

Traitement automatique des formes métriques des textes versifiés

Éliane Delente & Richard Renault

CRISCO EA 4255

Université de Caen Basse-Normandie

eliane.delente@unicaen.fr, richard.renault@unicaen.fr,

Résumé. L'objectif de cet article est de présenter tout d'abord dans ses grandes lignes le projet *Anamètre*¹ qui a pour objet le traitement automatique des formes métriques de la poésie et du théâtre français du début du XVIIe au début du XXe siècle. Nous présenterons ensuite un programme de calcul automatique des mètres appliqué à notre corpus dans le cadre d'une approche déterministe en nous appuyant sur la méthode métricométrique de B. de Cornulier ainsi que la procédure d'appariement des rimes et la détermination des schémas de strophes dans les suites périodiques et les formes fixes.

Abstract.

Automatic Processing of Metrical Forms in Verse Texts.

The purpose of this paper is to present the project *Anamètre*. The project proposes automatic processing of metrical forms in French poetry and drama from the early seventeenth to the early twentieth century. Then we present the calculation program of meters on our corpus using a deterministic approach relying on the "métricométric" method of B. de Cornulier. Finally, we present the procedure of matching rimes and determination of rhyme schemes in periodic sequences and specific forms.

Mots-clés : *métrique française, corpus poétique, calcul du mètre, appariement des rimes, détermination des schémas de strophes*

Keywords: *French metrics, poetic corpus, calculation of meter, matching rhymes, determining of rhyme schemes*

1 Introduction

Le projet en question, mis en place il y a huit ans, a pour objet le traitement automatique des formes métriques dans la poésie littéraire française et dans le théâtre, du début du XVIIe au début du XXe siècle. Nous présenterons le projet dans sa globalité ainsi que les différentes étapes de l'analyse automatique qui conduisent à calculer de manière satisfaisante les différents mètres des vers, à repérer les rimes, à déterminer les schémas strophiques, ainsi qu'à restituer la forme globale des textes (formes fixes : sonnet, ballade, terza rima, rondeau... et suites périodiques : suites de distiques, de quatrains, de sizains...).

À ce jour, le corpus analysé est constitué de 8900 poèmes et 75 pièces de théâtre pour 56 auteurs, ce qui représente un total de 540 000 vers intégrés dans une base de données qui regroupe les différents résultats de l'analyse métrique, rimique et strophique.

2 Métrique et informatique

En tant qu'objets construits, les textes versifiés sont intégralement structurés par des équivalences métriques portant sur les vers et les schémas de strophes au moyen des rimes. Pourtant, la métrique a très peu été investie par le traitement automatique. Notons cependant les travaux pionniers à cet égard de Roubaud (1986 et 1988) et Beaudouin & Yvon (1994) et Beaudouin (2002).

Le renouvellement de la métrique française dans les années 80 par B. de Cornulier fournit un cadre théorique et méthodologique au développement de procédures et programmes d'analyse des régularités métriques. Pourtant, cette méthode quantitative, fiable et reproductible est parfois perçue comme trop formelle pour un objet littéraire, alors qu'elle est à même de fournir aux études stylistiques des analyses métriques rigoureuses et argumentées.

Par ailleurs, ce type de traitement est aujourd'hui plus abordable. Les linguistes peuvent rapidement s'approprier des langages de scripts tels que Perl ou Python et l'émergence du langage de description de données XML ainsi que la normalisation des pratiques d'encodage encouragée par la TEI ont considérablement facilité la façon de traiter les données linguistiques.

¹ Le site web *Métrique en ligne* est associé au projet *Anamètre* : <http://www.crisco.unicaen.fr/verlaine/>

3 Objectifs et résultats attendus

3.1 Les exploitations strictement métriques

La grande nouveauté est que chaque unité rythmique est désormais balisée. On peut donc rapidement et aisément les extraire pour diverses exploitations :

- 1- Établir des dictionnaires de diérèses par auteur notamment et ainsi constituer une histoire de la versification à cet égard.
- 2- Observer systématiquement la composition des vers complexes (composé de 2 hémistiches), notamment par l'examen des propriétés des mots à la césure.
- 3- Étudier systématiquement les relations entre unités rythmiques : l'hémistiche, le vers, le module de strophe et la strophe.

Ces analyses, qui relèvent de l'organisation micro-textuelle, conduisent au second grand type d'exploitation qu'on peut appeler macro-textuelle.

3.2 Les exploitations textuelles

Le balisage des unités rythmiques permettra d'examiner l'organisation syntaxique et sémantique des textes versifiés :

- 1- les études statistiques en lexicométrie et textométrie gagneront en pertinence puisque les recherches lexicales se feront non plus sur le texte poétique traité en bloc mais selon son organisation rythmique – strophe, module de strophe, vers et hémistiches- jusqu'alors négligée faute d'outils.
- 2- les études sur le degré de convergence entre les expressions linguistiques et les unités rythmiques pourront se développer (propriétés sémantiques du second hémistiche par rapport au premier, du second module de strophe par rapport au premier, de la strophe)

L'enjeu est le développement d'une linguistique textuelle caractéristique du texte versifié.

4 Le projet dans ses grandes lignes

Le projet comporte trois volets :

- 1- la constitution d'une base de données de textes poétiques et théâtraux annotés, du début du XVIIIe au début du XXe siècle
- 2- la conception et la mise au point de programmes d'analyse métrique pour l'élaboration d'une base de données de relevés métriques générés automatiquement (un relevé métrique est une série d'analyses métriques portant sur un objet métriquement cohérent et autonome)
- 3- la constitution d'une base de données de relevés métriques faits manuellement par divers métriciens depuis les années 80 qu'il convient de normaliser et d'optimiser pour les rendre pérennes

Le développement qui suit porte uniquement sur le second volet.

5 Les étapes du traitement du corpus

5.1 Formatage initial du texte

Le texte fourni en entrée du traitement est un fichier au format XML-TEI qui contient un formatage minimal en sections et sous-sections <div>, strophes typographiques <lg>, et vers <l> pour un poème ; actes et scènes <div>, tours de parole <sp> et vers <l> pour les pièces de théâtre. La gestion des retraits des vers en cas de polymétrie est réalisée automatiquement après une première passe calculant la longueur métrique des vers. Exemple extrait d'un poème de Théodore de Banville (*Le Budget*, Occidentales, 1875, vers 5) :

(1) Formatage initial

<l>Malgré ses ailes d'aigle et son corps de lion,</l>

5.2 Découpage du texte en mots

La notion de mot est celle définie par les expressions régulières. Bien que le découpage en mots ne soit pas nécessaire à l'identification des voyelles de l'étape suivante, il est indispensable dès lors qu'il s'agit de mettre en rapport les mots du texte avec des ressources lexicales. Un algorithme déterministe ne peut pas décider si un mot doit être traité en diérèse ou en synérèse, ou si un mot commençant par un "h" doit être traité comme jonctif ou disjonctif. Le découpage en mots

de vers en est exclu puisqu'il n'appartient pas au domaine du mètre mais à celui de la rime.

(6) Calcul de la longueur métrique des vers

Malgré ses ailes d'aigle et son corps de lion,
 1 1 1 1 1 1 0 1 1 1 1 1 = 12

5.7 Calcul du mètre des vers

5.7.1 Distribution de propriétés pertinentes pour le calcul du mètre

Pour tous les vers dont la longueur métrique est ≤ 8 , le mètre correspond à la longueur métrique. Pour tous les vers dont la longueur métrique est > 8 , le calcul du mètre met en œuvre une procédure en trois temps : 1) Repérage de propriétés phonologiques (place des voyelles masculines et féminines dans le mot), morphologiques (distribution des proclitiques dans le vers), et syntaxiques (distribution des prépositions monosyllabiques). Ces propriétés (propriétés métricométriques²) sont pertinentes pour déterminer le profil métrique du texte puis le mètre des vers. Précisons que ce traitement n'utilise pas d'analyseur syntaxique mais seulement un algorithme fondé sur des règles d'identification contextuelle des proclitiques, des prépositions monosyllabiques, et de mots ambigus tels que *entre* et *contre* qui sont des prépositions monosyllabiques en contexte d'élision. Cet algorithme utilise une ressource lexicale constituée uniquement de mots grammaticaux (prépositions, déterminants, pronoms...). Une voyelle métrique reçoit un attribut de propriété dans les cas suivants : 1) si elle appartient à une préposition monosyllabique (propriété P), si elle appartient à un proclitique (propriété C), si la voyelle masculine est prétonique de mot (propriété M), si la voyelle féminine est post-tonique de mot (propriété F).

(7) Distribution des propriétés métricométriques

Malgré [ses] ailes d'aigle et [son] corps [de] lion,
 M [C] F [C] [P]

propriétés phonologiques :
 M = voyelle masculine prétonique
 F = voyelle féminine
 propriétés morphosyntaxiques :
 [P] = préposition monosyllabique
 [C] = proclitique

5.7.2 Détermination d'un profil métrique

La distribution régulière des propriétés repérées sur l'ensemble du texte permet d'en évaluer le profil métrique (6+6 par exemple pour un poème constitué uniquement de vers classiques de 12 voyelles). Ce profil métrique est nécessaire notamment au calcul du mètre des vers afin de départager le mètre 4+6 du 5+5 pour une même longueur métrique de 10 voyelles.

5.7.3 Calcul du mètre des vers complexes

Le profil métrique correspondant à chaque longueur métrique permet ensuite de préciser le mètre de chaque vers en tenant compte de la présence éventuelle des propriétés métricométriques dans les places pertinentes pour le calcul du mètre (places 6, 4 et 8, par exemple, pour les vers de 12 voyelles). À l'issue de cette procédure, une balise de césure <caesura /> est introduite entre les hémistiches.

(8) Calcul du mètre des vers (après détermination du profil métrique du poème)

Malgré ses ailes d'aigle | et son corps de lion, 6+6

| = césure
 6+6 = mètre du vers

² Cornulier, Benoît de, 1982, *Théorie du vers*, Éditions du Seuil, Paris
 Cornulier, Benoît de, 1982 1995, *Art Poétique*, Presses universitaires de Lyon.

5.8 Traitement des rimes

5.8.1 Appariement des vers en rimes

L'appariement (appel-écho) se fait à partir de la dernière voyelle tonique du vers. La relation rimique entre deux vers est établie s'il y a : 1) une équivalence phonétique de la voyelle tonique, 2) une équivalence graphique de la consonne finale, 3) une équivalence graphique des consonnes subséquentes (consonnes placées entre la voyelle tonique et le "e" des terminaisons féminines). L'appariement des vers est soumis à une contrainte de localité³ qui fait qu'un appel trouvera son écho dans un domaine proche. Les schémas de rimes permettent l'identification des strophes qui sont ensuite comparées avec des schémas de formes fixes (sonnet, ballade, rondeau...) ou des schémas de formes périodiques (distique, quatrain, quintil...) afin de déterminer la forme globale du poème.

(9) Appariement des vers en rimes

Malgré ses ailes d'aigle | et son corps de lion, 3
 ...
 Et je pense qu'avec un petit million 3

■	= segment phonologique de la rime
3	= n° d'ordre d'appariement de la rime

5.8.2 Détermination de la forme globale

Quatre procédures sont utilisées pour déterminer la forme globale : 1) Reconnaissance d'une forme fixe à partir des données quantitatives (nombre total de vers, nombre de strophes, nombre de vers par strophe), des schémas de rimes et des éventuelles répétitions totales ou partielles de vers. Cette procédure permet d'identifier ainsi les sonnets, les ballades, les rondeaux, les triolets, les terza rima, les pantoums, les sextines... 2) Reconnaissance d'une forme périodique à partir de la distribution régulière des schémas de rimes relativement au découpage du poème en strophes typographiques. Cette procédure permet d'identifier les suites régulières de distiques, quatrains, quintils, sizains... ainsi que les différentes combinaisons de schémas rimiques. 3) Reconnaissance d'une suite de distiques tels qu'on les trouve dans les pièces de théâtre. 4) Reconnaissance d'une suite de schémas de rimes distincts au moyen d'expressions régulières construites à partir d'un répertoire de base de schémas de rimes. Cette dernière procédure permet d'identifier des séquences de schémas rimiques dans les poèmes sans découpage typographique en strophes.

(10) Construction des strophes au moyen des schémas de rimes

Malgré ses ailes d'aigle et son corps de lion,	a
Il n'a pas du tout l'air farouche	b
Et je pense qu'avec un petit million	a
Nous pourrons lui fermer la bouche	b

■	= segment phonologique de la rime
a	= étiquette de la rime dans le schéma de strophe

6. Choix théoriques et méthodologiques

Notre approche diffère très nettement de celle de V. Beaudouin, tout d'abord, par l'ampleur et la nature du corpus traité. Le *Métromètre*, appliqué essentiellement au théâtre classique, n'épuise donc pas, loin s'en faut, la variété métrique et strophique de la poésie littéraire française puisqu'il ne traite guère que du seul mètre 6+6 et du seul schéma strophique (aa) en rimes suivies. *Anamètre* n'a pas de limitation de ce type puisqu'il couvre près de trois siècles d'écriture théâtrale et poétique dans le cadre du système métrique classique, offrant ainsi toutes les variétés métriques et strophiques caractéristiques de ce système. Mais il traite également les productions de la deuxième moitié du XIXe siècle et du début du XXe siècle qui obéissent à des régularités assouplies conduisant progressivement à l'abandon de la notion même de régularité métrique. Le corpus offre ainsi une perspective historique indispensable pour l'exploitation et l'interprétation des données.

³ Ruwet, Nicolas, 1981, "Linguistique et poétique : une brève introduction", *Le Français moderne*, 49:1, p. 1-19
 Cornulier, Benoît de, 1982, *Théorie du vers*, Éditions du Seuil, Paris

Le type d'approche est également très différent. Nos programmes ne procèdent pas à un découpage du vers en syllabes puisque les consonnes ne sont pas pertinentes pour le traitement du mètre. Ils ne phonétisent pas non plus les vers. Ce traitement se révèle en effet inutile ; puisque les régularités métriques reposent pour une large part sur la graphie, la problématique des régularités métriques et celle de la déclamation sont en grande partie indépendantes. Enfin, nous n'avons pas recours à des patrons métriques extérieurs au poème ; notre approche consiste à intégrer le principe d'équivalence contextuelle au moyen de la notion de profil métrique.

Afin d'évaluer la fiabilité des résultats du traitement, une comparaison systématique, vers par vers, avec le corpus analysé manuellement par B. de Cornulier est en cours. Mais un exemple peut d'ores et déjà nous servir d'indice de fiabilité. *Les Méfaits de la lune* est un poème appartenant à la production tardive de Verlaine et présentant une complexité métrique telle qu'elle rend problématique l'analyse humaine du poème et malgré cela, l'analyse automatique du poème est en accord avec celle de J. L. Aroui⁴.

7. Conclusion

Comme on peut s'y attendre, l'état actuel du traitement présente des limites, touchant notamment à l'interprétation de nouveaux mètres, dits de substitution, apparaissant à partir de 1860. En effet, certains de ces vers présentent une coupe possible en 4e et 8e positions. Ces vers sont tous potentiellement 4.4.4. mais pour certains, l'interprétation 8.4 ou bien 4.8 paraît plus évidente que l'interprétation 4.4.4. Seule une analyse syntaxique du vers, croisée à l'analyse métrique effectuée, pourrait permettre de décider quelle est l'interprétation rythmique la plus naturelle entre ces trois possibilités. À ce jour, l'analyse syntaxique du vers n'a pas encore été intégrée dans nos programmes qui proposent pour l'instant une interprétation correcte mais sous-déterminée.

⁴ Répertoire métrique des *Œuvres Poétiques complètes* de Paul Verlaine, Nantes, M.S.H. Ange Guépin, Centre d'Études Métriques, 1993.

Bibliographie

- Beaudouin V. (2000), *Rythme et rime de l'alexandrin classique, étude empirique des 80 000 vers du théâtre de Corneille et Racine*. Thèse de Doctorat. EHESS.
- Beaudouin V. (2002), *Mètre et rythmes du vers classique : Corneille et Racine*, Paris, Honoré Champion.
- Bobillot, J. P. (1991) : Recherches sur la crise d'identité du vers dans la poésie française _ 1873-1913 [4 vol.], Université Sorbonne Nouvelle, Paris 3.
- Bobillot, J. P. (1993) : « Entre mètre & non-mètre : le « décasyllabe chez Verlaine », *Revue Verlaine* n° 1, pp. 179-200, Musée Bibliothèque Rimbaud, Charleville-Mézières.
- Bobillot, J. P. (1994) : « De l'anti-nombre au quasi-mètre : le « hendécasyllabe » chez Verlaine », *Revue Verlaine* n° 1, pp. 66-86, Musée Bibliothèque Rimbaud, Charleville-Mézières.
- Cornulier, B. de (1979) : *Problèmes de métrique française*, thèse d'État, Université de Provence.
- Cornulier B. de (1982), *Théorie du vers : Rimbaud, Verlaine, Mallarmé*, Paris, Seuil.
- Cornulier B. de (1995), *Art poétique*, IUFM, Presses universitaires de Lyon.
- Cornulier B. de (1999), *Petit dictionnaire de métrique*, Université de Nantes.
- Cornulier B. de (2000a), "La place de l'accent, ou l'accent à sa place : position, longueur, concordance", dans *Le Vers Français : Histoire, théorie, esthétique*, textes recueillis par Michel Murat, Champion, Paris, p. 57-91.
- Cornulier B. de (2000b), « L'invention du « décasyllabe » chez Verlaine décadent. Le 4-6, le 5+5, le mixte, et le n'importe quoi », Actes du colloque *Verlaine à la loupe*, 11-18 juillet 1996, dir. J. M. Gouvard et S. Murphy, Paris, Honoré Champion.
- Cornulier B. de (2003), « Problèmes d'analyse rythmique du non-métrique » Semen.
- Cornulier, B. de (2009) : « Types de césures ou plutôt manières de rythmer le vers composé », *L'Information grammaticale* n° 121, Peters.
- Cornulier, B. de (2010) : *Notions d'analyse métrique*.
<http://www.normalesup.org/~bdecornulier/notmet.pdf>
- Delente, E. (à paraître) : « La dimension textuelle du rythme. Étude des *Poèmes Saturniens* jusqu'à *Bonheur* de Verlaine », dans *Cahiers du Centre d'Études Métriques* n° 7, université de Nantes.
- Delente E. (2003), "Distribution du syntagme adjectival épithète dans l'alexandrin verlainien", in *Le sens et la mesure. De la pragmatique à la métrique*. Hommages à Benoît de Cornulier, Textes réunis et édités par Jean-Louis Aroui, Paris, Ed. Honoré Champion, p. 399-414.
- Delente E. (2010), « Métrique et traitement automatique. Une question difficile : l'étude de la concordance », Actes du colloque *Linguistique et Littérature : Cluny, 40 ans après*, Besançon, 29-31 octobre 2007.
- Delente E. & Renault R. (2011) : « Annotation automatique des textes versifiés », Colloque « Patrimoine à l'ère du numérique » – 10-11 décembre 2009 à Caen, Schedae, prépublication n° 5, fascicule n° 1, p. 39-52.
- Dominicy, M. (1984) : « Sur la notion d'e féminin ou masculin en métrique et en phonologie », *Recherches linguistiques* n° 12, pp. 7-45.
- Gouvard J.-M. (2000), *Critique du vers*, Paris, Honoré Champion.
- Holowacz, W. (1997) : « Le décasyllabe dans l'œuvre de Verlaine », *Cahiers du Centre d'Études Métriques* de Nantes, Université de Nantes.
- Murat M. (éd.) (2000), *Le vers français : histoire, théorie, esthétique*, Paris, Honoré Champion.
- Roubaud J. (1986) et (1988) : « Dynastie : études sur le vers français, sur l'alexandrin classique » *Cahiers de poésie comparée* n°13 et n°16, publications des langues'O.
- Ruwet N. (1981), « Typography, Rhymes, and Linguistic Structures in Poetry », in W. Steiner (éd.) *The Sign in Music and Literature*, University of Texas Press, Austin, 242 p.
- Tobler, A. (1885) : *Le vers français ancien et moderne*, Slatkine, Genève, éd. 1972.