Vers long, vers court, lieux de différenciation lexicale et linguistique : le cas du distique élégiaque latin

Gérald Purnelle

Université de Liège – Gerald.Purnelle@ulg.ac.be

Abstract

Basing itself on an exemplary case in Latin poetry, this paper will ask two questions: does metrical form have an influence on the linguistic and lexical choices of a poet? Can short and long verses in a given poetry be differentiated using lexical, grammatical (Part of Speech frequency) and length of words criteria?

The corpus used in this research is based on a precise Latin poetic form, the elegiac couplet, made of a dactylic hexameter followed by a dactylic pentameter, two verses of different length and structure. It is composed of 9 works from 4 different authors and is divided in 48 subsets. Each text of the corpus is systematically parted in two subsets, hexameters and pentameters, so as to be able to oppose a formal parameter (verse length) to other parameters (author and text topics).

Résumé

À partir d'un cas exemplaire propre à la poésie latine, la communication pose deux questions liées à la forme métrique : celle-ci a-t-elle une influence sur les choix lexicaux et linguistiques opérés par le poète ? Les vers plus courts ou plus longs de l'arsenal métrique d'une poésie donnée se différencient-ils selon les critères du lexique, de la grammaire (fréquence des parties du discours) et de la longueur des mots ?

Le corpus étudié pour cette première recherche porte sur une forme poétique latine particulière, le distique élégiaque, formé de deux vers de structure et de longueur différentes, l'hexamètre et le pentamètre dactyliques. Les œuvres observées sont au nombre de 9 (divisées en 48 sous-ensembles) et relèvent de 4 auteurs différents. Chaque texte est divisé en deux sous-ensembles, les hexamètres et les pentamètres, afin d'opposer le paramètre de la forme (nature et longueur du vers) aux autres (essentiellement l'auteur et la thématique des textes).

Mots-clés: poésie latine, distique élégiaque, forme métrique, statistique lexicale, genre.

1. Problématique

Les facteurs qui induisent distance ou proximité entre les textes et que la statistique lexicale, appliquée à un corpus littéraire, fait généralement ressortir, sont bien connus : ils vont de la chronologie jusqu'à l'auteur, en passant par le genre et la thématique.

Parmi ces facteurs, le plus puissant est sans conteste le genre, comme il ressort de maintes études (cf., par exemple, Brunet 1988 et 2009). Les facteurs ne sont pas indépendants : la thématique et le genre (ou le sous-genre) sont assez étroitement interdépendants, et c'est au moins autant la première que le second qui distingue roman et poésie ou comédie et tragédie.

Quant au facteur « auteur », il ne peut être totalement indépendant que quand les autres paramètres sont neutralisés, comme dans le cas d'un corpus entièrement romanesque.

À la notion de genre est souvent liée celle de forme : la prose pour le roman, l'essai ou la correspondance s'oppose au vers pour la poésie, avec la possibilité d'exceptions (le poème en prose) et d'hybridité (le théâtre est en vers ou en prose).

Le paramètre « forme », sans existence dans le roman (sauf exceptions), interfère-t-il avec les autres ? Peut-il, par exemple se cumuler au paramètre « genre » ? Oui, comme le montre la distinction statistique entre comédies en vers et en prose chez Molière (cf. Brunet 2009). Il apparaît donc que dans un même genre, censé véhiculer les mêmes thématiques et donc le même lexique, celui-ci peut varier selon la forme. Or un trait capital distingue les deux formes majeures, le vers étant contraint et la prose, libre. Ce qui amène à une question précise : dans quelle mesure le vers contraint-il la langue (morphologique et lexicale) de l'auteur ?

En préliminaire à d'autres investigations, l'étude qui suit présente une première expérience portant sur l'impact de la forme-vers sur la distance intertextuelle. En attendant de passer à des corpus récents (la poésie française du 19^e et du 20^e), le choix s'est porté sur un exemple latin et se fonde sur quelques fichiers de la base de données textuelles lemmatisées du LASLA.

2. Le distique élégiaque

L'objet est le distique élégiaque et le corpus sélectionné est constitué des œuvres composées dans cette forme. La poésie élégiaque, d'origine grecque et adoptée par les Latins, se définit entièrement par sa forme métrique : tout poème est constitué d'une séquence de distiques élégiaques, cette paire de vers formant module et étant constituée d'un hexamètre dactylique et d'un pentamètre dactylique.

L'hexamètre est constitué de six pieds dactyliques ; le dactyle est un pied constitué d'une syllabe longue et de deux brèves (¯); chacun des quatre premiers pieds peut être remplacé par un spondée, constitué de deux syllabes longues ; le cinquième pied est (presque) toujours un dactyle ; le sixième est catalectique, c'est-à-dire qu'il ne présente que deux syllabes, un longue et une syllabe de longueur indifférente. Le schéma le plus long que peut réaliser un hexamètre est donc : ¯ ¯ / ¯ / ¯ / ¯ / ¯ / ¯ / ¯ et compte 17 syllabes (exemple : ārmă grăvī nŭměrō vĭŏlēntă quě bēllă părābăm, Ovide, Amores, 1,1,1) ; le plus court est ¯ / ¯ / ¯ / ¯ / ¯ / ¯ et en compte 13 (exemple : nātūrāe dēcūs mērcātō pērděrě cūltǔ, Properce, 1,2,5).

Le pentamètre est constitué de deux séquences correspondant à un début d'hexamètre, plus précisément à un hémistiche initial devant une césure penthémimère (c'est-à-dire intervenant après le 5° demi-pied) : ~~/~~/~; les substitutions de spondées aux dactyles sont possibles dans le premier hémistiche du pentamètre (~~/~/~ ou ~/~~/~), non dans le second. La forme la plus courte du pentamètre est donc ~~/~//~/~(12 syllabes ; exemple : cōntāctūm nūllīs āntě cǔpīdĭnĭbūs, Properce, 1,1,2) et la plus longue ~~/~//~/~(14 syllabes ; exemple : ēt căpūt īmpŏsĭtīs prēssĭt Ămōr pědĭbūs, Properce, 1,1,4).

Le pentamètre est donc en moyenne légèrement mais sensiblement plus court que l'hexamètre. On dispose ainsi, à l'intérieur des mêmes poèmes et donc des mêmes corpus, de deux formes qui se différencient par leur longueur.

3. Le corpus

Les poètes élégiaques latins sont au nombre de quatre. 52 des 116 poèmes de Catulle (87-54 av. J.-C.) sont rédigés en distique élégiaque et datent du milieu du 1^{er} s. av. J.-C. (v. 62 – 54).

La publication des quatre livres de Properce (v. 47 – b. 16 av. J.-C.) s'étale de 25 à 16 av. J.-C. Les deux premiers ont une thématique amoureuse, qui évolue vers des sujets divers dans le troisième et touche à l'explication des rites et mythes latins dans le quatrième.

De Tibulle (v. 54-19 av. J.-C.), seuls les deux premiers livres sont entièrement authentiques ; le premier est publié en -26 ou -25, le second est posthume. Les deux autres livres rassemblent des poèmes d'auteurs multiples, parmi lesques Lygdamus et Sulpicia. Il n'est pas sûr d'attribuer les autres à Tibulle (Le poème VII, « Panégyrique de Messala », étant tout entier en hexamètre, a été soustrait du corpus). La thématique des quatre livres est majoritairement amoureuse. Le $4^{\rm e}$ n'est pas élégiaque.

Ovide (43 av. – 17 ap. J.-C.) laisse plusieurs œuvres élégiaques : les *Héroïdes*, lettres fictives d'héroïnes mythologiques (– 19 ?), les *Amours* (*Amores*) (– 16-15), les *Medicamina faciei*, l'*Art d'aimer* (*Ars amatoria*) et les *Remèdes à l'amour* (*Remedia amoris*) (2 ap. J.-C.), les *Fastes* (*Fasti*), traité des fêtes romaines (v. + 8). Ses deux derniers recueils, les *Tristia* et les *Epistulae ex Ponto*, poèmes épistolaires écrits en exil, ne font pas partie de notre corpus.

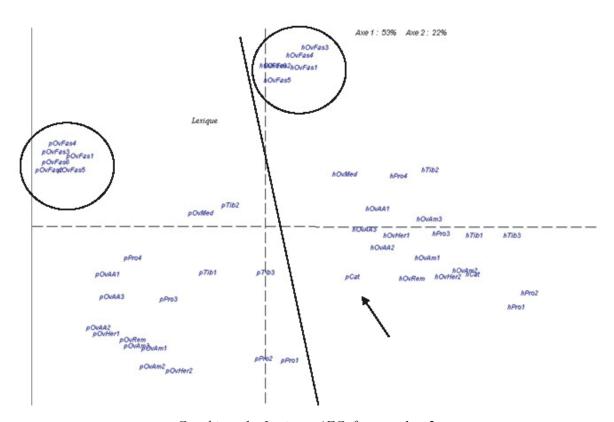
Ce corpus poétique réduit au maximum les facteurs autres que la forme métrique : il ne confronte pas prose et poésie (de toute manière, la littérature latine répartit radicalement les fonctions des formes : ainsi n'y a-t-il pas de théâtre en prose) ; l'empan chronologique est étroit ; le corpus est limité à un seul genre et un seul sous-genre (formel). Seuls demeurent comme facteurs concurrents de la forme l'auteur et la thématique. L'amour est le thème dominant, mais non exclusif, chez chacun des quatre poètes.

Afin d'affiner l'analyse en testant l'éventuelle homogénéité de chaque auteur ou de chaque œuvre, nous structurons notre corpus global selon la division de certaines œuvres en livres (Properce, Tibulle, Ovide hormis les *Medicamina* et les *Remedia*). En outre, les 7 dernières des 21 *Héroïdes* d'Ovide soulèvent des problèmes d'authenticité. Nous divisons donc ce souscorpus en deux : les 14 premières et les autres.

Chacun des 24 sous-ensembles textuels ainsi produits (recueils ou livres) a été divisé en deux fichiers, l'un contenant tous les hexamètres, l'autre les pentamètres. L'ensemble est constitué de 48 fichiers.

4. Statistique lexicale

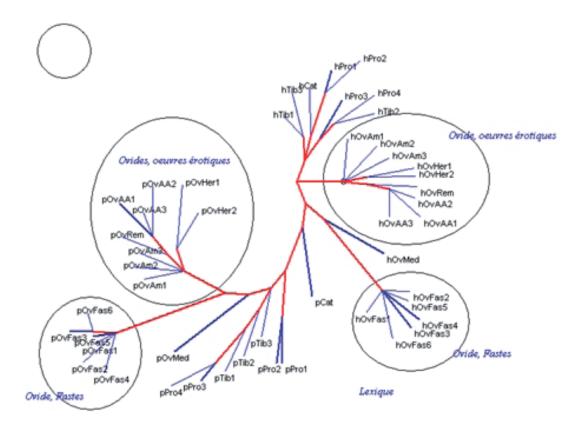
Les hexamètres et les pentamètres alternent régulièrement dans le cours de chaque poème. On peut donc s'attendre à ce que l'énonciation y sélectionne et dispose le lexique d'une façon qui ne diffère pas d'un texte composé dans un mètre unique (par exemple l'épopée latine, fondée sur l'hexamètre) ou d'un texte en prose.



Graphique 1 : Lexique, AFC, facteurs 1 et 2.

Il n'en est rien : le graphique 1 montre que le premier facteur de l'AFC oppose de façon extrêmement nette les sous-corpus de pentamètres (dont les noms commencent par p) et ceux d'hexamètres (h). Les poètes n'utilisent donc pas le même lexique dans les deux types de vers ; ou, pour nuancer : chaque mètre, hexamètre ou pentamètre, privilégie une certaine part du lexique global. Relevons toutefois la position différente des pentamètres de Catulle (pCat) sur laquelle nous reviendrons.

La thématique n'agit qu'en second facteur : dans la moitié supérieure figurent les *Fastes* d'Ovide et le livre 4 de Properce, dont la thématique n'est ni amoureuse ni érotique, mais religieuse et historique. La forme est un facteur plus puissant que la thématique elle-même. Le phénomène est confirmé par le graphique de l'analyse arborée :

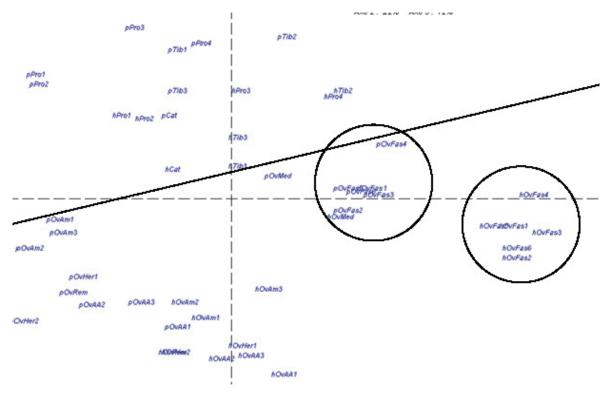


Graphique 2 : Lexique, analyse arborée

L'auteur n'intervient qu'en troisième facteur, comme le montre le graphique relatif aux 2° et 3° facteurs, qui oppose Ovide aux autres poètes (graphique 3); encore voit-on à nouveau que, pour les *Fastes*, la distinction hexamètre-pentamètre reste partiellement opérante.

Il n'y a donc qu'une exception à la partition du corpus par le premier facteur : les pentamètres de Catulle figurent dans la partie droite de l'AFC. Le poète le plus ancien paraît donc moins frappé par la tendance observée : la spécialisation du lexique selon la forme.

Ce dernier phénomène est étonnant : ce qui, dans un même genre (la poésie) distingue le lexique d'un texte se manifeste globalement ou, à tout le moins, à l'intérieur de parties d'une certaine étendue. Mais on ne s'attend pas à ce que le phénomène joue d'un vers à l'autre et systématiquement. Il convient donc de pousser plus avant l'investigation, en dépassant la dimension sémantique (et donc thématique) du lexique pour se tourner vers sa morphologie.



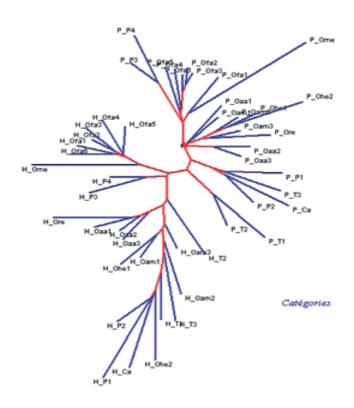
Graphique 3 : Lexique, AFC, facteurs 2 et 3.

5. Les catégories grammaticales

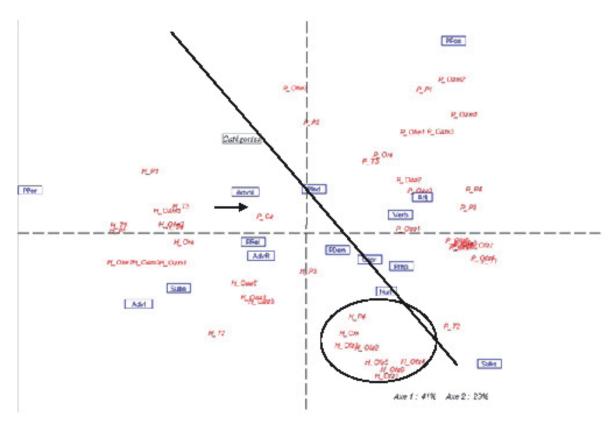
Le second test évalue la répartition des catégories grammaticales dans les 48 sous-ensembles. Celles-ci sont assez nombreuses dans le codage morphologique des fichiers du Lasla. Nous avons procédé à quelques regroupements pour des raisons quantitatives et détaillé les catégories suivantes : substantif, verbe, adjectif, numéral, pronom personnel + pronom réfléchi, pronom possessif + pronom possessif réfléchi, pronom démonstratif, pronom relatif + pronom interrogatif, pronom indéfini, adverbe + interjection, adverbe relatif + adverbe interrogatif + adverbe interrogatif, adverbe négatif, préposition, conjonction de coordination, conjonction de subordination — soit 15 colonnes d'effectifs pour chaque sous-ensemble.

L'analyse arborée montre à nouveau une distinction nette entre sous-ensembles hexamétriques et pentamétriques (graphique 4). On ne suppose pas à priori d'impact direct de la thématique sur la morphologie. Ce paramètre n'en est pas pour autant totalement indépendant. Ainsi, dans le graphique 5, la position excentrée des pronoms personnels (PPer) indique un fait qui structure pour une bonne part le graphique : ces pronoms sont spécifiques de la partie lyrique du corpus global (la poésie amoureuse), tandis que les 1^{re} et 2^e personnes seront moins fréquentes dans des textes étiologiques comme les *Fastes*. Par ailleurs, dans les textes où *ego*, *tu*, *vos* et *nos* sont fréquents, ils paraissent se placer préférentiellement dans les hexamètres. Le cumul des deux facteurs explique le groupement à gauche des parties H des textes amoureux, et à droite des parties P, mais aussi des parties H des *Fastes*.

Pour le reste, les catégories les plus spécifiques du pentamètre sont l'adjectif possessif (PPos : *meus, tuus, suus*), surtout chez Ovide, et, dans une moindre mesure, l'adjectif et le verbe (Adj, Verb). Quant au substantif (Subs), il est attiré autant par le pentamètre que par les *Fastes*, ce qui paraît indiquer une évolution chronologique chez Ovide, qui, en passant du corpus érotique aux *Fastes*, a rééquilibré la répartition des substantifs entre hexamètre et pentamètre.



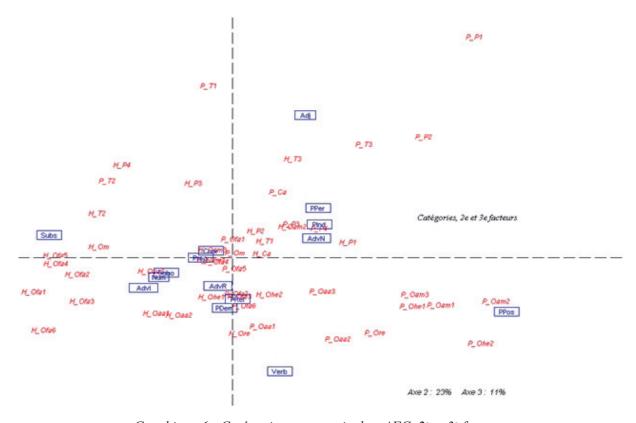
Graphique 4 : Catégories grammaticales, analyse arborée



Graphique 5 : Catégories grammaticales, AFC

Quant à l'hexamètre, il attire davantage les conjonctions de subordination et les adverbes (et interjections) (Subo et ADvI).

Il n'y a donc au total que certaines catégories bien précises qui sont en lien privilégié avec l'un ou l'autre mètre, et les facteurs métrique, thématique et stylistique se cumulent ou se concurrencent.



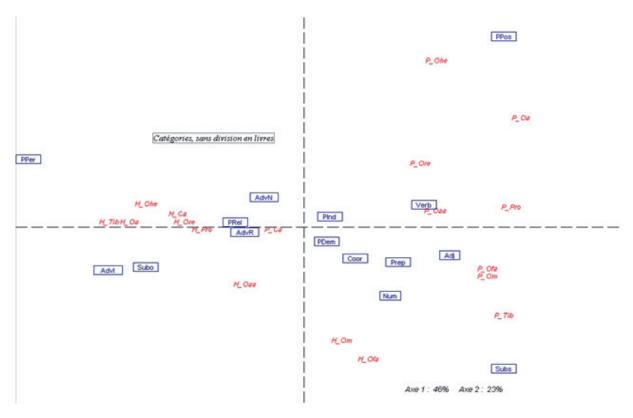
Graphique 6 : Catégories grammaticales, AFC, 2^e et 3^e facteurs

Pour le reste, le plan des 2° et 3° facteurs montre à nouveau que l'auteur joue comme troisième facteur, en opposant Ovide (en dessous) et les autres (graphique 6).

Tous les graphiques qui précèdent sont fondés sur la subdivision de certaines œuvres en livres, afin de tester l'homogénéité de ces œuvres, voire de chaque auteur, ou, à l'inverse, d'observer une éventuelle variation thématique (Properce) ou auctoriale (Tibulle) à l'intérieur d'une même « œuvre ». Les écarts éventuels entre parties d'une même œuvre ne sont pas à ce point grands qu'ils puissent indiquer des phénomènes significatifs. Il est donc possible de regrouper les parties d'œuvres, en ramenant ainsi les sous-parties du corpus à 18 (9 œuvres), et sans que les observations qui précèdent soient grandement modifiées (graphique 7) : les hexamètres sont à gauche avec les pentamètre de Catulle, les autres pentamètres à droite ; la position des catégories est inchangée

6. Contrainte et stylistique

En tant que forme métrique, le distique élégiaque exerce donc un effet certain sur la syntaxe de la phrase et, partant, sur la position des catégories grammaticales. Le phénomène est d'ordre essentiellement stylistique. Dans un poème entièrement écrit en hexamètres (par exemple le poème 64 de Catulle), la phrase court de vers en vers en déployant une longueur qui lui est propre, varie fortement et relève de l'énoncé. La phrase peut même s'achever ailleurs qu'en fin de vers (cf. Purnelle 2005).



Graphique 7 : Catégories grammaticales, AFC sur les œuvres réunifiées

Le distique, quant à lui, fonctionne comme un module autonome. Selon les poètes, les phrases qui dépassent une longueur d'un distique ou lui sont inférieures sont moins nombreuses que celles qui s'y inscrivent. Le poète élégiaque subit ou respecte une contrainte et dispose sa phrase en répartissant (en moyenne) ses catégories d'un vers à l'autre : dans ces phrases-distiques, le poète tend à placer de façon répétée les mêmes éléments syntaxiques aux mêmes endroits (hexamètre ou pentamètre). Stylistiquement, étant donné la relative liberté de l'ordre des mots en latin, le poète privilégie certaines catégories dans un vers et dans l'autre. Souvent, le verbe principal occupera le pentamètre et les subordonnées (avec leur conjonction) l'hexamètre (ce qui explique la position du verbe près de la ligne centrale du graphique). Enfin, un tic stylistique facilement observable amène les poètes élégiaques, singulièrement Ovide, à placer dans le pentamètre deux syntagmes nominaux constitués d'un substantif et de son épithète (ou de son adjectif possessif)¹; exemple : virginea tenuit cornua vara manu (« elle tint d'une main virginale les cornes recourbées », Ovide, Amores, I, 3, 24). Un trait stylistique comparable amène Ovide à placer souvent l'adjectif possessif (meus, tuus, suus) en fin de pentamètre, ce qui explique leur affinité: Conveniunt cumbae vela minora meae (« de moindres voiles conviennent à ma nacelle », Ovide, Ars amatoria, 1, 264).

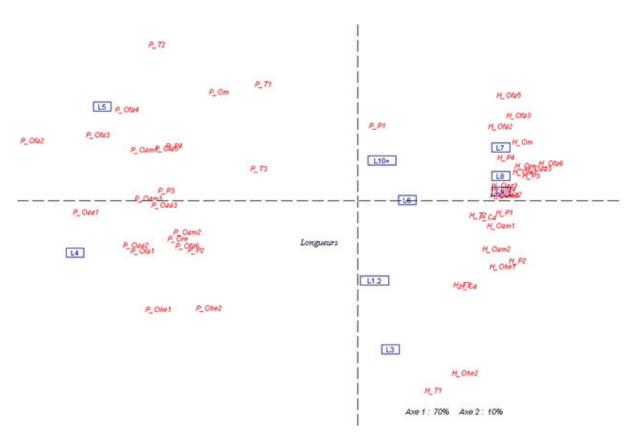
On notera à nouveau la position centrale des pentamètres de Catulle, qui indique une nette indifférence à l'égard de toutes les catégories : apparemment, Catulle compose son distique d'une façon plus libre que ses successeurs ; peut-être même ne l'écrit-il guère différemment de son hexamètre. Cela peut s'expliquer par la diversité de sa pratique poétique : il a composé des poèmes lyriques en vers logaédiques (poésie lyrique strophique), en vers iambiques et en hexamètres (poème 64). Aucun des trois élégiaques de l'époque d'Auguste n'a écrit de poésie lyrique (au sens formel du terme) ; seul Ovide pratiquera le poème totalement hexamétrique, avec les *Métamorphoses*. Encore sont-elles postérieures aux élégies amoureuses et contemporaines des *Fastes*.

Quelle que soit la fréquence de ces trois catégories dans l'hexamètre, de tels procédés stylistiques déterminent directement leur fréquence dans le pentamètre.

La plus ou moins grande affinité de l'hexamètre ou du pentamètre pour les quatre catégories fréquentes et lexicalement abondantes, substantif, verbe, adjectif et adverbe, détermine la différenciation lexicale des deux mètres : ne partageant pas à parts égales les substantifs et les verbes, hexamètre et pentamètre ne présentent pas un lexique homogène dans un même texte : la forme contraint le lexique à travers la morphologie — davantage que la thématique.

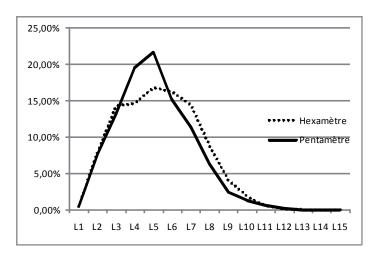
7. La longueur des mots

La longueur des mots est un critère plus radical encore pour tester la répartition du matériau linguistique d'un vers à l'autre (graphique 8). À nouveau, hexamètre et pentamètre se différencient : les mots de 4 et 5 lettres sont nettement privilégiés par le pentamètre, ceux de 7 à 9 lettres par l'hexamètre, les longueurs courtes et plus longues étant partagées.



Graphique 8 : Longueur des mots

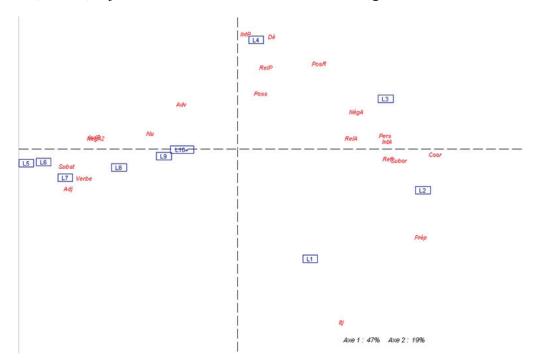
La longueur moyenne des mots est à peine supérieure dans l'hexamètre : de 5,2 (Catulle et *Héroïdes*) à 5,6 (*Medicamina*), pour 5,0 (*Héroïdes*) à 5,3 (Catulle) dans le pentamètre.



Graphique 9 : longueur des mots

Le graphique 9, produit à partir du corpus cumulé, montre qu'une courbe écrasée répartit les longueurs de 3 à 7 lettres dans l'hexamètre (variant de 14,3 à 16,7 %), alors que le pentamètre accuse un pic aux valeurs 4 et 5 (19,4 et 21,7 %)².

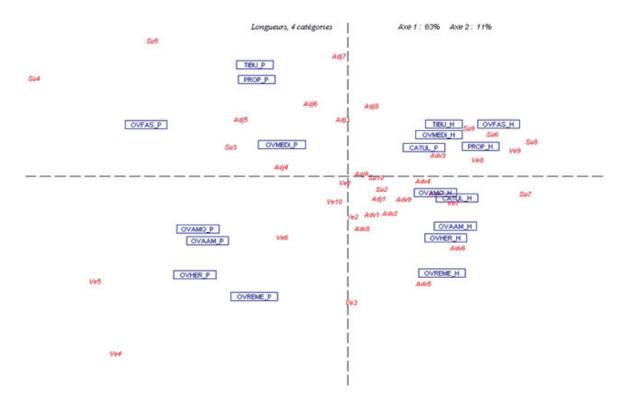
Pour expliquer cette répartition, il convient d'examiner la variation de longueur selon les catégories grammaticales. Les deux paramètres interfèrent nettement (graphique 10) : sans grande surprise, les mots-outils sont en moyenne plus courts, tandis que les longueurs des substantifs, verbes, adjectifs et adverbes s'échelonnent davantage de 5 à 15.



Graphique 10 : Catégories grammaticales et longueur des mots

À nouveau, Catulle se distingue : chez lui, les deux courbes se recouvrent à peu près parfaitement ; la composition de son pentamètre est plus souple et plus libre à l'égard des contraintes de cette forme. Il parvient à nourrir son pentamètre de mots de mêmes natures et de mêmes longueurs que l'hexamètre.

En conséquence, on restreindra le test à ces quatre catégories, qui sont à la fois porteuses du sens et susceptibles de varier le plus en longueur (de 2 à 15 caractères) (graphique 11).



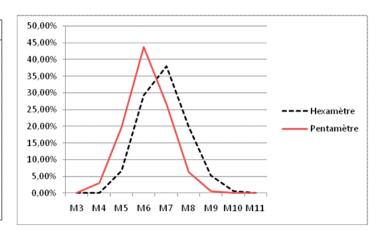
Graphique 11 : Longueur des substantifs, verbes, adjectifs et adverbes

Le graphique fournit l'explication au phénomène observé dans le graphique 7 : le lien entre pentamètre et longueurs de 4 ou 5 lettres est bien dû aux quatre catégories signifiantes. Tous les poètes privilégient les substantifs, verbes, adjectifs et adverbes de 4 et 5 lettres dans le pentamètre, réservant les longueurs inférieures et supérieures à l'hexamètre. Il est tentant d'expliquer cette différence par la longueur même des deux vers du distique : le pentamètre étant plus court, donner aux deux vers une densité sémantique aussi équivalente que possible impose au poète de raccourcir la longueur moyenne des mots placés dans le pentamètre. Mais l'explication ne peut être que partielle : sur le plan métrique, chacun des deux hémistiches de pentamètre ne devrait en rien constituer un espace plus étriqué qu'un premier hémistiche d'hexamètre, puisqu'il en procède. L'explication complémentaire est justement à chercher du côté de la structure du pentamètre : ce « collage » de deux débuts d'hexamètres, dont est fabriqué le pentamètre, induit une césure plus forte que celle de l'hexamètre.

Quoi qu'il en soit, la forme métrique et les contraintes qu'elle induit, différentes d'un vers à l'autre, sont un facteur supplémentaire de différenciation morphologique des deux vers du distique, qui s'ajoute ou même contribue aux tendances syntaxiques et stylistiques évoquées plus haut.

Le calcul des longueurs de vers en nombre de mots confirme ces observations. Le corpus complet révèle un nombre de mots en moyenne plus grand dans l'hexamètre.

	Hexamètre	Pentamètre
3 mots	0,01%	0,02%
4 mots	0,02%	2,97%
5 mots	6,66%	19,77%
6 mots	29,32%	43,67%
7 mots	37,98%	26,65%
8 mots	19,91%	6,32%
9 mots	5,36%	0,57%
10 mots	0,67%	0,03%
11 mots	0,07%	0,00%



La longueur propre (en syllabes) de l'hexamètre autorise donc deux tendances apparemment contradictoires : le placement de mots légèrement plus longs, mais aussi plus nombreux ; c'est évidemment un nouveau facteur expliquant la préférence de l'hexamètre pour les mots courts et, consécutivement, pour les mots outils tels que les conjonctions de subordination, pronoms relatifs ou pronoms personnels, à côté de substantifs, verbes ou adjectifs plus longs que ceux du pentamètre.

8. Conclusion et perspectives

S'il est démontré que, dans le cas du distique élégiaque, la forme propre de chacun des deux vers qui le constitue, c'est-à-dire, notamment, sa longueur, mais aussi sa césure, contribue à déterminer stylistiquement son contenu morphologique (et syntaxique) et, consécutivement, lexical, il ne faut pas se cacher ce que ce cas a de particulier, dès lors qu'il s'est agi de comparer, systématiquement, chaque texte avec lui-même : les « formes » comparées alternent dans les mêmes textes. Les circonstances étaient donc favorables pour « neutraliser » ou contrôler des paramètres tels que la thématique, la chronologie ou même l'auteur.

Il n'en reste pas moins établi que la forme peut être un facteur de détermination et de différenciation lexicales et morphologiques puissant, plus fort même que le genre, la thématique ou l'auteur. Somme toute, le seul facteur concurrent dont nous ayons observé un effet (limité) est chronologique, puisque Catulle, dont les poèmes précèdent ceux des autres auteurs de 30 à 70 ans, présente une autre pratique du distique — mais ceci est sans doute autant affaire de stylistique, et donc d'individualité auctoriale, que de chronologie.

Pour explorer plus avant les potentialités de notre conclusion — la forme contraint la morphologie qui contraint le lexique — et tester l'hypothèse de sa généralité, nous envisageons d'effectuer de semblables tests sur d'autres corpus, latins ou français, en les appliquant à d'autres formes : l'hexamètre, le trimètre iambique, les poèmes logaédiques, pour le latin ; les différentes mesures métriques du vers français, alexandrin, décasyllabe, octosyllabe. Pour le français, il sera même intéressant d'opposer le vers régulier, par exemple l'alexandrin, au vers libre.

Dans la plupart des cas, il ne s'agira plus de comparer un texte à lui-même en le scindant artificiellement en deux sous-ensembles (bien que des cas favorables puissent se présenter), mais d'opposer des sous-corpus textuels homogènes une œuvre, un auteur), fondés sur le critère de la forme métrique. Les enjeux sont donc tout à la fois les mêmes, et différents.

Bibliographie

Brunet (Ét.), 1988: « Hugocentric Tendencies or Can One Approach Hugo Counting Words », dans *Literary and Linguistic Computing*, vol. 3, n° 2, 1988, p. 79-93 (repris dans E. Brunet, *Comptes d'auteurs*, Tome I, *Études statistiques, de Rabelais à Gracq*, texts édités par Damon Mayaffre, Paris, Champion, 2009, pp. 105-130).

Brunet (Ét.), 2009 : « Muller le lexicomaître », dans *Mélanges offerts à Charles Muller pour son centième anniversaire*, Paris, 2009, pp. 99-119 (repris dans E. Brunet, *Ce qui compte*, Écrits choisis, Tome II, *Méthodes statistiques*, Textes édités par Céline Poudat, Paris, Champion, 2011, 331-351).

Purnelle (G.), 2005: « Mètre et syntaxe dans la pratique de trois poètes latins: Catulle, Virgile et Horace», Papers on Grammar IX 2, Proceedings of the Twelth International Colloquium on Latin Linguistics (Bologna, 9-14 june 2003), edited by Gualtiero Calboli, Herder Editrice, Roma, 2005.

Annexe : liste des sous-ensembles du corpus (p- = pentamètres, h- = hexamètres)

Catulle	hCat, pCat	
Ovide, Ars amatoria, livres 1, 2 et 3	hOvAA1, pOvAA1, hOvAA2, pOvAA2, hOvAA3, pOvAA3	
Ovide, Amores, livres 1, 2 et 3	hOvAm1, pOvAm1, hOvAm2, pOvAm2, hOvAm3, pOvAm3	
Ovide, Fasti, livres 1 à 6	hOvFas1, pOvFas1, hOvFas2, pOvFas2, hOvFas3, pOvFas3, hOvFas4, pOvFas4, hOvFas5, pOvFas5, hOvFas6, pOvFas6	
Ovide, Heroides 1 à 14	hOvHer1, pOvHer1	
Ovide, Heroides 15 à 21	hOvHer2, pOvHer2	
Ovide, Medicamina faciei	hOvMed, pOvMed	
Ovide, Remedia amoris	hOvRem, pOvRem	
Properce, livres 1, 2, 3 et 4	hPro1, pPro1, hPro2, pPro2, hPro3, pPro3, hPro4, pPro4	
Tibulle, livres 1, 2 et 3	hTib1, pTib1, hTib2, pTib2, hTib3, pTib3	