
Appendix Vergiliana: nova disputatio

Stephan Vonfelt

stephan.vonfelt@free.fr

ABSTRACT: *Who wrote the Appendix Vergiliana? The question has been controversial for centuries, and we try to lift a veil with new techniques. Our measurement, based on the rhythm and on textual characters transitions, reveals an heterogeneous set. The Culex, authenticated by the most ancien accounts, stands out the corpus and is attributed to Virgil. On the contrary, Ausone's poems are relegated out of Virgil's area. Regarding the Aetna, the lot remains dubious, as Donat judges.*

KEYWORDS: *stylometry, author attribution, rhythm, transition, character.*

RÉSUMÉ : *Qui a écrit l'Appendix Vergiliana ? La question est controversée depuis des siècles, et nous tentons de lever un voile à l'aide de techniques nouvelles. Nos mesures, fondées sur le rythme et les transitions entre les caractères textuels, révèle un ensemble hétéroclite et vraisemblablement hétérogène. Le Culex, authentifié par les témoignages les plus anciens, ressort du corpus et se voit attribué à Virgile. À l'inverse, les poèmes d'Ausone sont relégués hors de l'aire virgilienne. Quant à l'Aetna, son sort reste douteux comme le juge Donat.*

MOTS-CLÉS : *stylométrie, attribution d'auteur, rythme, transition, caractère.*

1. Préambule

Cet article doit beaucoup à Jean Soubiran, professeur émérite de l'Université de Toulouse. Lors d'une conversation improvisée, cet éminent latiniste évoqua ses travaux pointus sur la *Ciris*. Ce poème et le recueil de l'Appendix Vergiliana étaient-ils l'œuvre de Virgile ? J'avais trouvé un nouveau terrain pour mes statistiques textuelles.

2. Virgile et ses œuvres

Publius Vergilius Maro naquit en 70 avant J.-C. près de Mantoue, d'une ascendance étrusque et romaine. À l'école, Virgile reçut les influences de deux écrivains opposés par leur style, Lucrèce fidèle à la tradition d'Ennius, et Catulle porteur de la mode alexandrine. Formé à la rhétorique, mis

aussi aux mathématiques et à la médecine à Rome, il expérimenta la philosophie épicurienne lors d'un séjour à Naples.

À partir de 42 av. J.-C., Virgile composa sa trilogie maitresse. S'inspirant de Théocrite, Hésiode et Homère, il chanta successivement les pâturages, les campagnes, et les héros. *Les Bucoliques*, formées de huit cent vers, auraient été écrites en trois ans, les *Géorgiques* de deux mille vers en sept ans, et l'*Énéide* de dix mille vers en onze ans. Si sa poésie bucolique se respire dans l'instant, sa poésie didactique s'égrène dans l'année, tandis que sa poésie épique franchit les siècles. Ce faisant, le style s'élève, de l'humble, au moyen puis au sublime. Virgile mourut avant d'achever sa dernière œuvre, en 19 avant J.-C.

Virgile n'aurait-il écrit que des chefs-d'œuvre ? Des manuscrits inédits furent pourtant évoqués dans son testament. Au 1^e siècle, les témoignages précoces de Lucain, Stace et Martial associent le *Culex* à Virgile. Les grammairiens du 4^e siècle abondent dans ce sens : dans sa *Vie de Virgile*, Donat présenta le *Culex* au même titre que la trilogie, citant le *Catalepton*, les *Priapea*, les *Dirae*, la *Ciris*, et discutant l'authenticité de l'*Aetna* ; Servius maintint ce dernier poème, pour ajouter la *Copa*. Le Moyen Âge adjoignit les *Elegiae in Maecenatem*, le *Moretum*, et trois poèmes aujourd'hui attribués à Ausone : *De institutione viri boni*, *Est et non* et *De rosis nascentibus*. À la Renaissance, l'ensemble fut intitulé Appendix Vergiliana par Scaliger. Tribut de la gloire, cette inflation se vit contestée à l'époque moderne par plusieurs analystes. Nous tentons de lever un voile à l'aune de techniques nouvelles.

3. Statistiques textuelles

Le premier choix est celui des unités linguistiques : vocables lemmatisés, catégories grammaticales, sons élémentaires, sont privilégiés par les études. Mais ces variables de seconde main impliquent un étiquetage préalable, parfois douteux. Fondamentalement, un texte reste une succession de caractères, éléments tangibles et abondants, sources de statistiques fiables et significatives. Dans cette veine, citons Markov et ses processus aléatoires, Shannon et sa théorie de l'information, Brunet et son étude du vocabulaire français, Khmelev & Tweedie puis Jardino pour l'attribution d'auteur¹. Plus récemment, notre thèse a constaté l'efficacité des caractères par rapport aux catégories grammaticales et aux vocables². Notre étude suit cette voie étroite.

De ces unités, les statistiques textuelles retiennent traditionnellement les fréquences d'apparition : le texte est appréhendé comme un sac dont on pèse les billes, les yeux fermés sur l'agencement. Pour pallier cette lacune, une première méthode consiste à segmenter un texte en morceaux, puis à suivre les évolutions des fréquences. Mais le découpage est souvent délicat, a fortiori la comparaison de deux architectures différentes, ainsi que le soulignent Longrée, Luong & Mellet³. Prolongeant la marche vers l'infiniment petit, nous adoptons une division élémentaire qui suit le fil ténu de chaque occurrence. Voyant dans le rythme le fondement de l'art, nous étudions les transitions entre les caractères textuels.

¹ MARKOV A. (1913), SHANNON C.E. (1951), BRUNET E. (1981), KHMELEV D. & TWEEDIE F.J. (2001), JARDINO M. (2006).

² VONFELT S. (2008), chapitre 8.

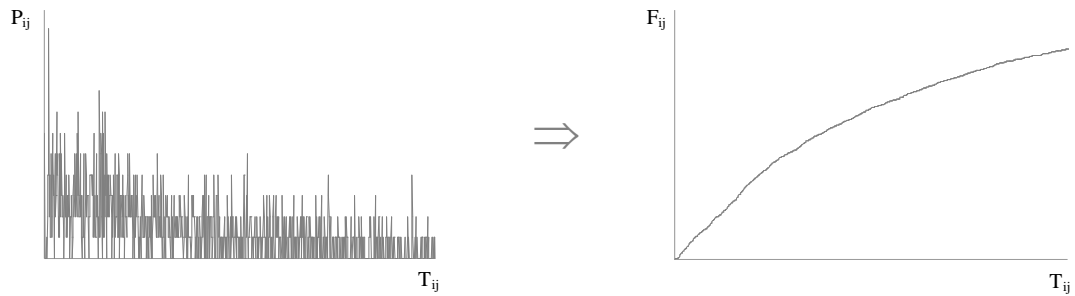
³ LONGRÉE D, LUONG X., MELLET S. (2004).

3.1 Caractérisation d'un texte

Pour fixer les idées, soit la séquence de lettres ABBAABBA, dont nous retenons les temps de transition T entre les lettres A et B. La matrice ci-dessous recense les valeurs correspondantes.

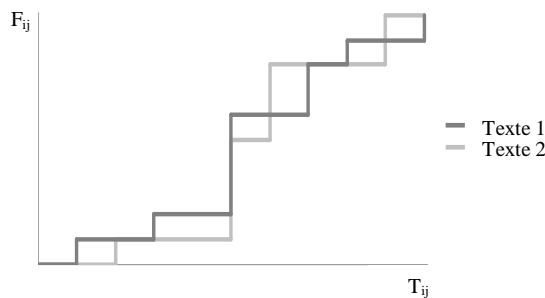
T	A	B
A	3, 1, 3	1, 1
B	1, 1	1, 3, 1

Parmi les transitions du caractère i au caractère j , les valeurs successives de T_{ij} ne présentent généralement aucune corrélation, et se résument à leur distribution P_{ij} ⁴. Ces dernières présentant un aspect chaotique, nous recourons aux répartitions F_{ij} qui les lissent par intégration, soit : $F_{ij}(t) = P_{ij}(T_{ij} \leq t)$.



3.2 Comparaison de deux textes

La comparaison porte dans un premier temps sur les transitions T_{ij} : suivant les valeurs croissantes des temps de transition T_{ijk} , les écarts quadratiques des répartitions entre deux textes sont pondérés par les effectifs n_{ijk} : $d_{12ij}^2 = \sum_k (n_{1ijk} + n_{2ijk}) (F(T_{1ijk}) - F(T_{2ijk}))^2 / (n_{1ij} + n_{2ij})$.



Ces écarts particuliers sont intégrés sur l'ensemble des transitions pour définir une distance intertextuelle : $d_{12}^2 = \sum_i \sum_j (n_{1ij} + n_{2ij}) d_{12ij}^2 / (n_1 + n_2) = \sum_i \sum_j \sum_k (n_{1ijk} + n_{2ijk}) (F(T_{1ijk}) - F(T_{2ijk}))^2 / (n_1 + n_2)$.

⁴ Sur le corpus de notre thèse, les corrélations sont nulles dans le cas des lettres, et légèrement perceptibles pour les espaces et la ponctuation : VONFELT S. (2008), chapitre 7.

Moyenne des écarts quadratiques entre les répartitions, d est comprise entre 0 et 1. C'est une pseudo distance⁵ : $d_{11} = 0$, $d_{12} = d_{21}$, et $d_{13} \leq d_{12} + d_{23}$ d'après l'inégalité de Cauchy-Schwarz. Formellement, nous noterons : $d_{12} = |F_1 - F_2|$.

La longueur est une caractéristique importante d'un texte : sous le joug du nombre, les écrits des uns et des autres se rendent à la langue commune. La distance décroît naturellement avec les effectifs⁶ : si $n = (n_1^{-1} + n_2^{-1})^{-1}$, d est de l'ordre de $n^{-0.5}$.

4. Application

4.1 Corpus

Le corpus comprend en premier lieu les œuvres de référence de Virgile : *Ecloga*, *Georgica*, *Aeneis*. Il recueille par ailleurs les œuvres de l'Appendix Vergiliana : *Dirae*, *Culex*, *Aetna*, *Copa*, *Elegiae in Maecenatem*, *Ciris*, *Priapea*, *Catalepton*, *Priapeum*⁷, *Moretum*, *De institutione viri boni*, *Est et non*, *De rosas nascentibus*.

Les éditions utilisées sont respectivement celles de Mynors et Clausen, de l'Université d'Oxford⁸. Les versions informatisées sont tirées de la Bibliotheca Augustana, de l'Université d'Augsburg⁹. L'intégralité des caractères est prise en compte : les espaces, la ponctuation, les lettres en distinguant minuscules et majuscules.

4.2 Mesures

4.2.1 Carte

Chaque œuvre du corpus peut être vue comme un point d'un espace de dimension $m-1$, m désignant le nombre d'œuvres (16 en l'occurrence). Pour imaginer cette géométrie complexe, les points sont projetés vers leur plan principal en diagonalisant la matrice des distances mutuelles¹⁰. Susceptible de trahir ponctuellement, cette carte traduit en bonne justice les masses en instance.

Les coordonnées des points sont données en annexe 1. Sur la carte, la trilogie de Virgile forme un groupe soudé. En revanche, l'Appendix Vergilina constitue un ensemble hétéroclite, vraisemblablement hétérogène : il paraît douteux que Virgile en soit l'auteur unique.

⁵ La nullité de la distance n'implique pas l'identité des textes : sans la propriété de séparation, cette mesure n'est pas une vraie distance, mais la limitation est ici sans incidence.

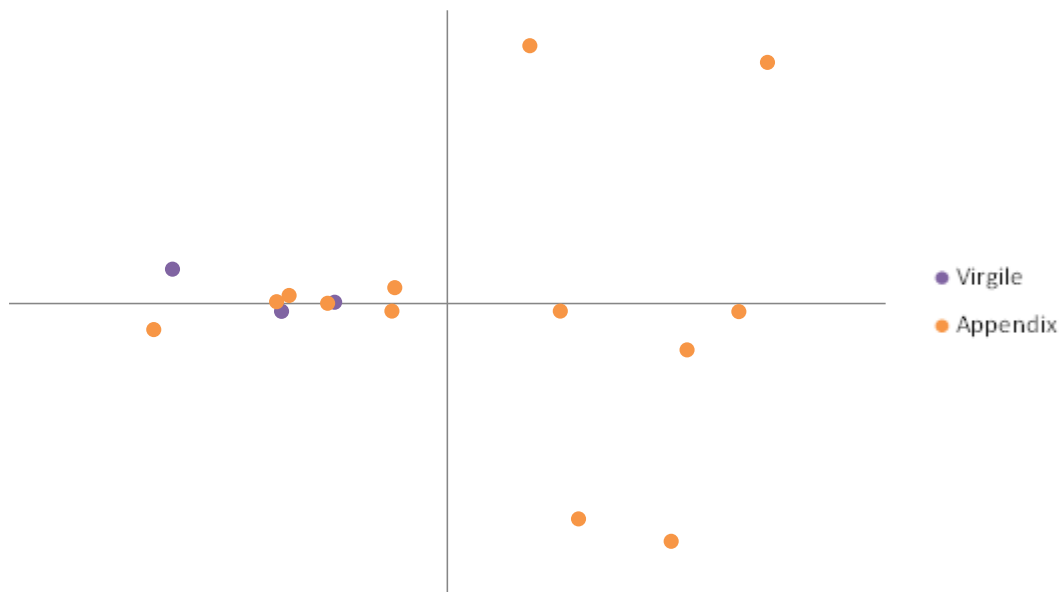
⁶ Dans le cas de variables continues, si deux échantillons de tailles n_1 et n_2 sont tirés d'une même population, $(n_1^{-1} + n_2^{-1})^{-1} d^2$ tend vers une loi indépendante de cette population : ANDERSON T.W. (1962). Nous admettons ce gabarit pour des variables discrètes.

⁷ Le *Priapeum* est inclus dans les *Priapea* dans certaines éditions.

⁸ VIRGIL (1969, 1966).

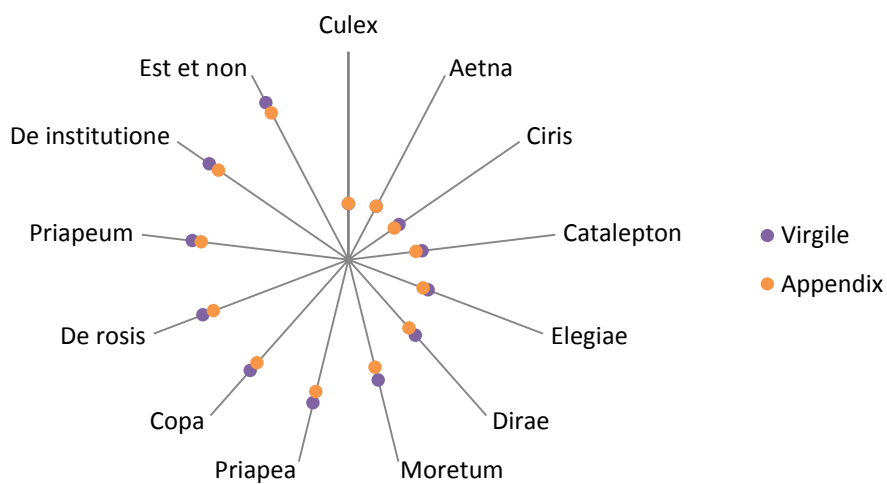
⁹ www.hs-augsburg.de/~harsch/Chronologia/Lsante01/Vergilius/ver_intr.html

¹⁰ La théorie est donnée dans BENZÉCRI J.P. (1973).



4.2.2 Attributions

Dans le détail, une œuvre de l'Appendix Vergiliana est attribuée à Virgile ou à l'Appendix (a priori d'un autre auteur) en comparant sa distance par rapport à ces deux sous-corpus. Afin que le jeu soit équitable, l'œuvre en question est boutée hors de l'Appendix pour le calcul de la distance interne.



Les distances données en annexe 2 sont représentées sur la figure, par valeurs croissantes. Œuvre la plus proche de Virgile, le *Culex* lui est attribuée, et l'*Aetna* en second est donné de peu à l'Appendix. À l'autre extrémité du cadran figurent les œuvres d'Ausone : *Est et non*, *De institutione viri boni*, *De rosis nascentibus*.

5. Conclusion

Selon nos mesures, l'Appendix Vergiliana forme un ensemble hétéroclite et vraisemblablement hétérogène. Le *Culex*, authentifié par les témoignages les plus anciens, ressort du corpus et se voit attribué à Virgile. À l'inverse, les œuvres d'Ausone sont reléguées hors de l'aire virgilienne. Quant à l'*Aetna*, son sort reste douteux comme le juge Donat.

Ce point de vue ne saurait clore la question. Nous espérons cependant avoir donné un éclairage nouveau sur l'Appendix Vergiliana.

Références

- ANDERSON T.W. (1962), « On the distribution of the two-sample Cramer-Von Mises criterion », *The Annals of Mathematical Statistics*, Beachwood, Institute of Mathematical Statistics, vol. 33, n° 3, p. 1148-1159.
- BENZÉCRI J.P. (1973), *L'analyse des Données*, Paris, Dunod.
- BRUNET E. (1981), *Le vocabulaire français de 1789 à nos jours*, Paris, Champion.
- JARDINO M. (2006), « Identification des auteurs de textes courts avec des n-grammes de caractères », *Actes des Journées internationales d'Analyse statistique des Données Textuelles*, Besançon, Presses Universitaires de Franche-Comté.
- KHMELEV D. & TWEEDIE F.J. (2001), « Using Markov Chains for Identification of Writers », *Literary and Linguistics Computing*, Oxford University Press, vol. 16, n° 4, p. 299-307.
- LONGRÉE D., LUONG X., MELLET S. (2004), « Temps verbaux, axe syntagmatique, topologie textuelle : analyse d'un corpus lemmatisé », *Actes des Journées internationales d'Analyse statistique des Données Textuelles*, Presses Universitaires de Louvain.
- MARKOV A. (1913), « Un exemple de recherche statistique sur le texte d'Eugène Onéguine illustrant la liaison des épreuves en chaînes », *Bulletin de l'Académie Impériale des Sciences*, vol. 7, Saint-Pétersbourg, p. 153-162.
- SHANNON C.E. (1951), « Prediction and Entropy of Printed English », *Bell Systems Technical Journal*, Hoboken, Wiley, vol. 30, p. 50-64.
- VIRGIL (1966), *Appendix Vergiliana*, W.V. Clausen & alii, Oxford University Press.
- VIRGIL (1969), *P. Vergili Maronis: Opera*, R.A.B Mynors, Oxford University Press.
- VIRGILE (1935), *La fille d'auberge* suivi des autres poèmes attribués à Virgile, traduction nouvelle par M. Rat, Paris, Garnier.
- VIRGILE (2015), *Œuvres complètes*, édition bilingue établie par J. Dion et P. Heuzé, Paris, Gallimard.
- VONFELT S. (2008), *La musique des lettres : variations sur Yourcenar, Tournier et Le Clézio*, Université de Toulouse.

Annexes

1 Carte

		X	Y
Virgile	<i>Ecloga</i>	-9,41	1,17
	<i>Georgica</i>	-3,86	0,03
	<i>Aeneis</i>	-5,67	-0,27
Appendix	<i>Dirae</i>	-1,79	0,53
	<i>Culex</i>	-5,84	0,06
	<i>Aetna</i>	-1,90	-0,27
	<i>Copa</i>	2,83	8,79
	<i>Elegiae in Maecenatem</i>	-5,41	0,27
	<i>Ciris</i>	-4,09	0,00
	<i>Priapea</i>	4,49	-7,36
	<i>Catalepton</i>	-10,05	-0,90
	<i>Priapeum</i>	7,66	-8,12
	<i>Moretum</i>	3,87	-0,27
	<i>De institutione viri boni</i>	9,98	-0,28
	<i>Est et non</i>	10,96	8,22
	<i>De rosis nascentibus</i>	8,21	-1,59

2 Attributions

	Virgile	Appendix
<i>Culex</i>	6,70	6,74
<i>Aetna</i>	7,24	7,23
<i>Ciris</i>	7,39	6,70
<i>Catalepton</i>	8,90	8,19
<i>Elegiae in Maecenatem</i>	10,25	9,60
<i>Dirae</i>	12,16	10,95
<i>Moretum</i>	14,93	13,31
<i>Priapea</i>	17,73	16,32
<i>Copa</i>	17,79	16,59
<i>De rosis nascentibus</i>	18,70	17,34
<i>Priapeum</i>	18,89	17,80
<i>De institutione viri boni</i>	20,29	18,95
<i>Est et non</i>	21,32	19,89