

M. ABIFARAH

## **Reconnaissance de l'auteur d'un texte d'après les caractères utilisés**

*Les cahiers de l'analyse des données*, tome 13, n° 1 (1988),  
p. 95-96

[http://www.numdam.org/item?id=CAD\\_1988\\_\\_13\\_1\\_95\\_0](http://www.numdam.org/item?id=CAD_1988__13_1_95_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/>

# RECONNAISSANCE DE L'AUTEUR D'UN TEXTE D'APRÈS LES CARACTÈRES UTILISÉS

## [LETTRES]

M. ABI FARAH\*

L'objet de la présente note est seulement de rendre compte de recherches en cours sur le thème qu'indique le titre.

### 1 Rappel de résultats antérieurs

Étienne BRUNET est, à notre connaissance, le premier auteur à avoir signalé qu'on obtenait des résultats interprétables, quant au genre littéraire, en soumettant à l'analyse des correspondances un tableau de contingence croisant un ensemble de textes (ou de groupes de textes) avec l'ensemble des signes typographiques employés (ou une partie de ceux-ci, par exemple les signes de ponctuation). Anis ABI FARAH a fait de semblables recherches sur des textes en langue arabe. On peut légitimement s'étonner du succès de ces analyses dont l'interprétation est assurément délicate: É. Brunet a montré que les mots les plus fréquents n'étaient pas seuls en cause, car les résultats subsistent même si ces mots sont éliminés du bilan. Il faut donc conclure que les œuvres sont caractérisées par les morphèmes (désinences, préfixes, suffixes...) dont l'usage massif détermine le profil d'utilisation des lettres.

### 2 Compte-rendu de nos expériences

Marcel Proust et Marcel Aymé ont fait l'objet d'un premier essai, avec 4 textes pour chaque auteur, pris à des endroits différents, respectivement dans *Un amour de Swann* et dans *Derrière chez Martin*: sur l'axe 1 la séparation est parfaite, comme le montre le schéma ci-dessous.

Aym4Aym2Aym1Aym3-----+-----Pro4--Pro2--Pro3-----Pro1

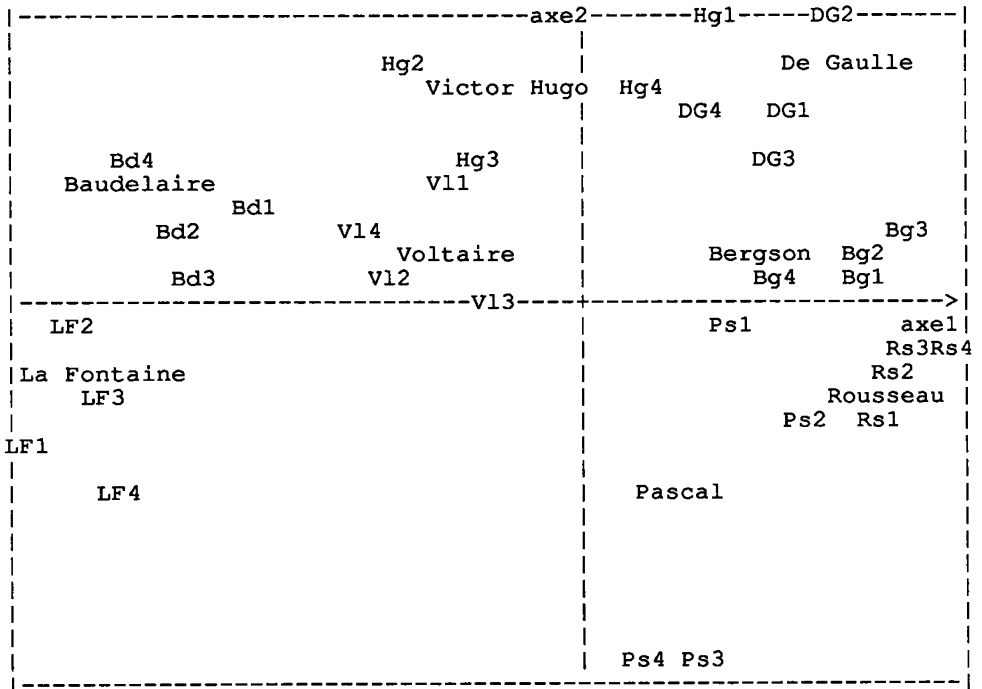
Un essai similaire, portant sur des textes courts (environ 2000 caractères) d'Émile Zola et Romain Gary a produit des résultats aussi nets.

Gar3-----Gar4Gar1Gar2-----+-----Zol4Zol3Zol1-----Zol2

---

(\*) 1 Rue des églantines, 95320, Saint-Leu la Forêt.

Nous avons ensuite abordé l'étude simultanée de plus de deux auteurs avec Baudelaire, Bergson, De Gaulle, La Fontaine, Rousseau et Voltaire représentés chacun par 4 textes de quelque 2000 caractères (sauf pour Baudelaire, 3 textes). L'analyse a été faite d'une part en dénombrant seulement les lettres, et d'autre part en dénombrant lettres et signes de ponctuation. Nous choisissons de présenter ici le plan (1,2) issu de l'analyse de la correspondance croisant des textes de 4000 caractères, issus de huit auteurs, avec l'ensemble des caractères (lettres et ponctuations). Même si l'interprétation en terme de morphèmes reste à faire, on ne peut nier que la discrimination obtenue est prometteuse.



### Références bibliographiques

Étienne Brunet: *Le Vocabulaire français de 1789 à nos jours*; 1824 pp., Slatkine, Genève-Paris, 1981; résumé par l'auteur dans: *L'analisi delle frequenze, problemi di lessicologia*; A cura di M. Fattori e M. Bianchi; Lessico Intellettuale Europeo XXVII; Centro di studio del CNR; Edizioni dell'Ateneo; Roma, 1982; brièvement analysé dans *CAD*, Vol VIII, n°1, pp. 123-125, 1983.

Anis Abi Farah: *La reconnaissance automatique des auteurs inconnus des textes arabes*; Thèse de doctorat ès-sciences; Université Pierre et Marie Curie; Paris, 1985.