

L'apport du numérique à l'analyse métrique et rythmique

Valérie Beaudouin

I3-SES, CNRS, Télécom ParisTech

Des humanités numériques littéraires ?

Cerisy, juin 2017

La question qui est au cœur de cette recherche est la suivante en quoi la métrique et le rythme du vers sont une composante à part entière du style ? Je pars d'un triple constat :

- 1) les textes en vers sont souvent analysés et traités comme s'ils étaient en prose ; la description de la versification est rarement reliée à ce que dit le poème ;
- 2) la compétence métrique se perd avec l'essor du vers libre et des formes poétiques non métriques tout au long du XX^e siècle. Le modèle n'est plus intégré dans nos mémoires internes, on a du mal à percevoir la dimension transgressive de certains vers de Rimbaud, Mallarmé... Le rythme de l'alexandrin n'est plus un donné.
- 3) Dans une perspective roubaldienne, le poème dit quelque chose sur sa forme : il faut pouvoir entendre.

Pourquoi passer par la machine ?

Le décompte manuel des vers est très long et fastidieux ; si la reconnaissance est simple pour l'œil humain, il devrait être possible de faire faire le travail par une machine. Dans la Vieillesse d'Alexandre (Roubaud, 1978), Jacques Roubaud propose une tentative de formalisation qui prépare déjà le passage à la machine.

Dans le cadre de ma formation en statistique, je découvre l'analyse des données à la française, dans la lignée des travaux de Benzécri, et un peu plus tard les méthodes de lexicométrie et de statistique textuelle (Benzécri, 1981; Lebart and Salem, 1994 ; Beaudouin, 2016) qui permettent d'exploiter de grands corpus : identification des champs lexico-sémantique, de manière inductive, à travers la question des cooccurrences. La machine sert à accéder à la sémantique des textes.

Je cherche donc à créer un pont entre le monde de la littérature et celui de la statistique, que tout éloigne.

Ma perspective est double. Premièrement, il s'agit de créer un outil informatique pour l'analyse métrique qui permette d'identifier le mètre et de décrire le rythme des vers, autrement dit d'observer la pratique effective du vers. Deuxièmement, il s'agit de mettre en relation l'approche métrique et les méthodes de statistique textuelle, afin de montrer les relations entre le rythme des vers et les champs lexico-sémantiques.

Se faisant, il s'agit de relier le domaine de la métrique et celui de la statistique textuelle, ce qui n'a jamais été fait.

Après une présentation rapide des corpus et des outils construits pour une analyse automatique du vers, j'expliquerai les présupposés théoriques du modèle qui a été implémenté dans les outils. Je présenterai ensuite trois ensembles de résultats : l'un sur les régularités

rythmiques qui ont pu échapper à la codification normative, l'autre sur l'évolution du vers et le dernier sur la relation entre rythme et sémantique¹.

Chaîne d'analyse du vers

Pour mettre en place une chaîne de traitement du vers, il faut : un corpus, des algorithmes, une base de données et des outils d'analyse statistique.

Nous avons tout d'abord besoin d'un corpus numérisé, bien documenté, bien étiqueté. En effet, il est essentiel de pouvoir bien indexer tous les éléments paratextuels (auteur, pièce, genre, année, acte, scène, personnage) mais aussi distinguer les vers de la prose. Évidemment, les noms de personnages ou les didascalies ne doivent pas être analysés comme des vers. Un codage avec les normes de la TEI pour annoter les textes est indispensable.

Il est également nécessaire de recomposer les vers prononcés par plusieurs personnages (stichomythie) pour éviter de mauvaises analyses métriques.

```
**** *Corneille *Mélite *comédie *1629
**ACTE_I
*SCÈNE_PREMIÈRE.
***ÉRASTE.
Je te l'avoue, ami, mon mal est incurable ;
Je n'y sais qu'un remède, et j'en suis incapable :
```

Mon corpus est constitué des œuvres de théâtre de Corneille et Racine, soit 46 pièces. Chaque pièce a en moyenne 1500 vers, et le corpus complet fait 75000 vers, dont 97% sont des alexandrins organisés par paires de vers rimant ensemble avec alternance de genre. On ne peut donc pas considérer dans le programme qu'il s'agit d'alexandrins, ce qui pourrait simplifier le problème (faire le décompte avec l'objectif d'arriver à 12).

Le vers du XVIIe est le plus réglé de toute l'histoire de la littérature, soumis à un arsenal de règles strictes portant sur sa longueur, son découpage interne, la césure... Il constitue un candidat idéal pour le traitement par la machine.

Un corpus complémentaire de 30 000 vers provenant de différents auteurs du XVIIIe et XIXe siècles (Voltaire, Hugo, Baudelaire, Mallarmé, Rimbaud) permet d'explorer l'altération progressive de la structure du vers.

La question de l'édition mérite qu'on s'y attarde. Idéalement, il faudrait travailler sur les éditions originales et avoir des outils adaptés à la langue de l'époque. Mais, dans les années 1990, les seules éditions numériques disponibles étaient pour Corneille, les éditions de Charles Marty-Laveaux et pour Racine, celles de P. Mesnard. Et surtout, les outils de traitement automatique de la langue étaient conçus pour le français contemporain. L'utilisation d'outils appliqués aux éditions modernisées constitue un compromis acceptable à condition de retourner aux éditions d'origine.

¹ Cette communication s'appuie sur les travaux de la thèse que j'ai soutenue en 2000 et qui a été publiée (Beaudouin and Yvon, 1996).

Ensuite, nous avons besoin d'un ensemble d'algorithmes capables d'analyser la structure métrique du vers et l'organisation de la rime.

Trois stratégies peuvent être utilisées pour construire l'outil

On peut partir de la représentation graphique du vers et construire des règles pour identifier les syllabes. Roubaud dans son programme *Dynastie* propose une formalisation précise du passage de la chaîne graphique à l'identification des syllabes métriques (Roubaud, 1986). Mais l'orthographe française est si compliquée et irrégulière que la tâche est ardue. En espagnol, cette approche a donné de très bons résultats comme le montrent les travaux de Gervas (Gervás, 2000). L'anamètre développé par Eliane Delente et Richard Renault va dans cette direction en identifiant les voyelles métriques (Delente and Renault, 2015).

Une seconde stratégie consiste à utiliser les méthodes d'apprentissage automatique. JG Ganascia, pionnier dans ce domaine, m'avait accueillie dans son laboratoire en 1993 pour tester leur système, *Charade* (Ganascia, 1987). Mais la graphie du français est particulièrement complexe comparée à celles de langues latines comme l'italien et l'espagnol, ce qui rend la modélisation fort difficile ; les systèmes d'apprentissage en sont encore à leur début.

La troisième stratégie que j'ai adoptée, consiste à utiliser des outils sophistiqués de TAL (traitement automatique des langues) : des systèmes de phonétisation (Yvon, 1996) et d'analyse syntaxique (Constant, 1991) conçus pour la langue et de les adapter aux spécificités du vers. Cela impliquait de détourner des outils industriels pour étudier le vers. J'ai pour cela travaillé en collaboration avec François Yvon, alors chercheur à Télécom ParisTech (Beaudouin and Yvon, 1996).

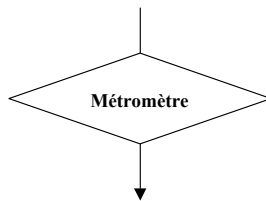
Pour l'analyse de la rime, j'ai utilisé les travaux sur les graphes et les outils d'analyse de réseaux sociaux pour étudier la relation de rime en posant que deux mots rimant ensemble sont en relation².

Le métromètre analyse la structure lexicale, syntaxique et prosodique du vers en découpant le vers selon ses positions métriques, et en décrivant chacune des syllabes au niveau phonétique, lexical, syntaxique et prosodique comme on peut le voir sur la figure 1.

² Je renvoie au chapitre 3 de (Beaudouin, 2002) consacré à la rime.

De cette nuit, Phénice, as-tu vu la splendeur ?

Racine, *Bérénice*, vers 302.



Représentation multidimensionnelle :

| | | | | | | | | | | | | |
|---------------------------|-----|-----|-------------------|-------|-----|-------|---|-----|-----|-----|---------|-------|
| Syllabes métriques | d \ | s ´ | t \f ³ | n ¥ i | f e | n i s | a | t y | v y | l a | sp l â~ | d œ r |
| Voyelles métriques | \ | ´ | \f | ¥ i | e | i | a | y | y | a | â~ | œ |
| Repérage des fins de mots | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 |
| Catégories syntaxiques | 7 | 6 | 6 | 0 | 4 | 4 | 1 | 6 | 2 | 6 | 0 | 0 |
| Marquage accentuel | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 |
| Nb syllabes | 12 | | | | | | | | | | | |

Clef de lecture du tableau : La cinquième syllabe métrique correspond à la séquence phonémique /f e/, elle n'est pas située en fin de mot (*cf.* repérage des fins de mots = 0) et appartient à un nom propre (*cf.* catégorie syntaxique = 4).

Repérage des fins de mot :

| | |
|---------------------|--------------------|
| 0 = syllabe interne | 1 = syllabe finale |
|---------------------|--------------------|

Catégories syntaxiques (selon l'analyseur syntaxique Sylex)

| | |
|-------------------------------|-------------------------|
| 0 = Nom | 5 = Sigles |
| 1 = Verbe | 6 = Déterminant, pronom |
| 2 = Adjectif, participe passé | 7 = Préposition |
| 3 = Adverbe | 8 = Conjonction |
| 4 = Nom propre | 9 = Pronom relatif |

Marquage accentuel

| | |
|--|--------------------------------------|
| 0 = syllabe ne portant pas d'accent de mot | 1 = syllabe portant un accent de mot |
|--|--------------------------------------|

Figure 1. Fonctionnement du métromètre et représentation métrique

Enfin, nous avons besoin d'une base de données et d'outils statistiques.

Pourquoi une base de données? Il est essentiel de pouvoir intégrer dans la même base les résultats obtenus par différents processus d'analyse : la structure métrique produite par le métromètre, les réseaux de rime, le résultat des outils de fouille de texte (champs lexicaux sémantiques). L'objectif de l'intégration est de pouvoir établir des ponts entre les différents niveaux linguistiques. Les outils d'interrogation de la base de données et les outils statistiques sont utilisés pour explorer et tester des hypothèses.

Résumons la chaîne de traitement. On a un corpus étiqueté en entrée, des algorithmes qui produisent des sorties, une base de données qui intègre tous les résultats, des traitements pour explorer des hypothèses.

³ ef : désigne un e muet « faible », susceptible d'élosion selon son environnement.

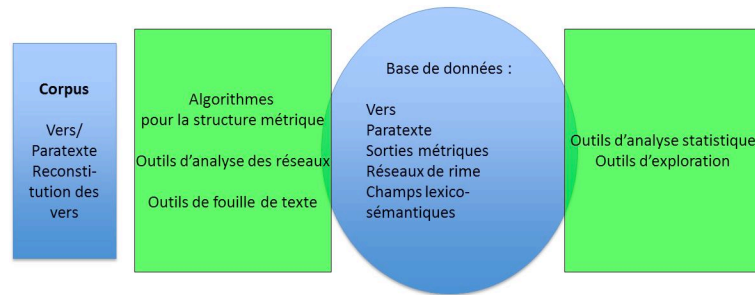


Figure 2. La chaîne de traitement

La démarche adoptée date de la fin des années 1990 et utilise les technologies de pointe de l'époque. Si c'était à refaire aujourd'hui, voici les orientations que j'adopterai. En ce qui concerne le corpus, j'utiliserai l'annotation XML-TEI et j'utiliserai des éditions à la fois en orthographe d'époque et en orthographe modernisée avec alignement des deux versions. En termes d'outils, j'opterai pour des outils open source (dépendre d'outils privés rend l'avenir incertain). D'autres options permettraient d'affiner l'analyse : utiliser une représentation graphique ET phonétique des syllabes métriques (puisque le vers est fait pour un « œil-oreille », Roubaud), marquer la structure dramatique, syntaxique et métrique pour étudier concordances et discordances entre les niveaux. Enfin, il serait intéressant de trouver un système qui permette de repérer des structures métriques dans la prose.

Présupposés théoriques inscrits dans les outils

Un outil informatique n'est jamais neutre : il repose toujours sur des hypothèses théoriques qu'il est nécessaire d'explicitier. Dans notre cas, trois d'entre elles méritent d'être explicitées : la conception du rythme, le modèle du vers, la relation entre langue et métrique.

Une conception du rythme

Dans un article de 1951, Benvéniste réexamine l'étymologie du mot rythme (Benvéniste, 1966) : la plupart des définitions du rythme s'appuient sur une interprétation inexacte de l'étymologie, qui associe le concept au mouvement régulier des flots. Or le verbe *couler* (ὄρει~ν) dont dérive *rythme* (ὄρυθμός) en grec, n'a jamais été utilisé pour désigner le mouvement de la mer. Benvéniste revient aux usages premiers du terme. Le mot *rythme* en grec serait utilisé par Démocrite pour désigner la « forme », « en entendant par là la forme distinctive, l'arrangement des parties dans un tout » [p. 330]. Il désigne des configurations toujours susceptibles de changer, d'où l'importance de l'étymologie, *couler*. Platon modifie le sens du terme et lui donne son sens actuel : « configuration des mouvements ordonnés dans la durée » [p. 335].

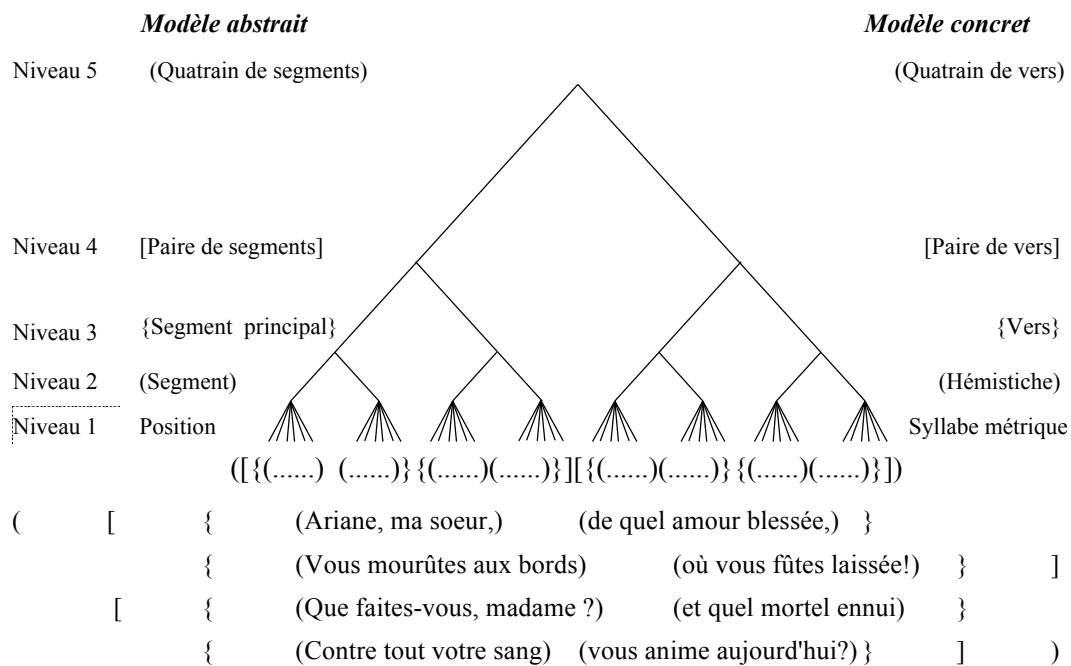
Pour Paul Fraisse qui travaille sur la perception du rythme, le concept de rythme articule les notions de périodicité et de structure (Fraisse, 1974).

Nous retiendrons donc du mot *rythme* qu'il désigne la répétition de formes ou configurations dans la séquence temporelle. Le rythme se situe entre le même et le différent, entre la répétition à l'identique et la distribution aléatoire, entre la marche militaire et le désordre, pour reprendre les éléments de définitions de la théorie du rythme de Roubaud et Lusson.

Théorie du rythme

Je m'appuie sur la théorie du rythme développée par Roubaud et Lusson (Lusson et Roubaud, 1974; Roubaud, 1991) et en particulier sur le modèle hiérarchisé du vers (figure 3), qui permet : 1) de comprendre les articulations entre les différents niveaux d'analyse du vers : mètre et rime ; 2) de mettre en résonance différentes dimensions d'analyse : niveau phonétique, lexical, sémantique et prosodique Ce modèle hiérarchisé permet aussi de voir comment le mètre s'articule avec la syntaxe, mais aussi avec la structure dramatique.

La question des alignements et désalignements entre structures métriques, syntaxiques et dramaturgiques est passionnante à envisager.



Phèdre, vers 253-256

Figure 3. Modèle hiérarchisé du vers

Langue ordinaire, langue des vers

Nous avons utilisé des outils de linguistique conçus pour la langue en les adaptant au vers. Les travaux de Milner ont été très utiles pour penser cette adaptation (Milner, 1974; Milner and Regnault, 1987).

Son approche est extrêmement élégante : pour lui, quelques lois règlent le traitement du e muet, de la liaison et de l'accent dans la langue et dépendent d'une unité linguistique qui est le mot phonologique. Par exemple, en langue, le e muet tombe devant voyelle (loi 1) et tombe en fin de mot phonologique (loi 2). Ces mêmes lois s'appliquent au vers en considérant que le vers tout entier constitue un mot phonologique. A l'intérieur du vers, d'après la première loi, tous les e muets tombent devant voyelle et en fin de vers, le e muet tombe, d'après la deuxième loi. Nous avons ainsi quelques règles simples pour le traitement du e muet et de la liaison.

Pour l'accent, dont l'existence même en français est très controversée, j'ai fait la synthèse de différents travaux pour proposer un marquage de la dernière position tonique des mots pleins (nom, verbe, adjectif, adverbe).

L'outil est porteur d'une hypothèse forte. Dans notre cas, le vers est considéré comme une unité de langue et tous les marquages utilisés pour le décrire sont linguistiques (phonétiques, lexicaux, morpho-syntaxiques et prosodiques). Aucune hypothèse métrique n'est posée a

priori, pas même sur la fin de vers ou l'hémistiche. Je cherche à faire émerger la figure rythmique du vers à partir de marqueurs linguistiques, avec l'hypothèse que l'ordonnement de la langue dans le mètre suffit à constituer le rythme. Le motif rythmique du vers est réalisé par un certain arrangement des parties du discours et des accents le long des positions métriques.

Je n'ai donc utilisé que des marquages linguistiques pour caractériser les syllabes métriques : je ne présuppose pas l'existence d'une césure, je la laisse émerger des données.

Venons en à quelques résultats

Figure de l'alexandrin classique

L'alexandrin est communément décrit comme un vers de 12 syllabes métriques avec une césure coupant le vers en deux hémistiches.

Puisque Vénus le **veut**, de ce sang déplorable
 Je péris la dern**ière**, et la plus misérable.
 Racine, Phèdre, acte I, scène III

La césure est décrite par une variété de règles. Certains traités disent que le e muet, les prépositions et les pronoms sont interdits. Elle est définie de manière floue par Boileau :

Que toujours, dans vos vers, // le sens coupant les mots
 Suspende l'hémistiche, // en marque le repos.

L'idée de repos est considérée comme centrale et doit coïncider avec une frontière syntaxique.

La répartition des différents éléments linguistiques, confirme ce qui est prescrit dans les traités mais aussi des phénomènes qui n'ont jamais été relevés.

Regardons à présent la répartition de quelques traits linguistique dans le vers. Et commençons par les voyelles et par le e muet, phonème instable qui disparaît à la fin d'un mot phonologique ou quand il est suivi par un mot commençant par une voyelle.

Je le vis, je rougis, je pâlis à sa vue.
 Un trouble s'éleva dans mon âme éperdue.
 Racine, Phèdre, acte I, scène III

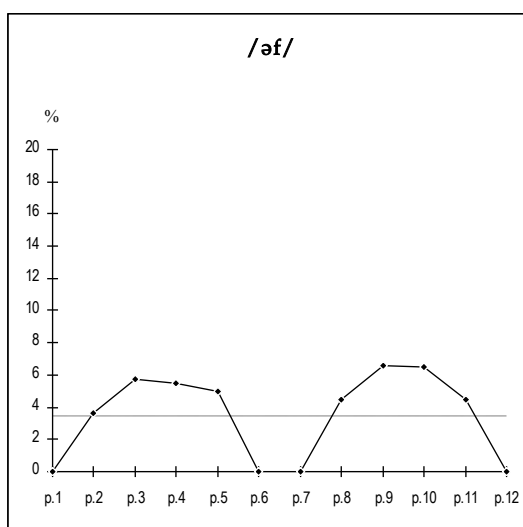
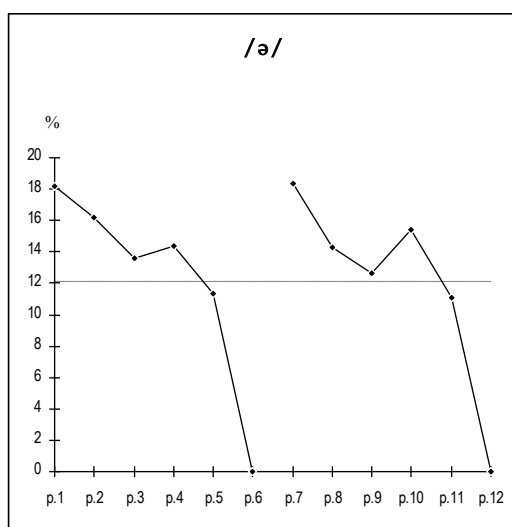


Figure 4. Répartition des e muets

La courbe confirme ce que nous savons déjà : le e muet (faible, partie droite de la figure) est absent en P6 et P12, mais aussi en P1 et en P7. Cela signifie que l'hémistiche ne peut jamais s'achever sur une position métrique non tonique. Il n'y a aucune exception à cette règle au XVIIe siècle.

Des vers comme ceux de Rimbaud n'auraient jamais pu être écrits au XVIIe :

Périssez ! puissance, justice, histoire, à bas !
 Et de braise, et mille meurtres, et les longs cris
 Rimbaud, Qu'est-ce pour nous Mon Cœur...

La répartition des voyelles sur les 12 positions métriques montre aussi la symétrie entre le premier et second segment métrique, qui témoigne d'une forme d'autonomie de chaque segment métrique

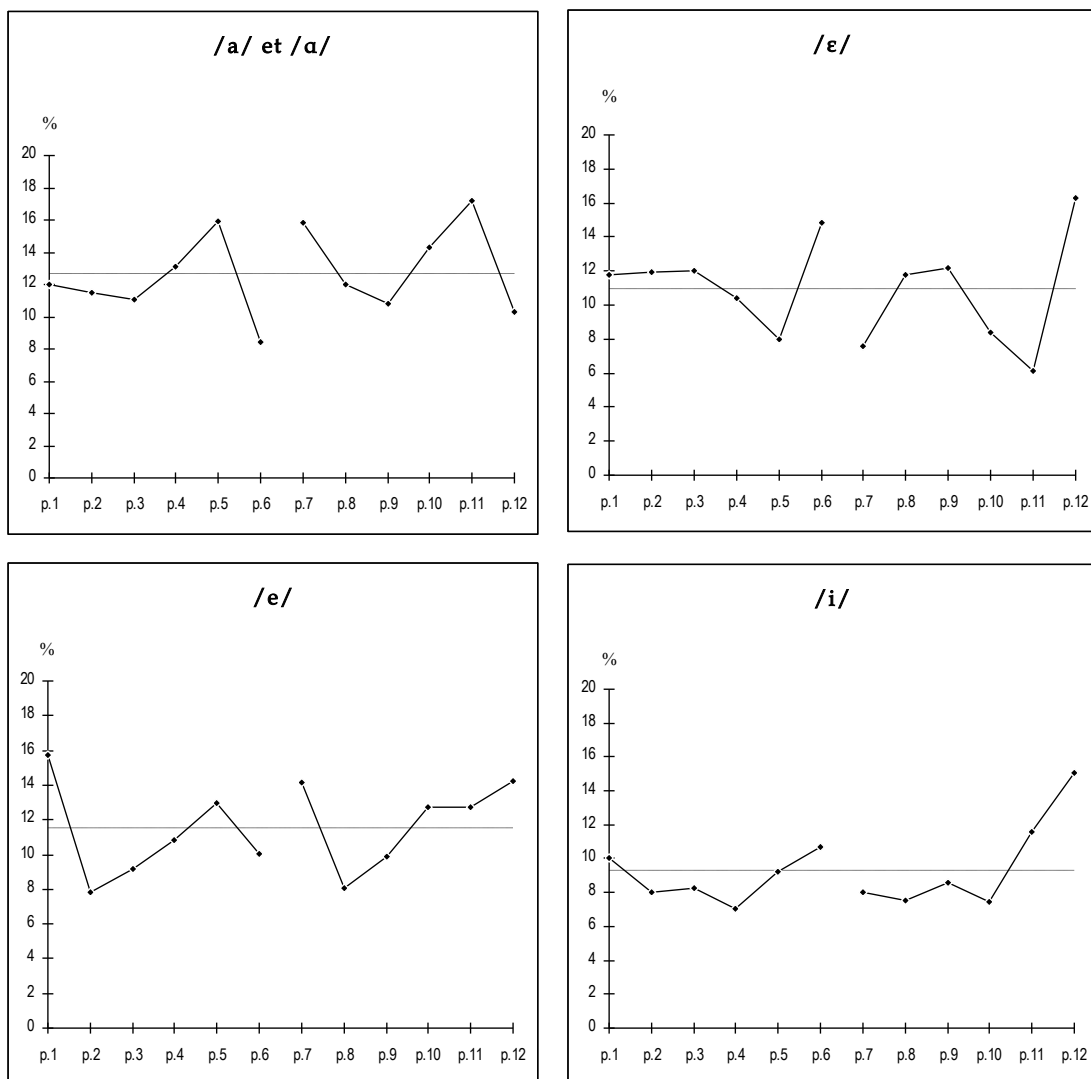


Figure 5. Répartition de quelques voyelles

Quand on regarde la longueur des syllabes (en considérant le nombre de phonèmes), on observe un phénomène étonnant. La première syllabe de chaque hémistiche est très légère (moins de deux phonèmes en moyenne) tandis que la dernière est particulièrement épaisse (au

moins trois phonèmes). Ceci montre un écart avec le modèle abstrait du vers qui est constituée de deux séquences de six syllabes CV (Consonne Voyelle).

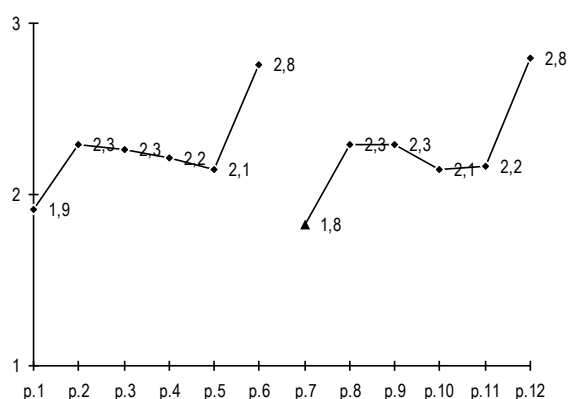


Figure 6. Nombre moyen de phonèmes par syllabe métrique

La répartition des fins de mots et des parties du discours confirme les effets de symétrie entre les deux hémistiches. La sixième et la douzième positions coïncident toujours avec la fin d'un mot, jamais un mot ne traverse l'hémistiche (Figure 7). Un vers comme celui de Rimbaud s'éloigne violemment du modèle classique :

Nous la voulons ! **Ind**ustriels, princes, sénats,
Rimbaud, Qu'est-ce pour nous Mon Cœur

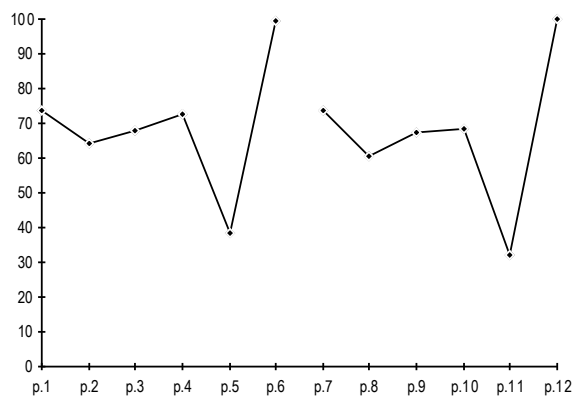


Figure 7. Répartition des fins de mot

La distribution des mots pleins (noms, verbes, adjectifs et adverbes) montre qu'ils occupent systématiquement les 6^{ème} et 12^{ème} positions. Les mots grammaticaux sont absents des fins d'hémistiches. On trouve moins de 10 exceptions, l'une d'entre elles que l'on doit à Corneille est particulièrement frappante :

Il monte à son retour, il frappe à la porte : **elle**
Transit, pâlit, rougit, me cache en sa ruelle,
Corneille, *Le Menteur*, vers 621-622.

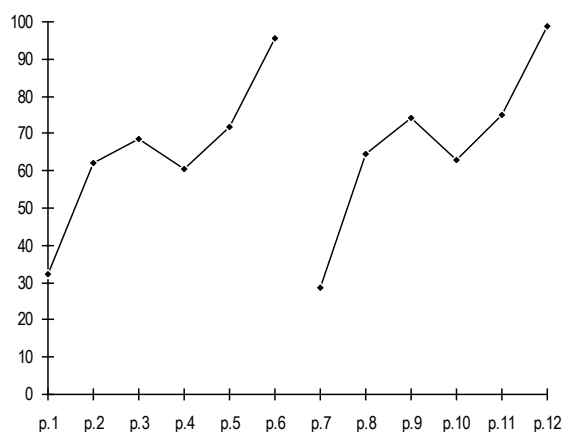


Figure 8. Répartition des « mots pleins » (nom, verbe, adjectif, adverbes)

Enfin, la répartition des marquages accentuels le long du vers nous donne les résultats les plus intéressants. La règle appliquée est la suivante : la dernière syllabe pleine de mots appartenant aux catégories noms, verbes, adjectifs adverbess porte un accent (« 1 »).

Je le vis, je rougis, je pâlis à sa vue.

0 0 1 0 0 1 0 0 1 0 0 1

H1_001001 H2_001001

Un trouble s'éleva dans mon âme éperdue.

0 1 0 0 0 1 0 0 1 0 0 1

H1_010101 H2_001001

A première vue on voit une symétrie entre les deux courbes. La dernière syllabe de l'hémistiche est toujours accentuée tandis que ses voisines 5 et 7, 11 et 1 sont très rarement accentuées. L'accent est renforcé par la faiblesse des positions qui l'entourent.

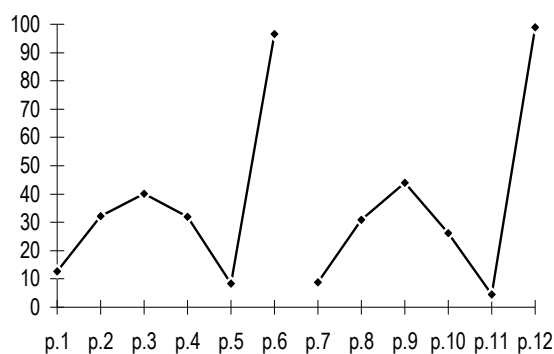


Figure 9. Répartition des « accents »

Si l'on examine la séquence des marquages accentuels sur chaque hémistiche, il apparaît que la majorité des hémistiches présentent un rythme de type 010101 avec des accents sur les positions paires ou de type 001001 avec des accents sur les multiples de trois, les formes alternatives ou irrégulières représentant cependant plus d'un tiers des cas.

La comparaison des deux hémistiches montre que la forme rythmique du second est beaucoup plus régulière (71% de vers de types 010101 ou 001001) que celle du premier (61%), comme si en s'approchant de la rime, le vers gagnait en clarté rythmique. Ceci confirme des résultats

de Gasparov dans d'autres langues: le rythme s'affirme au fil du vers avec moins d'accents mais aux places stratégiques (Gasparov, 1987).

Cette asymétrie montre que l'alexandrin n'est pas la simple concaténation de deux segments métriques. Comme l'écrit Roubaud, il est à la fois un et deux segments métriques.

| Premier hémistiche | % | Second hémistiche | % |
|--------------------|-------|-------------------|-------|
| H1_001001 | 24 | H2_001001 | 32 |
| H1_010101 | 37 | H2_010101 | 38 |
| H1_autres | 39 | H2_autres | 30 |
| Effectifs | 74917 | Effectifs | 74917 |

Tableau 1. Répartition des formes rythmiques des hémistiches

Pour conclure, le premier phénomène marquant est l'étonnante symétrie de la distribution des marquages, quel que soit le marquage, entre les deux hémistiches (si on observe un pic de distribution sur la 3^{ème} position du premier hémistiche, on en verra un autre sur la 3^{ème} du second hémistiche). Les allures des courbes sont identiques, ce qui montre que le segment métrique constitue une entité à part entière.

Le deuxième est le mécanisme de convergence et d'amplification entre les marquages qui proviennent d'ordres linguistiques différents. En fin de segment métrique, la syllabe est toujours accentuée, correspond toujours à un mot plein, toujours à la fin d'un mot, est beaucoup plus riche en phonèmes... La synchronisation des marquages produit un effet d'amplification.

Le troisième phénomène tient au contraste entre la position éminente de la fin d'un segment métrique et ses voisines qui présentent un profil toujours dépressif. Le marquage accentuel de la fin du segment métrique est renforcé par la faiblesse des positions voisines (5^{ème} et 7^{ème} position à l'hémistiche, 11^{ème} et 1^{ère} du vers suivant en fin de vers).

Si l'unité métrique centrale est l'hémistiche, comment l'unité du vers est-elle assurée ? : elle l'est par les mécanismes complexes d'ajustement à la césure (deux segments métriques de six positions ne peuvent former un vers que si à la jonction on a une césure classique, définie principalement par l'interdiction du e muet post-vocalique devant consonne), par le fait que la deuxième partie du vers a un rythme beaucoup plus structuré que la première et évidemment par la rime qui délimite le vers.

Evolution de l'alexandrin

Sur l'évolution de l'alexandrin, je renvoie au très beau livre de J. Roubaud, la Vieillesse d'Alexandre (Roubaud, 1978) qui montre comment s'est peu à peu altéré l'édifice (la métaphore architecturale est filée tout au long de l'essai) de l'alexandrin (dont les principaux artisans sont Hugo, Baudelaire et Nerval d'une part Mallarmé, Lautréamont et Rimbaud de l'autre). D'autres travaux se sont intéressés aux transgressions métriques dans le vers comme ceux B de Cornulier (Cornulier, 1982) ou de J.-M Gouvard (Gouvard, 2000).

Certains résultats permettent simplement de confirmer les travaux de Jacques Roubaud, en montrant en particulier que les transgressions portent exclusivement sur l'hémistiche et pas sur la fin du vers, dont la nature n'est pas altérée. Ainsi, la part des mots-outils (déterminants, prépositions, pronoms, etc.) en sixième position, qui est nulle à l'époque classique, ne cesse d'augmenter (atteignant 12% des vers chez Mallarmé) tandis qu'elle reste quasiment nulle en fin du vers (figure 10). Cette figure montre le changement de relation entre la syntaxe et le mètre au XIXe : la syntaxe prend de l'autonomie à l'intérieur du vers, mais le respect de l'unité du vers est respectée, ce qui dure très longtemps, puisque le vers libre, en dépit de ses déclarations d'affranchissement par rapport à la contrainte du vers continuera de respecter, comme le montre Roubaud, « la concordance des frontières syntaxiques et métriques » (Roubaud, 1978, 2000) p. 127.

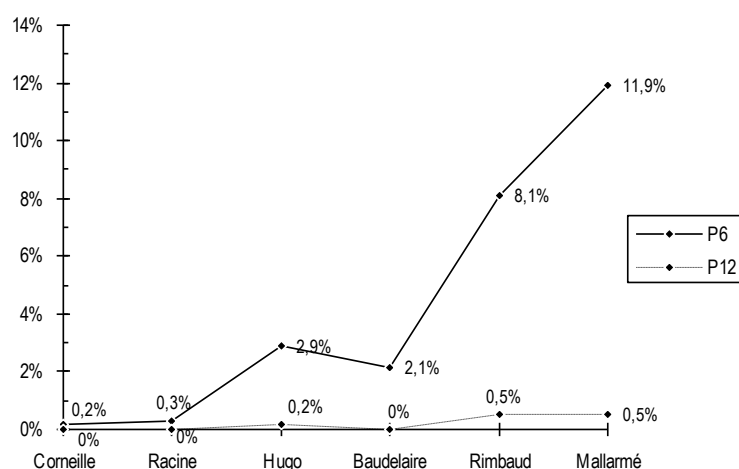


Figure 10. Proportion de mots-outils en 6ème et 12ème positions

L'étude temporelle fait émerger un phénomène étonnant : le vers s'éloigne du modèle du mètre classique en introduisant des e muets sur des positions interdites comme la sixième position. Mais, en même temps, et cela est moins connu, le vers accentue aussi la distance avec le modèle de la langue en introduisant de plus en plus de e muets comptés alors que ces derniers s'élident en langue. Ces vers saturés de e muets de Baudelaire en sont un exemple extrême.

Laissent piteusement leurs grandes ailes blanches
 e e e e e
 Baudelaire, Les Fleurs du Mal, L'Albatros

Au fil des années, on assiste à un double écart, par rapport à la langue et par rapport à la métrique, ce qui ne laisse pas présager d'un avenir serein.

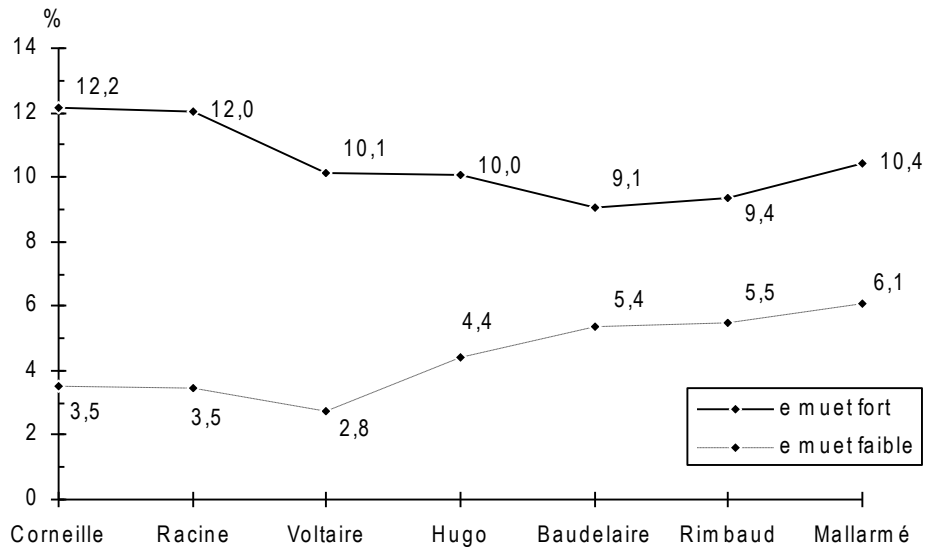


Figure 11. Évolution des e muets comptés dans le vers

Rythme et champs lexico-sémantiques

Pour finir, je reviens donc à ma question initiale, quelle relation entre la structure rythmique et les champs lexico-sémantiques ?

Pour l'identification des champs lexico-sémantiques, j'ai utilisé le logiciel Alceste, un outil de fouille de texte développé par Max Reinert (Reinert, 1993, 1983), dans la lignée des travaux de Benzécri. Cet outil a été réécrit et modernisé par Pierre Ratinaud dans le logiciel Iramuteq (Ratinaud, 2009).

Cet outil s'inscrit dans la tradition de l'analyse des données à la française, école développée par Benzécri, avec l'idée de proposer une boîte à outil générique qui puisse être appliquée à tout type de tableau de données (et être utile aux sociologues, littéraires, médecins etc.). Les constituants principaux de la boîte à outils sont l'analyse des correspondances et la classification hiérarchique. Reinert a développé un algorithme très robuste pour segmenter les textes.

L'idée de base dans le logiciel Alceste est tout d'abord de découper le corpus en fragments de texte de taille homogène (par exemple chaque fragment est composé de huit vers), puis l'algorithme de classification permet de découper en deux le corpus de façon à ce que les fragments de chaque classe soient le plus homogène possible en termes de vocabulaire utilisé et le plus différent des fragments de l'autre classe. Le processus est réitéré sur la classe la plus importante, et ainsi de suite. Chaque classe (qui regroupe une proportion variable de fragments) peut alors être interprétée comme un champ lexico-sémantique. Je m'appuie pour cela sur les travaux de François Rastier (Rastier, 1989).

Avec ces outils de statistique textuelle, je distingue au sein des pièces de théâtre deux grands univers lexico-sémantiques, l'un relevant du registre de l'amour et l'autre du registre de la mort. En croisant chacun de ces registres avec la structure rythmique des vers, une corrélation entre le contenu et la forme de ces vers apparaît. Le registre de la mort est lié à des vers plus réguliers (dominante de rythmes 010101 et 001001 ; par exemple 30% d'hémistiches de type 001001, contre 25% dans l'autre classe). Inversement, le registre de l'amour est caractérisé

par des vers moins structurés et plus irréguliers. Cela correspond à des effets de genre (la tragédie a un mètre plus régulier que la comédie) et d’auteur (Racine marque une nette préférence pour le rythme 001001).

Pour faire abstraction des effets de genre et d’auteur, j’ai cherché à identifier au sein du corpus tragique racinien, les pièces qui présentent le profil lexico-sémantique le plus proche. Pour cela j’ai utilisé Alceste pour produire une classification sur les pièces de Racine qui permet de distinguer quatre champs lexico-sémantiques ; j’ai ensuite évalué la répartition de ces quatre univers dans chaque pièce (cf. Figure 12). *Les Plaideurs*, unique comédie, a un profil très distinct des autres ; les deux pièces du début, *La Thébàide* et *Alexandre le Grand* ont encore un profil proche des pièces de Corneille et de part leur thématique, les pièces tardives de Racine (après un silence de plus de 10 ans), *Esther* et *Athalie*, s’inscrivent dans une autre tradition. J’ai donc sélectionné les tragédies d’*Andromaque* à *Phèdre*, celles qui ont le profil le plus proche et qui présentent une grande homogénéité du vocabulaire dès que l’on fait abstraction des noms de lieux et de personnages.

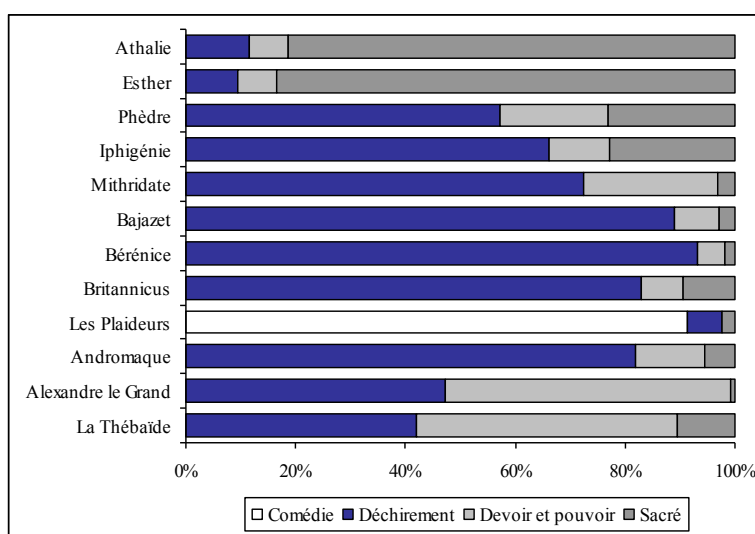


Figure 12. Profil lexico-sémantique des pièces de Racine

Sur ce sous-corpus, en excluant de l’analyse les mots grammaticaux, et en prenant comme unité d’analyse la paire de vers rimant ensemble, la classification nous donne deux classes, chacune étant caractérisée par les éléments suivants :

| Dialogue-amour (59%) | Récit-mort (41%) |
|---|--|
| Pronoms 1ère et 2nde pers. | Pronoms 3ème pers. |
| Impératifs | Noms propres |
| Interpellations | |
| Interjections | |
| Ici et maintenant | Hors scène, ailleurs |
| Mise en scène des difficultés de la relation (se taire/parler ; rester/partir ; vivre/mourir) | Espace et temporalité hors scène, référence à événements passés ou à venir |

La part des hémistiches réguliers, qu’ils soient de type anapestique ou iambique est plus forte dans les vers appartenant au champ lexico-sémantique « récit-mort ». Les vers irréguliers dominent quand les personnages s’affrontent sur scène via le dialogue.

| | dialogue-amour (59 %) | récit-mort (41 %) |
|----------|-----------------------|-------------------|
| H_001001 | 31% | 34% |
| H_010101 | 37% | 38% |
| H_autres | 31% | 29% |
| Total | 100% | 100% |

Je montre sur ces pièces une opposition entre les vers correspondant à des scènes narratives ou descriptives et des scènes de dialogue et de confrontation. Cette différence est redoublée au niveau métrico-rythmique. Le vers d'exposition ou de narration est beaucoup plus régulier dans son rythme que celui des scènes où les personnages s'affrontent, vers souvent haché (stichomythies) et moins régulier.

Ce type de résultat n'a jamais été identifié car il s'agit de différences significatives au sens statistique mais pas forcément perceptibles à l'œil-oreille. Je montre comment la métrique fait partie intégrante de la sémantique : elle joue le rôle d'une ligne de basse qui transmet quelque chose de la densité émotionnelle de la situation. Plus on est dans le récit, plus le vers se calme ; plus on est dans l'action, plus il est marqué par les aspérités de l'ici et maintenant de la scène.

Enfin en analysant ces mêmes tragédies, je montre un usage stylistique de la rime propre à Racine. Les rimèmes (terminaison graphico-phonétique) les plus spécifiques de chaque pièce correspondent aux sonorités finales des noms des personnages, même quand on enlève du décompte les occurrences des noms propres. Léo Spitzer a bien montré comment chez Racine l'usage des noms propres à la place du pronom, y compris pour parler de soi et de son interlocuteur contribuait à mettre à distance la violence émotionnelle de la situation et à produire ce qu'il désigne comme un « effet de sourdine » (Spitzer, 1970) (par exemple, Bérénice s'adresse à Titus : « Sans que jamais Titus puisse voir Bérénice »). Cela ne peut être un effet du hasard, surtout quand on regarde avec quel soin Racine s'est efforcé d'associer à ses personnages des confidents dont les noms ont les mêmes sonorités finales (Hermione/Cléone, Bérénice/Phénice, Mithridate/Arbate etc.) à l'exception notable de Phèdre dont la singularité ne permet aucun écho.

Conclusion

Au terme de ce parcours, j'ai pu montrer que la structure rythmique du vers est un phénomène multimodal qui engage les différents niveaux linguistiques (phonétique, lexical, syntaxique et prosodique), avec des effets de résonance entre les différents niveaux. Cette structure rythmique interagit avec les autres composantes stylistiques : la forme du vers change selon les genres, les auteurs et les formes discursives (récit/dialogue).

Les outils informatiques ont été conçus et mobilisés pour répondre à la question initiale des relations entre rythme et sens, ce qui nous rappelle que les humanités numériques ne pourront avoir de valeur que si elles sont mises au service de problématiques provenant des humanités. En termes de méthode, cette expérience d'analyse des aspects formels du vers avec des outils s'est faite sous le signe de la diversité. Tout d'abord, par la diversité des outils mobilisés :

outils de Traitement Automatique de la Langue, outils d'analyse de graphe et de réseaux ; d'autre part par la diversité des composantes linguistiques analysées (phonétique, syntaxique, prosodique, lexicale, sémantique). Ensuite par la diversité des questions examinées : identification des syllabes métriques, figures des vers, traitement de la rime, analyse lexicosémantique. Cette diversité trouve son sens dans le modèle hiérarchisé du rythme qui permet de faire dialoguer toutes ces composantes. Le pendant informatique de ce modèle est la base de données des vers enrichie par de nombreux traits affectant la syllabe, l'hémistiche, le vers, la paire de vers, la pièce entière. Cette représentation informatique prend en compte différentes dimensions constitutives du rythme. Elle permet aussi un enrichissement par l'introduction de nouveaux traits ou marquages, qui peuvent résulter de nouveaux traitements. Cette représentation, dans une base de données, autorise alors toutes formes d'explorations et de mises en relation.

Il faudrait à présent étendre l'investigation à l'ensemble du théâtre du XVIIe : en quoi la forme rythmique des vers peut-elle contribuer à expliquer que certains auteurs aient sombré dans l'oubli tandis que d'autres non. Par delà le théâtre se pose la question des spécificités du vers selon les genres : quels sont les traits du vers de théâtre par rapport au vers de poésie ? L'outil et les travaux menés sur le vers français devraient être étendus aux vers d'autres pays européens : la circulation des formes poétiques a été si importante en Europe qu'il serait passionnant de pouvoir identifier les trajectoires des formes poétiques, les mécanismes de traduction d'une langue à l'autre de phénomènes et de règles métriques. Il reste aussi à inventer des outils pour traiter des formes poétiques contemporaines, ce qui constitue un chantier d'une bien plus grande complexité.

Bibliographie

- Beaudouin, V., 2002. Mètre et rythmes du vers classique - Corneille et Racine. Slatkine-Champion, Paris.
- Beaudouin, V., 2016. Retour aux origines de la statistique textuelle : Benzécri et l'école française d'analyse des données, in: Mayaffre, D., Poudat, C., Vanni, L., Magri, V., Follette, P. (Eds.), *Statistical Analysis of Textual Data - JADT 2016*. Nice, pp. 17–27.
- Beaudouin, V., Yvon, F., 1996. The Metrometer : a Tool for Analysing French Verse. *Literary & Linguistic Computing* 11, 23–32.
- Benvéniste, E., 1966. Les niveaux de l'analyse linguistique, in: *Problèmes de Linguistique Générale*. Gallimard, Paris, pp. 118–131.
- Benzécri, J.-P. et coll., 1981. *Pratique de l'analyse des données, Linguistique et lexicologie*. Dunod, Paris.
- Constant, P., 1991. *Analyse syntaxique par couche*. Doctorat ENST.
- Cornulier, B. de, 1982. *Théorie du vers*. Rimbaud, Verlaine, Mallarmé. Éditions du Seuil, Paris.
- Delente, É., Renault, R., 2015. *Projet Anamètre : le calcul du mètre des vers complexes*. *Langages* 199, 125. doi:10.3917/lang.199.0125
- Fraisse, P., 1974. *Psychologie du rythme*. PUF, Paris.
- Ganascia, J.-G., 1987. CHARADE: A rule system learning system, in: *Proceedings of the 10th International Joint Conference on Artificial Intelligence IJCAI*. pp. 345–347.
- Gasparov, M.L., 1987. A Probability Model of Verse (English, Latin, French, Italian, Spanish, Portuguese). *New metrics* 21, 322–358.

- Gervás, P., 2000. A Logic Programming Application for the Analysis of Spanish Verse, in: First International Conference on Computational Logic, Logic Programming Implementations and Applications Stream. Imperial College, London.
- Lebart, L., Salem, A., 1994. Statistique textuelle. Dunod, Paris.
- Lusson, P., Roubaud, J., 1974. Mètre et rythme de l'alexandrin ordinaire. Langue française 41–53.
- Milner, J.-C., 1974. Réflexions sur le fonctionnement du vers français. Cahiers de poétique comparée vol I, pp 2-21, 283-301.
- Milner, J.-C., Regnault, F., 1987. Dire le vers. Court traité à l'intention des acteurs et des amateurs d'alexandrins. Seuil, Paris.
- Rastier, F., 1989. Sens et textualité. Hachette, Paris.
- Ratinaud, P., 2009. IRaMuTeQ : implémentation de la méthode ALCESTE d'analyse de texte dans un logiciel libre. Modélisation Appliquée aux Sciences Humaines et Sociales (MASHS2009).
- Reinert, M., 1983. Une méthode de classification descendante hiérarchique : application à l'analyse lexicale par contexte. Les cahiers de l'analyse des données Vol VIII, 187–198.
- Reinert, M., 1993. Les “mondes lexicaux” et leur “logique” à travers l'analyse statistique d'un corpus de récits de cauchemars. Langage et société 5–39.
- Roubaud, J., 1986. DYNASTIE : études sur le vers français, sur l'alexandrin classique, Première partie. Cahiers de poétique comparée 47–109.
- Roubaud, J., 1991. T.R.A.(M,m) (question d'une poétique formelle, I). Cahiers de poétique comparée, Mezura 26 p.
- Roubaud, J., 2000. La vieillesse d'Alexandre. Editions Ivrea, Éditions Ramsay, (éd. François Maspero 1978), Paris.
- Spitzer, L., 1970. L'effet de sourdine, in: Études de Style. Gallimard, Paris, pp. 208–335.
- Yvon, F., 1996. Prononcer par analogie : motivation, formalisation et évaluation. Thèse de l'ENST.