

# L'extraction automatique des motifs dramaturgiques dans les séquences de deux scènes

Marc Douguet  
 marc.douguet@paris-sorbonne.fr  
 OBVIL, Université de Paris-Sorbonne, France

La structuration intrinsèque du texte théâtral en une suite de scènes réunissant chacune un nombre défini de personnages offre des potentialités infinies pour l'analyse quantitative. Ce domaine a déjà fait l'objet de nombreuses études, notamment dans la lignée des travaux de S. Marcus, qui montre qu'une pièce de théâtre peut être modélisée sous la forme d'une matrice binaire où chaque personnage est codé comme une suite de 0 et 1 selon qu'il est absent ou présent dans les scènes successives qui divisent le texte (Marcus 1973).

Jusqu'à présent, l'analyse des matrices dramatiques s'est notamment concentrée sur le nombre de scènes où apparaît un personnage, son nombre d'entrées, les différentes relations de coprésence qu'il entretient avec d'autres personnages, etc. (Marcus 1972; Lafon 1990; Ilsemann 1997) Ces paramètres permettent de qualifier de manière globale les personnages et l'intrigue en appliquant à la dramaturgie des outils empruntés à l'analyse statistique ou à l'analyse des réseaux. Cependant, selon cette approche, l'enchaînement des scènes et l'ordre dans lequel elles sont disposées ne sont pas pris en compte. Ils le sont en revanche dans les travaux de M. Dinu sur la stratégie des personnages, qui cherchent à calculer la probabilité de la réalisation d'une configuration de personnages donnée en fonction de celles qui la précèdent (Dinu 1984).

Nous voudrions ici présenter une nouvelle approche qui, comme celle M. Dinu, prend en compte la progression du texte, mais adopte le point de vue surplombant qui est celui du dramaturge au moment où il dresse le plan de sa pièce et s'intéresse avant tout aux techniques d'écriture dramatique. Nous introduisons pour cela la notion de *motif dramaturgique*, défini comme une séquence d'entrées et de sorties de personnages, abstraction faite de l'identité de ces derniers. Deux suite de scènes ne faisant intervenir aucun

personnage commun ou appartenant à des pièces différentes peuvent donc instancier le même motif.

Prenons comme exemple la matrice dramatique de *Mélite* de Corneille :

Actes	I	II	III	IV	V
Scènes	1 2 3 4 5	1 2 3 4 5 6 7 8	1 2 3 4 5 6 1	2 3 4 5 6 7 8 9 10	1 2 3 4 5 6
TIRCIS	1 1 1 0 1	0 0 0 1 0 0 1 1	<b>0 1 1 1 0</b> 0 0 0 0 0 0 0 0 0 0	0 0 0 0 1 1 1	0 0 0 0 1 1 1
ERASTE	1 1 1 0 0	1 1 1 0 1 1 1 0	<b>0 0 0 0 0 0 0 0 0 0 0 0 1 1</b>	0 1 1 0 1 0	1 1 0 0 0 1 1
MÉLITE	0 1 0 0 0 0	1 0 0 0 0 1 1 0	<b>0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0</b>	0 0 0 0 1 1 1	0 0 0 0 1 1 1
CHLORIS	0 0 0 1 1	0 0 0 1 0 0 0 0	<b>0 0 0 1 1 1 0 1 1 1 0 0 0 0 0 0</b>	1 0 0 1 0 1 1	1 0 0 1 0 1 1
PHILANDRE	0 0 0 1 1	0 0 0 0 0 1 0 0	<b>1 1 0 0 0 1 0 0 0 0 0 0 0 0 1 1 0</b>	0 0 0 0 1 0 0	0 0 0 1 0 0 0
CLITON	0 0 0 0 0 0 0 0 0 0	1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	1 0 1 0 0 0 0 0	0 1 0 0 0 0 0
LA NOURRICE	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	1 0 0 1 0 0 0 0 0 0	0 1 1 0 0 0 1
LISIS	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	1 1 0 0 0 0 0 0	1 0 0 0 0 0 0

Les scènes III, 1-3, III, 3-5 et IV, 7-9 instancient un même motif, représenté en gras, qui peut être exprimé sous la forme d'une sous-matrice

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$$

ou sous la forme plus lisible d'une séquence codifiée de caractères où A désigne le personnage qui apparaît en premier, et B celui qui apparaît en second, les scènes étant délimitées par "/" et les personnages, à l'intérieur d'une scène, par "-" :

$$A/A-B/B$$

La comparaison de ce motif avec un autre motif de trois scènes et deux personnages permet de comprendre l'intérêt de cette méthode. Ce motif, que l'on rencontre dans *Mélite*, II, 1-3, est le suivant :

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

soit

$$A/A-B/A$$

Notre hypothèse est qu'indépendamment de la longueur de chaque scène composant le motif et de la répartition de la parole entre les personnages, chaque motif présente toujours, dans la diversité de ses réalisations, des enjeux comparables. En l'occurrence, les deux motifs que nous venons de citer illustrent deux techniques de composition bien distinctes. Soit un personnage reste en permanence sur le plateau, soit le dramaturge renouvelle entièrement le personnel dramatique, profitant de la scène centrale pour faire se croiser les personnages qui se relaient sous les yeux des spectateurs. Dans le premier cas, un personnage sert de pivot à l'action, et en rencontre successivement

plusieurs autres. Ce faisant, il guide l'attention du public. Dans le second, le point de vue du spectateur dépasse celui de l'ensemble des personnages : il en sait et en voit plus que chacun d'eux individuellement.

L'étude des motifs dramaturgiques devient encore plus intéressante quand on regroupe ensemble les personnages qui, à l'intérieur d'un motif, ont une distance scénique nulle (Marcus, 1972), c'est-à-dire sont toujours présents et absents en même temps. Ainsi, *Mélie*, I, 1-3 repose sur le motif

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

Or malgré le fait qu'il y a deux personnages présents en permanence, et non un seul, ce motif n'est pas fondamentalement différent de

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

Dans les deux cas, une confrontation centrale est entourée par deux scènes présentant un caractère plus "privé", qui permettent de comparer l'attitude d'un personnage ou groupe de personnages avant et après une rencontre avec un autre personnage ou groupe de personnages et de faire entendre le discours qu'ils tiennent en l'absence (et bien souvent au sujet) de ce dernier.

Méthodologiquement, cette approche est comparable à l'étude des motifs syntaxiques développée dans le cadre de la stylistique computationnelle (Boukhaled 2015), si ce n'est que chaque élément du motif se définit non par le choix unique d'une classe de mot, mais par le choix multiple d'un ensemble de personnages. Elle possède donc *a priori* les mêmes applications, notamment la qualification du style d'un auteur (entendu ici comme style de composition, et non comme style d'écriture, et dont l'unité de base n'est pas le mot, mais la scène).

Nous avons constitué une base de données comportant tous les motifs de moins de huit scènes extraits d'un corpus de plus de 200 pièces de théâtre françaises écrites entre 1630 et 1680. Le choix de cette période se justifie par son importance dans l'histoire du théâtre français, et par le renouvellement des techniques dramaturgiques qu'elle permet d'étudier (passage du théâtre "baroque" au théâtre "classique"). Pour interroger cette base, nous avons développé un [moteur de recherche](#) qui, pour un motif donné, renvoie sa fréquence d'apparition sur l'ensemble du corpus,

son évolution, ainsi que sa répartition par genre et par auteur.

Il ne s'agira pas ici de présenter les détails techniques de ce travail, mais de livrer les premières conclusions auxquelles nous sommes parvenus. Une des pistes possibles pour exploiter cette base de données est d'étudier l'évolution et la répartition de l'ensemble des motifs possibles pour un nombre donné de personnages et un nombre donné de scènes. Pour cette première étude, nous nous intéresserons aux motifs composés de deux scènes. La combinatoire est facile à produire. Si l'on regroupe les personnages qui sont présents et absents en même temps, seuls trois personnages ou groupes de personnages peuvent intervenir (un quatrième personnage aurait nécessairement une distance scénique nulle avec un autre), et quatre motifs sont possibles:

A/A-B  
A-B/A  
A/B  
A-B/A-C

On constate, dans notre corpus, la disparition progressive du motif A/B (aucun personnage commun entre deux scènes contiguës), qui s'explique par l'exigence croissante de continuité dramatique. Un autre phénomène est plus inattendu et n'a pas encore, à notre connaissance, été vraiment étudié : deux scènes sont plus souvent reliées par une entrée (A/A-B) que par une sortie (A-B/A). Autrement dit, les personnages ont tendance à entrer séparément, mais à sortir de manière groupée, ce que nous tenterons d'expliquer en convoquant la notion de "tension dramatique". Enfin, on constate une légère augmentation puis une légère baisse des entrées et des sorties simultanées (A-B/A-C), dont nous essaierons de rendre compte en mettant ce phénomène en corrélation avec l'apparition d'une technique d'enchaînement plus complexe reposant sur un tuilage des entrées et des sorties à l'échelle de trois scènes successives : pour relier deux dialogues différents, les dramaturges préfèrent de plus en plus intercaler une scène de transition à trois personnages (A-B/A-B-C/A-C) plutôt que faire entrer et sortir deux personnages en même temps sans que ceux-ci ne se parlent (A-B/A-C).

Ces considérations générales nous mèneront à étudier le détail de la répartition des quatre motifs par genre et par auteur. Enfin, nous distinguerons les scènes qui appartiennent au même acte et celles qui se situent de part et d'autre d'un entracte. Les entractes sont en effet rarement pris en compte dans l'analyse des matrices dramatiques, alors qu'ils jouent un rôle essentiel. Une des règles de la dramaturgie classique

voudrait par exemple qu'un personnage présent à la fin d'un acte ne le soit pas au début du suivant (Aubignac 1657). Une étude statistique vérifie en partie cette règle, mais montre aussi qu'elle est très fréquemment enfreinte.

Toutes ces questions visent à définir une norme d'écriture (pour une époque, un genre, un auteur). Mais les outils numériques ne se limitent pas à l'analyse quantitative : ils enrichissent également l'analyse littéraire en permettant de découvrir à l'intérieur d'un vaste corpus des cas particuliers, qu'il serait quasiment impossible de repérer sans procéder à une extraction automatique. On se penchera ainsi sur les pièces où l'un des quatre motifs est sur- ou sous-représenté, en se demandant quels sont les enjeux de cette anomalie pour la conduite de l'action et l'effet produit sur le spectateur.

## Bibliographie

**Aubignac, abbé d'** (1657). *La Pratique du théâtre*. Paris: Sommaville.

**Boukhaled, A., Frontini, F. et Ganascia, J.-G.** (2015). Une Mesure d'intérêt à base de surreprésentation pour l'extraction des motifs syntaxiques stylistiques. *Actes de la 22e conférence sur le Traitement Automatique des Langues Naturelles*.

**Dinu, M.** (1984). Entropy and Prediction in the Study of Theatre. *Poetics* 13, pp. 57-70.

**Ilsemann, H.** (1995). Computerized Drama Analysis. *Literary and Linguistic Computing* 10(1), pp. 11-21.

**Lafon, D.** (1990). *Le Chiffre scénique dans la dramaturgie moliéresque*. Ottawa/Paris: Presses de l'Université d'Ottawa/Klincksieck.

**Marcus, S.** (1973). *Mathematische Poetik*. București/Frankfurt am Main: Editura Academiei/Athenäum Verlag.

**Marcus, S.** (1972). Stratégie des personnages dramatiques. *Sémiologie de la représentation*. Bruxelles: Complexe, pp. 73-95.