parole.

[18] L'analyse des correspondances, comme d'autres types d'analyse factorielle, entraîne une perte d'informations, car les distances sont représentées dans un espace bidimensionnel. La quantité d'informations représentée sur chaque axe est indiquée en pourcentage sur le graphique.

[19] L'AFC fournit la contribution de chaque variable sur chaque dimension. Pour notre étude, nous prenons seulement en compte la contribution sur les deux premières dimensions, c'est-à-dire celles qui sont représentées dans le graphique. Cette contribution est calculée en suivant l'implémentation proposée par le module FactoMineR.

[20] Dans le cadre de cet article, nous avons choisi de nous limiter à l'analyse des cinq séquences les plus associées à chaque Sganarelle. Il est toutefois possible de prendre en considération toutes les séquences dont la contribution dépasse la moyenne et de déterminer les textes qui leur sont associés. En outre, des méthodes statistiques telles que le *chi-square* permettraient d'identifier les différences, significatives du point de vue statistique, en ce qui concerne la distribution des séquences dans les différents textes.

[21] Ces données sont fournies par Frédéric Glorieux : <http://obvil.paris-sorbonne.fr/corpus/moliere/moliere>.

[22] Nous soulignons la séquence et nous signalons en gras le syntagme que nous souhaitons mettre en évidence.

[23] L'ordre des mots permet en général d'aller du connu (le thème « ce dont on parle ») vers le nouveau (le rhème qui correspond à l'apport d'informations, « ce qu'on dit du thème »).

[24] Le nom est inventé par Molière à partir du verbe italien *ingannare* « tromper » et du suffixe « -elle » que l'on trouve dans le nom de plusieurs *zanni* de la *commedia dell'arte* (Polichinelle, Brigueulle, etc.). Voir [Forestier et Bourqui 2010, I, 1227].

[25] [Forestier et Bourqui 2010, I, 1255].

[26] La ponctuation des textes de notre corpus est celle des éditions originales des pièces, les seules publiées avec l'aval de Molière.

[27] Pour une mise au point sur le statut des humanités numériques par rapport à la distinction traditionnelle entre « sciences de la culture » et « sciences de la nature », voir [Ganascia 2015].

[28] Nous nous fondons sur la distinction établie par Pierre Larthomas : le *rythme* désigne « tout effet de répétition », le *tempo*, « la plus ou moins grande rapidité avec laquelle la pièce est jouée », et le *nombre* est déterminé par « la longueur des différents membres de la phrase et des effets produits par leur juxtaposition ». [Larthomas 1972/2001, 72, 309, 311].

[29] La « prose cadencée » est définie comme une « prose constituée d'une alternance de vers irréguliers non rimés » [Forestier et Bourqui 2010, II, 1623].

[30] Nous avons approfondi cette question dans le cadre d'une étude qui porte sur les « raisonneurs » chez Molière. L'analyse a permis de mettre en évidence l'usage particulier de la ponctuation dans *Le Misanthrope*, comédie mondaine par excellence. Cette étude, intitulée « Être ou ne pas être raisonneur : étude du lexique et des motifs syntaxiques dans les comédies de Molière », a été présentée, avec la collaboration de Jean-Gabriel Ganascia, dans le cadre du colloque « Des humanités numériques littéraires ? », organisé par Daniel Alexandre, Milad Doueïhi et Marc Douguet, le 21 juin 2017 au Centre international culturel de Cerisy.

Works Cited

Benzécri 1982 Benzécri, Jean-Paul. *Histoire et préhistoire de l'analyse des données*. Paris, Dunod (1982).

- Bernet 1983** Bernet Charles. *Le Vocabulaire des tragédies de Jean Racine. Analyse statistique*. Genève-Paris, Slatkine-Champion (1983).
- Biber et Conrad 2009** Biber, Douglas, Conrad, Susan. *Register, Genre and Style*. Cambridge, Cambridge University Press (2009).
- Bloch 2009** Bloch Olivier. *Molière: comique et communication*. Pantin, Le Temps des Cerises (2009).
- Boukhaled 2016** Boukhaled, Mohamed Amine. *On Computational Stylistics: Mining Literary Texts for the Extraction of Characterizing Stylistic Patterns*. Thèse de doctorat (sous la direction de J.-G. Ganascia). Université Pierre et Marie Curie (2016).
- Brunet 2004** Brunet Étienne. "Où l'on mesure la distance entre les distances?". *Texto !* (2004) [en ligne]. Rubrique "Dits et inédits." http://www.revue-texto.net/Inedits/Brunet/Brunet_Distance.html [dernière consultation le 14/07/17].
- Cichocki 2006** Cichocki Wladyslaw. "Geographic Variation in Acadian French: What can Correspondence Analysis Contribute Toward Explanation?". *Literary and Linguistic Computing*, 21 (4): 529-541 (2006).
- Conesa 1983** Conesa Gabriel. *Le Dialogue moliéresque. Étude stylistique et dramaturgique*. Paris, Presses universitaires de France (1983).
- Forestier et Bourqui 2010** Forestier, Georges, Bourqui, Claude (éd.). Molière, *Œuvres complètes*. Paris, Gallimard, Bibliothèque de la Pléiade, 2 tomes (2010).
- Fournier 1999** Fournier Nathalie. "La position des groupes prépositionnels dans l'énoncé, dans *Le Misanthrope* et *George Dandin*". In *Phrases: syntaxe, rythme, cohésion du texte*. Neveu, Frank (éd.). Paris, SEDES (1999).
- Frontini, Boukhaled et Ganascia 2015a** Frontini, Francesca, Boukhaled, Mohamed Amine, Ganascia, Jean-Gabriel. "Molière's Raisonners: a Quantitative Study of Distinctive Linguistic patterns". In *Corpus Linguistics 2015 — Abstract Book*. Lancaster: 114-117 (2015). <http://ucrel.lancs.ac.uk/cl2015/doc/CL2015-AbstractBook.pdf> [dernière consultation le 14/07/17].
- Frontini, Boukhaled et Ganascia 2015b** Frontini, Francesca, Boukhaled, Mohamed Amine, Ganascia, Jean-Gabriel. "Linguistic Pattern Extraction and Analysis for Classic French Plays". Présentation aux Journées ConSciLa, Paris (2015). <http://lipn.univ-paris13.fr/~charnois/conscilaGenres/resumes/frontini.pdf>.
- Frontini, Boukhaled et Ganascia 2017** Frontini, Francesca, Boukhaled, Mohamed Amine, Ganascia, Jean-Gabriel. "Mining for Characterizing Patterns in Literature Using Correspondance Analysis. An Experiment on French Novels". *Digital Humanities Quarterly Proceedings of the Göttingen Dialogue for Digital Humanities 2015*. 11-2 (2017).
- Ganascia 2015** Ganascia, Jean-Gabriel. "The Logic of the Big Data Turn in Digital Literary Studies", *Frontiers in Digital Humanities*, 2-7 (2015). <http://dx.doi.org/10.3389/fdigh.2015.00007>.
- Kylander 1995** Kylander, Britt-Marie. *Le Vocabulaire de Molière dans les comédies en alexandrins*. Göteborg, Acta Universitatis Gothoburgensis (1995).
- Labbé et Labbé 2001** Labbé Dominique, Labbé Cyril. "Inter-Textual Distance and Authorship Attribution. Corneille and Molière" *Journal of Quantitative Linguistics*. 8-3: 213-231 (2001).
- Larthomas 1972/2001** Larthomas, Pierre. *Le Langage dramatique: sa nature, ses procédés*. Paris, Presses universitaires de France (2001).
- Lê et al. 2008** Lê Sébastien, Josse, Julie, Husson François. 2008. "FactoMineR: An R Package for Multivariate Analysis". *Journal of Statistical Software*, 25-1: 1-18 (2008).
- Muller 1967** Muller, Charles. *Étude de statistique lexicale. Le vocabulaire du théâtre de Pierre Corneille*. Paris, Larousse (1967).
- Quiniou, Cellier, Charnois et Legallois 2012** Quiniou Solen, Cellier Peggy, Charnois Thierry, Legallois Dominique. "What About Sequential Data Mining Techniques to Identify Linguistic Patterns for Stylistics?". *International Conference on Intelligent Text Processing and Computational Linguistics (CICLing'12)*, New Delhi, India: 166-177 (2012).
- Ramsay 2008** Ramsay Stephen. "Algorithmic Criticism". In Schreibman, Susan, Siemens, Ray (éd.). *A Companion to Digital Literary Studies*. Oxford, Blackwell Publishing Professional: 477-492 (2008).
- Schmid 1994** Schmid, Helmut. "Probabilistic Part-of-Speech Tagging Using Decision Trees". *Proceedings of International Conference on New Methods in Language Processing*, Manchester, UK (1994).

Schmid 1995 Schmid Helmut. "Improvements in Part-of-Speech Tagging with an Application to German". *Proceedings of the ACL SIGDAT-Workshop*. Dublin, Ireland (1995).

Schöch 2016 Schöch Christof. "Spitzer on Racine, digitally revisited". *The Digital Literary Stylistics Workshop co-located with Digital Humanities 2016*. Cracovie (2016). <https://zenodo.org/record/61434#.V9pja5N96T9> [dernière consultation le 14/07/17].

Stein 2003 Stein Achim. *TreeTagger part-of-speech tags* (2003). <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/french-tagset.html> [dernière consultation le 14/07/17].

Viprey et Ledoux 2006 Viprey, Jean-Marie, et Ledoux Claude-Nicolas. "About Labbé's intertextual distance". *Journal of Quantitative Linguistics*, 13 (2-3): 265-283 (2006).



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Molière's Sganarelles: one name, several syntaxes? [fr]

Élodie Bénard <elodiebenard27_at_gmail_dot_com>, Université Paris-Sorbonne

Francesca Frontini <francesca_dot_frontini_at_univ-montp3_dot_fr>, Université Paul Valéry Montpellier

Abstract

Quantitative studies in the syntax of Molière are not very frequent. The present paper proposes a syntactic analysis of Molière's characters that relies on the extraction of part of speech sequences and their filtering using correspondence analysis. In particular, the methodology is applied to the study of the evolution of the character of Sganarelle, appearing in several of Molière's plays in prose and verse, and often reflecting the prototype of the "bourgeois de Paris", who isn't attuned to the new society of the young reign of Louis XIV, the "mondains". The analysis of the results is carried out with a register perspective, with the aim of identifying syntactic patterns that characterize the different Sganarelles with respect to the type of context in which they find themselves in the different plays. The results show how the evolution of the character can be traced from his first appearance to his later uses by Molière.

Note on Translation

For articles in languages other than English, DHQ provides an English-language abstract to support searching and discovery, and to enable those not fluent in the article's original language to get a basic understanding of its contents. In many cases, machine translation may be helpful for those seeking more detailed access. While DHQ does not typically have the resources to translate articles in full, we welcome contributions of effort from readers. If you are interested in translating any article into another language, please contact us at editors@digitalhumanities.org and we will be happy to work with you.

1

Works Cited

- Benzécri 1982** Benzécri, Jean-Paul. *Histoire et préhistoire de l'analyse des données*. Paris, Dunod (1982).
- Bernet 1983** Bernet Charles. *Le Vocabulaire des tragédies de Jean Racine. Analyse statistique*. Genève-Paris, Slatkine-Champion (1983).
- Biber et Conrad 2009** Biber, Douglas, Conrad, Susan. *Register, Genre and Style*. Cambridge, Cambridge University Press (2009).
- Bloch 2009** Bloch Olivier. *Molière: comique et communication*. Pantin, Le Temps des Cerises (2009).
- Boukhaled 2016** Boukhaled, Mohamed Amine. *On Computational Stylistics: Mining Literary Texts for the Extraction of Characterizing Stylistic Patterns*. Thèse de doctorat (sous la direction de J.-G. Ganascia). Université Pierre et Marie Curie (2016).
- Brunet 2004** Brunet Étienne. "Où l'on mesure la distance entre les distances?". *Texto !* (2004) [en ligne]. Rubrique "Dits et inédits." http://www.revue-texto.net/Inedits/Brunet/Brunet_Distance.html [dernière consultation le 14/07/17].
- Cichocki 2006** Cichocki Wladyslaw. "Geographic Variation in Acadian French: What can Correspondence Analysis Contribute Toward Explanation?". *Literary and Linguistic Computing*, 21 (4): 529-541 (2006).
- Conesa 1983** Conesa Gabriel. *Le Dialogue moliéresque. Étude stylistique et dramaturgique*. Paris, Presses universitaires de France (1983).

- Forestier et Bourqui 2010** Forestier, Georges, Bourqui, Claude (éd.). Molière, *Œuvres complètes*. Paris, Gallimard, Bibliothèque de la Pléiade, 2 tomes (2010).
- Fournier 1999** Fournier Nathalie. "La position des groupes prépositionnels dans l'énoncé, dans *Le Misanthrope* et *George Dandin*". In *Phrases: syntaxe, rythme, cohésion du texte*. Neveu, Frank (éd.). Paris, SEDES (1999).
- Frontini, Boukhaled et Ganascia 2015a** Frontini, Francesca, Boukhaled, Mohamed Amine, Ganascia, Jean-Gabriel. "Molière's Raisonners: a Quantitative Study of Distinctive Linguistic patterns". In *Corpus Linguistics 2015 — Abstract Book*. Lancaster: 114-117 (2015). <http://ucrel.lancs.ac.uk/cl2015/doc/CL2015-AbstractBook.pdf> [dernière consultation le 14/07/17].
- Frontini, Boukhaled et Ganascia 2015b** Frontini, Francesca, Boukhaled, Mohamed Amine, Ganascia, Jean-Gabriel. "Linguistic Pattern Extraction and Analysis for Classic French Plays". Présentation aux Journées ConSciLa, Paris (2015). <http://lipn.univ-paris13.fr/~charnois/conscilaGenres/resumes/frontini.pdf>.
- Frontini, Boukhaled et Ganascia 2017** Frontini, Francesca, Boukhaled, Mohamed Amine, Ganascia, Jean-Gabriel. "Mining for Characterizing Patterns in Literature Using Correspondance Analysis. An Experiment on French Novels". *Digital Humanities Quarterly Proceedings of the Göttingen Dialogue for Digital Humanities 2015*. 11-2 (2017).
- Ganascia 2015** Ganascia, Jean-Gabriel. "The Logic of the Big Data Turn in Digital Literary Studies", *Frontiers in Digital Humanities*, 2-7 (2015). <http://dx.doi.org/10.3389/fdigh.2015.00007>.
- Kylander 1995** Kylander, Britt-Marie. *Le Vocabulaire de Molière dans les comédies en alexandrins*. Göteborg, Acta Universitatis Gothoburgensis (1995).
- Labbé et Labbé 2001** Labbé Dominique, Labbé Cyril. "Inter-Textual Distance and Authorship Attribution. Corneille and Molière" *Journal of Quantitative Linguistics*. 8-3: 213-231 (2001).
- Larthomas 1972/2001** Larthomas, Pierre. *Le Langage dramatique: sa nature, ses procédés*. Paris, Presses universitaires de France (2001).
- Lê et al. 2008** Lê Sébastien, Josse, Julie, Husson François. 2008. "FactoMineR: An R Package for Multivariate Analysis". *Journal of Statistical Software*, 25-1: 1-18 (2008).
- Muller 1967** Muller, Charles. *Étude de statistique lexicale. Le vocabulaire du théâtre de Pierre Corneille*. Paris, Larousse (1967).
- Quiniou, Cellier, Charnois et Legallois 2012** Quiniou Solen, Cellier Peggy, Charnois Thierry, Legallois Dominique. "What About Sequential Data Mining Techniques to Identify Linguistic Patterns for Stylistics?". *International Conference on Intelligent Text Processing and Computational Linguistics (CICLing'12)*, New Delhi, India: 166-177 (2012).
- Ramsay 2008** Ramsay Stephen. "Algorithmic Criticism". In Schreibman, Susan, Siemens, Ray (éd.). *A Companion to Digital Literary Studies*. Oxford, Blackwell Publishing Professional: 477-492 (2008).
- Schmid 1994** Schmid, Helmut. "Probabilistic Part-of-Speech Tagging Using Decision Trees". *Proceedings of International Conference on New Methods in Language Processing*, Manchester, UK (1994).
- Schmid 1995** Schmid Helmut. "Improvements in Part-of-Speech Tagging with an Application to German". *Proceedings of the ACL SIGDAT-Workshop*. Dublin, Ireland (1995).
- Schöch 2016** Schöch Christof. "Spitzer on Racine, digitally revisited". *The Digital Literary Stylistics Workshop co-located with Digital Humanities 2016*. Cracovie (2016). <https://zenodo.org/record/61434#.V9pja5N96T9> [dernière consultation le 14/07/17].
- Stein 2003** Stein Achim. *TreeTagger part-of-speech tags* (2003). <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/french-tagset.html> [dernière consultation le 14/07/17].
- Viprey et Ledoux 2006** Viprey, Jean-Marie, et Ledoux Claude-Nicolas. "About Labbé's intertextual distance". *Journal of Quantitative Linguistics*, 13 (2-3): 265-283 (2006).

