

AVANT-PROPOS

Cet ouvrage s'adresse à ceux qui, pour leurs recherches, leurs travaux d'études, leur enseignement, doivent décrire, comparer, classer, analyser des ensembles de textes. Il peut s'agir de textes littéraires, scientifiques (bibliométrie, scientométrie, recherche documentaire), économiques, sociologiques (réponses aux questions ouvertes dans des enquêtes socio-économiques, entretiens divers en marketing, psychologie appliquée, pédagogie, médecine), de textes historiques, politiques...

On a tenté de faire le point sur les développements de la *statistique textuelle*, domaine de recherche vivant dont les contours exacts sont difficiles à établir tant est large l'éventail des disciplines concernées, et aussi celui des applications possibles. Les chapitres qui suivent voudraient, tout en présentant l'acquis de ce champ disciplinaire, témoigner de cette richesse d'approches, de méthodes et de domaines.

L'ouvrage reprend, en intégrant des développements récents, certains exemples du manuel *Analyse statistique des données textuelles* publié par les mêmes auteurs en 1988. Le champ des applications précédemment limité aux traitements de *questions ouvertes* a été considérablement élargi de même que l'éventail des méthodes proposées. L'ensemble, profondément remanié, inclut de nouveaux chapitres qui traitent des structures a priori et de l'analyse discriminante textuelle, thèmes qui dépassent largement l'optique essentiellement descriptive de l'ouvrage antérieur.

Plusieurs lectures devraient être possibles selon la formation du lecteur, et selon notamment ses connaissances en mathématique et statistique. Une lecture technique, complète, pour une personne ayant dans ces matières une formation équivalente à une maîtrise de sciences économiques, aux écoles d'ingénieurs ou de commerce. Une lecture pratique, d'utilisateur, pour les personnes spécialisées dans les divers domaines d'application potentiels.

Les démonstrations strictement mathématiques ne figurent pas dans le texte. On renvoie à chaque fois le lecteur curieux d'en connaître les détails à des publications ou ouvrages plus spécialisés lorsque ceux-ci sont facilement accessibles. En revanche, la part belle est faite à la définition des concepts, à la mise en oeuvre des procédures, aux règles de lecture et d'interprétation des résultats. Le glossaire en fin d'ouvrage aidera le lecteur à préciser le contenu des notions ou des conventions de notation les plus importantes.

L'ensemble doit beaucoup à des collaborations et des cadres de travail divers : au sein du département Economie et Management, de l'Ecole Nationale Supérieure des Télécommunications (Télécom Paris) et de l'URA820 du Centre National de la Recherche Scientifique (Traitement et Communication de l'Information) de cette même Ecole ; au sein du Laboratoire "Lexicométrie et textes politiques", URL 3 de l'Institut national de la langue française (INaLF) et de l'Ecole Normale Supérieure de Fontenay-Saint-Cloud.

Nous remercions également les autres chercheurs ou professeurs auprès desquels nous avons puisé collaboration et soutien, ou simplement eu d'intéressants débats ou discussions. Citons, sans être exhaustif, C. Baudelot (ENS, Paris), M. Bécue, (UPC., Barcelone), L. Benzoni (Télécom Paris), E. Brunet (INaLF, Nice), S. Bolasco (Univ. de Salerne), L. Haeusler (Cisia, Paris), G. Hébrail (EDF, Clamart), D. Labbé (CERAT, Grenoble), A. Lelu (Univ. Paris VIII), M. Reinert (Univ. Toulouse Le Mirail).

Nous sommes heureux d'adresser ici nos remerciements à Gisèle Maïus, des éditions Dunod, pour l'accueil qu'elle a réservé à cet ouvrage.

L. L., A. S.
Paris, Janvier 1994