

## **LES VOIES DE LA CRÉATION D'UNE BASE DE DONNÉES POÉTIQUE GÉNÉRALE**

**Levente Seláf**

*Eötvös Loránd University*

levente.selaf@gmail.com

Stephen Ramsay, dans son essai intitulé *Reading Machines*, raconte le récit édifiant de la découverte des règles poétiques «anagrammatiques» des poèmes saturniens par Ferdinand de Saussure<sup>1</sup>. Dans l'interprétation de Ramsay, cette découverte a été rendue possible par un changement de perspective radical : Saussure a été amené à lire les poèmes d'une manière complètement différente de celles qui ont été expérimentées, essayées auparavant. Pour Ramsay, l'approche oulipienne de la poésie ou la 'pataphysique du docteur Faustroll inventée par Alfred Jarry sont des tentatives pareilles de regarder le monde et, plus spécialement, le matériel poétique dans une nouvelle optique ou d'en proposer une «lecture déformante». Pour lui, la *distant reading* n'en est qu'une autre. Or nous croyons que même si ce n'est pas forcément le regard jeté sur un poème d'un angle nouveau qui rendra possible sa nouvelle interprétation (comme s'il s'agissait d'une révolution scientifique), mais sa confrontation avec des contraintes et des règles de composition qui n'ont jamais été proposées pour son interprétation pourra effectivement donner de nouvelles pistes pour sa meilleure compréhension. Ramsay a raison de voir une forte parenté dans l'approche oulipienne de la littérature et les méthodes d'analyse littéraire assistées par l'ordinateur. Dans ce qui suit, nous voudrions proposer quelques réflexions sur les méthodes potentielles de l'analyse comparée des corpus poétiques, tout en nous inspirant des mêmes sources que Ramsay : l'Oulipo, la poétique générative et la poésie médiévale. Pour des raisons pratiques nous lançons cette réflexion concernant le

---

<sup>1</sup> Stephen Ramsay, *Reading Machines. Towards an Algorithmic Criticism*, Urbana, University of Illinois Press, 2011, pp. 46-57.

traitement automatique de la poésie à partir de l'analyse métrico-poétique des poèmes médiévaux. Au stade actuel des humanités numériques, il vaut la peine de regarder quelles étaient les solutions précédentes de la définition d'un corpus poétique et de l'identification des règles poétiques derrière leur création, et surtout, quelles sont les voies possibles (et les difficultés) devant l'analyse comparatiste assistée par ordinateur.

## 1. Le répertoire poétique comme outil de description

Dans le cas de la poésie médiévale, la méthode de la *distant reading*, donc grosso modo l'analyse statistique sur une grande quantité de textes, avait des antécédents fort honorables. La création des répertoires poétiques au XX<sup>e</sup> siècle fut un premier grand résultat facilitant l'étude quantitative des poèmes. À une époque où l'informatique ne procurait pas encore l'aide nécessaire à la création des bases de données, les premiers répertoires métriques et poétiques ont déjà décrit leurs spécimens à partir de nombreuses caractéristiques, permettant leur comparaison et le repérage des règles poétiques obligatoires et accidentelles de leur création. Donc la «machination de la poésie» (en but de son analyse profonde) a commencé dans ce domaine bien avant la création des outils informatiques qui nous servent à les traiter à l'heure actuelle.

L'histoire de l'évolution des répertoires poétiques est longue, et elle a connu plusieurs étapes. Le premier répertoire de grande valeur, et d'ailleurs le plus prestigieux parmi tous, est celui d'István Frank, consacré à la poésie lyrique des troubadours occitans<sup>2</sup>. Il a servi de modèle aux entreprises suivantes, autant par sa précision et sa richesse que par le prestige du matériel avec lequel il a travaillé. Du point de vue formel, la poésie des troubadours était la plus riche parmi les traditions poétiques européennes et celle-ci a exercé une influence durable sur plusieurs littératures ; il allait donc de soi que les autres tentatives de description profonde de corpus poétiques devaient prendre pour base le travail de Frank. Ces répertoires ont toujours eu comme objectif, avoué ou caché, l'identification des règles déterminant la composition poétique, et ainsi, dans une perspective plus large, une définition globale des contraintes formelles poétiques ou de l'essence de la poésie. Désir vain, mais fort respectable. De ce point de vue, rien ne distingue les analyses qu'on peut fonder sur ces informations de celles qui sont automatisées, rendues possibles par l'aide des ordinateurs, à partir de textes prétraités.

---

<sup>2</sup> István Frank, *Répertoire métrique de la poésie des troubadours*, Paris, H. Champion, 1966.

## 2. La description des poèmes : entre minutie et standardisation

La préparation d'un outil de recherche qui permette d'acquérir une vision globale de tous les éléments d'un corpus poétique pose deux problèmes. D'une part, le rassemblement (ou le choix) de tous les poèmes y appartenant et, d'autre part, la collection de toutes les caractéristiques qui déterminent les règles poétiques implicites et explicites de leur création et réception, ainsi que toutes les données qui permettent de les situer dans le temps et dans l'espace, dans le contexte d'un œuvre, d'une littérature, d'un groupe littéraire, etc. Mais le problème se pose justement dans la hiérarchie et la succession de ces deux étapes qui se supposent et qui se déterminent. Comment savoir quels sont les critères et les éléments de poéticité à enregistrer sans connaître le corpus intégral? Et peut-on croire à l'unité exclusive d'un tel corpus si nous choisissons les critères de description au préalable?

Très rares sont encore les corpus poétiques qui seraient entièrement numérisés. Pour la poésie médiévale, la volonté de faire des bases de données de texte intégral cause des problèmes particuliers, à cause de la multiplicité des sources manuscrites et de la variance inhérente aux copies multiples. Cette variance peut être reflétée par les outils informatiques, mais au prix de beaucoup de travail<sup>3</sup>. Or au cours des dernières soixante années, les répertoires poétiques ont permis et proposé une analyse formelle très fine des poèmes, basée sur une *close reading* très attentive. Ils ont accumulé une quantité énorme de données. Pour l'instant, il existe surtout des bases de données élaborées suivant les méthodes de la création des répertoires poétiques, donc sans contenir le texte des poèmes et, par conséquent, sans moyen d'automatiser le dépouillement des textes à la recherche de leurs principes poétiques. Si l'on a l'ambition de mener des recherches comparatistes des corpus définis par des critères linguistiques, temporels et/ou formels, il faudra appliquer des critères de description identiques.

Dans le passé, à la phase de création des répertoires poétiques, la volonté d'ouverture et d'interopérabilité a toujours été anticipée par la réduction du matériel analysé : la définition primitive du corpus a été résolue en général par l'unité linguistique et par des grilles temporelles de la périodisation qui suivaient les standards de l'histoire littéraire. Les questions poétiques n'entraient en jeu que dans un deuxième temps pour séparer le grain de la paille,

<sup>3</sup> Bien sûr, la création d'un logiciel capable de transcrire en alphabet courant les signes graphiques des scripta médiévales n'est pas impossible, mais le taux des fautes serait assez grand dans un premier temps. Voir à ce propos les articles dans le domaine de la paléographie digitale, par exemple, sur <<http://www.digipal.eu>> [dernière requête : 11/11/2019].

pour formuler l'identité poétique de tous les poèmes finalement inclus dans le corpus. Mais déjà les premiers choix opérés (langue et date) coupèrent une partie de l'altérité, de la diversité de la matière traitée et contribuèrent donc à la simplification, à l'homogénéisation du corpus. Et la non-fiabilité de ces critères (pour le Moyen Âge, surtout, celui de la datation, mais dans quelques cas extrêmes aussi celui de la langue ; pensons aux poèmes occitans conservés dans des chansonniers français avec une forme linguistique très altérée) pose déjà un problème : le prétraitement évince des cas qu'il serait difficile à décrire ou bien qui pourraient modifier le paradigme poétique d'une tradition telle qu'on la perçoit en l'absence de ces cas particuliers<sup>4</sup>.

L'approche des répertoriés à leur matière pouvait être très différente selon la définition préalable du corpus étudié. Nous proposons de parler dans le cas d'un corpus strictement défini de « corpus clos ». Dans un cas pareil, la matière d'analyse est choisie avec une telle rigidité que son unité poétique est déjà garantie par ce choix préalable : certaines règles fortes de la composition sont présentes dans chaque élément ; c'est le fondement de l'identité du corpus. L'un des meilleurs exemples du « corpus clos » est celui de la poésie des troubadours, tel qu'on le voit dans le répertoire de Frank (mais la tragédie grecque classique, l'alexandrin dans le théâtre classique de langue française ou le tanka japonais médiéval seraient pareils)<sup>5</sup>. Ces règles ont été connues et adoptées avec des variations locales dans la plupart des traditions poétiques européennes. Frank a pris soin de filtrer et d'écarter à l'avance du corpus des troubadours occitans les poèmes qui n'observent pas toutes les règles de versification élémentaires du *trobar*, et qui auraient pu rendre confus les résultats d'analyse<sup>6</sup>.

Nous pouvons mettre en opposition avec ce type de matériel le « corpus ouvert ». Il est possible d'utiliser le terme soit au cas où *tous les poèmes* d'une période et d'une langue données sont pris en compte, sans exclusion, soit quand le corpus est défini d'une manière négative, par l'absence d'une contrainte poétique qui peut déterminer un « corpus clos ». Nous connaissons des exemples de tous les deux. *Le Répertoire de la Poésie Hongroise*

<sup>4</sup> Ce n'est pas une falsification des données, mais il faut en tenir compte. Afin de comprendre l'évolution des formes poétiques, sont justement essentielles les deux phases qui sont opposables avec l'âge d'or, les origines et le déclin, donc au-delà et en-deçà les frontières plus ou moins arbitraires de l'histoire littéraire.

<sup>5</sup> La *canso* (et la chanson des trouvères) comme forme poétique est vraiment particulière ; elle a des règles de composition structurales très raffinées et complexes, qui permettent une grande variabilité des formes à partir de composants prédéfinis.

<sup>6</sup> C'est pour cette raison que Frank a publié une longue annexe à son répertoire, en y énumérant tous les poèmes occitans qu'il a écartés de son analyse, soit à cause de leur forme qui indique clairement qu'ils n'étaient pas chantés, soit à partir de critères linguistiques ou de datation.

*Ancienne* (RPHA) a eu comme vocation de recueillir tous les poèmes hongrois des origines à 1600 (y compris poèmes en prose, fragments de quelques mots, poèmes de versification gréco-latine, vraiment tout)<sup>7</sup>. Le corpus ainsi défini est le résultat très fragmentaire d'une histoire poétique de plusieurs siècles, avec des textes composés à partir de beaucoup de règles différentes, y compris plusieurs hapax qui seraient écartés si une règle de composition serait privilégiée au détriment des autres. Pour donner un exemple de la seconde catégorie, nous pouvons évoquer le *Nouveau Naetebus* (NN), base de données en construction depuis 2004. Le NN contient la description des poèmes strophiques en ancien français qui n'ont pas été composés à partir des mêmes règles poétiques que la chanson courtoise des trouvères<sup>8</sup>. La brisure de la contrainte d'*homogonie* a regroupé beaucoup de textes par ailleurs très différents. On a pu interpréter cet écart de diverses manières : comme un manque d'harmonie, une maladresse (donc une complexité réduite selon la supposition traditionnelle des philologues), une conséquence de l'usage d'un dialecte particulier ou encore comme le signe de l'absence d'une mélodie accompagnant le poème.

La définition négative (l'absence d'une règle importante pour les trouvères) n'était pas suffisante pour affirmer que les poèmes qui violent cette règle partagent également d'autres caractéristiques communes<sup>9</sup>. La conception théorique de cette base de données destinée à recueillir tous les « poèmes strophiques non-lyriques » de langue française antérieurs à 1400 devait accomplir deux exigences : d'une part, celle de décrire au mieux possible le corpus spécifique pour comprendre la cohésion du matériel recueilli et, d'autre

<sup>7</sup> Iván Horváth et al., *Répertoire de la poésie hongroise ancienne : manuel de correction d'erreurs dans la base de données*, Paris, Le Nouvel Objet, 1992; voir, également, <<https://rpha.oszk.hu>> [dernière requête : 11/11/2019].

<sup>8</sup> Gotthold Naetebus, *Die nicht-lyrischen Strophenformen des Altfranzösischen*, Leipzig, Hirzel, 1891, en ligne : <<http://tinyurl.com/jrhad8t>> [dernière requête : 11/11/2019]. Plus précisément, une règle, celle de l'*homogonie*, presque toujours observée dans la poésie chantée, est ignorée dans le corpus pris en considération par le modèle du *Nouveau Naetebus*, le relevé compilé par Gotthold Naetebus à la fin du XIX<sup>e</sup> siècle, un répertoire poétique de type ancien, créé bien avant celui de Frank. L'*homogonie* signifie l'identité des genres de rimes (masculine ou féminine) à la même position dans les strophes successives. Il y a aussi une zone grise, les poèmes *hétérogoniques* certainement chantés (d'origine essentiellement anglo-normande), intégrés dans le corpus des trouvères. La problématique de l'*hétérogonie* n'a pas de relevance pour notre propos actuel ; celles et ceux qui s'y intéressent sont priés de consulter la base de données *Nouveau Naetebus* : <<http://www.nouveaunaetebus.elte.hu>> [dernière requête : 11/11/2019]. Les poèmes lyriques des trouvères sont répertoriés dans un manuel préparé à l'image du livre de Frank, compilé sans l'assistance de l'ordinateur ; voir Ulrich Mölk et Friedrich Wolfzettel, *Répertoire métrique de la poésie lyrique française des origines à 1350*, Munich, Wilhelm Fink, 1972.

<sup>9</sup> Ajoutons aussi que la composition strophique assurait déjà en elle-même une série de contraintes très fortes, écartant du *Nouveau Naetebus* une bonne partie de la production poétique française du Moyen Âge (romans à rimes plates, chansons de geste en laisses, etc.).

part, l'exigence de permettre la comparaison de ce corpus avec d'autres traditions poétiques qui ont été construites potentiellement à partir des mêmes règles ou contraintes (encore peut-être inconnues ou méconnues).

### 3. Les tentatives d'harmonisation des répertoires poétiques

Les réflexions théoriques sur la coordination des répertoires et des bases de données poétiques ont un passé relativement long. Nous nous contentons de renvoyer ici à quelques études sur la question. Déjà le manuel du RPHA, paru en 1992 à Paris, a été conçu comme collaboration au projet *Corpus Poeticarum*, idée du linguiste russe Evgueni Polivanov, qui voulait recueillir la poésie de toute l'humanité. L'inspiration a été fournie par les membres du Centre de Poétique Comparée de Paris, dont Jacques Roubaud, également membre de l'Oulipo. À la fin des années 1990, Dominique Billy et Thierry Glon ont proposé des conseils pour l'élaboration d'un répertoire poétique qui couvrirait toute la poésie médiévale<sup>10</sup>. Dans un texte écrit en 2003, publié en 2008, nous avons continué dans ce même sillon<sup>11</sup>. Par la suite, plusieurs colloques ont été consacrés à la problématique de l'harmonisation des bases de données de la poésie médiévale entre 2005 et 2013 à Budapest, et plus tard à Madrid<sup>12</sup>. Dans ce cadre nous avons expérimenté l'attachement et la coordination partielle de deux bases de données poétiques : le *Nouveau Naetebus* et le RPHA. L'homogénéisation de certains critères de description a permis de faire des recherches à la fois dans les deux bases de données, pour certains champs au moins (structure syllabique, formule de rime). Cette expérience (le *Megarep*) avait pour but et pour résultat de démontrer la parenté formelle de textes appartenant à deux littératures éloignées et par leur langue et dans le temps et dans l'espace<sup>13</sup>. La solution technique utilisée était fondée sur le protocole *Open Search* et celle-ci a demandé une intervention d'ordre technique minimale, sans l'obligation de restructurer les bases de données restées

<sup>10</sup> Dominique Billy et Thierry Glon, «Vers un répertoire métrique général des strophes du moyen âge», dans *Métriques du Moyen Âge et de la Renaissance*, éd. Dominique Billy, Paris, Harmattan, 1999, pp. 305-315.

<sup>11</sup> Cf. le dernier chapitre de Levente Seláf, *Chanter plus haut. La chanson religieuse vernaculaire au Moyen Âge (essai de contextualisation)*, Paris, Champion, 2008.

<sup>12</sup> Le matériel des conférences de Budapest est consultable partiellement sur le site de la revue *Ars metrica*, <<https://arsmetrica.elte.hu/>> [dernière requête : 11/11/2019]. Concernant le projet ERC intitulé *Postdata*, voir <<http://postdata.linhd.es/project/>> [dernière requête : 11/11/2019].

<sup>13</sup> Voir à ce propos Elena González-Blanco Garcia et Levente Seláf, «*Megarep: A comprehensive research tool in Medieval and Renaissance poetic and metrical repertoires*», in *Humanitats a la xarxa: món medieval. Humanities on the web: the medieval world*, éd. Lourdes Soriano, Helena Rovira et al., Bern-Berlin-Vienne, Peter Lang, 2014, pp. 333-344.

autonomes<sup>14</sup>. Même dans cette forme minimaliste, la conjointure de deux « corpus ouverts » a prouvé que la coordination des bases de données poétiques fait sens. Récemment, c'est autour d'Elena Gonzalez-Blanco qu'un groupe s'est formé afin de promouvoir l'harmonisation des répertoires poétiques, désormais non seulement pour les périodes du Moyen Âge et de la Renaissance<sup>15</sup>. Cette tentative de la jointure des bases de données est fondée sur le modèle *Linked Data*, mais on ne sait pas encore quand aboutira le projet.

#### 4. La description idéale : vers la TEI et au delà

Il faudra décider du modèle de base de la coordination des répertoires poétiques, et surtout de leur unification. Si l'on veut créer une plateforme unique qui permette de mener des recherches simultanées dans plusieurs corpus (peu importe si les bases de données conservent leur autonomie ou bien si elles sont fondées dans une base commune), il est nécessaire d'accomplir un travail d'harmonisation pour rendre possible une analyse et une description identique de toutes les pièces. La comparaison de corpus poétiques très différents exige de réserver de la place à des contraintes poétiques non encore identifiées. Une grande difficulté de la coordination est que la profondeur de la description des poèmes dans les bases de données différentes qui existent déjà n'est pas forcément identique, et il est tout sauf certain que les éléments poétiques relevés dans une base de données soient aussi enregistrés dans les autres.

Une solution pour le traitement identique des poèmes serait de les numériser et de les étiqueter à partir des mêmes principes. La langue de codage TEI-Unicode pourra servir comme cadre à l'enregistrement de tous les éléments des poèmes qui sont significatifs du point de vue poétique<sup>16</sup>. Mais la TEI-Unicode n'est qu'un cadre abstrait, assez souple, dont le manuel ne fixe

<sup>14</sup> Levente Seláf et Péter Király, «Federated search engine of European Poetical databases. A proposal». Article de blog sur le site *Kirunews*, en ligne : <[http://kirunews.blog.hu/2014/10/26/federated\\_search\\_engine\\_of\\_european\\_poetical\\_databases](http://kirunews.blog.hu/2014/10/26/federated_search_engine_of_european_poetical_databases)> [dernière requête : 11/11/2019].

<sup>15</sup> Helena Bermúdez-Sabel, Mariana Curado Malta et Elena Gonzalez-Blanco, «Towards Interoperability in the European Poetry Community: The Standardization of Philological Concepts», dans éd. Jorge Garcia *et al.*, *Language, Data, and Knowledge: First International Conference, Proceedings*, Cham, Springer, 2017, pp. 156-165.

<sup>16</sup> *P5: Guidelines for Electronic Text Encoding and Interchange*, Version 2.9.1, URL2. Pour les principes actuels du codage des textes versifiés voir le site de TEI : <<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/VE.html>> [dernière requête : 11/11/2019]. La langue de codage TEI-Unicode, la plus courante actuellement dans le domaine de la philologie numérique, n'a pas été inventée et élaborée en vue de la description poétique et rhétorique des textes ; néanmoins, grâce à la collaboration des spécialistes de la littérature, elle a également évolué dans cette direction. Elle a été complétée au fur et à mesure lors des ateliers et congrès, par les ajouts des philologues qui avaient besoin de nouveaux éléments pour le travail de codage.

pas la définition de tous les procédés rhétoriques et métriques ; elle permet d'ajouter des valeurs et des critères individuels nécessaires pour l'analyse des poèmes, mais le travail de définition reste à accomplir pour beaucoup de caractéristiques spéciales des poèmes médiévaux. On ne peut pas espérer de l'emploi de cette langue de codage la résolution de toutes les difficultés concernant l'analyse : son utilisation permet de travailler systématiquement et avec rigueur, mais ne remplace pas la collection intégrale et la définition précise de chaque contrainte et caractéristique poétique<sup>17</sup>.

Les initiatives récentes d'unification des traits poétiques enregistrés dans les répertoires et bases de données sont fondées sur l'accumulation des caractéristiques poétiques déjà reconnues par des analyses et enregistrées comme figures rhétoriques, métriques ou stylistiques dans les traités de poétique historiques. En partant de l'idée d'un système de description assez complexe, et en ajoutant au fur et à mesure les traits poétiques d'autres poèmes, d'une tradition poétique différente, aux informations déjà contenues dans la base de données unifiée. Une autre voie nous serait offerte par les poétiques génératives. Si l'on prévoit la création d'un système de contraintes poétiques qui englobe systématiquement toutes les possibilités de la création poétique, sans se soucier au préalable si elles ont été réellement expérimentées ou non, nous avons de fortes chances de ne rien laisser de côté : toutes les figures de style et de rhétorique, tous les éléments de la versification seront ainsi captés et enregistrés. L'idée n'est pas nouvelle : nous renvoyons au tableau des contraintes de l'Oulipo, les illusoires « tables queneleieff » inventées en 1974, des grilles de contraintes généralisables permettant la prise en compte des structures potentielles d'un quelconque texte poétique. Conformément à l'esprit de l'Oulipo, cette approche permettrait d'inclure dans la base de données désirée non seulement les poèmes composés à partir de contraintes existantes, mais aussi ceux qui seront composés à partir de contraintes non encore expérimentées. C'est une solution idéale (ce qui ne veut pas dire qu'elle soit réalisable).

---

<sup>17</sup> En plus la numérisation des textes poétiques reste un travail dur et long et ses pratiques sont loin d'être évidentes et identiques (pensons à la différence entre transcription et étiquetage des manuscrits individuels, d'une seule version du texte ou d'un texte critique), et l'automatisation de ce processus n'est pas résolue. Même si l'étiquetage est très précis, il sera nécessaire de créer des logiciels puissants, capables de l'analyse des données récoltées à partir des critères préétablis. Une analyse (humaine ou numérique) des textes demandera dans chaque cas la vérification des données textuelles à la recherche des contraintes poétiques encore non-inclues dans la langue de codage.

## 5. Conclusion

Pour conclure je voudrais présenter une analogie historique fort semblable aux initiatives actuelles de la systématisation de l'analyse des poèmes : un cas spécial où les visées descriptive et générative ont été jointes. Guilhem Molinier, poète occitan postérieur à l'époque glorieuse des troubadours, est devenu théoricien de la poésie quand il devait trouver les moyens de décrire la production poétique de ses prédécesseurs. Mais il avait aussi l'ambition par son travail de montrer les voies de la création future, pour ladite « école poétique de Toulouse ». Les *Leys d'amors* (1323-1338) est un traité de poétique qui avait pour ambition première de fixer les règles de composition de la poésie des troubadours, mais l'ouvrage est devenu à la fois une somme des règles et un traité de poétique expérimentale, élevant son auteur au statut de précurseur des répertoireurs et des poéticiens des XX<sup>e</sup> et XXI<sup>e</sup> siècles.

Un ouvrage tellement hybride peut nous servir d'une réelle source d'inspiration. Le caractère composite de ce traité de poétique a provoqué les jugements ambivalents de Joseph Anglade, éditeur du texte, qui n'a pas trouvé très louables les efforts de Guilhem Molinier, esprit d'une époque « décadente »<sup>18</sup>:

Le rédacteur des *Leys* nous apprend que le Consistoire, en lui confiant ce travail, a voulu rassembler tout ce qui jusque-là avait été dispersé. [...] Il a sans doute inventé l'idée des différents mariages qui ont donné tant de filles, neveux et nièces à Barbarisme, Solécisme et Allebolus ; les divisions et subdivisions d'une matière qui en comporte tant devaient représenter la fin du fin de la science de son temps ; toutes les figures de pensée, de mots, tous les schémas et tous les *colores rhetorici* que le Moyen Âge avait catalogués et classés se trouvent réunis dans ce livre quatre des *Leys*<sup>19</sup>.

C'est un traité qui veut être normatif pour la grammaire, le vocabulaire et la poésie, donc très contraignant, avec appel à l'autorité des anciens troubadours et à la tradition rhétorique. Néanmoins, il est également très inventif et surtout libre à exploiter des formes poétiques auparavant inconnues. L'ouvrage devait servir de manuel aux poètes contemporains, avec parfois des conseils de composition pratiques. Anglade remarque ainsi à propos du second chapitre du cinquième livre : « c'est une recette pour écrire des vers au moyen d'un dictionnaire de rimes »<sup>20</sup>.

<sup>18</sup> *Las Leys d'amors. Manuscrit de l'Académie de Jeux Floraux, Tome IV. Études, notes, glossaire et index*, éd. Joseph Anglade, Toulouse, Privat, 1920.

<sup>19</sup> *Ibidem*, p. 80.

<sup>20</sup> *Ibidem*, p. 81.

Il y a encore d'autres traits originaux de cette compilation :

Guilhem Molinier a traité la rythmique latine comme la grammaire latine ; il lui a emprunté les éléments qui pouvaient lui servir pour sa métrique de la langue vulgaire et a pris à cette métrique qui depuis plusieurs siècles avait fait ses preuves, les compléments qu'il n'aurait pas trouvés dans les traités latins. Il en a sans doute aussi inventé une partie, en se fondant peut-être sur des modèles de son temps, où devaient fleurir les fantaisies métriques qui éclosent à toutes les périodes de décadence<sup>21</sup>.

Et surtout, on peut voir dans les *Leys d'amors* le reflet de l'application d'une pratique de composition combinatoire. Molinier a enregistré parmi les règles de compositions des contraintes potentielles qu'il n'a jamais rencontrées auparavant, et qu'il était probablement le premier à mettre en œuvre :

Mais il a probablement gardé quelques-uns de ses exemples (peut-être quelques-unes de ses définitions) à la versification rythmique de son temps ; ces emprunts ne sont pas sans doute les plus heureux, pas plus que les modèles que son imagination a dû lui suggérer dans le domaine immense des combinaisons strophiques. Certaines de ses réflexions, même très discrètes, prouvent qu'il s'est rendu compte plus d'une fois de l'excès de subtilité – et de puérilité – de quelques-unes de ces inventions, qu'elles lui fussent propres ou qu'il les ait empruntées. Dans les vraies « exemples » de poésie que contiennent les *Leys* (*coissirs*, chansons, pièces lyriques complètes), Molinier n'est tombé dans aucune des ridicules combinaisons rythmiques qu'il a complaisamment énumérées dans son œuvre ; et là aussi le métricien a fait preuve de goût, et même d'un goût de vrai poète dans la série des *coissirs* et dans le poème de la *Contemplacio de la Crotz*<sup>22</sup>.

Contrairement à Anglade, nous pensons que cette hybridité de l'approche poétique de Guilhem Molinier n'est pas à blâmer. C'est plutôt un exemple à suivre, même si, du point de vue d'Anglade, son imitation signifie l'aveu que notre époque est décadente, elle aussi. Il est bien possible que l'élaboration d'une base de données universelle de la poésie devrait se passer ainsi, d'une manière tâtonnante, exploratoire ; par une approche fondée sur la bonne connaissance de la matière poétique existante mais aussi par l'exploration de la potentialité des règles de composition : un peu à la manière de l'Oulipo.

Reçu : 19/07/2019  
Accepté : 3/10/2019

---

<sup>21</sup> *Ibidem*, p. 89.

<sup>22</sup> *Ibidem*, p. 90.



LES VOIES DE LA CRÉATION D'UNE BASE DE DONNÉES POÉTIQUE GÉNÉRALE

RÉSUMÉ : Cet article propose une réflexion sur les méthodes potentielles de l'analyse comparative des corpus poétiques, assistée par ordinateur. On distingue les « corpus clos », dont l'unité est garantie par la présence de certaines règles de composition dans chaque élément, des « corpus ouverts », où les critères chronologiques ou linguistiques comptent, sans prendre en considération les contraintes poétiques (ou bien le corpus est défini par l'absence d'une contrainte poétique spécifique). La pratique courante est l'étiquetage avec le code TEI, mais une description des poèmes dans la plus grande profondeur possible reste rare. Comme solution alternative, idéalisée, de poésie générative, on propose la construction d'une grille de description semblable aux « tables queneleieff » de l'Oulipo ; un exemple précoce de ce type d'approche est le traité de poésie *Leys d'amors* de Guilhem Molinier (XIV<sup>e</sup> siècle), qui contient des éléments de poésie combinatoire.

MOTS-CLÉ : Répertoires poétiques. Bases de données poétiques. Poésie générative. TEI. Oulipo. *Leys d'amors*.

THE WAYS TO CREATE A GENERIC DATABASE OF POETRY

ABSTRACT: The article offers a reflection on the potential methods of digital comparative analysis of poetical corpora. Two types of corpora are opposed: the closed ones, defined by some common poetical patterns shared by all poems, and the open ones, based exclusively on chronological or linguistic criteria, or the absence of a specific poetical constraint. Currently the TEI-encoding of the poems is the most frequent practice of description, but very uneven in the deepness of analysis. As an idealistic alternative the creation of a global, generative poetical set of poetical tools, like the «Queneleieff tables» of the Oulipo, is proposed; an early example of this approach is to be found in Guilhem Molinier's 14<sup>th</sup> century poetical treatise, the *Leys d'amors*.

KEYWORDS: Poetical repertories. Poetical databases. Generative poetics. TEI. Oulipo. *Leys d'amors*.

