

Préface

Et le Verbe s'est fait Nombre...

Il y a dans l'activité qui consiste à traiter les mots comme des nombres - opération de base de la statistique textuelle - un a priori qui ne manquera pas d'apparaître à certains comme outrageusement réducteur voire même sacrilège. Surtout si l'on en croit Victor Hugo : *Car le mot, c'est le Verbe, et le Verbe c'est Dieu...*

Il suffit de lire ce livre et surtout d'en appliquer les principes à ses propres enquêtes pour se convaincre du contraire. Avec ses graphes d'analyse factorielle, J.P. Benzécri a rendu les individus à la statistique : longtemps ignorés à force d'être confondus dans de vastes agrégats ou pulvérisés dans des formules inférentielles qui s'intéressent d'abord aux relations entre des grandeurs abstraites (revenu et consommation, salaire et diplôme...), les individus effectuent leur rentrée sur la scène statistique sous la forme de points dans un nuage. Les positions respectives qu'ils occupent au sein de ce nuage démontrent d'abord qu'ils diffèrent tous les uns des autres. Les distances et les proximités qu'ils entretiennent avec les modalités des variables considérées permettent ensuite de comprendre en quoi chacun diffère de l'autre : par ses goûts, ses opinions politiques, son âge, son sexe, la marque de sa voiture, la profession de son père... mais la statistique est encore une histoire sans parole.

L'une des contributions majeure de la statistique textuelle est précisément d'animer tous ces graphes en donnant la parole à chacun de ces individus. Grâce à Lebart et Salem, les fameux points-individus ne sont plus muets, ils parlent. Vole alors en éclats la traditionnelle mais artificielle distinction entre le quantitatif et le qualitatif. Les méthodes ici présentées permettent de mettre en relation les propriétés sociales ou personnelles des individus telles que les saisit l'enquête statistique avec les textes par lesquels ces mêmes individus répondent aux questions qu'on leur pose sans en réduire le moins du monde l'information. Les nuances les plus subtiles de l'expression sont conservées : le singulier et le pluriel, la majuscule et la minuscule, l'usage du "je", du "on", du "nous". La formule le dit bien : *s'exprimer* c'est d'abord se livrer soi-même au-dehors. Chaque forme lexicale tire alors son sens d'un triple registre : celui que lui donne celui qui la prononce, celui que lui confère la place qu'elle occupe dans l'espace dessiné par

toutes les autres formes lexicales énoncées par le même individu, celui, enfin, qu'elle tient de la place qu'elle occupe dans l'espace dessiné par toutes les autres formes énoncées par tous les autres locuteurs. Le sens jaillit des différences de profil.

Cet ouvrage a le mérite de déborder largement le cadre de l'analyse de contenu ou du traitement statistique des questions ouvertes dans les enquêtes. Il fait le point sur l'état de développement d'un chantier particulièrement foisonnant depuis dix ans. Il expose les dernières découvertes. Elles sont nombreuses et riches d'application dans les domaines les plus divers : stylométrie, recherche documentaire, modèles prévisionnels. Comment attribuer un texte à un auteur ou à une période ? Combien d'auteurs ont contribué à la rédaction du livre de la Bible attribué au prophète Isaïe ? Peut-on comparer des comportements exprimés dans des textes écrits dans des langues différentes sans les traduire ni les coder ?

C'est souvent aux confins des disciplines instituées que l'invention scientifique est la plus féconde. Lorsque deux statisticiens tout particulièrement sensibilisés aux problèmes que l'on rencontre dans les sciences humaines se réunissent autour d'un ordinateur pour élaborer les principes et les outils d'une statistique textuelle, ils occupent le coeur d'un carrefour scientifique vers lequel convergent tout naturellement des linguistes, d'autres statisticiens bien sûr mais aussi les spécialistes d'analyse du discours, d'analyse de contenu, d'analyse des textes littéraires, de recherche documentaire et d'intelligence artificielle. A ce noyau dur de producteurs de théories et d'outils est venu petit à petit s'agréger un univers polyglotte d'utilisateurs aux formations diverses : sociologues, littéraires, stylomètres, historiens, géographes, politologues, médecins, éthologues, psychologues, publicitaires, etc.

On peut savoir gré à l'ouverture d'esprit des deux auteurs (et de leurs associés !), à leur générosité intellectuelle et humaine pour avoir su accueillir autour de leur disque dur un nombre croissant de producteurs et d'utilisateurs dont ils ont souvent stimulé l'inventivité. Il suffit pour s'en convaincre de feuilleter les actes des deux journées internationales qu'ils ont suscitées, avec d'autres, à Barcelone en 1990 et à Montpellier en 1993. Ou de goûter, chez soi, le charme inattendu de nouveaux logiciels.

Au-delà de la collection de principes et d'outils statistiques présentés dans les pages qui suivent, n'oublions pas que la nature même de la matière travaillée - le texte - confère à l'entreprise des dimensions à la fois culturelles, internationales et universelles car comme le disait si bien Victor Hugo ...

Christian Baudelot