

Informatique & Bible, asbl - Belgique
Rue de Maredsous, 11 B5537 Denée - Belgique
Tél:+32(0)82.69.96.47 Fax:+32(0)82.22.32.69
cib@cibmaredsous.be



Interface n° 98/69 2e trimestre 1998

Si la Bible m'était comptée

Du 19 au 21 février 1998, le Frère R.-Ferdinand Poswick a participé aux 4èmes Journées Internationales d'Analyse Statistique des Données Textuelles, organisées par l'Université de Nice Sophia Antipolis. Voici sa contribution.

Il y a différentes manières d'analyser un texte. L'une d'elle - amplement facilitée par l'informatique - consiste à compter le nombre de phrases, de mots, de lettres qu'il contient. C'est à cette analyse statistique des textes qu'étaient consacrées ces 4èmes Journées Internationales.

Y participaient d'éminents spécialistes comme Etienne Brunet, Michel Juillard et Xuan Luong de l'Université de Nice Sophia Antipolis ainsi qu'André Salem de l'Université de Paris III. Le Père Robert Busa, sj parla des onze millions de mots lemmatisés que comprend l'oeuvre de Saint Thomas d'Aquin. N. Nacourly de l'Université du Chili et Ernesto San Martin de l'UCL analysèrent les fréquences statistiques des prépositions dans la Bible des Septante.

Pour sa part, R.-Ferdinand Poswick a fait le point sur les connaissances sur la manière dont le comptage s'est fait en ce qui concerne la Bible.

Chiffres, nombres et le calcul dans la Bible

Concernant les chiffres, nombres et le calcul dans la Bible, l'orateur rappelle l'avis de Jens Hyrup selon lequel "ni l'Ancien ni le Nouveau Testament ne sont le produit de cultures qui aient une tradition mathématique supérieure au niveau normal des mathématiques populaires. Mais l'un et l'autre Testament furent le produit de cultures en contact avec les traditions mathématiques éprouvées et savantes." Ces cultures sont celles de la tradition suméro-babylonienne, de l'Égypte et de la Grèce. Mais la nature même de la Bible comme texte religieux, littéraire et historique n'appelle pas vraiment un discours sur les nombres.

Les plus vieilles attestations d'usage de nombres dans les cultures qui ont généré la Bible sont liées aux pratiques mésopotamiennes dans lesquelles les grands nombres sont exprimés sur une base de 60. L'écriture hiéroglyphe égyptienne qui se développe durant les mêmes périodes possède un système de numération plus complet permettant de compter jusqu'au million. Plus tard, les "papyri" égypto-grecs utilisent la notation grecque. Les 22 lettres de l'alphabet linéaire phénicien sur le modèle duquel seront bâtis les alphabets hébreu, araméen, protoarabique, grec, étrusque, romain... ne sont pas antérieures au 14ème siècle av. J.C.

Les Hébreux emprunteront d'abord le système de notation araméen. Les lettres alphabétiques hébraïques semblent cependant n'avoir été utilisées qu'à partir du 2ème siècle avant J.C. pour le calcul ou la numération. La diaspora juive des mêmes époques utilise d'ailleurs la numération alphabétique grecque pour représenter les chiffres.

Tout comme la fonction scribale qui la développe, la numération est liée au sacré et le nombre 7, dès les temps les plus anciens, une relation au divin. La valeur symbolique du nombre 7 se retrouve dans les textes suméro-accadiens comme l'Épopée de Gilgamesh ou l'Histoire du Déluge babylonienne et donne à ce chiffre le sens de "totalité".

C'est sur ces bases et dans cette ligne que se développera la "gématrie" dans le Talmud, les Midrashim et, finalement, les littératures cabalistiques. Ce type de spéculation dont les

pythagoriciens étaient friands et que l'on retrouvera dans tous les courants allégoriques et mystiques de commentaire de la Bible, se retrouve très explicitement dans l'usage de la Bible grecque des Septante. Ces spéculations seront reprises jusqu'à nos jours. Mais elles peuvent mener à des aberrations d'autant plus perverses qu'elles sont présentées avec l'apparence d'une rigueur mathématique dont le profane peut difficilement percevoir les failles.

Tel est le cas pour le best-seller récent de Michael Drosnin, *La Bible: le code secret* qui prétend trouver dans le "texte" hébraïque de la Bible des révélations explicites sur des événements contemporains.

Bien plus sérieuse est l'observation patiente d'une utilisation des nombres dans la structures de certaines unités littéraires à l'intérieur du texte biblique. Le P. Langlamet montre que certains textes ont été structurés de façon numérique et qu'ils peuvent représenter, à ce titre, un "exercice scribal".

La Bible comptée

L'école des scribes "massorètes" (compteurs, dénombreurs) a été l'une des premières, dans un souci de fidélité de la transmission, à effectuer des comptes minutieux des mots et des lettres de toutes les unités littéraires rassemblées dans la Bible hébraïque.

Gérard-E. Weil a étudié leur travail tout au long de sa vie. Il nous apprend que les vieilles listes de décompte de versets de la Torah (Pentateuque) se trouve dans un manuscrit du Caire vers 875-900 de notre ère. L'auteur-compteur y note un total de 5.843 versets. D'autres décomptes donnent respectivement 5.845, 5.844, 5.841 et 5.855 versets.

Les lettres elles-mêmes ont été dénombrées: selon les décomptes de l'ordinateur, on en trouve 304.848 (alors que les colophons des scribes massorètes en annoncent 400.945).

Une somme du comptage sur le texte massorétique de la Bible a été publiée en 1989 par Francis I. Andersen et A. Dean Forbes sous le titre *The Vocabulary of the Old Testament*. Les statistiques et comptages présentés là sont également basés sur un enregistrement du codex B19a de Leningrad. Non seulement les mots sont comptés, mais aussi tous les types d'éléments de la langue: particules, verbes, noms, pronoms.

Un premier type d'affichage en table donne les totaux des faits observés par ensemble. Vient ensuite une Concordance de tous les mots en ordre alphabétique avec le compte pour chaque unité et toutes les références des occurrences; une Table de répartition des mots qui adviennent plus de 20 fois dans le texte. Chaque Table présente également ce que les auteurs nomment une "mesure d'incidence" soit le nombre d'occurrences par 10.000 mots. Par exemple, dans la Bible hébraïque, il y a 5.025 noms sur 10.000 mots, ce qui signifie "qu'on peut s'attendre à ce que les noms constituent la moitié de presque tous les textes".

Malgré l'importance de cette version pour l'étude actuelle de la Bible, nous ne possédons pas encore d'outils comparables pour le texte des Septante dont on sait aujourd'hui qu'il reflète souvent un état du texte hébreu de la Bible antérieur aux fixations massorétiques et dont a retrouvé des témoins à Qumrân.

Il faut utiliser des comptages anciens réalisés selon une méthode non informatisée ou recourir aux outils électroniques disponibles. Toutefois les résultats de l'emploi de ces derniers doivent être pris avec précautions car a) le texte enregistré peut être différent; b) la façon d'isoler des unités peut varier selon la conception linguistique ou grammaticale; c) les fonctions de recherche ou de calcul peuvent être conçues selon des principes différents.

L'importante synthèse de A. Dean Forbes rappelle que l'histoire de l'application de méthodes statistiques à des textes bibliques commence par une recherche sur l'authenticité des Epîtres de Paul à partir de méthodes proposées dès 1939 par G.U. Yule pour déterminer si l'auteur de l'imitation de Jésus Christ est Thomas à Kempis ou Jean Gerson.

La principale méthode utilisée est celle d'une évaluation de la longueur des phrases et de leur distribution dans une oeuvre. Ou encore celle d'une mesure des habitudes inconscientes d'un auteur quand il écrit (particules, morphologie, syntaxe). En tête de son article, Forbes donne quelques critères pour un travail de statistique valable sur un corpus littéraire.

Nous avons nous-mêmes proposé, directement ou indirectement, quelques pistes d'utilisation statistique du matériau biblique. L'édition de la Concordance de la Bible de Jérusalem proposait en appendice quelques listes résultant de tris exclusifs avec le vocabulaire qui constituait un écart significatif d'usage dans l'Ancien et le Nouveau Testament pour 75 lemmes les plus fréquemment présentés en ordre de fréquence décroissante. Cette méthode, appliquée à la comparaison d'autres ensembles à l'intérieur du Corpus, permettrait des observations de champs sémantiques caractéristiques et spécifiques. Ainsi, à titre d'exemple, les 10 premiers lemmes n'intervenant jamais dans le Nouveau Testament, tel que traduit par la Bible de Jérusalem sont: bétail (200), bélier (170), sicle (152), vallée (138), frontière (128), graisse (97), attaquer (94), fosse (88), mâle (80), shéol (78). Et les 10 premiers lemmes n'intervenant jamais dans l'Ancien Testament: apôtre (80), baptiser (67), synagogue (67), crucifier (49), spirituel (27), baptême (26), centurion (23), publicain (22), tribun (18), denier (16). Deux champs sémantiques différents, deux époques, deux contextes culturels mis en évidence par simple tri exclusif.

Les Colloques de l'Association Internationale Bible et Informatique ont présenté régulièrement des pistes de recherche dans ce sens. Des piles de listings au CIB-Maredsous donnent les statistiques d'usage de tous les livres de la Bible hébraïque et de la Bible grecque. Plusieurs tentatives de publications ont échoué devant le caractère limité du public intéressé.

En conclusion, il ne faut pas cacher que l'application de la statistique à la Bible vient seulement de commencer de par l'association possible du texte avec la puissance de calcul de l'ordinateur. Ce petit bilan permet de mesurer les efforts et essais déjà réalisés. Mais il révèle le manque d'un outil majeur encore à proposer dans le domaine biblique et pour lequel certains principes et d'autres expériences d'application des statistiques au domaine littéraire ont donné des bases solides.

Une chose est certaine, la statistique restera avant tout un outil heuristique permettant de faire surgir des phénomènes. Elle servira d'autant mieux le chercheur que celui-ci connaîtra avec précision, sous d'autres modalités de connaissances, l'objet de sa recherche.

R.-Ferdinand POSWICK, osb

