

J.-P. BENZÉCRI

Programme de recherches en stylométrie

Les cahiers de l'analyse des données, tome 13, n° 1 (1988),
p. 97-98

http://www.numdam.org/item?id=CAD_1988__13_1_97_0

© Les cahiers de l'analyse des données, Dunod, 1988, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

PROGRAMME DE RECHERCHES EN STYLOMÉTRIE

[PROGRAMME]

J.-P. BENZÉCRI

Quelque intérêt que présentent en elles-mêmes les recherches stylométriques déjà réalisées grâce à l'ordinateur, nous pensons qu'elles valent plus encore comme la preuve que des recherches bien plus vastes peuvent être, dès maintenant, entreprises. D'où le titre de la présente note.

Étant admis qu'à la base de toute étude est un tableau de correspondance, croisant un ensemble I de textes avec un ensemble J de traits caractéristiques qui y sont dénombrés, nous considérerons successivement I, J et les conclusions à attendre du labeur du statisticien. Pour être bref, on désignera par leurs sigles les articles du présent cahier d'après lesquels nous argumenterons.

L'étude [FRÉQ. CAT. LATIN], J. DENOOZ nous paraît avoir démontré qu'une seule étude statistique pouvait embrasser l'ensemble des grandes œuvres d'une littérature. Même si, d'autre part, des études partielles, consacrées à un genre ou à une période, demeurent nécessaires, il sied de leur donner un cadre.

Pour ample que soit le domaine couvert, il faut se garder de prendre pour individus des textes trop longs, des agrégats hétérogènes. Dans [HEXA. DACTYL. LATIN], Étienne ÉVRARD nous en avertit en ces termes:

Quand il s'agit d'analyse quantitative, si l'on traite l'ensemble de l'œuvre comme un bloc indifférencié, on masque nécessairement les particularités locales, s'il y en a. C'est la raison pour laquelle j'ai traité séparément non seulement les œuvres d'un même auteur, mais mêmes leurs subdivisions.

Nous savons que la statistique sans ordinateur impose souvent de prendre des échantillons de longueur fixée et assez grande: l'analyse factorielle n'a pas de telles contraintes; il est loisible de descendre en taille en ne s'arrêtant qu'au niveau où les fragments pulvérisés perdent tout caractère. L'expérience seule nous peut dire quel est ce niveau. On voit dans [MÉT. ARISTOTE] que Chr. RUTTEN a été bien avisé de descendre jusqu'au chapitre; et dans [GENÈSE], I. KHARCHAF, au niveau du verset, a obtenu des résultats intéressants un spécialiste aussi compétent que notre regretté collègue Gérard WEIL.

Si l'on analyse une œuvre, ou un recueil d'œuvres, en forme de dialogue, un type particulier de subdivision mérite d'être considéré: le *rôle*, ou ensemble des paroles attribuées dans une œuvre à un personnage déterminé. On atteindrait ainsi une image globale de la société figurant dans les comédies de Molière, ou dans les dialogues de Platon.

Quant à l'ensemble J des traits dénombrés, on sait que de belles études ont été faites en dénombrant des mots ou des formes fléchies: le recueil *Pratique de l'A. des D. en Linguistique* en offre plus d'un exemple. En dénombrant des formes, comme en dénombrant des lettres ou des parties du discours, on touche à la syntaxe et aux morphèmes non sémantiques. Un légitime souci de comparabilité peut contraindre à accepter une nomenclature des catégories grammaticales, tout autre qu'innocente! Cependant, s'écartant des voies frayées, Gérard LECOMTE a mis à profit le système morphologique de la langue arabe, et, non content de recenser les formes, il a tenu compte des fonctions, notamment pour les emplois des cas. Les analyses de [STYL. ARAB.] nous paraissent montrer la pertinence des distinctions ainsi introduites.

Le linguiste qui recourt à la stylométrie, ou plus généralement à la statistique, cherche d'ordinaire la réponse à une question spécifique: datation relative de textes ou de fragments, attribution d'une œuvre, subdivision éventuelle d'un texte d'abord reçu pour homogène. Nous reconnaissons l'importance de telles questions, mais ferons plusieurs réserves.

De même qu'il n'y a pas de diagnostic sans physiologie de l'homme sain, on ne répondra valablement aux questions de datation, d'attribution, de propriété qu'après avoir étudié les marques statistiques de la diversité spatiotemporelle des œuvres, sur de nombreux cas qui ne font pas problème. Cette étude générale nous paraît ouvrir les plus belles perspectives sur l'économie de la production littéraire; thème qui surpasse en intérêt toutes les énigmes particulières.

L'analyse multidimensionnelle doit être regardée comme une discipline expérimentale, en ce sens que le chercheur doit considérer les données disponibles sous tous leurs aspects, accepter les suggestions qui s'offrent, poursuivre la collecte si les données manquent pour conclure; se rappeler que, même s'il y a des lois générales, Étienne ÉVRARD dit sagement

qu'il n'y a pas d'équivalence automatique entre un caractère quantitatif et tel trait relatif à l'histoire littéraire ou à l'interprétation doctrinale. Ce sont chaque fois des cas d'espèce, qu'il convient d'examiner en particulier.

إذا لم يكن أمضى من السيف حامل
فلا تقطع ، إن ، الكف ، لا السيف تقطع

S'il n'y a, plus incisif que le sabre, un sabreur

Point de coup: car c'est la main, non le sabre qui coupe! (Al Buhturî)