



HAL
open science

Outils et poésie : un tour d'horizon

Eliane Delente, Richard Renault

► **To cite this version:**

Eliane Delente, Richard Renault. Outils et poésie : un tour d'horizon. Langages, 2015, Traitement automatique des textes versifiés : problématiques et pratiques, 3 (199), pp.5-22. 10.3917/lang.199.0005 . hal-04072792

HAL Id: hal-04072792

<https://hal.science/hal-04072792>

Submitted on 28 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Éliane Delente

CRISCO EA 4255, Université de Caen Basse-Normandie

Richard Renault

CRISCO EA 4255, Université de Caen Basse-Normandie

Outils et métrique : un tour d'horizon¹.

1. Introduction

Ce numéro porte sur les différents aspects du traitement automatique des régularités rythmiques des textes versifiés. Il concerne la poésie et le théâtre, en France mais aussi en Europe ; en effet, les développements des standards et des pratiques d'encodage des corpus littéraires suscitent des problématiques communes et cela indépendamment des langues et des systèmes métriques.

En France, les études métriques ont connu ces trente dernières années un renouvellement important à partir de la création du *Centre d'Études Métriques* de Nantes en 1984, première équipe de recherche vouée à la métrique française, et à partir de la création de la collection *Métrique française et comparée* par les éditions Champion en 2000. Le colloque consacré au vers français qui s'est déroulé en 1996 à la Sorbonne (Université Paris-IV) et dont les contributions ont été réunies et éditées par M. Murat chez Champion en 2000, constitue un état de l'art de l'époque qui peut utilement servir de point de départ à la présente réflexion, près de quinze ans plus tard.

Ces événements ont, à eux seuls, dynamisé la recherche en métrique française et comparée et assuré une diffusion nécessaire de ces travaux, ce qui témoigne suffisamment de la vitalité de la recherche en ce domaine. On notera cependant que ce champ disciplinaire n'a guère été investi par les linguistiques de corpus ou par les analyses textuelles menées à l'aide d'outils informatiques (voir cependant Viprey, 2002 et Brunet, 1992). On peut avancer une cause à cette timidité, qui n'est pas nouvelle et qui relève d'une attitude naïve à l'égard de cette approche de la discipline : de même que, par le passé, on a pu reprocher aux analyses formelles et linguistiques, telles que celles développées par R. Jakobson et N. Ruwet d'être mal appropriées à leur objet, on peut penser, aujourd'hui, qu'un balayage aveugle effectué par une machine ne saurait rendre compte des multiples aspects de la création poétique. On peut cependant adopter une démarche raisonnable consistant à reconnaître et exploiter la matérialité textuelle du poème ce qui a pour effet de mettre en évidence une sorte de paradoxe. En linguistique de corpus, l'approche informatisée permet une recherche de régularités assurant une fonction structurante. Or, les textes versifiés représentent, à cet égard, un matériau idéal pour le traitement automatique puisque les propriétés métriques qui les structurent sont périodiques et systématiques. Le traitement automatique exige de concevoir la métrique comme une discipline à part entière, ce dont témoigne le recueil d'articles récemment édité par Küper (2011). Entièrement consacré à l'étude de différents systèmes métriques, il ne contient encore que deux études portant sur le traitement automatique des textes versifiés. Pourtant, l'organisation métrique du texte poétique, par sa systématique, se prête idéalement à ce type d'approche rapide et fiable. Étendue à de larges corpus, celle-ci constitue une base indispensable à l'analyse sémantique et pragmatique du texte, à même de favoriser les interactions entre linguistes et

¹ Précédents numéros de revue consacrés à la poésie et/ou à la métrique : *Poétique générative*, Langages, n° 51, (1978) ; *Métrique française et métrique accentuelle*, Langue Française, n° 99, (1993) et *Linguistique et poésie : après Jakobson*, Langue Française, n° 110, (1996).

littéraires.

Dans le domaine français, les premiers travaux en la matière ont été réalisés par J. Roubaud, qui a participé à l'élaboration de l'outil *Dynastie* (Roubaud, 1986 et 1988), premier programme d'analyse du vers. Le *Métromètre* (Beaudouin & Yvon, 1994, Beaudouin, 2002) a été élaboré dans une perspective semblable. Aujourd'hui, ces recherches de pionniers, discutables à certains égards, font place à de nouvelles approches.

1.1. Originalité de l'approche dans le contexte actuel

Les études littéraires réservent un accueil de plus en plus favorable à l'informatique, notamment par le développement de l'édition électronique, des études statistiques (stylométrie, textométrie...) et par l'élaboration de bases de données de textes littéraires. On observe ensuite que le développement des standards et des logiciels facilite l'appropriation de l'outil informatique par les non-spécialistes. D'une part, l'émergence du langage XML² et des outils de manipulation des textes encodés en XML permet d'harmoniser les données grâce à la TEI (Sperberg-McQueen & Burnard, 1994 et Burnard, 1997). D'autre part, le partage et la transformation des données sont facilités par les langages de haut niveau de type Perl ou Python qui rendent la programmation informatique plus accessible aux linguistes et métriciens.

Au regard des recherches actuelles dans le champ du traitement automatique des textes versifiés, ce numéro fait le point sur l'état des études métriques qui touchent essentiellement à l'élaboration d'outils d'analyse automatique des formes métriques sans exclure la constitution de corpus électroniques annotés et l'élaboration de répertoires métriques et de bases de données.

1.2. La constitution de corpus électroniques

Un bref coup d'œil chez nos voisins européens les plus proches met en évidence l'utilité d'une telle entreprise pour le domaine français³. La plus grosse banque de données textuelles française, Frantext, offre au téléchargement une partie de son corpus à partir du portail du CNRTL⁴. Le corpus poétique ainsi mis à disposition présente de nombreux inconvénients pour une exploitation métrique automatique. Tout d'abord, les textes poétiques sont faiblement représentés et les textes versifiés ne font pas l'objet d'un critère de sélection rigoureux et le vers n'apparaît pas toujours sur une seule ligne. Le portail du CNRTL offre toutefois un recueil de Leconte de Lisle pour lequel une transposition XML peut se faire facilement puisque le formatage strophique, les retraits dans les cas de polymétrie et la majuscule de début de vers sont préservés. Par ailleurs, l'annotation minimale au moyen de balises XML conformes à la TEI n'encode pas l'unité de base du traitement métrique automatique – le vers – qui est identifié dans le corpus Frantext uniquement par un retour à la ligne. Ce traitement permet en effet de préserver l'organisation syntaxique, niveau d'analyse privilégié par

2 XML (*Extensible Markup Language* dérivé du SGML : Standard Generalized Markup Language) : langage de description de données dont l'unité fondamentale pour les données textuelles est un élément constitué d'une portion de texte et de balises avec des attributs décrivant des propriétés.

3 L'Italie dispose d'une base de données de textes en prose et en poésie de Dante à Pirandello, disponible sur le site <http://www.ecstore.it/negozio/bcliz30.htm> et du dictionnaire de concordances poétiques, le *Vocabolario della poesia italiana del Novecento* (1995) de S. Giuseppe. L'Espagne offre deux grands corpus interrogeables séparément ou conjointement : le CORDE, corpus diachronique (jusqu'à 1966) et le CREA, corpus actuel (depuis 1967) disponibles sur les sites <http://corpus.rae.es/cordenet.html> et <http://corpus.rae.es/creanet.html>. L'Allemagne dispose de la « *Deutsche Lyrik in Reclams Universal-Bibliothek* » qui inclut les œuvres de cinq cents poètes du XVIe siècle au XXe siècle, consultables sur le site <http://glp.chadwyck.com/>. Enfin, la « *English Poetry Database* » publiée par Chadwyck-Healey est exclusivement consacrée à la poésie anglaise, de la période anglo-saxonne (VIIIe siècle) au début du XXe siècle. Elle comprend les œuvres de 1350 poètes et elle est consultable sur le site <http://www.letrs.indiana.edu/epd/>. La base de données *English Poetry, Second Edition*, éditée par les bibliothèques de l'Université du Texas reprend et complète la précédente en offrant 183 000 poèmes du VIIIe siècle au XXe siècle, représentant ainsi 2 700 poètes sur le site http://collections.chadwyck.com/html/ep2/help/ep2_ind.htm.

4 <http://www.cnrtl.fr/corpus/frantext/>

Frantext. Enfin, l'orthographe et la ponctuation, notamment des œuvres du XVIIe siècle, ont été modernisées. Les effets pervers de ce choix, qui n'est d'ailleurs pas imputable à la seule base Frantext, rendent les textes impropres au repérage des rimes et à l'analyse ponctométrique⁵. Beaudouin (2002), utilisant les éditions électroniques de la base Frantext, signale que la forme moderne du subjonctif *dise* supprime la rime *die* : *s'ennuie* dans les vers suivants (Corneille, *Attila*, vers 1-2) :

- (1) *Ils ne sont pas venus, nos deux rois ? Qu'on leur dise
Qu'ils se font trop attendre, et qu'Attila s'ennuie ;*

Il arrive même qu'une « casserole » rende le vers faux. Le site du projet *Molière 21* propose une édition électronique de la comédie de Quinault *L'amant indiscret ou le maître étourdi*⁶ dans laquelle la forme lexicale ancienne *castrole*, modernisée en *casserole*, devient un mauvais chaudron pour un alexandrin de 13 voyelles métriques (Acte I, scène 3) :

- (2) *Saumon, brochet, turbot, alose, truite, et sole
Soit fris au court-bouillon, en ragoût, en casserole,*

D'autres corpus de textes préparés mais insuffisamment fiables, tels que le projet Gutenberg⁷, le corpus Enclitt⁸, le corpus du CNAM (ABU)⁹ et enfin Wikisource¹⁰ présentent tous des inconvénients semblables : l'orthographe et la ponctuation sont modernisées, l'intégralité du texte fait défaut, les sources éditoriales ne sont pas toujours mentionnées et enfin, les œuvres ne présentent pas toutes les garanties de vérification du texte lorsqu'il s'agit d'ouvrages scannés. Aujourd'hui, on peut disposer de corpus plus facilement exploitables, du moins sur le plan du balisage XML. Le site *Théâtre classique*¹¹ propose un corpus important de pièces de théâtre du XVIIe siècle.

La France ne dispose donc, à ce jour, d'aucun site dédié à l'archivage des textes versifiés ni de portail recensant toutes les ressources fiables qui soient directement exploitables pour le traitement automatique. Ça et là pourtant, des laboratoires universitaires ont pu développer des versions électroniques de l'œuvre intégrale de tel ou tel poète mais le défaut d'un lieu centralisateur nuit à la visibilité de ces travaux. Ce problème, qui devient un enjeu crucial de la recherche, trouvera à terme une solution grâce notamment aux travaux du consortium *Corpus écrits* de l'ILF. Seul le site Gallica, portail de la BNF, met à disposition des documents numérisés¹² mais qui nécessitent bien souvent un pré-traitement important afin de les rendre utilisables dans un format XML. Alors que la numérisation d'un document est déclarée comme fiable à 99,99%, voire 100%, il arrive souvent que ce texte comporte encore des erreurs de numérisation que seul un traitement métrique permet de révéler : un mot pour un autre, de longueur syllabique différente, produisant un vers faux ou bien un mot pour un autre en fin de vers, bloquant la reconnaissance de la rime. Mais le traitement métrique ne pourra jamais révéler des erreurs du type *clans* à la place de *dans* puisque la substitution n'affecte en rien les régularités métriques.

La création d'un corpus de textes poétiques et théâtraux versifiés répond donc à un réel besoin. L'existence d'un tel corpus aurait pour effet de dynamiser les recherches métriques, poétiques, littéraires et de conforter la convergence entre les recherches linguistiques et les travaux littéraires.

5 Voir Cornulier (1995 : 265-266).

6 http://moliere.paris-sorbonne.fr/base.php?L'amant_indiscret_ou_le_Maître_Étourdi

7 <http://www.gutenberg.org/browse/languages/fr>

8 <http://www.unicaen.fr/ufr/homme/linguistique/ressources/EncLitt/portail.php>

9 <http://abu.cnam.fr/>

10 <http://fr.wikisource.org/wiki/Wikisource:Accueil>

11 <http://www.theatre-classique.fr/>

12 Voir également le site du projet *Internet Archive* : <https://archive.org/>

La constitution de tels corpus réclame une réflexion portant sur la problématique éditoriale, l'analyse critique des sources et les méthodes mises en œuvre. Le problème de l'annotation se trouve aujourd'hui au centre de plusieurs projets¹³ qui s'intéressent à la question de la granularité de l'annotation des textes et au choix des formats et des standards. Le projet italien *Poeml*¹⁴ (Renello, 2007 et 2010) de l'université de Bologne est un des tout premiers projets à avoir traité les textes versifiés au format XML, avec des annotations encodant des unités métriques.

1.3. L'élaboration d'outils d'analyse automatique des formes métriques

Nombreux sont les projets d'analyse automatique des textes versifiés qui ont vu le jour ça et là depuis les années 80. K. Bobenhausen en propose un recensement¹⁵ selon la langue des corpus : langues anciennes, anglais, allemand, français, italien, espagnol, portugais, russe... Ces projets ont toutefois des orientations très différentes :

- analyse des formes métriques (mètres, rimes, strophes)¹⁶
- traitement statistique (stylistique fondée sur les calculs de fréquence)
- outils d'édition (éditeur de texte, police de caractères...)
- outils de génération et de traduction de poèmes (en conformité avec des contraintes métriques données)

Quelques-uns ne font plus l'objet de développements, d'autres ont donné lieu à l'élaboration d'outils qui ne sont plus exploitables aujourd'hui. Quant aux projets en cours, on observe, par-delà la différence des langues traitées, une convergence à la fois dans les objectifs, les méthodes et les moyens utilisés, ce qui justifie l'opportunité de ce numéro. Ces outils produisent bien souvent des descriptions formalisées et des analyses qui alimentent des bases de données métriques, comme c'est le cas pour les outils d'analyse du mètre anglais (Plamondon, 2006 et Kavanagh, 2007) conçus à l'origine pour renseigner une base de données de corpus poétiques.

1.4. Élaboration de bases de données

L'objet de telles bases de données est de mettre à disposition des corpus de textes versifiés mis en relation avec les informations métriques, aujourd'hui produites par les programmes d'analyse mais dont la pratique est déjà ancienne chez les métriciens. Les tout premiers relevés métriques n'étaient pas le résultat d'un traitement automatique mais offraient une description systématique qui garantissait la stabilité des traits d'analyse. Avant même l'intégration dans des bases de données, le métricien utilisait déjà un outil informatique minimal, sous forme de tableur, pour la manipulation des propriétés encodées. L'extension de cette pratique combinée à la diversité des logiciels et des formats, n'ont pas facilité l'exploitation puis l'homogénéisation des données. L'émergence et la large diffusion des formats ouverts et des standards rend aujourd'hui possible cette harmonisation, tandis que l'échange et le partage des données garantissent leur pérennité. Le traitement de telles données massives ouvre de nouvelles perspectives par l'observation de phénomènes mesurables, facilitant ainsi leur comparaison en synchronie comme en diachronie.

13 Mentionnons les deux consortiums *Corpus écrits* de l'ILF et le projet *Cahier* de l'UMR 6576 (CESR) de l'université de Tours. Citons également le projet Base de Français Médiéval de l'équipe icar3, UMR 5191 : <http://icar.univ-lyon2.fr/>

14 Voir le site : <http://www2.lingue.unibo.it/dese/poeml/poeml.html>

15 Disponible à l'adresse : <http://home.versanet.de/~kb-369326/> mais il n'est malheureusement pas à jour.

16 Les formes métriques sont des formes dont le caractère métrique ne repose pas sur des propriétés objectives des constituants étudiés mais sur l'équivalence sensible de ces formes avec les formes voisines.

2. Les Objectifs

2.1. Objectifs scientifiques

L'intérêt majeur de ces projets est de fournir aux métriciens et aux stylisticiens, et de manière plus générale, aux littéraires et aux linguistes, des outils d'analyse automatique des propriétés caractéristiques des textes versifiés.

Dans le domaine métrique, ces outils facilitent tout d'abord le travail du métricien en permettant des analyses à grande échelle du mètre, de la rime et des strophes, ce qui a pour effet de donner une impulsion aux études métriques déjà existantes et de permettre une vue d'ensemble de la métrique¹⁷. Par ailleurs, l'établissement de bases de données facilite considérablement une réflexion menée dans une perspective historique, en rendant accessible l'examen de périodes charnières pendant lesquelles le changement progressif du système métrique devient observable. Un des formats possibles de ces analyses est le relevé ou répertoire métrique tel qu'il a été développé par le *Centre d'Études Métriques* de Nantes¹⁸.

Pour le français, on peut évoquer des exploitations strictement métriques. Quelques exemples :

1- établissement de dictionnaires de diérèses par auteur. Les diérèses pratiquées par Théophile de Viau (XVIIe) le sont-elles encore par Victor Hugo, Baudelaire ou Verlaine ? Une histoire de la versification à cet égard devient ainsi envisageable.

2- examen systématique de la composition des vers par extraction des mots à la césure et de leurs propriétés.

3- examen systématique des relations entre constituants rythmiques (hémistiche, vers, module de strophe et strophe). On sait par exemple que la cohésion syntaxique est moins forte dans le premier hémistiche que dans le second. On pourra alors comparer les constituants rythmiques conclusifs par rapport aux constituants rythmiques non conclusifs. Ces descriptions, qui relèvent de l'organisation micro-textuelle, conduisent au second grand type d'exploitation qu'on peut appeler macro-textuelle.

Dans le domaine textuel, à la croisée de la structure rythmique et de l'organisation sémantique des textes versifiés, les études textométriques et stylométriques gagneront en finesse, en rigueur et en pertinence. Faute d'outils, les statistiques ont souvent été appliquées à un document qui n'offrait pas de prétraitement adapté aux spécificités métriques de ces textes. On ne peut pourtant négliger le risque encouru par des études statistiques grossières qui se pratiqueraient dans l'ignorance des différents niveaux de structuration métrique. Le développement des outils permettra des recherches lexicales, non plus à l'échelle du texte poétique appréhendé comme un tout non structuré, mais selon son organisation rythmique, jusqu'alors négligée.¹⁹ Selon les constituants métriques, on pourra rechercher la fréquence de tel lexème ou de tel motif syntaxique (patron syntaxique) par strophe, module de strophe, vers et hémistiches (Exemple : Le N1 qui V / et le N2 qui V : *Le cœur qui se repose / et le sang qui s'affaisse* Verlaine).

Il peut paraître utopique, aujourd'hui, de s'attacher à appréhender globalement l'articulation des différents niveaux de structuration des textes versifiés. Mais le traitement automatique peut permettre des approches certes incomplètes, partielles ou locales, mais déjà très précieuses, telles que celles mises en œuvre par Donow (1970) qui propose un programme statistique pour l'analyse stylistique de la poésie (mais également de la prose). L'originalité du traitement des caractéristiques rythmiques des vers est de corrélérer notamment le nombre de mots par vers, le nombre de syllabes par mot et la place de l'accent. Une approche semblable est celle de Barquist & Shie (1991) qui

17 Voir le projet abouti de D. Robey portant sur 160 000 vers : « Sound and Metre in Italian Narrative Verse » à l'adresse <http://www.italianverse.reading.ac.uk/>

18 Voir section 3.4 et l'article de N. Hervé dans ce numéro.

19 Voir les programmes de traitement statistique ErMaStat qui relèvent les assonances, les allitérations et les répétitions de mots notamment dans l'ensemble du texte analysé (épopées rimées du moyen-haut allemand) : <http://www.uni-protokolle.de/nachrichten/id/34089/>

proposent une étude statistique de *Beowulf* (poème épique de la littérature anglo-saxonne composé entre la première moitié du VIIe siècle et la fin du premier millénaire) ; Les auteurs notent le rôle privilégié que les consonnes post-vocaliques jouent dans l'allitération, et établissent une corrélation entre le pourcentage de traits distinctifs communs à ces consonnes et le caractère sémantiquement plus ou moins euphorique ou dysphorique de la réalité décrite²⁰. Dans une perspective comparable, on pourrait s'interroger sur les divergences entre les vers courts et les vers longs, sur le plan de l'organisation syntaxique ou sémantique.

Dans le domaine français, le développement de l'analyse automatique de textes poétiques et théâtraux aura pour effet non seulement de tester les hypothèses avancées dans les travaux récents mais également de développer la recherche actuelle en métrique. Si l'analyse métricométrique développée dans Cornulier (1982) et (1995) permet l'étude de la concordance entre la mesure et l'énoncé, elle présente également des limites. En effet, le repérage de la distribution de certaines propriétés prosodiques et morpho-syntaxiques dans le vers permet aisément de relever des cas limites d'enjambement. Mais l'interprétation de ces cas exigerait que l'on connaisse, pour chaque poète, la force moyenne des enjambements. Il serait possible également de tester de nouvelles hypothèses portant sur ces cas, plus nombreux mais plus difficiles à cerner, où l'articulation de la mesure et de l'énoncé est à mi-chemin entre la concordance maximale et la forte discordance, en distinguant les époques et les poètes. Voici un exemple d'interrogation que l'analyse automatique permettra de pratiquer systématiquement : *La fin d'un premier hémistiche peut-elle correspondre à un adjectif épithète quand il est antéposé au nom qui le suit immédiatement, ce qui implique que ce nom ouvre l'hémistiche suivant ? Chez quel poète ? A quelles conditions²¹ ?*

Pour nécessaire qu'elle soit, cette analyse de la concordance se limite à examiner la cohésion locale en posant des questions du type : "À telle époque, chez tel poète, quelles sont les conditions d'association entre une expression métrique donnée et un syntagme donné ?". Or, ces conditions dépassent largement les problèmes de cohésion locale et pour cela constituent aujourd'hui un vaste chantier. L'analyse automatique permettrait d'entreprendre une étude sur les conditions d'association entre l'énoncé et le vers, l'énoncé et le module de strophe, l'énoncé et la strophe. Il sera alors possible de tester une hypothèse telle que : *La contrainte de concordance s'exerce prioritairement sur les expressions métriquement conclusives : l'hémistiche 2 dans un 6-6v par exemple, le vers 2 dans un module de strophe ab par exemple, le module 2 dans une strophe abab par exemple.*

Dans cette perspective, l'analyse automatique de la structuration syntaxique n'offre pas encore toutes les garanties de fiabilité souhaitables. Cette difficulté peut néanmoins être contournée. En France, on a eu recours ça et là à des méthodes qui consistent à repérer, pondérer et quantifier les signes de ponctuation de fins de vers, vus comme des indicateurs des constituants syntaxiques (phrase, proposition, syntagme...) dans les textes versifiés (Roubaud, 1988, Cornulier, 1995 : 265). Au sein du projet *Anamètre*²², un programme a été élaboré, qui permet d'accomplir ces tâches automatiquement et de mesurer ainsi la concordance, notamment de la structure strophique et de la structure syntaxique-sémantique afin d'évaluer la cohérence sémantique des différents constituants rythmiques, de manière certes grossière mais néanmoins significative. L'articulation entre les structures métrique et linguistique, insuffisamment étudiée aujourd'hui, nécessite de croiser les informations rythmiques avec les informations syntaxiques. Il s'agit là d'un chantier appelé à se développer. L'enjeu est donc l'ouverture d'un champ d'étude qui n'est qu'à ses balbutiements : une linguistique textuelle caractéristique du texte versifié.

20 Pour une étude semblable en France, voir les travaux de P. Guiraud, notamment (1953 et 1960).

21 Cf. Delente (2003).

22 Le projet *Anamètre* est développé par E. Delente et R. Renault au CRISCO, Université de Caen.

2.2. Objectifs didactiques

Une des applications de ces recherches est de favoriser l'apprentissage et la maîtrise de la versification²³, par la diffusion des ressources développées dans le cadre de programmes scientifiques telles qu'un dictionnaire de diérèses²⁴, un dictionnaire de rimes attestées ainsi que des applications (en ligne ou non) utilisant les algorithmes traitant de la distribution accentuelle, des phénomènes d'élision et de synalèphe (Gervás, 2000), de la césure, du mètre, des rimes et des formes strophiques. Une autre application possible est l'assistance à la génération automatique de poèmes (Oliveira, 2009), voir le projet ASPERA dans Gervás (2001), Mamede, Trancoso, Auraujo & Viana (2004),²⁵ parfois combinée à la traduction de poèmes dans le respect des contraintes rythmiques, comme par exemple la traduction de sonnets italiens en pentamètres iambiques anglais (Greene, Bodrumlu & Knight, 2010).

2.3. Objectifs éditoriaux

Certains projets, comme RPO à Toronto (Representative Poetry Online), développent des outils pour fournir une description métrique de textes poétiques dans une base de données²⁶. D'autres développent des outils d'édition XML²⁷ ou encore permettent l'authentification et la datation de textes anciens en utilisant des critères métriques tels que l'hiatus et la distribution des syllabes selon leur quantité (Walton, 1990).

3. Des problématiques communes

Les différents projets de traitement automatique des formes métriques, par-delà la diversité des langues, se trouvent confrontés à des problématiques communes, relatives à la forme graphique, au recours ou non à des ressources externes, au type d'approche, déterministe ou probabiliste, et à l'élaboration de relevés métriques et de bases de données.

3.1. Le traitement de la forme graphique

S'appliquant à une chaîne de caractères, on peut envisager deux approches extrêmes du traitement : d'une part, traiter directement les séquences de caractères comme unités de traitement minimal (noyau syllabique, syllabe, more) ; d'autre part, établir les unités minimales de traitement à partir des propriétés phonologiques du vers (voyelle, consonne, frontière de syllabe, accent...) préalablement encodées, dans une ressource indépendante pour la plupart des cas. Ces deux approches vont traiter différemment la forme graphique soit comme une simple chaîne de caractères, soit comme une chaîne de caractères segmentée en mots. La première approche peut être associée à un traitement minimaliste, traitement qui s'impose pour les systèmes d'écriture dans lesquels la correspondance graphie/son est bi-univoque (Gervás, 2001). Ce type de traitement est

23 Voir par exemple :

- *Poetiks* de Garrard & Head propose un traitement en ligne interactif à visée didactique dont la première étape calcule le mètre en réglant la distribution de l'accent avec une procédure d'ajustement, consultable sur le site <http://www.poetiks.com/>.
- *AnalysePoems* à l'adresse : <http://homes.chass.utoronto.ca/~mplamond/digitalwork/phonetics.jpg>,
- *Metricalizer 2* dans ce numéro et sur le site : <http://www.poetron-zone.de/metricalizer.php>.
- *Scandroid* à l'adresse : <http://oak.conncoll.edu/cohar/Programs.htm#>

24 Voir *Anamètre* et le site *Métrique en ligne* qui lui est dédié : <http://www.crisco.unicaen.fr/verlaine/>

25 Voir également les travaux et développements de J. P. Balpe autour de la poésie numérique et le travail de génération de sonnets aléatoires à partir des sonnets de R. Quenault « Cent mille milliards de poèmes ».

26 <http://rpo.library.utoronto.ca/>

27 Voir *Tustep*, outil de manipulation de données textuelles en XML à l'adresse : http://www.tustep.uni-tuebingen.de/tustep_eng.html.

essentiellement d'ordre quantitatif (détermination du nombre d'unités métriques sans les distinguer selon leurs propriétés phonologiques) et se révèle parfaitement adapté à une approche probabiliste (voir B. Brard et S. Ferrari dans ce numéro).

Quant à la seconde approche, souvent justifiée par la nécessité de recourir à une ressource, elle est notamment privilégiée dans les métriques accentuelles, et illustrée en métrique française par le *Métromètre* (Beaudouin & Yvon, 1994, Beaudouin, 2002). Elle présuppose une segmentation en mots afin, soit de procéder à une phonétisation plus ou moins complète avec un jeu complexe de règles²⁸, soit de comparer les mots du vers à ceux d'une ressource externe, pour récupérer la place de l'accent lexical²⁹. Comme pour la première approche, il s'agit, dans les deux cas, d'identifier les unités métriques, mais cette fois de façon indirecte.

Une troisième approche intermédiaire se situe entre la sobriété d'un traitement direct à partir des graphèmes et la contrainte de recourir à une ressource externe pour l'identification des unités minimales. Le projet *Anamètre* d'E. Delente et R. Renault³⁰ ne retient de la forme graphique que les éléments métriquement pertinents : repérage des graphèmes correspondant aux noyaux syllabiques. Ici, la syllabation n'est pas nécessaire mais elle peut l'être dans certaines langues telles que le japonais pour l'identification des mores (Kozasa, 1997). Les approches qui n'ont pas besoin de segmentation en mots pour l'identification des unités minimales peuvent cependant en avoir besoin par la suite ; c'est le cas pour le projet *Anamètre* qui ne peut traiter le caractère jonctif ou disjonctif d'un mot sans faire intervenir la notion d'unité lexicale puisque le comportement du "e" instable³¹ à la frontière de mot ne peut être traité strictement par des règles phonologiques. Il en va de même pour le traitement des diérèses.

3.2. Ressources externes

Le recours à des ressources est incontournable. Ces ressources peuvent être de deux types : internes ou externes. Les ressources internes sont élaborées conjointement à un algorithme afin de permettre le contrôle des données (bien entendu, une petite quantité de données linguistiques peuvent être intégrées dans l'algorithme sans qu'il y ait lieu pour autant de parler de ressources). Mais elles peuvent être des objets déjà construits à partir de ressources disponibles. Par exemple, le projet *Anamètre* utilise un dictionnaire de diérèses à partir de la version XML du dictionnaire Littré³² et un dictionnaire des mots jonctifs/disjonctifs. Les ressources externes sont des données conçues en dehors du projet en cours, comme par exemple un dictionnaire précisant la distribution de l'accent de mot. Les ressources sont principalement de deux types, lexicales ou grammaticales. Les ressources lexicales interviennent dans le traitement initial pour l'identification des noyaux syllabiques dans les programmes de Pil'shchikov & Starostin (2011), de M. Plamondon (2006), et de C. O. Hartman³³. Elles interviennent également pour des traitements ultérieurs (calcul du mètre, de la rime (Kavannagh, 2007)).

Comme pour les ressources lexicales, le traitement des unités grammaticales peut se faire soit en utilisant une ressource externe déjà constituée (TreeTagger, Cordial...), soit en développant en interne des analyseurs syntaxiques utilisant des règles contextuelles, réduites au strict nécessaire. Quelle que soit l'approche choisie, ces ressources permettent de catégoriser les mots, information utile à divers égards ; les recherches lexicométriques s'appuient souvent sur la catégorie

28 Dans le programme *Xaliyl*, la phonétisation de la chaîne graphique est imposée par les habitudes orthographiques de l'écriture arabe, voir Kouloughli (2010). Voir également l'article de D. Fusi dans ce numéro.

29 Cas illustré par le programme *Metrical Analysis of Russian Verse*, développé par I. Pilščíkov & A. Starostin et le *Metricalizer 2* de K. Bobenhausen. Voir leur article dans ce numéro.

30 Voir notre article dans ce numéro.

31 En français, toutes les voyelles sont stables sauf le "e" dit muet qui, selon les contextes, est maintenu ou non. Ex. : le mot « fenêtre » comporte deux occurrences de "e" instable, en gras.

32 <http://www.littre.org/>

33 Projet Scandroid : <http://oak.conncoll.edu/cohar/Programs.htm>

grammaticale, et le calcul du mètre composé français, (vers de plus de 8 voyelles) requiert le repérage des déterminants, des pronoms clitiques et de certaines prépositions.

Les principales problématiques liées à l'utilisation de ressources, qu'elles soient lexicales ou grammaticales, sont, d'une part, l'évaluation des contraintes dues au recours à des ressources externes, et d'autre part, le temps et l'investissement requis par le développement du traitement en interne. Les ressources externes peuvent être incomplètes et inappropriées à la nature du corpus. Par ailleurs, il convient de tenir compte de la nature des licences de diffusion des ressources externes utilisées, qui peuvent ne pas être compatibles avec celles du projet.

3.3. Approche déterministe et approche probabiliste

Comme pour tout traitement automatique de nature linguistique, la question se pose du choix entre une approche déterministe et une approche probabiliste. L'approche déterministe repose nécessairement sur un jeu plus ou moins conséquent de règles, plus ou moins descriptives et plus ou moins indépendantes du corpus traité. Ces règles peuvent être définies indépendamment des corpus ou bien générées par apprentissage (Beaudouin, 2002). En revanche, cette approche implique une description exhaustive qui peut obliger de prévoir un jeu complémentaire de règles d'ajustement et de recourir à des ressources complémentaires pour une analyse complète du poème.

Dans le cadre d'une approche probabiliste, un formatage minimal du texte en entrée et un petit nombre de règles pour l'identification des unités pertinentes, rendent possible le calcul de la longueur métrique probable. Dans cette approche, les vers présentant une ambivalence métrique et les vers non conformes à un patron métrique mais apparentés à celui-ci ne perturbent pas l'analyse qui retiendra l'interprétation la plus plausible par rapport au contexte métrique défini par le poème. Par contre, une telle approche ne permet pas d'enrichir l'annotation comme peut le faire l'approche par règles puisque l'algorithme ne vise qu'un seul aspect du traitement.

Le choix entre ces deux approches est largement déterminé par la nature du corpus analysé et par les objectifs visés. Si le corpus de textes analysés couvre une période très large, réunissant des textes anciens et modernes, une approche probabiliste présente l'avantage de s'affranchir des particularités graphiques d'un état de langue donné. Si l'objectif vise un seul aspect de l'analyse (mètres, rimes ou strophes), l'approche probabiliste se révèle plus économe en conception algorithmique mais quantitativement plus pauvre en information tandis que l'approche déterministe permet d'enrichir les annotations en ajoutant des informations qui pourront faire l'objet d'un traitement statistique sur un état stable d'un système métrique donné ou dans une perspective historique. Ces informations permettent également de conserver la trace des singularités propres à chaque poète.

Dans l'évaluation de ces approches, déterministe ou probabiliste, il faut encore prendre en compte ce qui relève de l'analyse réservée à l'humain et ce qui relève de l'analyse réalisée par les programmes. Certains programmes nécessitent d'intégrer une part d'analyse métrique dans les données initiales ; il en va ainsi de *Xaliyl* (Kouloughli, 2010), programme de traitement de la métrique arabe : l'auteur intègre la frontière d'hémistiche dans les données initiales. D'autres ont opté pour des données d'entrée plus pauvres, réduites à la seule disposition typographique encodée pour les textes en XML et produisent alors des résultats sollicitant une participation interprétative de l'utilisateur. C'est le cas pour le projet *Anamètre* dans lequel la frontière d'hémistiche n'est pas donnée d'entrée mais entièrement déterminée par le programme. Les résultats nécessitent alors une vérification et, pour les cas indécidables, le programme se refuse à calculer le mètre mais fournit à l'utilisateur toutes les informations le conduisant à une interprétation.

3.4. L'élaboration de relevés métriques et de bases de données

Un relevé métrique est une série d'analyses métriques portant soit sur une partie de poème, soit sur

un poème entier, pourvu que l'objet analysé soit métriquement cohérent et autonome. L'élaboration de relevés métriques répond à plusieurs besoins :

- faciliter la consultation et la comparaison des données massives.
- offrir une vue synthétique de la métrique d'un auteur ou d'une époque.
- permettre d'appréhender la métrique dans une perspective historique et de repérer et de décrire des évolutions dans la pratique métrique.

En France, sous l'impulsion des travaux de B. de Cornulier depuis les années 80, les métriciens ont progressé dans la description formelle des poèmes et ont systématisé leur analyse par l'élaboration manuelle de relevés métriques. Une trentaine de relevés métriques a été élaborée depuis les années 80 à partir d'une codification proposée par B. de Cornulier. Néanmoins, les variations de notation observées d'un auteur à l'autre rendent parfois difficile leur interprétation et l'exploitation des résultats. Un autre problème, discuté dans l'article de N. Hervé, et connu de longue date par les métriciens, est la difficulté de distinguer, dans la pratique de leurs codifications, la part analytique de la part interprétative.

Pour une exploitation optimale, ces relevés métriques exigent de normaliser les différentes notations et d'établir un lien entre les différentes bases de données, l'une portant sur l'analyse globale des poèmes, l'autre sur les vers. Toutefois ce lien implique une nomenclature de référencement des poèmes. La constitution d'une base de données générale de relevés métriques apparaît comme une réponse raisonnable à ces difficultés. Une telle base de données est en cours de développement dans le cadre du projet *Anamètre* sur le site *Métrique en ligne*, qui vise à dépasser ces limites en établissant une normalisation et une optimisation des différents encodages.

Désormais, grâce à l'informatique, ces relevés métriques peuvent être générés automatiquement. Ils fournissent soit une analyse codifiée d'un ensemble de poèmes, soit une analyse codifiée des vers simples et composés en garantissant la stabilité des traits d'analyse permettant ainsi un tri de données massives grâce à des méthodes d'interrogation simples ou complexes.

4. Présentation des articles

Le présent numéro est composé de six articles centrés sur le traitement automatique de textes versifiés de différentes langues et sur l'élaboration de relevés métriques. Certains abordent également la question de la constitution de corpus électroniques.

La première contribution se démarque des suivantes en ce que l'auteur aborde la question des relevés métriques sans génération automatique mais en intégrant néanmoins cette problématique dans sa réflexion. N. Hervé montre tout l'intérêt scientifique de tels relevés établis manuellement, en évaluant les retombées d'une automatisation de ces tâches sur la pratique des métriciens. La codification indispensable à la description des données, qu'elle soit manuelle ou automatique, relève de problématiques communes. La principale difficulté consiste à faire la part de la description stricte et de l'interprétation. L'auteur avance que la pratique du relevé métrique automatique oblige à une partition plus nette entre analyse et interprétation en produisant des formules essentiellement analytiques. De même, le relevé métrique automatique apporte une réponse à la tension, courante dans la pratique des relevés métriques manuels, entre la nécessité de codifier les régularités systématiques d'un corpus donné et la tentation d'y intégrer les singularités liées à un poète ou à un poème. Selon N. Hervé, la constitution d'une base de données générale de relevés métriques fournit un cadre favorable à la stabilisation des normes d'encodage. Le relevé métrique décrit dans cet article (œuvre de Clément Marot) présente, par ailleurs, la particularité de fournir une analyse sur une œuvre qui se situe à la croisée du système métrique médiéval et du système métrique classique.

Dans l'article suivant, D. Fusi présente une plateforme d'analyse multilingue pour le traitement

métrique des formes poétiques, d'où une architecture en modules totalement indépendants, certains obligatoires, d'autres facultatifs. Cette plateforme s'applique à plusieurs langues, (grec ancien et moderne, latin et italien), pour lesquelles l'auteur développe une approche générale et comparative de la correspondance graphie/sons, où la chaîne de caractères (traités comme des unités d'analyse premières) se voit prise en charge afin d'obtenir une interprétation phonologique. L'analyse des textes nécessite un phonétiseur, un analyseur syntaxique sommaire pour l'identification des mots et un analyseur métrique exploitant les données fournies par les outils précédents. L'application de ces programmes permet ainsi de construire une base de données. L'approche de D. Fusi sépare donc rigoureusement la collecte des données et leur interprétation. Les programmes, comparables en cela à ceux de Beaudouin (2002), consistent à collecter le plus possible d'informations de niveaux différents (phonologique, phonétique, syntaxique, métrique) et à laisser le métricien construire les requêtes les mieux adaptées à son objet.

L'article suivant présente la dernière version (2011) d'un programme de traitement informatique en ligne de la métrique accentuelle de l'allemand, *Metricalizer*², élaboré par K. Bobenhausen & B. Hammerich³⁴. La première étape a pour objet l'identification des constituants phonologiques (syllabe, phonème, accent et durée) indépendamment de la nature versifiée du texte à partir d'algorithmes inspirés de ceux de la synthèse de la parole. Ces données linguistiques servent de base pour un calcul statistique des mètres possibles à partir desquels *Metricalizer*² propose l'interprétation la plus probable. L'article introduit une réflexion sur la notion de « complexité métrique » et en propose une évaluation automatique. Cette notion de complexité métrique, impliquant une discordance entre structure linguistique et structure métrique constitue un aspect bien connu du discours poétique par les effets de sens qu'il suscite.

I. Pilščík & A. Starostin³⁵ se proposent d'aborder les problèmes théoriques et les différentes solutions envisagées pour le traitement automatique de la métrique de la poésie russe. Dans une première partie, les auteurs défendent une approche probabiliste et ce, pour deux raisons : d'une part, chaque vers fait l'objet, en principe, de plusieurs interprétations métriques et d'autre part, la poésie russe se caractérise par la coexistence de plusieurs systèmes métriques (vers syllabo-tonique, vers tonique et *dolnik*). Par ailleurs, certains vers, relativement à la distribution accentuelle, sont irréguliers en contexte métrique régulier mais insuffisamment irréguliers pour exiger une analyse métrique différente. Ces conditions rendent donc impossible l'utilisation d'algorithmes se basant sur une hypothèse au profit d'une autre. La seconde partie de l'article décrit la procédure adoptée. Le schéma métrique des vers nécessite une catégorisation pour la distribution de l'accent de mot (non prédictible comme en allemand par exemple, cf. Bobenhausen dans ce même numéro) à partir d'une ressource externe et permet ainsi la génération de toutes les formes accentuelles du vers. La pertinence métrique de ces formes accentuelles est ensuite évaluée, par une procédure probabiliste, selon leur conformité au répertoire métrique et selon le contexte³⁶. La troisième partie est consacrée à la description du *dolnik* et des caractéristiques qui le distinguent des mètres toniques purs et des mètres syllabo-toniques «classiques» (iambes, trochée, dactyle, anapeste, amphibraque).

La métrique française est l'objet des deux dernières présentations.

B. Brard & S. Ferrari proposent un traitement automatique du calcul de la longueur métrique des vers. Leur approche permet d'analyser des textes du XVe siècle à nos jours dans leur graphie originale, puisque le traitement n'intègre ni syllabation ni phonétisation des vers. L'approche est justifiée par la complexité de la relation graphie/sons de l'orthographe du français et par la prise en compte de plusieurs états de langue. En choisissant une approche probabiliste, les auteurs font le pari que la part d'indétermination dans le calcul de la longueur métrique sera réglée par le poids du

34 Les éditeurs de ce numéro tiennent à remercier Pascale van Praet et Maxi Krause (Université de Caen) pour la traduction de cet article.

35 Leur article a été traduit de l'anglais par E. Delente et R. Renault.

36 L'approche contextuelle du mètre d'un vers est partagée par bon nombre des programmes ici présentés. Cf. Bobenhausen, Delente & Renault, Brard & Ferrari.

contexte. Un jeu de règles réduit permet l'identification des noyaux syllabiques. Le calcul des longueurs métriques consiste à laisser en suspens les deux facteurs d'indétermination en français, l'*e* instable et les semi-consonnes (diérèse), dont les occurrences seront minimisées par le poids du contexte. Les performances de l'outil ont été vérifiées à la fois empiriquement sur un corpus électronique de quelques milliers de vers, et méthodiquement, au moyen d'une procédure élaborée à partir de modèles hypothétiques.

Dans la dernière contribution, E. Delente & R. Renault présentent le calcul des mètres complexes élaboré dans le cadre du projet *Anamètre*. Les auteurs se démarquent de B. Brard & S. Ferrari par une approche déterministe recourant à des ressources lexicales spécifiques au projet (dictionnaire de diérèses et dictionnaire de mots jonctifs) qui permet l'identification des noyaux syllabiques et le calcul de la longueur métrique. La détermination du mètre des vers simples (≤ 8 voyelles) correspond à la longueur métrique. Quant à la détermination des vers composés (> 8 voyelles), elle exige qu'à la longueur métrique soit associée la distribution de propriétés morphophonologiques et syntaxiques du vers, pour identifier les hémistiches. Le repérage automatique de ces propriétés s'effectue sans recourir à un analyseur syntaxique. Les auteurs prennent en compte également le fait qu'une longueur métrique peut correspondre à plusieurs mètres. La prise en compte du contexte dans le calcul du mètre se fait par la détermination préalable d'un profil métrique pour le poème. Le traitement automatique du mètre sera illustré par des poèmes dont l'analyse métrique est problématique.

p.-s. Les éditeurs de ce numéro tiennent à rendre hommage à D. E. Kouloughli décédé en 2013. Il devait figurer parmi les contributeurs de ce numéro et nous avait fait parvenir le résumé de son article dans lequel il envisageait de présenter le programme *Xaliyl*, logiciel capable de produire une analyse métrique automatique d'une pièce de vers arabes classiques, de longueur quelconque. Sa connaissance de la métrique et de la poésie arabe, ses travaux en traitement automatique des langues et sa pratique des corpus de textes en langue arabe lui donnaient toute sa place dans ce numéro.

Références

- Barquist C. R. & Shie D. L. (1991), « Computer Analysis of Alliteration in Beowulf. Using Distinctive Feature Theory », in *Literary and Linguistic Computing*, Volume 6, Issue 4, Oxford University Press, pp. 274-280.
- Beaudouin V. & Yvon F. (1994), "Le Métromètre et les douze positions alexandrines". *Mezura*, n°32, Publications Langues'O, 33 p.
- Beaudouin V. (2002), *Mètre et rythmes du vers classique : Corneille et Racine*. Lettres numériques. Paris, Champion, coll. « Lettres Numériques », 620 p.
- Brunet E. (1992), "Rhymes in Victor Hugo's *Les Feuilles d'automne* and *Les chansons des rues et des bois*" in *Empirical Studies of the Arts*, Vol.10(2), pp.183-193.
- Burnard L. (1997), « The Text Encoding Initiative's Recommendations for Encoding of Language Corpora : Theory and Practice », <<http://users.ox.ac.uk/~lou/wip/Soria/>>.
- Cornulier B. de (1982), *Théorie du vers. Rimbaud, Verlaine, Mallarmé*, Travaux Linguistiques, Seuil, Paris.
- Cornulier B. de (1995), *Art Poétique : Notions et problèmes de métrique*, Presses Universitaires de Lyon, printemps 95.
- Donow H. S. (1970), « Prosody and the computer: A text processor for stylistic analysis », Spring Joint Computer Conference, pp. 287-295.
- Gervás, P. (2000), « A Logic Programming Application for the Analysis of Spanish Verse », in *Proceedings of the First International Conference on Computational Logic*, Imperial College, London, UK, pp. 1330-1344.
- Gervás P. (2001), « An expert system for the composition of formal Spanish poetry », *Journal of Knowledge-Based Systems*, vol. 14, pp 181-188.
- Greene E. Bodrumlu T. & Knight K. (2010), « Automatic Analysis of Rhythmic Poetry with Applications to Generation and Translation », in *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 524–533, MIT, Massachusetts, USA, 9-11 October 2010.
- Guiraud P. (1953), *Langage et versification d'après l'œuvre de Paul Valéry*, Klincksieck, Paris.
- Guiraud P. (1960), *Problèmes et méthodes de la statistique linguistique*, PUF, Paris.
- Kavanagh F. (2007), *Analysis of a phonetic and rule based algorithm approach to determine rhyme categories and patterns in verse*, Technical report n° 2007/23, Department of Computing, Faculty of Mathematics, Computing and Technology, The Open University, UK.
- Kouloughli D. E. (2010), « Traitement automatique de la métrique arabe : réalisations et perspectives » in *Bulletin d'études orientales, La métrique arabe au XIIIe siècle après al-Halil*, numéro tome LIX.
- Kozasa T. (1997), « Moraic Tetrameter in Japanese Poetry ». Ms. California State University, Fresno.
- Küper, C. (éd.) (2011), *Current Trends in Metrical Analysis*, Littera, Vol./Bd.2, Peter Lang.
- Mamede N. Trancoso I. Auraujo P. & Viana C. (2004), « An Electronic Assistant for Poetry Writing », *Congrès Advances in artificial intelligence : IBERAMIA 2004*, vol. 3315, pp. 286-294.
- Murat M. & Molinié G. (2000), *Le vers français. Histoire, théorie, esthétique*. Col. Métrique française et comparée, Champion, Paris.
- Oliveira H. (2009), « Automatic generation of poetry : an overview », Technical report, University of Coimbra.

- Pilščíkov I. & Starostin A. (2011), « Automated Analysis of Poetic Texts and the Problem of Verse Meter », in C. Küper (éd.) *Current Trends in Metrical Analysis*, Peter Lang, Berne, pp. 133-140.
- Plamondon M. (2006), « Virtual Verse Analysis: Analysing Patterns in Poetry » in *Literary & Linguistic Computing*, Volume 21 Supplementary Issue, pp. 114-127.
- Renello G. P. (2007), « POEML: a new verse encoding structure? », *LYNX Numéro spécial*, pp. 81-91.
- Renello, G. P. (2010a), *Machinae. Studi sulla poetica di Nanni Balestrini*, Bologna, CLUEB II Castello di Atlante
- Renello, G. P. (2010b), « PoeML: appunti per la codifica del verso libero », *Misure critiche* N.S. IX: 1. 183-194.
- Roubaud J. (1986 et 1988), « Dynastie : études sur le vers français, sur l'alexandrin classique » *Cahiers de poétique comparée* n°13 et n°16, publications des langues'O.
- Sperberg-McQueen C.M. & Burnard L. (éds) (1994), *Guidelines for Electronic Text Encoding and Interchange (TEI P3)*, 2 vols. Chicago and Oxford : Text Encoding Initiative.
- Viprey J. M. (2002), *Analyse textuelles et hypertextuelles des « Fleurs du Mal »*, Champion, Paris.
- numéro spécial *Linx* (2007) : *Corpora et Questionnements du littéraire*, sous la direction de Denise Malrieu, université Paris X-Nanterre.
- Walton E. W. (1990), « The letters of Euripides », in *Computing and the classics*, vol. VI, n° 2.