Paragrammes, paraphones et statistiques littéraires

Éric Lysøe

▶ **To cite this version:**

Éric Lysøe. Paragrammes, paraphones et statistiques littéraires : Quelques remarques préliminaires. Lynx, 2007, n° spécial : “ Corpora et questionnement du littéraire ”, p. 33-55. hal-00381028

HAL Id: hal-00381028
https://hal.science/hal-00381028

Submitted on 5 May 2009

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.

Paragrammes, paraphones et statistiques littéraires : quelques remarques préliminaires

La publication tardive des recherches de Saussure sur les anagrammes[1] a conduit, on le sait, le monde de la critique littéraire à réviser bon nombre de ses pratiques. On se souvient sans doute des travaux de Julia Kisteva centrées sur la « sémiologie des paragrammes » et développés par la suite dans un ouvrage au titre pour le moins révélateur : *La Révolution du langage poétique*. On n'a pas oublié non plus, sur un registre quelque peu différent, le stimulant essai de François Rigolot, *Poétique et onomastique*. Il n'en reste pas moins que, même si les échos en sont demeurés un temps sensibles, ces réflexions et les débats qu'elles ont suscités ont très vite cessé d'intéresser bon nombre de chercheurs. Un peu comme Saussure lui-même qui, petit à petit, s'était laissé emporter par le doute, on se mit renoncer à l'hypothèse hypogrammatique. Et c'est ainsi que dès 1986, Jan Baetens commençait à s'interroger sur la « Postérité littéraire des anagrammes »[2].

Sans chercher à reprendre l'ensemble du dossier, ni même à convaincre le lecteur de la « matérialité », de la « réalité » de ces « mots sous les mots », on se contentera ici d'aborder le problème sous l'angle des statistiques littéraires, dans le simple but de proposer une méthode de mise à jour automatisée susceptible de saisir les phénomènes dans le mouvement même du poème. Pareille perspective impose de limiter étroitement le champ des observations. On peut concevoir en effet que la procédure de dissémination des signes procède d'au moins quatre niveaux distincts : graphique, phonétique, syntaxique et sémantique. Pour s'en tenir aux limites qui lui on été assignées, la présente étude n'abordera toutefois que les deux premiers : ces paragrammes et *paraphones*[3] qui, dans bien des cas, entretiennent d'étroites relations d'interdépendances. Car l'objectif ne consistera pas seulement à souligner les difficultés posées la dissémination aux spécialistes de l'informatique littéraire, mais encore de faire entrevoir les enjeux et les problèmes d'un questionnement dynamique du texte. Traiter d'éléments déterminants du langage poétique suppose en effet qu'on se fonde sur le déroulement temporel du vers, sur les phénomènes de récurrences obligées (mètres, rimes, refrains) qui lui servent d'assise : autant d'éléments qu'une problématique science des anagrammes ne saurait entièrement oblitérer.

1 Voir Jean Starobinski, « Les Anagrammes de Saussure », *Mercure de France*, n°350, février 1964, p. 243-263 ; « Le Nom caché », *in* Centre international d'études humanistes, *L'Analyse du langage théologique : le nom de Dieu*, Paris, Aubier, 1969 et surtout *Les Mots sous les mots*, Paris, Gallimard, « Le Chemin », 1971.

2 *Poétique*, n°66, 1986, p. 218-232.

3 Il paraît légitime de ne parler d'anagramme qu'à propos de mots obtenus par réorganisation de toutes les lettres d'un autre mot, comme c'est le cas plus loin, dans l'exemple tiré de Du Bellay, de l'allusion à « olivier » par le biais de « l'ivoire ». Le paragramme évoque quant à lui une dérivation plus souple, fondée sur la dissémination de lettres d'un mot dans une séquence de mots. Dans la mesure même où cette forme de jeu de mots repose sur des éléments graphiques, il a paru nécessaire de forger le néologisme « paraphone »pour rendre compte du même mouvement de dispersion, mais concernant cette fois les seuls aspects sonores du plan de l'expression.

Définition d'un protocole d'analyse

Mais commençons par le début et par la possibilité qu'offrent les statistiques de mettre en évidence la présence ou l'absence de paragramme ou de paraphones. L'hypothèse la plus économique qu'on puisse formuler à ce propos consiste à poser que si un texte par sa structure impose l'apparition de certaines lettres en certaines régions du texte, il doit par le fait même bouleverser l'ordre de fréquence habituel des graphèmes ou des phonèmes, tel qu'on peut l'observer dans un corpus plus vaste, produit à la même époque et si possible dans le même genre, voire à partir la même forme – sonnet, rondeau, ballade, etc. À partir de cette idée simple, il est loisible d'élaborer un protocole d'analyse qui réponde aux hypothèses formulées par Saussure et se laisse décrire comme la succession de deux stades opératoires de caractère bien différent.

Phase 1 : statistiques

La première phase, purement statistique, consiste à appliquer une procédure de comparaison qui offre de mettre en regard tel poème (P_x) avec un corpus de référence (C_o), puis d'évaluer les résultats obtenus par le biais de tests statistiques. La composition de ce corpus de référence est évidemment un élément déterminant. À l'heure actuelle, on peut utiliser faute de mieux, et pour rendre compte de textes récents, telles ou telles données préexistantes : entre autres, pour les paragrammes, celles que propose l'équipe de « Lexique » avec la base « Surface »[4] et, pour les paraphones, celles qui se dégagent des travaux de François Wioland[5]. On peut penser cependant que lorsque le projet D.A.N.T.E. aura été mené a bien, il deviendra possible de travailler sur un poème – dans un premier temps sur un sonnet – en le comparant à un ensemble plus rigoureusement défini en fonction de critères de langue, d'époque et de style.

La procédure de comparaison peut se limiter à l'observation de la fréquence particulière de telle ou telle lettre ou *monogramme* : $\lambda^1(x)$. L'expérience montre cependant, tout comme les travaux de Saussure, qu'il est souvent utile de travailler également sur des *bigrammes* : $\lambda^2(x)$. Dans les lignes qui suivent, les moyens techniques existants n'ont permis de travailler que sur des couples de lettres consécutives ou, si l'on préfère, sur des bigrammes de distance nulle (λ^2_o, par exemple : « ES »). On sait cependant que les observations de Saussure montrent que, bien souvent, la procédure de dissémination conduit à prendre en compte des couples dont les éléments sont séparés par une ou plusieurs lettres. Les bases de données futures devraient permettre de travailler sur des bigrammes de distance variable : λ^2_1 (ExS), λ^2_2 (ExxS), λ^2_3 (ExxxS-, etc.)

Par ailleurs, et pour des raisons de temps, le seul test qui a pu être appliqué sur les résultats obtenus est le calcul d'un *indice de singularité* : $\varsigma(x)$. Celui-ci a été obtenu à partir de la formule suivante où $E^r(x)$ équivaut à l'effectif observé pour tel élément x (lettre ou bigramme), et $E^t(x)$ à l'effectif théorique que devrait atteindre ce même élément :

$$\varsigma(x) = \frac{[E^r(x) - E^t(x)]^2}{E^t(x)}$$

On aura reconnu la formule de base du test du χ^2. Il faut noter toutefois que le nombre obtenu rend compte de toute forme de singularité, qu'il s'agisse d'excédents ou de déficits. Or, puisqu'ici seuls les excédents méritent d'être retenus, on a choisi de multiplier $\varsigma(x)$ par -1 chaque fois que l'effectif observé est supérieur à l'effectif théorique. L'opération produit le

4 On peut élécharger cette base à l'adresse suivante : www.lexique.org (site consulté la denrière fois le 1er novembre 2005)

5 François Wioland, « Étude statistique des phonèmes et diphonèmes dans le français parlé », Revue acoustique, n°16, 1971, p. 258-262 ; « Estimation de la fréquence des phonèmes en français parlé », Travaux de l'institut de phonétiques de l'Université de Strasbourg, n°4, 1972, p. 177-204.

seul indice qui sera examiné par la suite, $\varsigma'(x)$:

$$\varsigma'(x) = [E^r(x) > E^t(x)] * \varsigma(x)$$

Phase 2 : Validation

Cette première série d'opérations réalisée, reste à valider les résultats obtenus en ramenant à quelques formes « anormales » l'ensemble des éléments faisant apparaître un ς' négatif. Dans l'attente de règles plus fiables, on a choisi de se conformer ici, en tout cas dans un premier temps, à trois critères restrictifs directement déduits des propositions de Saussure. À la différence de la phase 1, cet ensemble de « règles » ne relève d'aucune nécessité absolue. Il est simplement le fruit des prémisses qu'on s'est ici données : admettre *a priori* la validité des hypothèses de Saussure. Il est clair que l'expérience conduira à affiner, voire à remplacer par d'autres les trois principes qu'on s'est imposé de la sorte et dont on discutera amplement le bien fondé dans les pages qui suivent.

Règle de distribution. Le coefficient de singularité n'est qu'un indicateur. Un élément apparaissant une ou deux fois dans le texte peut fort bien, compte tenu de la faible taille du poème, apparaître comme relativement exceptionnel. C'est ici que les propositions de Saussure se découvrent toute leur valeur heuristique. Le linguiste établit, on s'en souvient, que le phénomène de dissémination se limite approximativement à l'ensemble que constitue le vers. Il note tout d'abord qu'« une voyelle n'a le droit de figurer dans le saturnien que si elle a sa contre-voyelle [c'est-à-dire une voyelle identique] dans un endroit quelconque du vers »[6]. Constatant ensuite l'existence d'excédent ou de déficit dans tel ou tel vers donné, il est amené à vérifier que les irrégularités se trouvent compensées dans le vers suivant. On peut en déduire l'existence de seuils et ne retenir comme valides que des éléments formés de signes apparaissant en moyenne au moins une fois dans chaque vers. Pour être utilisée avec une réelle chance de succès, une lettre x doit donc satisfaire à la règle suivante où N_v représente le nombre de vers du poème :

$$E^r(x) \geq N_v$$

Règle de configuration. Saussure, on le sait, met en évidence l'existence de « mannequins » qui permettent d'inscrire le mot-thème en filigrane de telle ou telle série de vocables. Ces mannequins commencent par la ou les première(s) lettres du mot et s'achèvent par la ou les dernière(s) lettres du mot. Différentes expressions de l'Énéide offrent de la sorte de retrouver les éléments initiaux et terminaux de « Priamides ». Par exemple :

```
PRIMA  QVIES
PR       ES
   MA    I
PRIAM I(D)ES

PERQVE   PEDES
P R        DES
P R (IAMI) DES
```

Si l'on veut bien suivre le linguiste sur ce point, on ne devrait donc tenir compte que des mots-thèmes dont on peut repérer la présence sous-jacente sous forme de mannequin.

Règle tacite formulation. De même, les travaux de Saussure – comme ceux d'ailleurs, sur un tout autre registre, de poéticiens comme Michael Riffatterre[7] – invitent à ne retenir en priorité que les termes n'apparaissant pas sous leur forme ordinaire dans l'énoncé.

6 Saussure « Le Nom caché », éd. cit., p. 7.

7 Voir Michael Riffaterre, *Semiotics of Poetry*, Bloomington, Indiana University Press, 1978 ; trad. fr. de Jean-Jacques Thomas : Sémiotique de la poésie, Paris, Seuil, « Poétique », 1983.

Un premier exemple : *Sonnets à ma grammaire*

La procédure de mise à jour ainsi définie, il est très facile d'en tester l'efficacité sans prendre le risque d'inventer un prétendu paragramme. Il suffit en effet de prendre pour objet d'étude un poème dont l'auteur lui-même a livré la clef. Au dire de son créateur, l'introducteur de l'OULIPO en Italie, chaque vers du premier des *Sonnets à ma grammaire* renferme toutes les lettres d'un certain mot de huit lettres, composé de sept lettres différentes et désignant une partie du corps féminin :

Ton parian torpide dans ses clivages rit,
Ou, clinamen torride, rissole ton voisement,
Quand le cycliste orphique dans ton vibrage est mis,
Mû par l'espoir tonique d'un clystère d'amant.

Mais si tu es clastique, quitte ce lot rassis ;
D'un cothurne panique, lindor cautérisant,
Hors du klystron morbide, où tu seras crépi,
Sans klaxonner débride ton tic et ton orient.

Vers un cloaque hirsute insinue donc ta rage ;
Au cliquetis d'une flûte, au flash d'un oribus,
Tu cligneras des yeux à l'orbite fœtale,

Tu clisseras tes vœux. Ironisant ton stage,
Clamant ton rouspetème, hissant ton omnibus,
Sus, Sisyphe à la crème, rêve d'être crotale[8] !

En comparant la fréquence des lettres composant ce poème avec celle, toute théorique, que permet de calculer la base Surface on obtient une première série de résultats. Six des sept lettres composant le mot « caché » se retrouvent bien parmi les lettres anormalement fréquentes, mais la septième – le « L », qui occupe en réalité la seconde position dans le mot – ne présente pas de ς' révélateur. À sa place, le calcul conduit à isoler un « N » apparemment « parasite » :

Tableau I

	Campagnoli		Surface			
Lettre	E(x)	%	E_0	%	Δ	ς'
E	55	10,93	74,56	14,74	3,80	5,13
S	**48**	**9,54**	**42,44**	**8,39**	**-1,16**	**-0,73**
A	36	7,16	38,49	7,61	0,45	0,16
I	**41**	**8,15**	**37,27**	**7,37**	**-0,79**	**-0,37**
T	**46**	**9,15**	**36,51**	**7,22**	**-1,93**	**-2,47**
N	**40**	**7,95**	**35,76**	**7,07**	**-0,89**	**-0,50**
R	**36**	**7,16**	**32,93**	**6,51**	**-0,65**	**-0,29**
U	30	5,96	30,71	6,07	0,10	0,02
O	**37**	**7,36**	**26,97**	**5,33**	**-2,03**	**-3,73**
L	23	4,57	26,78	5,29	0,72	0,53
D	18	3,58	19,18	3,79	0,21	0,07
C	**19**	**3,78**	**15,67**	**3,10**	**-0,68**	**-0,71**

Les lettres sont classées selon l'ordre de fréquence dans la base Surface. Les deux colonnes « Campagnoli » donnent : 1) le nombre d'occurrences dans le poème, 2) la fréquence exprimée en pourcentage. La colonne Δ correspond à la différence entre la fréquence en pourcentage dans la base Surface (%S) et la fréquence

8 Ruggero Campagnoli, *Sonnets à ma grammaire*, Bologne, Clueb 2002, p. 5.

en pourcentage dans le sonnet étudié[9]. Conformément aux règles énoncées ci-dessus, le tableau se limite aux lettres dont les effectifs réels sont supérieurs ou égal au nombre de vers du poème (≥ 14).

Si l'on demande à un générateur d'anagrammes de déterminer les possibilités de constructions de mots à partir des sept lettres retenues après cette première sélection, on n'est qu'à peine surpris de la réponse : « citrons », ou encore, en laissant à la machine la possibilité de redoubler certaines éléments : « inscrit-on », « nos conscrits ont ri », « Rinçons nos tricots », etc. Nous sommes assez loin du terme anatomique qu'il convenait de retrouver et que lecteur a sans doute déjà deviné : « clitoris ». Ce mot-thème se trouve pourtant clairement évoqué à travers certains débuts de mots (« CLIvages », « CLInamen TORride, RISsole », etc.). L'explication d'un tel échec est triple : la recherche s'effectue sur un corpus extrêmement réduit ; le mot en lui-même utilise les lettres les plus fréquentes du corpus de référence et surtout redouble l'une de ces lettres – le « i » – sans que pour autant sa fréquence soit relativement plus élevée que les autres monogrammes sur-représentés. La réponse automatique ne peut être qu'approximative. On notera cependant qu'elle permet de dessiner un « mannequin » dans lequelle le « n » ne semble jouer aucun rôle, du moins à en croire les premiers vers[10] :

```
Ton parian torpide dans ses clivages rit,
TOn   RI n TOR I     nS S S C I    S RIT
TO    RI  (TOR I)     S     CLI    S

Ou, clinamen torride, rissole ton voisement,
O   C In   n TORRI    RISSO   TOn  OIS   nT
    CLI      TO(RI)   RIS
Quand le cycliste orphique dans ton vibrage est mis,
   n     C C IST  OR  I       S TOn  I R     ST  IS
           CLI T  OR  I       S
```

L'ensemble demeure néanmoins trop peu satisfaisant et une solution pour réduire sensiblement les risques d'erreurs consiste à travailler cette fois non plus sur de simples lettres, mais sur des bigrammes. Un second relevé devrait permettre non seulement d'infirmer ou de confirmer, mais encore de compléter les résultats précédents. De fait, en recourant de nouveau aux données de la base Surface, on obtient une série de résultats cette fois déterminant. En se limitant aux bigrammes de fréquence égale ou supérieure à 7 ($N_v/2$), on obtient le tableau suivant :

Tableau II

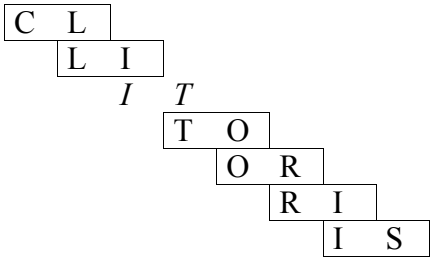
	Campagnoli		Surface		%S-%C	
	E (x)	%	E_o	%	Δ	ς'
ON	**13**	**3,359**	**7,81**	**2,019**	**-1,34**	**-3,44**
IS	**12**	**3,101**	**4,96**	**1,283**	**-1,82**	**-9,97**
TO	**11**	**2,842**	**1,74**	**0,448**	**-2,39**	**-49,47**
AN	**11**	**2,842**	**5,60**	**1,446**	**-1,40**	**-5,22**
CL	**10**	**2,584**	**0,44**	**0,114**	**-2,47**	**-207,68**
OR	**9**	**2,326**	**2,36**	**0,610**	**-1,72**	**-18,70**
RI	**8**	**2,067**	**2,59**	**0,670**	**-1,40**	**-11,29**
QU	*8*	*2,067*	*5,53*	*1,428*	*-0,64*	*-1,11*
LI	**7**	**1,809**	**1,84**	**0,476**	**-1,33**	**-14,42**
UE	*7*	*1,809*	*3,84*	*0,993*	*-0,82*	*-2,59*

9 On notera, mais c'est évidemment un hasard, que les douze lettres les plus représentées dans le poème permettent de composer des énoncés du genre : « un clitoris adulé », « saint clitoris ». Jugé à cette aune, le mot-thème de la langue française serait « sainte » qui rassemble les 6 lettres les plus fréquentes de la base Surface.

10 Sauf à croire que l'élément anatomique qui fascine le poète est lié de manière ou d'autre à la ville de Turin (Torino).

NT	7	1,809	8,05	2,079	0,27	0,14
ES	7	1,809	12,04	3,110	1,30	2,11

Si l'on ne retient que les bigrammes constitués par des lettres de fréquence égale ou supérieure à 14 et dont le ς' est négatif, on peut ramener cette première liste à huit bigrammes : ON, IS, TO, AN, CL, OR, RI, LI. Les résultats observés paraissent d'autant plus révélateurs qu'ils font apparaître des effectifs supérieurs ou égaux à ceux du « ES », couple de lettres ordinairement le plus présent dans les corpus français. Les six bigrammes dont la valeur de ς' est la plus significative – proche de -10 ou largement inférieure – permettent par ailleurs de recomposer l'ensemble du mot selon une logique d'emboîtement en tout point conforme aux observations de Saussure :



Seul le couple « IT » dont on ne compte que trois occurrences n'obéit pas au processus de dissémination qui affecte tous les autres couples de lettres composant le mot-thème, de sorte qu'il est presque impossible de ne pas détecter ici la présence du nom recherché.

Il reste qu'on repère également – et comme dans le cas du relevé limités aux lettres – la présence d'éléments apparemment étrangers au mot-thème : ON, AN, qui perturbent le travail de reconnaissance. On laissera ici aux beaux (ou aux mauvais) esprits le loisir de jouer avec ces deux syllabes résiduelles. Nul doute que certains se plaisent à imaginer le poète en disciple du fils de Judas, cet Onân dont la Genèse décrit la répugnance à doter son frère d'une descendance. D'autres ne manqueront pas de voir dans ces pages le résultat de quelque onanisme intellectuel. Une chose est sûre, il paraît difficile de prétendre à partir de cet exemple que la lecture anagrammatique du poème puisse se réduire à un seul et unique mot. Dans le cas présent, on peut certes poser que si le signifiant « clitoris » a fait l'objet d'une dissémination volontaire, là où le(s) mot(s) qu'on peut composer avec les deux syllabes résiduelles – en les combinant éventuellement avec les autres – a (ont) subi une dissémination dont l'origine est à rechercher dans l'inconscient. Mais cette double articulation reste elle-même insuffisante. On peut en effet imaginer que, sur le principe de la matrice poétique, telle que la définit Riffaterre dans *Semiotics of Poetry*, l'hypogramme se réduise à un mot particulièrement chargé en connotations ambivalentes. Mais il n'est pas impossible également qu'il se compose de deux ou trois mots pour esquisser un système de contradiction que le poète s'efforcerait de résoudre. Le phénomène supposerait alors que ces mots sous les mots ne puissent se saisir dans une perception uniquement statique, mais qu'il faille tenir compte du mouvement du texte et de tous les éléments d'un vers à l'autre, d'une strophe à l'autre peuvent entrer en résonance. Voilà bien en tout cas ce que vont tendre à établir les quelques observations qui suivent.

Le sonnet LXII de *L'Olive* et la question du corpus de référence

Appliquée au sonnet LXII de *L'Olive* de Du Bellay la procédure décrite ci-dessus donne des résultats immédiats :

Qui voudra voir le plus precieux arbre,
Que l'orient ou le midy avoüe,
Vienne, où mon fleuve en ses ondes se joüe :
Il y verra l'or, l'ivoire, et le marbre.
Il y verra les perles, le cinabre
Et le cristal : et dira que je loüe
Un digne object de Florence, et Mantoue,
De Smyrne encor', de Thebes, et Calabre.
Encor' dira que la Touvre, et la Seine,
Avec' la Saone arriveroient à peine
A la moitié d'un si divin ouvrage :
Ne cetuy là qui naguere a faict lire
En lettres d'or gravé sur son rivage
Le vieil honneur de l'une et l'autre lire[11].

Cette fois en effet le relevé des fréquences par lettres suffit à découvrir un hypogramme éventuel : « Olivier »:

Tableau III

	Du Bellay		Corpus de référence		%C-%D	
Lettres	E^r	%	E^t	%	Δ	ς'
E	**79**	**19,70**	**14,74**	**59,09**	**-4,97**	**-6,71**
S	16	3,99	8,39	33,63	4,40	9,25
A	28	6,98	7,61	30,50	0,62	0,20
I	**31**	**7,73**	**7,37**	**29,54**	**-0,37**	**-0,07**
T	20	4,99	7,22	28,93	2,23	2,76
N	27	6,73	7,07	28,34	0,33	0,06
R	**39**	**9,73**	**6,51**	**26,10**	**-3,22**	**-6,38**
U	21	5,24	6,07	24,33	0,83	0,46
O	**25**	**6,23**	**5,33**	**21,37**	**-0,90**	**-0,62**
L	**30**	**7,48**	**5,29**	**21,22**	**-2,19**	**-3,63**
D	13	3,24	3,79	15,20	0,55	0,32
C	11	2,74	3,10	12,42	0,35	0,16
P	4	1,00	2,90	11,63	1,90	5,01
M	6	1,50	2,85	11,41	1,35	2,57
V	**16**	**3,99**	**1,48**	**5,94**	**-2,51**	**-17,02**

On peut bien sûr voir dans ce relevé le simple effet du hasard. Il reste que le mot mis à jour par le calcul constitue de façon particulièrement évidente le mot-thème du poème. « Olivier » définit évidemment « le […] precieux arbre » que le poète dit pousser là « où [son] fleuve en ses ondes se joüe ». Cet olivier angevin – arbre tout poétique à la gloire de l'idéale (?) Olive – s'offre évidemment comme un contrepoint « français » au laurier de Pétrarque, l'arbre de Laura. Il permet à Du Bellay de broder sur le thème qui s'annonçait dès la première pièce du volume :

11 Joachim du Bellay, *L'Olive*, éd. d'Olivier Millet et *al.*, Paris, Champion, 2003, p.

Je ne quiers pas la fameuse couronne[12]…

Pareil incipit fait évidemment allusion aux lauriers dont Laura coiffe le poète italien et ses échos se font entendre jusque dans le sonnet 62 où, dès le deuxième vers, les jeux des sonorités associe cet olivier imaginaire à Laure :

Que l'orient ou le midy avoüe…

« L'orient » permet d'entendre [lɔʀ], voire [lɔʀj] soit, le [o] en moins[13], le début de « laurier ». Et le jeu se poursuit au vers 4. « Il y verra, dit en effet le poète, l'or, l'ivoire et le marbre ». Or si « l'or » est une allusion phonétique traditionnelle à la l'égérie de Pétrarque et à l'arbre qui lui est associé, « l'ivoire » est la parfaite anagramme d' « Olivier ». Le mot donne ainsi lieu à toute une série de dérivations. Puisque que « l'ivoire » revient à « l'y voir », ce quatrième vers sert à mettre en place la formule « Il y verra », laquelle, modifiée en« et dira », confère son armature au poème. Le jeu *laurier* (italien) vs *olivier* (français) justifie même *in fine* l'opposition entre les cinq poètes « méditerranéens » évoqués par des noms de ville – Pétrarque, le poète de Florence étant ainsi nommé le premier – et quatre poètes « gaulois » mentionnés par le biais de fleuves, Du Bellay, le chantre de la Loire angevine étant manifestement le cinquième, l'équivalent français du premier poète italien, Pétrarque.

Ces résultats apparemment indiscutable posent néanmoins un problème. Car le corpus de référence utilisé ci-dessus est en fait une fois de plus celui que propose la base Surface. Un premier test, élaboré sur un ensemble de textes de la Renaissance donnait des résultats bien moins convaincants :

Tableau IV

	Du Bellay		Corpus XVI^e^			
Lettre	E^r^	%	E^t^	%	Δ	ς'
E	**81**	**20,20**	**61,43**	**15,32**	**-4,88**	**-6,23**
R	**39**	**9,73**	**34,61**	**8,63**	**-1,09**	**-0,56**
U/V	**37**	**9,23**	**28,49**	**8,14**	**-1,37**	**-1,07**
I /J	**34**	**9,23**	**32,65**	**8,14**	**-1,08**	**-0,58**
L	**30**	**7,48**	**17,63**	**4,40**	**-3,08**	**-8,67**
A	*28*	*6,98*	*27,33*	*6,81*	*-0,17*	*-0,02*
N	*27*	*6,73*	*26,90*	*6,71*	*-0,03*	*0,00*
O	**25**	**6,23**	**26,43**	**6,59**	**0,36**	**0,08**
T	20	4,99	27,47	6,85	1,86	2,03
S	16	3,99	32,37	8,07	4,08	8,28

Les deux lettres parasites, A et N, qui se détachent de l'ensemble peuvent être négligées tant leur ς' est proche de zéro. Il n'en demeure pas moins que le « O », loin d'être excédentaire, accuse un déficit assez net (0,36 %). Il est clair toutefois que ce résultat décevant est directement fonction de la composition du corpus de référence – trop limité et hâtivement composé[14]. Une étude de la distribution des « O » menée avec le logiciel Hyperbase sur l'ensemble des textes retenus montre en effet que deux séries de données sont nettement plus importantes que les autres : celles qui proviennent de textes dramatiques (les tragédies de Garnier, notées « Garnier » sur le schéma ci-dessous) et celles qui proviennent de textes en

12 Sonnet I, *ibid*., p.

13 On sait que cette assimilation du [o] au [ɔ] est courante dans la poésie du XVI^e^ siècle. Laure renvoie communément à « l'or ».

14 Les ouvrages numérisés en mode textes disponibles sur Internet ne sont pas suffisamment nombreux ni transcrits selon des procédures assez uniformes (modernisation complète, partielle, respect absolu du texte original, etc.) pour permettre l'établissement d'un corpus réellement satisfaisant.

prose (*La Confession catholique du sieur de Sancy* d'Aubigné, notée « Aubigné Conf. »).

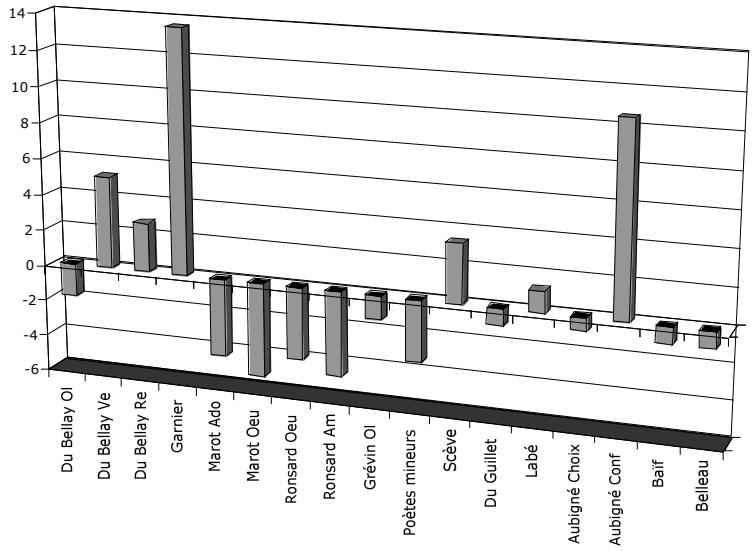


Figure 1

Or autre sondage, cette fois sur les « U » et les « V », donne des résultats bien différents et apparemment indépendants des caractéristiques génériques du corpus :

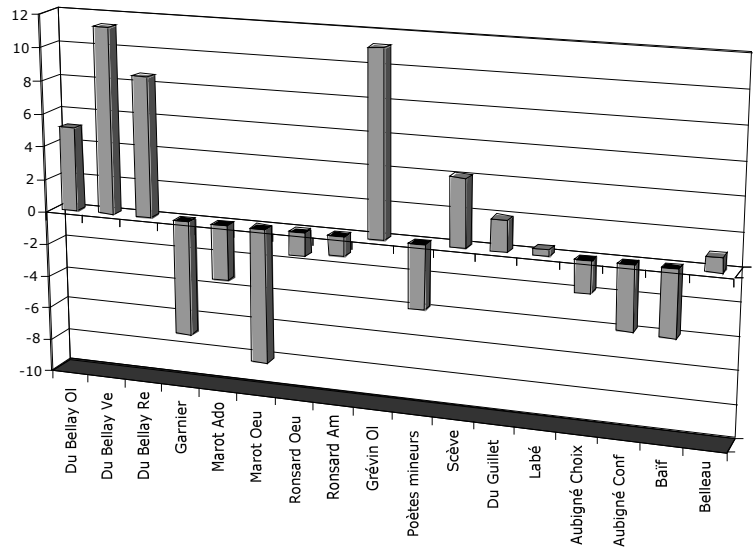


Figure 2

On peut supposer que certaines tendances propres aux genres – le nombre de formes verbales en -oi-, par exemple – faussent certains résultats – entres ceux qui concernent la fréquence des « O ». Une première tentative pour résoudre ce problème a consisté à comparer *L'Olive* à un corpus restreint à la poésie du XVI^e^ siècle. Mais ,là encore, le recueil du poète angevin a

révélé, du moins à en croire les résultats affichés par Hyperbase, un déficit de O particulièrement important. On s'est donc résolu à tenter une comparaison interne, limitée au seul volume de Du Bellay avant de confronter les résultats obtenus avec ceux déduits de la base Surface. Et cette fois, malgré d'importantes divergences, les deux relevés ont conduit à confirmer la sur-représentation des lettres O, L, I/J, V/U, E et R. Comme le montre le tableau suivant, parmi les 11 lettres dont les effectifs réels sont supérieurs à N_v, les lettres d'« olivier » sont bien celles qui possèdent le ς' le plus significatif :

Tableau V

	Sonnet LXII		*L'Olive*		%O-%S	
Lettres	E^r	%	E^t	%	Δ	ς'
V/U	**37**	**9,18**	**11,44**	**2,84**	**-6,34**	**-57,11**
O	**25**	**6,20**	**9,76**	**2,42**	**-3,78**	**-23,82**
L	**30**	**7,44**	**13,36**	**3,32**	**-4,13**	**-20,71**
R	**39**	**9,68**	**29,49**	**7,32**	**-2,36**	**-3,07**
E	**83**	**20,60**	**69,39**	**17,22**	**-3,38**	**-2,67**
I/J	**34**	**8,44**	**28,48**	**7,07**	**-1,37**	**-1,07**
N	27	6,70	24,16	6,00	-0,70	-0,33
A	28	6,95	25,57	6,35	-0,60	-0,23
T	20	4,96	29,76	7,39	2,42	3,20
S	16	3,97	25,26	6,27	2,30	3,39

Les diverses approximations auxquelles il a ainsi fallu se rendre montrent bien à quel point la composition du corpus de référence est une opération d'importance capitale. Lorsque le projet D.A.N.T.E. aura atteint son premier objectif, il sera sans doute facile de constituer un corpus de référence qui permette d'obtenir des résultats assurément plus fiables.

Mais les résultats obtenus à partir du texte de Du Bellay sont également révélateurs d'un phénomène d'un tout autre ordre. Si « olivier » donne « l'ivoire », privé de « O », il permet – entre autres – de composer le signifiant « livre ». Et c'est bien sur le motif de l'œuvre que s'achève le poème en évoquant à la fois les activités de lecture (« lire ») et de créération (« lyre », orthographié « lire »[15]). Dès lors, si l'on veut bien poursuivre ce jeu de déconstruction, on peut remarquer qu' « olivier » dépourvu de « V » donne « Loire ». Le mot en vient ainsi à faire écho aux fleuves français mentionnés à la fin du poème. Dans le même temps, il suggère l'existence, cette fois sur le registre géographique, d'un troisième sens de « lire »: celui du « petit Liré » que le sonnet XXXI des *Regrets et antiquités de Rome* associe, on le sait, directement à la Loire :

> Plus me plaist le sejour qu'on basty mes ayeux,
> Que des palais Romains le front audacieux,
> Plus que le marbre dur me plaist l'ardoise fine:
> Plus mon Loyre Gaulois, que le Tybre Latin,
> Plus mon petit Lyré, que le mont Palatin,
> Et plus que l'air marin la doulceur Angevine.

On comprend dès lors que les paragrammes se distribuent dans le poème selon une logique caractéristique dont il importe de saisir les règles distributionnelles. Le « poids anagrammatique », calculé à partir de l'effectif total par vers de l'ensemble des lettres entrant dans la composition de l'hypogramme révèle pat exemple une série de variations dont il n'est pas déraisonnable de penser qu'elles sont de manière ou d'autre liées au déploiement du texte et

15 Sur le modèle de l'italien « lira », l'expression « l'une et et l'autre lire » étant un calque du cliché italien « l'una et l'altra lira » qu'on rencontre notamment dans le sonnet CCXLVII d'*Il Canzoniere* de Pétrarque – poème qui soit dit au passage évoque quatre villes dont Mantoue et Smyrne.

son adéquation à telle ou telle forme particulière. Le schéma ci-dessous montre par exemple que les lettres d' « Olivier » se concentrent dans le premier quatrain – dont les vers d'ouverture et de clôture marquent des pics assez nets. Mais c'est pour voir par la suite leurs effectifs décroître sensiblement. Un nouveau sommet marque néanmoins le passage au système des tercets (vers 9) avant que le vers 14 traduise un ultime accroissement d'effectifs. L'ensemble du sonnet se découvre de la sorte une forme générale qui semble se déduire de celle du premier quatrain en ce qu'il s'ouvre et se ferme sur deux vers nettement excédentaires.

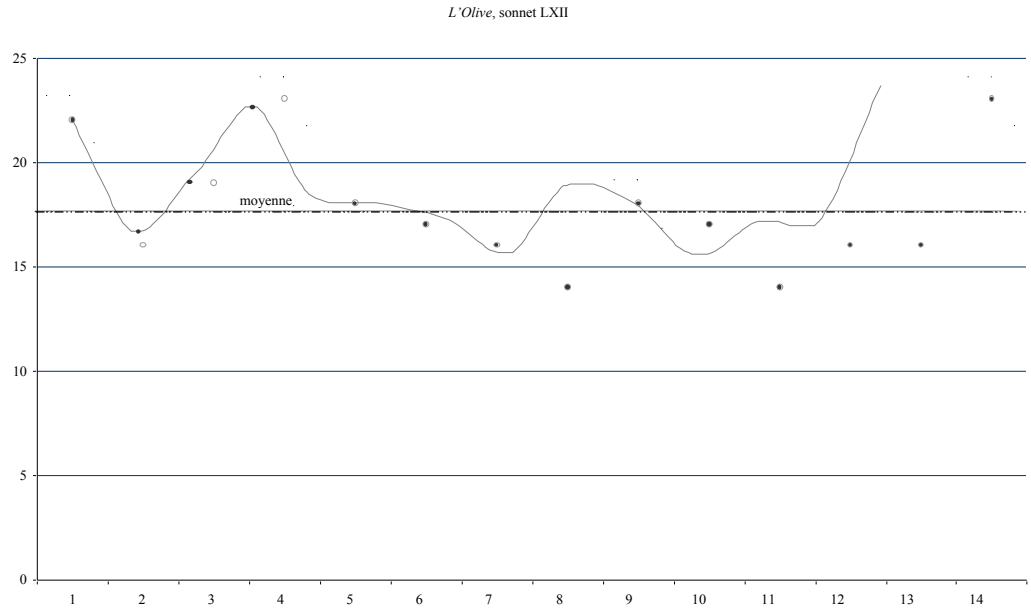


Figure 3

Correspondances et abondance de biens

Le célèbre poème de Baudelaire « Correspondances » invite pousser un peu plus loin la réflexion en ce domaine. Une première procédure, appliquée aux seules lettres, conduit à repérer quatre éléments excédentaires : « S », « O », « C », « M ». Les autres monogrammes exceptionnellement présents, « E », « N », « R », « D », « P » présentent un ς' si proche de 0 (de – 0,0003 pour E à – 0,04 pour R) qu'on est tenté de les négliger. Mais en se fondant sur quel critère ? Nul doute qu'un statisticien parvienne à proposer tel ou tel test pour permettre de conforter l'analyste dans le « bon » choix. Il n'en est pas moins vrai que, si les lettres de ς' inférieur à -1 permettent de reconstituer le mot « cosmos », l'ensemble des lettres de ς' négatif donnent l'impératif (?) : « comprends »:

Tableau VI

	Baudelaire		Surface		%S-%B	
Lettre	E^r	%	E^t	%	Δ	ς'
E	79	14,74	78,84	14,74	-0,00	-0,00
S	**67**	**12,50**	**44,87**	**8,39**	**-4,11**	**-10,91**
A	32	5,97	40,69	7,61	1,64	1,86
I	28	5,22	39,41	7,37	2,14	3,30
T	36	6,72	38,60	7,22	0,50	0,18
N	38	7,09	37,81	7,07	-0,02	-0,00

R	36	6,72	34,82	6,51	-0,21	-0,04
U	19	3,54	32,46	6,07	2,52	5,58
O	**37**	**6,90**	**28,51**	**5,33**	**-1,57**	**-2,53**
L	28	5,22	28,31	5,29	0,07	0,00
D	21	3,92	20,27	3,79	-0,13	-0,03
C	**22**	**4,10**	**16,56**	**3,10**	**-1,01**	**-1,78**
P	16	2,99	15,52	2,90	-0,08	-0,02
M	**25**	**4,66**	**15,22**	**2,85**	**-1,82**	**-6,28**

On reconnaîtra sans peine que le poème de Baudelaire s'accommode assez bien de ces deux hypogrammes. Il s'accorde de façon exemplaire avec l'image du cosmos, le terme étant entendu ici, comme en grec, sous sa double acception : il s'agit d'une part de l'ordre – et notamment l'ordre du monde – et, d'autre part, de la parure – et notamment la parure des femmes. L'incipit et l'explicit semblent d'ailleurs dessiner une expansion de ce double sens, en ce qu'ils évoquent d'un côté la Nature comme « un temple » et de l'autre des parfums directement associés à l'image du corps féminin, éléments *cosmétiques* qui, comme tels, se révèlent être secrètement liés au *cosmos*. Le mannequin COxxxS se repère d'ailleurs à plusieurs reprises dans le poème et semble alterner régulièrement avec la finale « -SMOS » :

```
Correspondances
CO    S O     C S
CO(SMO)S

La Nature est un temple où de vivants piliers
          S       M    O           S       S
          S       M    O           S
Laissent parfois sortir de confuses paroles;
   SS        O S SO        CO   S S    O  S
                           CO   S (M)  O  S
L'homme y passe à travers des forêts de symboles
   OMM      SS          S   S  O   S    S M O  S
                                        S M O  S
Qui l'observent avec des regards familiers.
      O S          C   S       S   M     S
                   C        (O)S   M(O)  S
```

Il faut admettre cependant que le verbe « comprends » ne contredit en rien cette première observation sur « cosmos ». Le poème est bien un appel à la saisie des mystérieuses lois de l'univers et, dans la mesure même où le mannequin censé figurer les deux mots peut-être identique, COxxxS, l'observateur ne dispose d'aucune règle pour trancher avec certitude.

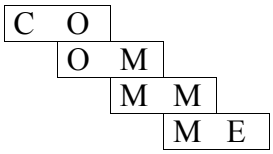
On pourrait espérer qu'un travail sur les bigrammes permette d'opter pour l'une ou l'autre des lectures. Il se trouve qu'il n'en est rien. Appliqué aux couples de lettres constituant le poème, la procédure fait apparaître deux nouveaux hypogrammes éventuels, apparemment sans rapport avec les précédents :

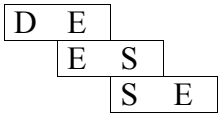
Tableau VII

	Baudelaire		Surface		%S-%B	
Bigr.	E^r	%	E^t	%	Δ	ς'
ES	**25**	**5,95**	**13,07**	**3,11**	**-2,84**	**-10,88**
DE	**15**	**3,57**	**9,58**	**2,28**	**-1,29**	**-3,06**
CO	**12**	**2,86**	**3,94**	**0,94**	**-1,92**	**-16,48**
LE	12	2,86	10,58	2,52	-0,34	-0,19
OM	**10**	**2,38**	**2,11**	**0,50**	**-1,88**	**-29,43**
NT	10	2,38	8,74	2,08	-0,30	-0,18
SE	**9**	**2,14**	**4,58**	**1,09**	**-1,05**	**-4,27**
AN	9	2,14	6,08	1,45	-0,70	-1,41
ON	9	2,14	8,49	2,02	-0,12	-0,03
MM	**8**	**1,90**	**1,23**	**0,29**	**-1,61**	**-37,42**

ME	8	1,90	4,87	1,16	-0,74	-2,00
ET	7	1,67	4,25	1,01	-0,65	-1,77

En ne retenant que les couples de lettres aux ς' les plus significatifs (inférieurs ou égals à -2), on obtient selon le principe découvert avec le premier des *Sonnets à ma grammaire* deux ensembles apparemment distincts :





Si le premier terme est immédiatement identifiable et renvoie aux multiples comparaisons qui sous-tendent le poème, le second est plus inattendu. Il peut s'agir de l'expansion de l'article indéfini pluriel, « des » ou encore d'une contraction de « déesse », voire de « déesses », signifiant dont le mannequin semble repérable en plusieurs endroits du texte :

```
Correspondances
     ES    D    ES
          D(EESS)ES
La Nature est un temple où de vivants piliers
        E ES      E   E    DE       S     E S
                           DE(E)    S(S)  E S
Laissent parfois sortir de confuses paroles;
   SSE         S S      DE      SES      ES
                        DE(E)   S S      ES
L'homme y passe à travers des forêts de   symboles
      E     SSE       E S DES    E S DE   S     ES
                          DES(S) E S DE(E)S(S)  ES
Qui l'observent avec des  regards familiers.
        SE  E     E  DES   E   DS       E S
                     DEESSE    D(ES)S   ES
```

Nous voilà donc cette fois non pas avec un mot-clef lentement et péniblement mis à jour, mais devant une collection d'hypogrammes possibles. Or ceux-ci constituent autant d'éléments qui, loin de se contredire, semblent au contraire s'épauler l'une l'autre. La parure cosmique est évidemment celle que les Grecs attribuaient en priorité aux déesses, et notamment à Aphrodite, le principe de l'amour. On se souvient, par exemple, des termes dans lesquels les *Hymnes homériques* font l'éloge de la superbe fille d'Ouranos :

> Sur son cou délicat
> sur ses seins blancs comme l'argent
> elles [les Saisons] ont disposé [*ekosmeon*, du verbe kosmeô] des colliers d'or,
> ceux dont elles se parent [*kosmeisthên*, du verbe kosmeô]
> elles-mêmes, les Saisons
> au diadème d'or quand elles vont
> à l'aimable assemblée des dieux
> et à la maison du père[16].

16 Hésiode, Théogonie et autres poèmes suivi des Hymnes homériques, trad. De Jean-Louis Backès, Paris, Folio classique, 2001, p. 295. Texte original : « hormoisi chruseoisin ekosmeon, hoisi per autai/ Hôrai kosmeisthên chrusampukes, hoppot' ioien/ es choron himeroenta theôn kai dômata patros ». (Eis Aphroditên, v. 11-13)

Mais l'hypogramme « Déesses » peut également renvoyer au mystère par excellence, celui auquel Baudelaire renvoie son lecteur par le biais de l'énigmatique « comprends ». C'est en effet sous cette appellation vague, « les Déesses » que le monde classique désigne Déméter et Perséphone, les divinités tutélaires des mystères d'Eleusis[17].

On voit dès lors se dessiner un ensemble cohérent qui associe le Cosmos, la ou les Déesses et le mystère de la compréhension. Car identifier le cosmos à la femme ou à la déesse permet de rendre compte de la façon qu'a le premier vers d'identifier la Nature à un lieu de culte. Le mot « temple » entre dans de nombreuse métaphore pour désigner la femme, la vertu de la femme ou même cette partie caractéristique du corps féminin dans laquelle la vertu est censée se loger. On se souvient de ce passage de Justine où l'héroïne évoque son sexe comme le « temple de la nature » :

> Le troisième me fit monter sur deux chaises écartées, et s'asseyant en dessous [...] il me fit pencher jusqu'à ce que sa bouche se trouvât perpendiculairement au temple de la nature[18].

La question de l'implicite

Est-ce à dire que le « comme » doit quant à lui être éliminé, du simple fait qu'il apparaît explicitement à sept reprises dans le poème ? Faut-il, en un mot, appliquer de façon aveugle la règle de tacite formulation énoncée plus haut ? Si l'on veut bien abandonner un instant Baudelaire, on est, c'est vrai, guère surpris de vérifier que l'épigramme « Du beau tetin » se caractérise par la fréquence remarquable les quatre bigrammes entrant dans la composition du substantif placé en titre et répété à maintes reprises par le poète. Et il en va bien sûr de même avec le contre-blason « Du laid tétin ». Faut-il pour autant renoncer à conjuguer les autres bigrammes remarquables avec les syllabes de ce mot-thème ? N'y a-t-il pas du « teint » dans ce tétin, partant quelque chose qui menace de s'y éteindre ? De même, pour revenir à Baudelaire, il est sans doute imprudent de renoncer à l'hypogramme « comme » sous le seul prétexte qu'il apparaît explicitement dans le sonnet. Car cette forme en partie seulement hypogrammatique permet de saisir une autre isotopie du texte, celle de l'art poétique. Non seulement le Cosmos « est comme » la ou les Déesse(s) – non seulement le corps féminin tend à s'identifier au cosmos, fut-ce par le biais des cosmétiques, mais l'univers tout entier se découvre étroitement corrélée à ces ornements rhétoriques que sont les images. Chez Aristote, le mot cosmos désigne les ornements du style et plus particulièrement les épithètes[19]. Or les sept conjonctions comparatives de « Correspondances » sont loin d'être distribuées au hasard. Si la première et la dernière – qui engendrent symétriquement deux expansions articulées autour d'une relative – les cinq autres, disposées au centre de la structure permettent de rattacher à chaque fois un trait sémantique matérialisé par un adjectif (vaste, frais, doux, verts) à un nom (nuit, clarté, chairs, hautbois, prairies).

> COMME de longs échos qui de loin se confondent
> Dans une ténébreuse et profonde unité,
> Vaste COMME la nuit et COMME la clarté,
> Les parfums, les couleurs et les sons se répondent.
>
> Il est des parfums frais COMME des chairs d'enfants,
> Doux COMME les hautbois, verts COMME les prairies,
> – Et d'autres, corrompus, riches et triomphants,

17 Voir par exemple Platon, Le Banquet, 180 d.

18 Sade, *Justine ou les malheurs de la vertu*, éd. de Béatrice Didier, Paris, Livre de poche, 1973, p. 48. Un peu plus haut, la jeune femme parle des deux « autels de la nature » et de leurs « péristyles » (p. 47).

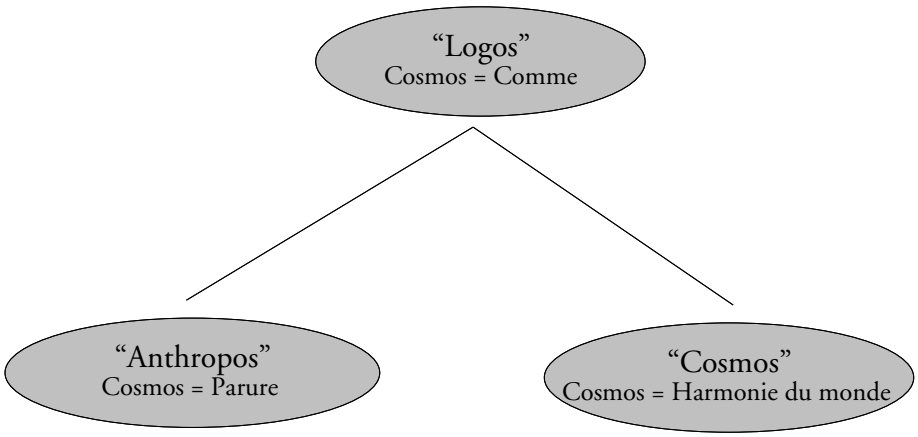
19 Voir par exemple, *Rhétorique*, Livre III, chapitre 7, § 2, texte établi et traduit par Médéric Dufour et André Wartelle, Paris, Les Belles Lettres, 1980, p. 54.

Ayant l'expansion des choses infinies,
COMME l'ambre, le musc, le benjoin et l'encens,
Qui chantent les transports de l'esprit et des sens.

Le « comme » s'impose donc comme une interprétation « française » du cosmos rhétorique des grecs. Il est la parure dont le premier quatrain[20] simule l'apparition aux vers 2 et 3 par un système d'enchaînement. Le mannequin COS(m)OS se prolonge en « -omme », finale identifiée dans un premier temps avec l'« homme », et destinée par la suite à produire les sept « comme » qu'on va voir se succéder en 10 vers :

```
Laissent parfois sortir de confuses paroles;
   SS        O S SO        CO   S S    O  S
L'homme y passe à travers des forêts de symboles
   OMME       E       E    E   O E    E   M O E
```

Ainsi, la mise à jour d'hypogrammes éventuels permet-elle de découvrir que le sonnet de Baudelaire s'organise autour de trois significations de cosmos qui chacune appartiennent à l'une des trois isotopies que le groupe μ a invité à reconnaître dans le poème[21]. Le cosmos, entendu comme ordre du monde répond à la simple parure féminine ou encore se réduit à l'échelle de l'homme pour transcrire le mystère du monde en termes de comparaisons. Il relève ainsi aussi bien de l'ordre du « Cosmos » – entendu ici dans le sens que donnent au termes les auteurs de *Rhétorique de la poésie* – que de celui de l'« Anthropos » ou encore du « Logos. Il offre de reconstituer de la sorte le modèle triadique qui, selon les critiques liégeois, rend compte du fonctionnement même du texte poétique :



Approche phonétique

La même analyse pourrait être faites avec les Déesses. D'Aphrodite à Déméter et Perséphone, celles-ci permettent en effet de reconstituer le passage de l'Anthropos au Cosmos et à ses mystères par le simple jeu du nombre, c'est à dire par l'apparition de ces déterminants pluriels qui émaillent régulièrement le texte. De même que la conjonction « comme », les « des »

20 Quatrain qui repose, lui, sur le principe de la métaphore et non sur celui de la comparaison
21 Groupe μ, Rhétorique de la poésie, Paris, Seuil, « Point », 1990, p. 96.

omniprésents renvoient en effet tout à la fois au mannequin « DxESxxES » et à la pluralité des significations que met en branle le poème. Ce sont bien eux qui donnent tout leur *sens* au dernier vers du sonnet à travers l'évocation des « transports de l'esprit et des sens ». Toutefois, plutôt que de développer une nouvelle lecture de ce poème, il importe sans doute de prendre un dernier exemple afin d'aborder enfin l'aspect purement phonétique du processus de dissémination.

Appliqué au poème de Baudelaire « La Mort des amants » et rapporté aux éléments de statistique phonétique fournis par François Wioland, la procédure décrite plus haut permet de dégager 7 phonèmes dont les effectifs sont égaux ou supérieurs à 14 (N_v), mais dont 3 seulement présentent un ς' négatif: [R], [l] et [d]

Tableau VIII

	Baudelaire		Corpus Wioland		%W-%B	
phonèmes	Er	%	Et	%	Δ	ς'
R	**37**	**11,35**	**25,43**	**7,80**	**-3,55**	**-5,27**
l	**24**	**7,36**	**20,21**	**6,20**	**-1,16**	**-0,71**
d	**21**	**6,44**	**14,02**	**4,30**	**-2,14**	**-3,48**
ε	18	5,52	18,26	5,60	0,08	0,00
a	15	4,60	24,78	7,60	3,00	3,86
e	15	4,60	17,60	5,40	0,80	0,39
i	14	4,29	17,28	5,30	1,01	0,62

Le relevé des ς' les plus révélateurs permet par ailleurs de faire apparaître 3 phonèmes qui, pour avoir des effectifs moins élevés, présentent cependant un degré de singularité particulièrement fort :

Tableau IX

	Baudelaire		Corpus Wioland		%W-%B	
phonèmes	Er	%	Et	%	Δ	ς'
ø	7	2,15	1,96	0,6	-1,55	-13,01
œ	6	1,84	1,63	0,5	-1,34	-11,72
o	12	3,68	6,19	1,9	-1,78	-5,44

En considérant [ø], [œ] et [ə] comme des formes de l'archiphonème [E], et – ce qui est sans doute plus discutable – [o] et [ɔ] comme des variantes de [O], on peut adjoindre deux nouveaux phonèmes à ceux du tableau VIII :

Tableau X

	Baudelaire		Corpus Wioland		%W-%B	
phonèmes	Er	%	Et	%	Δ	ς'
[o], [ɔ]	16	4,91	10,43	3,2	-1,71	-2,97

	Baudelaire		Corpus Wioland		%W-%B	
[ø], [œ][ə]	23	7,06	14,67	4,5	-2,56	-4,73
R	37	11,35	25,43	7,8	-3,55	-5,27
d	21	6,44	14,02	4,3	-2,14	-3,48
l	24	7,36	20,21	6,2	-1,16	-0,71

On met ainsi au jour une séquence : [odəl...R], dans laquelle on est tenté de reconnaître une approximation sonore de [bodəlɛR]... Et, de fait, plusieurs vers semblent broder autour du nom du poète et permettre de reconstituer son patronyme – [(b)odœ/əlɛR] – , voire, tout à la fois, son nom et son prénom – [ʃaʀlœ/ə (bod)œ/əlɛR] –

[n u z o ʀ õ d ɛ l i p l ɛ̃ d o d œ ʀ l e g ɛ ʀ]
[o d œ l ɛ ʀ]

[y z ã t a l ã v i l œ r ʃ a l œ r d ɛ r]
[l œ r ʃ a l œ r d ɛ r]
[ʃ a r l œ œ l ɛ r]

Mais il faut aller plus loin. Les rimes du poèmes, du moins en s'en tenant aux 7 premiers vers, font apparaître clairement un mannequin phonétique, [bo..ɛR] à travers la séquence légÈRES ⇨ tomBEAUX ⇨ étagÈRES ⇨ BEAUX. Dans le même temps, le second quatrain procède à un développement tout à la fois phonétique et sémantique des diphonèmes [dœ], [dø], ou [də], par la répétition de /deux/ et sa transformation en une série paradigmatique : deux, doubles, jumeaux. Ainsi toute la première partie du poème paraît se développer à partir des éléments saillants du nom : [bo] + [də] + [ɛR]. Le système des rimes au vers 8 marque une dégradation du processus, « jumeaux » venant subitement faire écho à la série tomBEAUX, BEAUX, flamBEAUX par une rime « pauvre » inattendue : jumEAUX. On est fondé par conséquent à identifier la suite du poème à un processus de délitement du nom. Peu à peu, les éléments de « Baudelaire » qui constituaient une part importante des premiers vers tendent à disparaître : le dernier [œ] se trouve au vers 7, le dernier [b] au vers 9, le dernier [o] au vers 11, le nombre de [ɛ] diminue singulièrement. Avec les tercets, le singulier – « un soir », « un éclair unique » – s'introduit brutalement dans un monde qui avait été d'abord régi par le pluriel – « des lits », « des odeurs », « des divans », etc. – puis par une sorte de « duel » – « « deux cœurs », « doubles lumières », etc. Si la fusion amoureuse est destructrice, c'est essentiellement parce qu'elle anéantit le poète – l'élément masculin –, laissant comme seul survivant du drame un ange, un être asexué ou, peut-être encore, celle que le poète des *Fleurs du mal* se plut longtemps à considérer comme un être idéal : Aglaé Sabatier, dite « La Présidente ». Car telle est bien la muse éthérée dont le poète chante les vertus séraphiques :

> Ange plein de gaieté, connaissez-vous l'angoisse ?
> [...] Ange plein de bonté, connaissez-vous la haine ?
> [...] Ange plein de santé, connaissez-vous les Fièvres ?
> [...] Ange plein de beauté, connaissez - vous les rides ?
> [...] Mais de toi je n'implore, ange, que tes prières ,
> Ange plein de bonheur, de joie et de lumières !

Si effectivement le nom de Baudelaire s'inscrit en filigrane de « La Mort des amants », on peut se demander (1) pourquoi ni le [b] – dont il faut cependant noter le ς' - 0,56 –, ni le ɛ n'ont pu être détecté par l'analyse automatisée ; (2) pourquoi le [o] et le [ə] ont dû, pour paraître significatifs, être associés à des phonèmes voisins. La réponse peut tenir une fois de plus dans le corpus de référence. Si le balisage du sonnet européen, tel qu'il est prévu dans le

projet D.A.N.T.E., inclut des éléments phonétiques, il sera sans doute plus facile de constituer des ensembles de références et surmonter ainsi l'obstacle que constitue la transcription d'un large corpus. Il ne faut assurément pas pour autant tout attendre de l'affinement des données statistiques. Il convient en effet de prendre en considération deux aspects essentiels de l'analyse : d'une part – comme on l'a déjà souligné –, le déploiement du processus de dissémination dans l'espace du poème ; d'autre part, le rôle joué par les caractéristiques de la métrique. Il est clair qu'un [bo] a plus de poids à la rime qu'en n'importe quel autre endroit du vers. Mais on ne peut négliger non plus les césures, les accents, les enjambements, etc. L'introduction de coefficients dans le calcul devrait permettre de résoudre certaines difficultés à condition, bien sûr, de disposer d'un corpus de référence qui soit uniformément transcrit selon les mêmes règles. C'est dire si le travail qui reste à abattre est dans tous les sens du terme important.

*

Si l'on ajoute que la dissémination touche également – et peut-être selon certains rythmes, selon certaines règles dynamiques des éléments sémantiques ou syntaxiques comme on a pu en avoir plusieurs exemples dans les lignes qui précèdent, on estimera sans doute que tout ou presque reste à faire. Ne peut-on imaginer qu'un poème joue à superposer ou à faire se succéder plusieurs modes de manifestation de l'hypogrammes, phonétique en tel vers, graphique en tel autre, sémantique ailleurs ? Comment, dans ce cas, l'ordinateur pourrait-il réussir un jour à tout mettre à jour– et ce d'autant plus que l'hypogramme est souvent (?) un nom propre, partant un signifiant difficilement décelable. ? Comment, d'un tout autre point de vue, pourrait s'étonner du faible taux de réussite des tests qui précèdent ? Car si la recherche en grande partie manuelle n'a permis d'appliquer le protocole d'analyse qu'à une centaine de poèmes, elle n'a aboutit à des résultats utilisables que dans quinzaine de cas. Sans doute l'informatisation de la démarche permettrait de pratiquer un nombre plus grand, partant plus révélateur d'expériences. Sans pour autant désespérer de l'évolution de la technique, il faut s'attendre toutefois à ce que les statistiques et leurs applications informatiques ne permettent d'accéder à un nombre limité de « clés ». Espérons seulement que celles-ci ne nous conduiront pas à nous ranger dans cette catégorie que dénonce Julien Gracq lorsqu'il écrit dans *En lisant, en écrivant* :

> Que dire de ces gens qui, croyant posséder un clef, n'ont de cesse qu'ils aient disposé votre œuvre en forme de serrure ?

Éric Lysøe
Institut de recherche en langues et littératures européennes (EA 3437)