

Techniques et lois bibliométriques

Par : ARAB Abdelhamid

ABSTRACT

In order to seize objectively the phenomena of written communication, Bibliometrics resort to different techniques : Direct citation, counting, Bibliographic coupling, co-citation analysis and studies of obsolescence. Furthermore, the numerous quantitative studies were culminated by emergence of bibliometric laws namely, Bradford's law, Lotka's law and Zipf's law.

Introduction :

La substitution du concept "bibliométrie" à celui de "la statistique bibliographique" en 1969, fut un tournant dans l'histoire de la quantification de l'écrit.

Depuis, la bibliométrie cherchait à fixer ses limites en tant que discipline métrique ayant pour objet l'application des méthodes mathématiques en vue d'expliquer les phénomènes qui régissent l'écrit. Pour ce faire, la bibliométrie fait appel à des analyses quantitatives d'où l'émergence des techniques ainsi que des lois bibliométriques.

I- Les techniques bibliométriques :

Le mot "bibliométrie" est un terme générique qui décrit une série de techniques qui cherchent à étudier les phénomènes de la communication écrite. Les dernières années ont assisté à une multitude d'études bibliométriques, menées dans une perspective d'optimiser le processus de quantification de l'information écrite.

Ces études se présentaient sous diverses formes et dans des axes de recherche variés à savoir, études d'évaluation, description des caractéristiques des collections de documents, recherche des phénomènes sociologiques, économiques et épistémologiques de l'écrit etc...

Quant aux techniques bibliométriques utilisées, elles sont diverses, dont les plus usitées sont :

- La quantification directe de citations
- L'accouplement bibliographique
- L'analyse de co-citations
- L'étude d'obsolescence

I.I- La quantification directe de citations :

En 1927, dans leur article intitulé "College Libraries and Chemical Education", P.L.K. Gross

et E.M. Gross (1), semblent avoir donné naissance aux conceptions de dépouillement et d'analyse de références bibliographiques parues dans la littérature scientifique.

Depuis, plusieurs recherches ont été menées sur l'analyse de citations concernant les auteurs, les périodiques ou les articles de périodiques. Parmi ces études, nous citons celles de E. Garfield (2) intitulées : "The 300 Most cited Authors, 1961-1976, including co-Authors at last", publiée en 1977. Et "Most cited Articles of the 1960s" en 1980. Aussi celle de B.M. Gupta et M.P.K. Nagpal (3) intitulée : "Citation Analysis and its Implications", éditée en 1979.

La quantification de citations est une technique qui détermine le degré d'utilisation ou de citations, d'un document, d'un auteur, d'un journal etc... durant une période donnée.

Il est clair que les citations sont des indicateurs objectifs de l'utilité d'un document ou de la productivité d'un auteur. Par conséquent, un auteur ou un périodique fréquemment cité est plus utile ou plus productif que celui qui est moins cité.

La quantification de citations a connu un essor considérable avec l'invention des néologismes "facteur d'impact" et "index d'immédiateté".

Dans une oeuvre intitulée : "Citation indexing : its theory and applications in science, technology and humanities", publiée en 1979, E. Garfield (4) utilisa le concept "facteur d'impact" qu'il définit comme : "un ratio entre le nombre de fois où une revue est citée pendant une période donnée et le nombre total d'articles cités publiés par cette revue durant cette période".

La formule pour calculer le facteur d'impact est le suivant :

Nombre de citations concernant le périodique
Nombre d'articles cités, publiés par le périodique

Quant à l'index d'immédiateté, il se calcule en divisant le nombre de citations d'un article qu'une revue publie pendant une période donnée, par le nombre total d'articles publiés par cette revue.

L'index d'immédiateté qui sert pour calculer la fréquence d'utilisation, une base de sélection des articles, se calcule en fonction de la formule suivante :

Nombre de citations d'un article durant une année
Nombre total d'articles cités, publiés par la revue

I.2- Accouplement bibliographique :

Le concept "Accouplement bibliographique" fut suggéré par R.M. Fano (5) en 1956 dans son article "Documentation in action".

En 1963, M.M. Kessler (6) dans "Bibliographic coupling between scientific papers", testa et concrétisa l'application du principe de l'accouplement bibliographique. Kessler postula que des périodiques scientifiques sont réciproquement en relation significative lorsqu'ils ont plus d'une référence en commun, comme l'illustre la fig. ci dessous :

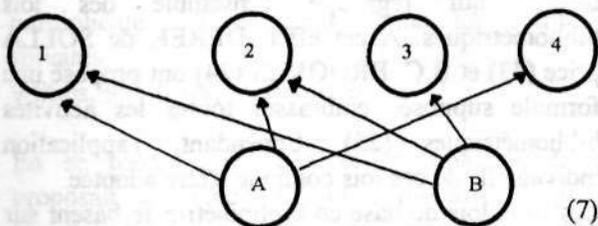


fig. n°1 : Accouplement bibliographique.
(1) (2) (3) (4) : documents cités
(A) (B) : documents sources

Les lettres encadrées représentent les sources d'articles, quant aux cercles numérotés, ils représentent les références citées.

D'après la fig. n°1, seuls les documents 1 et 2 sont cités par les documents sources A et B. Ce qui signifie que les documents A et B sont en relation réciproque, puisqu'ils ont deux références en commun.

Les plus ambitieuses applications bibliométriques ont été menées par N. Price et S. Schimovich (8)

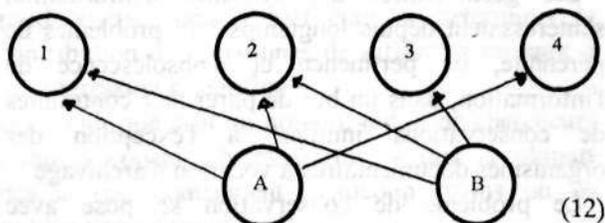
(9), dont la première fut publiée en 1968, et la deuxième en 1971.

Ces études ont été menées dans une perspective de dégager un système de classification automatisé. En effet, les auteurs ont été satisfaits des résultats obtenus. (10)

I.3- Analyse de co-citations :

Le concept "co-citation" fut découvert en 1973 par Marshakova et Smal, indépendamment l'un de l'autre. (11)

Le principe de co-citation se base sur l'existence d'un rapport entre deux références ou plus, citées ensemble.



(1) (2) (3) (4) : documents cités
(A) (B) : documents sources
fig. n°2 : illustration du principe co-citation.

A cite les documents 1, 2 et 3. Donc les documents 1 et 2, 1 et 3, 2 et 3 sont co-cités.

De même pour B qui cite les documents 1, 3 et 4 et que les documents 1 et 3, 1 et 4, 3 et 4 sont aussi co-cités.

Dans la fig. n°2, on remarque que les documents 1 et 3 sont doublement co-cités, car ils sont co-cités par A et B; par contre les documents 1 et 2, 2 et 3 ont une seule co-citation puisque chaque paire est citée seulement par A. Les documents 1 et 4, 3 et 4 aussi avec une co-citation, puisque chaque paire est citée seulement par B. Eu égard à la ressemblance des fig. n°1 et 2, il convient de différencier entre l'accouplement bibliographique et la co-citation.

A vrai dire la première technique agit sur la relation entre les documents sources, tandis que la seconde mesure la relation entre les documents cités.

La technique d'analyse de co-citations fut appliquée dans des cas multiples. Elle fut utilisée à titre d'exemple par E. Garfield dans une étude parue dans le n°20 (1979) de current contents, sous

le titre de : "Controversies over opiate receptor research typify problems facing awards committees". (13)

Elle fut utilisée aussi par H. Small dans son article publié par journal of the ASIS, n°24, 1973 sous le titre de : "Co-citation in the scientific littérature : A new Mesure of the Relation Ship Between Two documents". (14) Ainsi que par D.W. White et B.C. Griffith dans : "Authors as Markers of intellectual space : co-citation in studies of science, Technology and Society", (15) publié en 1982 dans le n°38 de journal of Documentation.

I.4- Etudes d'obsolescence :

Les gestionnaires des systèmes d'information s'intéressaient depuis longtemps aux problèmes de pérennité, de pertinence et d'obsolescence de l'information, dans un but de parer aux contraintes de conservations inutiles, à l'exception des organismes documentaires à vocation d'archivage.

Le problème de conservation se pose avec beaucoup plus d'acuité dans les bibliothèques spécialisées et à caractère scientifique, compte tenu du développement continu et des progrès sans cesse des sciences et des techniques, ce qui engendre une sorte de sénilité et désuétude.

Le développement d'une méthode scientifique, facile et sûre pour mesurer l'obsolescence serait une aide précieuse pour la prise de décision d'élagage des fonds d'un organisme documentaire.

Les études d'obsolescence prennent généralement l'une des deux formes :

I- L'observation de la circulation des documents des collections de la bibliothèque.

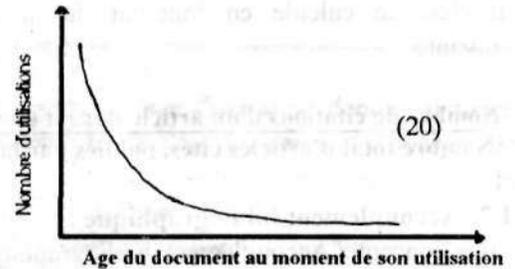
2- L'étude de citations. (16)

Les deux approches convergent finalement sur des résultats similaires. Une fois les documents rangés selon leurs âges, au moment de leur utilisation, les documents les plus récents se dégagent avec un taux d'utilisation plus élevés, tandis que les plus anciens reçoivent un taux beaucoup plus réduits.

L'obsolescence fut fréquemment exprimée dans le contexte de "half-life" de la littérature scientifique (17). Un terme emprunté à la physique nucléaire ; il fut employé par E.R. BURTON et R.W. KEBLER (18)(19) dans un article paru en 1960 dans le n°11 de "American documentation" intitulé : "The Half-

life of some scientific and technical littérature". Il s'agit d'une comparaison de citations relatives à un ou plusieurs documents et de décroissement de leur impact dans le temps. Les études d'obsolescence semblent avoir un impact considérable quant à la prise de décision concernant la sélection, la conservation et l'utilisation des documents dans les organismes documentaires.

Le principe d'obsolescence est ainsi illustré :



2. Les lois bibliométriques :

Les lois bibliométriques sont des expressions statistiques qui cherchent à décrire les écrits scientifiques par des moyens mathématiques. (21)

Si chaque loi s'applique à un phénomène spécifique de la communication écrite, l'aspect quantitatif des propriétés de l'écrit, demeure un facteur commun à toutes les lois bibliométriques (22).

Des efforts ont été déployés pour l'élaboration d'une théorie qui regroupe l'ensemble des lois bibliométriques. A cet effet, DEREK de SOLLA price (23) et B.C. BROOKES (24) ont proposé une formule supposée embrasser toutes les activités bibliométriques (25). Cependant, l'application individuelle de ces lois continue d'être adoptée.

Les trois lois de base en bibliométrie se basent sur différents types d'information ou de données bibliométriques (26). Si la loi de BRADFORD se base sur la distribution des publications, généralement périodiques, dans une discipline donnée, ou sur des articles d'une collection de périodiques ; celle de LOTKA se base sur la productivité des auteurs dans une discipline ou un domaine donné. Quant à la loi de ZIPF, elle est fondée sur le principe du moindre effort quant au choix des mots lors de la production intellectuelle.

2.1. La loi de BRADFORD :

S.C. BRADFORD fut un conservateur de la bibliothèque du musée des sciences de London (27).

Il commença lui aussi ses recherches par des nombres. Huit années après la publication de LOTKA sur la fréquence de distribution relative à la productivité scientifique, plus exactement en 1934, BRADFORD publia dans le n°137 de la revue "Engineering", un article intitulé : "Sources of information on specific subjects". (28)

Lorsqu'il procéda à un recensement des articles sur la lubrification et la géophysique appliquée, BRADFORD remarqua que la distribution de la plupart des articles publiés dans les divers périodiques est régie par un même modèle. Suite à cette remarque, il découvra sa fameuse loi de distribution élaborée dans le chapitre IX "Le chaos documentaire" de son livre "Documentation" publié en 1948 (29). Commencant par les titres des périodiques qui contenaient plus d'articles sur un sujet donné, il divisa l'ensemble des titres en 3 groupes (30) :

1°	9 titres	429 articles
2°	59 titres	499 articles
3°	258 titres	404 articles

Par le biais de cette division, BRADFORD avait découvert une régularité dans le calcul du nombre de titres de chacune des 3 catégories en utilisant le multiplicateur 5. Ce qui donne les résultats suivants:

9 titres, 9x5 titres, 9x5x5 titres

En se basant sur cette observation, Bradford proposait sa loi de distribution qui affirme que : "Lorsque les périodiques scientifiques sont arrangés par ordre décroissant de leur productivité sur un sujet, les articles en question se trouvent répartis entre un noyau de périodiques plus particulièrement consacrés au sujet et autres zones ou divisions, contenant approximativement le même nombre d'articles". (31)

Quant aux champs d'application de cette loi, elle fut appliquée en astronomie (32) (33), en sciences de l'information (34), en biologie marine (35) (36), en agriculture tropicale (37) (38) et en médecine vétérinaire. (39) (40) (41)

2.2- La loi de Lotka :

Qui est Lotka?

A.J. Lotka naquit en Pologne (42) de parents missionnaires américains, il vivait longtemps en Grande Bretagne, puis arriva aux USA en 1902. Durant l'année 1909 il travailla pour U.S. Patent office. C'était à ce moment qu'il avait rencontré E.W. Hulme.

Lotka, chimiste et mathématicien occupa le poste de président de American statistical society. Il travailla plusieurs années à Metropolitan life Insurance Company où il fut surnommé "père de l'analyse démographique". (43) En 1926, il publia dans le vol.16, n°12 de journal of the Washington Academy of Sciences, un article intitulé : "statistics : the frequency distribution of scientific productivity", dont l'objet était de déterminer la contribution des personnes de différents niveaux à la progression scientifique.

Il conclut que seul un nombre réduit de chercheurs publient considérablement, tandis que la plupart des autres, contribuent beaucoup moins ou ne publient pas du tout. (44)

Plus tard l'extrapolation de Lotka fut honorée, en devenant une loi bibliométrique qui porte sur la productivité scientifique.

A cet effet, Lotka propose une formule ayant pour objet de mesurer la productivité scientifique en fonction du principe du carré inversé (45), c'est à dire, si 100 auteurs contribuent avec 1 article chacun, 25 auteurs contribuent avec 2 articles, 11 auteurs avec 3 articles et 6 auteurs avec 4 articles. Et ce, conformément à la formule suivante : $1/n^2$.

Quoique la loi de Lotka fut basée sur l'étude de la littérature en chimie et en physique, elle fut éprouvée aussi dans les autres domaines.

Elle fut appliquée dans les sciences humaines par L. Murphy (46), dans la bibliothéconomie par E.A. Schorr (47), dans la géographie par W.O. Aiyepoku (48), en anthropologie par E.A. Rogge (49) et en informatique par T. Radhukrishmann et R. Kernizan. (50)

2.3- La loi de Zipf :

Georges Kingsley Zipf professeur de philosophie (51) à l'université de Harvard, fut un des précurseurs pour le traitement statistique des phénomènes linguistiques. Il fut considéré comme un des spécialistes de la bibliométrie. Dans un ouvrage de 1929 (Relative frequency as a

determinant of phonetic change...) et dans d'autres études, Zipf a établi une série de règles générales de fréquence. Il a montré entre autres que la "complexité des phonèmes est inversement proportionnelle à leur fréquence..." (52)

Après sa théorie générale du langage parue dans "Selected studies of the principale of relative frequency" en 1929 (53), il publia en 1945 (54) son oeuvre capitale intitulée : "Human behavior and the principale of least effort", dans laquelle il établit le principe dit du moindre effort. Dans ce principe, Zipf a montré qu'il est plus facile de choisir et d'utiliser des mots familiers, et inversement dans le cas contraire. Ce qui fait que les mots faciles sont beaucoup plus forgés, autrement dit, ils ont une fréquence d'utilisation plus élevée.

Pour illustration, Zipf étudia un corpus de 29 899 mots qu'il arrangea dans un ordre décroissant de fréquences d'occurrences. Il assigna ensuite un rang (r) à chaque mot. Il multiplia enfin la valeur numérique de chaque rang (r) par sa fréquence (f) pour obtenir le produit (c).

Par conséquent, la formule de la loi de Zipf est la suivante :

$$r.f = c \quad (55)$$

Dans le domaine des bibliothèques et celui de science de l'information, les principes de la loi de Zipf furent appliqués dans l'indexation automatique dont les pionniers furent H.P. Luhn (56) et P.B. Baxendale (57) qui utilisèrent l'ordinateur dans la sélection des mots ou groupes de mots qui reviennent fréquemment dans un texte ou dans un document. Ces mots ou groupes de mots témoignent en réalité du contenu des documents. Autrement dit, ils sont des descripteurs qui servent à l'indexation.

I Rang (r)	II Fréquences (f)	III Produit de I et II (rxf = c)
10	2 653	26 530
20	1 311	26 220
30	926	27 780
40	717	28 680
50	556	27 800
100	265	26 500
200	133	26 600
300	84	25 200
400	62	24 800
500	50	25 000
1000	26	26 000
2000	12	24 000
3000	8	24 000
4000	6	24 000
5000	5	25 000
10000	2	20 000
20000	1	20 000
29 899	1	29 899

Illustration de la distribution des mots selon la loi de Zipf.

Source : Zipf, G.K. *Human Behavior ans the Principale of least Effort : An Introduction to Human Ecology.*- New York : Hafner, 1969.- p.24.

CONCLUSION :

A l'instar des autres disciplines métriques, la bibliométrie s'intéresse à la mesure qui, comme le disait autrefois Paul Otlet, est une forme supérieure que prend la connaissance.

C'est grâce à la mesure et seulement à la mesure, que les phénomènes de l'écrit peuvent être saisis objectivement.

En effet, les résultats obtenus par des études bibliométriques

sont plus significatifs, plus raisonnables et plus objectifs.

Car ils émanent de la logique mathématique et de la rigueur statistique.

Cependant, les phénomènes de l'écrit sont aussi complexes que le sens commun ne l'en pense. Parce que quelques soient les efforts déployés pour agir quantitativement sur les phénomènes de l'écrit, la fixation des limites des lois qui les régissent, demeure encore dans la controverse.

Bibliographie :

- (1) Ikpaahindi, Linus .- An overview of Bibliometrics : Its Measurements, Laws and Their Applications. In. "Libri.", vol. 35, n° 2, 1985, p.166.
- (2) ibid.
- (3) Ibid.
- (4) Ibid.
- (5) Ibid.
- (6) Ibid.
- (7) Ibid.
- (8) Price, N. ; Schiminovich, S.- A Clustering experiment : first step toward a computer generated calssification scheme. Information storage retrieval. 4, 1968 : 271-280.
- (9) Schiminovich, S.- Automatic calssification and retrieval of documents by means of a bibliographic pattern discovery algorithm. Information storage retrieval. 6, 1971, 417-435.
- (10) Ikpaahindi, Linus, op. cit. p. 167.
- (11) Ibid., p. 168.
- (12) Ibid.

- (13) Ibid. p. 169.
- (14) Ibid.
- (15) Ibid.
- (16) Wallace, Danny P.- A Solution in Search of a problem : *Bibliometrics and Libraries*.
In : "Library Journal", vol. 112 B, n° 8, 1987, p. 44.
- (17) Ibid.
- (18) Ibid.
- (19) Peritz, Bluma C.- On the careers of **Terminologies: the case of Bibliometrics**.
In : "Libri", vol. 34, n° 3, p. 235.
- (20) Wallace, Danny P., op. cit., p. 45.
- (21) Ikpaahindi, Linus, op. cit., p. 169.
- (22) Schrader, Alvin M.- Teaching Bibliometrics.
In : "Library Trends", vol. 30, n° 1, 1981, p. 15.
- (23) Price, Derek de Solla.- A general theory of bibliometrics and other cumulative advantage processes. *Journal of the ASIS*, n° 27, 1976, pp. 292-306.
- (24) Brookes, B.C.- Theory of Bradford Law. *Journal of the Documentation*, 33, 1977, pp. 180-209.
- (25) Ikpaahindi, Linus, op. cit., p. 169.
- (26) Narin, Francis; Moll, Joy K.- Bibliometrics.
In : "Ann. Rev. Infor. Sci. Technol.", vol. 12, n° 8, 1977, p. 44.
- (27) Broadus, Robert N.- Early approaches to Bibliometrics.
In : "Journal of the American Society for Information Science", vol. 38, n° 2, 1987, p. 128.
- (28) Ibid.
- (29) Ibid.
- (30) Ikpaahindi, Linus, op. cit., p. 170.
- (31) Ibid.
- (32) Meadows, A.J.- The Citation characteristics of astronomical literature. *Journal of documentation*, n° 23, 1967, pp. 28-33.
- (33) Gross, O.V.- Citation characteristics of astronomical literature. *Journal of documentation*, n° 25, 1969, pp. 344-347.
- (34) Donahue, J.C.- A Bibliometrics analysis of certain information science literature. *Journal of the ASIS*, n° 23, 1972, pp. 313-317.
- (35) Freemand, C.- Citation analysis of the literature of biology. *Australian library journal*, n° 23, 1974, pp. 67-71.
- (36) Freemand, C.- Bradford bibliographics and the literature of marine science. *Australian Academic and research libraries*, n° 5, 1974, pp. 65-71.
- (37) Lawani, S.M.- Periodical literature of tropical and subtropical agriculture. *Unesco bull. lib.*, n° 6, 1972, pp. 88-93.
- (38) Lawani, S.M.- Bradford's law and the literature of agriculture. *International library review*, n° 5, 1973, pp. 341-350
- (39) Garfield, E.- Journal citation studies 35. *Veterinary Journals : what they cite and vice versa*. *Current contents*, n° 13, 1982, pp. 5-13.
- (40) Ikpaahindi, L.N.- Journal use by Nigerian Veterinary practitioners : the national veterinary research Institute. *Q Bull. IAALD*, n° 27, 1982, pp. 16-21.
- (41) Houston, W.- The Application of Bibliometrics to Veterinary science primary Literature. *Q Bull. IAALD*, n° 28, 1983, pp. 6-13.
- (42) Broadus, Robert N., op. cit., p. 128.
- (43) Ibid.
- (44) Ibid.
- (45) Ikpaahindi, Linus, op. cit., p. 171.
- (46) Murphy, L.- Lotka's and librarianship. *Journal of the ASIS*, n° 24, 1973, pp. 461-462.
- (47) Schorr, E.A.- Lotka's law and librarianship. *Journal of the ASIS*, n° 26, 1975, pp. 189-190.
- (48) Aiyepetu, W.C.- The Productivity of Geographical Authors : A case study from Nigeria. *Journal of documentation*, n° 32, 1976, pp. 105-117.
- (49) Rogge, E.A.- A Look at academic anthropology. *American anthropologist*, n° 78, 1976, p.835.
- (50) Radhukrishnan, T. and Kernizan, R.- Lotka's law and computer science literature. *Journal of the ASIS*, n° 30, 1979, pp. 51-54.
- (51) Ikpaahindi, Linus, op. cit., p. 172.
- (52) Broadus, Robert N., op. cit., p. 128.
- (53) Ibid.
- (54) Estivals, Robert.- La bibliométrie Bibliographique.- Lille III : service de reproduction des thèses, 1971.
In : *Lettres : Paris I* : 1971.
- (55) Ikpaahindi, Linus, op. cit., p. 172.