CHAPTER - III
RESEARCH METHODOLOGY

3.1 INTRODUCTION

The methodological strategy is the arrangement of conditions for collection and analysis of data in a manner that aims to combine relevance to the research purpose. It is the conceptual structure within which research is conducted. The methodological strategy contains

a) a clear statement of the research problem
b) Methods to be used in processing and analysing data.
c) procedures and techniques to be used for gathering information
d) the objective of the problem to be studied
e) the population to be studied and

3.2 STATEMENT OF THE PROBLEM

Problem formulation simply implies the formulation of the objectives of proposal or study, that is, the presentation of a very specific statement of what the researcher wants to do. The first step in the problem formulation is to specify the problem, creatively and objectively; because the formulated problem should not be misunderstood. The first essential condition for the formulation of a problem is a systematic immersion in the subject matter through first hand observation. Once the problem is identified and selected for research, the researcher should find out various sources to collect information about the problem.

The present study aims at analysing the research publications of in “A bibliometric study on journal of Tamil studies published for 1972 to 2012”. In academic work publication is the chief means of communicating research,
a primary means of recognition and reward and hence a social process. Therefore it is through the publication the scientists receive professional recognition and esteem as well as promotion, advancement and funding for future research. Publication is so central to productivity in research that the work becomes ‘a work’ only when it takes a conventional, physical (that is published) form, from which can be received, assessed and acknowledged by the scientific community. Hence publication is a social norm in a public sense and serves as a tool for the betterment of the individuals after publication only. It can be called a research and can be fixed or judged and acknowledged by the Tamil scientists in the Tamil society.

From the above discussion it could be seen vividly that bibliometric analysis is an important tool in analysing any discipline. By keeping this view in mind, the researcher intends to undertake the study on research publications by the Tamil scientists in ‘A bibliometric study on journal of Tamil studies published for 1972 to 2012’, The present study focuses attention on bibliometric study in terms of the pattern of publication, authorship, growth rate, area of research concentration and author productivity.

3.3 THE OBJECTIVES OF THE STUDY

The following are the important objectives of the study:

- To ascertain the authorship and collaboration pattern of scientists.
- To determine the degree of collaboration.
- To find out acknowledgement & appendix in the articles.
- To find out distribution of authors by gender
- To find out most prolific authors, most productive institution and author productivity.
- To find out the contribution of articles by pages
- To find out the distribution of authors by its membership.
To find out year-wise publication of articles and references.

3.4 FORMULATION OF HYPOTHESES

The researcher has formulated the following testable hypotheses; and from the statistical analysis of the collected data their validity is tested, and they are either validated or falsified.

1. There is a significant level of variation in paper publication in publications of Tamil.
2. There has been an increasing trend in collaborative author in recent years.
3. The implications of Lotka’s law are related to author productivity in paper publications.
4. Journal of Tamil Study authors have a tendency to cite latest literature relevant to their field.

3.5 DATA COLLECTION

The article publications in the journal ‘Tamil Studies’ by the Tamil scientists in Tamil were taken as source for the present study. The papers published from 1972 to 2012 by the Tamil scientists in Tamil. The bibliographical details of publication were entered in the catalogue cards. Finally the cards were arranged in different ways with a view to identify the Tamil scientists of Tamil.

3.6 METHODOLOGY

The data has been compiled from journal articles for each cited reference; the following data has been noted:

a) Total number of authors
b) Type of document
3.7 STATISTICAL TOOLS

After the collection of the required data pertaining to the study the collected data were tabulated and subject to statistical analysis.

3.8 THE SIGNIFICANCE OF THE STUDY

The study has been conducted with the journal articles of Journal of Tamil Studies. The study focuses on publishing trend, impact factor, and authorship pattern, types of articles, institutional collaborations of authors, affiliated institutions of authors, countries of contributing authors, keyword analysis and referencing pattern.

3.9 UNIVERSE AND SAMPLE

Collection of journals in Tamil studies in Tamil subject is taken as Universe and the journal ‘Tamil studies’ for the period of 1972-2012 is taken as sample.

3.10 SCOPE

The study covers 896 articles in volumes from 1 to 82 of the journal ‘Tamil Studies’ published in the years from 1972-2012. This journal allows any number of citations to be included in the reference section of the article.

3.11 FRAMEWORK OF ANALYSIS

The collected data were presented in the simple tables and these tables were systematically analysed with the help of statistical techniques.
3.12 STYLE OF BIBLIOGRAPHY

The prescriptions provided in the ‘Chicago Style Manual’ have been followed with small variations in general, maintaining the uniformity throughout except the rendering of Indic names in particular. Instead of reversing the Indic names, the natural sequence of occurrence has been taken in.

3.13 STATISTICAL TOOLS AND TECHNIQUES USED

3.13.1 RELATIVE GROWTH RATE (RGR)

The researcher has applied the relative growth rate and doubling time model by Mahapatra to examine the growth rate of papers published and weightage has been given to the scientists of engineering science.

The relative growth rate is increased in the number of publications or pages per unit of time. A specified period of interval can be calculated from the following equations.

\[ R (1-2) = \frac{W_2-W_1}{T_2-T_1} \]

Where \( R (1-2) \) is mean relative growth rate over the specified period of interval.

\[ W_1 = \log W_1: \text{(Natural log of initial number of Publications/pages)} \]

\[ W_2 = \log W_2: \text{(Natural log of final number of Publications/pages)} \]
T2 – T1 = the unit difference between the initial time and final time.

The relative growth rate for both publications and pages can be calculated separately. Therefore,

\[ R (a) = \text{Relative growth rate per unit of publications, per unit of time (Year);} \]

\[ R (b) = \text{Relative growth rate per unit of pages, per unit of time (year).} \]

### 3.13.2 Doubling Time

From the calculation, it is found that there is a direct equivalence existing between the relative growth rates and doubling time. If the number of publication/pages of a subject double during a given period, then the difference between the logarithm of the numbers at the beginning and at the end of the period must be number 2. If one uses a natural logarithm, this difference has a value of 0.693. The corresponding doubling time for publications and pages can be calculated by using the following formula.

\[ \text{Doubling Time (Dt)} = \frac{0.693}{R} \]

Therefore, Doubling time for Publications Dt (a) = 0.693

\[ \frac{0.693}{R} \]
Doubling time for Pages $D_t$ $(b)$ = \(0.693\)

---

3.13.3 DEGREE OF COLLABORATION

To examine the extent of research collaboration of scientists of engineering science, Subramaniam’s formulas are adopted for present study.

\[
C = \frac{N_m}{N_m + N_s}
\]

$C$ = Degree of collaboration of scientists

$N_m$ = Number of multiple authored Papers

$N_s$ = Number of single authored papers

3.13.4 PRIORITY INDEX AND ACTIVITY INDEX

Priority Index (PI) has been calculated to properly normalize the size of a Journal Article and the size of the subject field so that cross national comparisons can be done for these “Frontier” areas of research on the Journal of Current Science.

Priority Index is computed by the following formula:

\[
\frac{N_{ij}}{N_{io}} \times \frac{N_{oj}}{N_{oo}} \times 100
\]

where,

$N_{ij}$ = the number of publications of Journal $i$ in subfield $j$

$N_{io}$ = the number of publications of Journal $i$ is in all
subfields of the major fields

\[ N_{oj} = \] the number of publications of all Journals viz.,
the total world output in subfield

\[ N_{oo} = \] the number of publications in all sub fields of
those major fields

This index is identical to AI proposed (by Frame 1977) and
subsequently used among others by Schubert and Braun (Schubert 1986 and
Carpenter et al. 1988, Carpenter 1988). The value of PI = 100 indicates that
research priority of a Journal for a given subfield corresponds precisely to
the average of all Journals.

PI = 100 indicates average priority,
PI > 100 indicates higher than average priority and
PI < 100 lower than average priority.

It should, however, be kept in mind that (by virtue of definition of PI), no
Journal can have high or low priority in all sub fields. From the values of PI,
we can compare

(1) The priorities of a given Journal to different subfields in a given
time span
(2) The priorities of different Journals to a given subfield in a given
time span
(3) The priority to a given subfield in different time spans.

3.13.5 BIBLIOMETRIC LAWS

Van Raan (1993) stated that “Bibliometric analyses performed at the
macro level (e.g., a whole country) yield at best, general assessments of
fields as a whole, like how good a country's performance is in Physics,
Chemistry, Psychology or Immunology, without a breakdown to the individual research groups or programs.” Bibliometric laws come handy for such studies covering a field as a whole or a nation’s output. The three fundamental laws which laid the formation of Bibliometrics are

1) Bradford's law of scattering of scientific papers
2) Zipf's law of word occurrence
3) Lotka's inverse square law of scientific productivity

3.13.6 BRADFORD'S LAW OF SCATTERING

It is quite interesting to note that S.C. Bradford was a librarian also, besides being a chemist. S. C. Bradford, chemist and librarian at the Science Museum of South Kensington (London), showed great interest in documentation. Bradford (1934) published a short article which, as R. Bailon-Moreno, et al., (2005) expressed, could be considered as one of the beginnings of bibliometric studies, in which he compiled a bibliography on Lubrication, from 1931 to 1933 and Applied Geophysics from 1928 to 1931.

Bradford (1977) found out a pattern as to how literature in a subject was distributed in journals and expressed that, “If scientific journals are arranged in order of decreasing productivity of articles on a given subject, they may be divided into a nucleus of periodicals more particularly devoted to the subject and several groups or zones containing the same number of articles as the nucleus, when the numbers of periodicals in the nucleus and succeeding zones will be as 1 : n : n^2.”

3.13.7 BRADFORD'S FORMULA

R (n) = N Log n/s (1<=n<=N)
where $R(n)$ = Cumulative total of items contributed by the sources of rank 1 to n.

$N$ = the total number of contributing sources.

$S$ = constant characteristic of the literature.

Therefore,

$$R(N) = N \log \frac{N}{S}$$

is the total number of items contributed by N sources.

The statistical regularity pointed out by Bradford's law provides an objective means of determining zones of relative richness or value to a given kind of literature.

A simple model to account for Bradford's law was proposed by Naranan (1970). The most significant first step in the model was to recognize the Bradford's law as equivalent to a simple power law distribution of articles in journals. Specifically $J(p)$ the number of journals carrying exactly $p$ articles is of the form of $J(p) = Kp$ ($K$ being constant and $= 270$) to explain the power law relation. Naranan was inspired by a model proposed by Fermi in 1949 in Cosmic Ray Astrophysics. Naranan investigated statistical laws governing distribution of number of citations to articles, citation to journals etc. and found them also to be power laws similar to $J(p) = Kp$.

3.13.8 ZIPF’S LAW

Professor George Kingsley Zipf (1902-1950) belongs to the field of linguistics and was at Harvard University. Zipf's law is an empirical law set up for Linguistics in the early 1930s by the Harvard Linguistic. This heralded the power law $q(n) \propto 1/n$, now commonly known as Zipf's law. According to the Zipf’s law “the frequency $q$ of occurrences of some event (such as of a word in a text sample) is inversely proportional to its rank $n$.”
Actually, G. K. Zipf (1935, 1949) originally described a broad statistical regularity of natural languages and proposed two complementary empirical laws of word frequencies, as highlighted by Landini (2000), namely:

1. “The rank-frequency law. The procedure to estimate this relation is very simple. The words in a text are sorted by decreasing frequency and a rank number is assigned to each word. For words with the same frequency, the sub-sorting and ranking is arbitrary. The plot of log (frequency) versus log (rank) approximates a straight line of slope.”

2. “The number-frequency law. … The plot of log (frequency) versus log (number of words with the same frequency) approximates a straight line of slope 0.5. While the rank-frequency law tends to occur for the high frequency words (although not necessarily for the first few ranking positions), the number-frequency law is observed for the low frequency words.”

A look into the Zipf’s rank-frequency law reveals as expressed by the more general power-law function $q(n) = c n^{-b}$ is with the scaling constant $c = q(1)$ and the exponent $b$ close to unity ($b = 1$ in the original Zipf’s expression).

Zipf’s law can be effectively used in the generation of semiautomatic or automatic indexes useful for an information retrieval system. Its use has increased tremendously with the emergence of natural language indexing of textual matter especially in electronic form. Several studies aimed at finding out the pattern of frequency distribution of descriptors of a thesaurus and the distributions of indexing terms are available. Zipf’s law provides a measure of the richness in vocabulary of an author. This technique can be used for deciding the correct authorship of disputed works. The frequencies of occurrence of favorite words are identified to decide the author conclusively.
The law is also used for identifying words of higher frequency in a foreign language in the context of teaching that language to a target group. Such words of higher frequency identified are taught first in the instructional programmes for foreign language students.

3.13.9 LOTKA’S LAW

Lotka's (1926) proposition led to a whole gamut of studies on scientific productivity. Such productivity studies gained momentum in the post-second world war period. This, in fact, culminated in the rise of a new discipline called Scientometrics. It is defined as the 'study of the measurement of scientific and technological progress'. It provides an understanding of the structure of scientific activity, the disciplines being researched, the organisations involved, the strength and deficiency in the scientific groups and their communication channels and at different levels of aggregation. It follows a trajectory of econometrics in the use of quantitative data, concepts and models and extensive use of mathematical and statistical techniques of modeling and data analysis.

Scientific productivity studies have been made from different perspectives. Impact of social change on scientific productivity, relationship of publication output on scientific recognition, identification of elites in different disciplines and, occurrence of discoveries in different cultures are some of the approaches made in this line.

Price (1965) who had traced the development of science in Babylon and plotted the growth of big science from little science had observed that Lotka's law applied equally well to the productivity of scientists in the 17th as well as in the 20th century. This meant that a majority of publications emanated from a handful of people.
3.13.10 PRICE’S FUNDAMENTAL LAW OF SCIENCE

Price’s celebrated lectures on “Little Science and Big Science” reviewed some earlier works by Francis Galton; J.M.Cattell; and A.J.Lotka and presented a notable “feeling that most of the great scientists are still with us and that the greater part of scientific work has been produced within living memory, within the span of the present generation of scientists”. He considered an exponential time trend as the appropriate model to fit for data on number of scientists. He calls this principle of exponential growth as the “fundamental law of any analysis of science”.

Let \( y_t \) = number of scientists during a period \( t \) (\( t \) may be just 1 year or a span of say, 30 or 45, years).

\[
y_t = e^{a' + b't}
\]

\[
\log y_t = a' + b't
\]

Let \( a' = \log a \) and \( b' = \log b \).

Then \( \log y_t = \log a + t \log b \)

Or \( y_t = a * b^t \) --------------------------------2

In (2) if \( b > 1 \) the exponential curve is rising over time (+ve growth) and if \( b < 1 \), curve is falling down (-ve growth). (2) may also be written as

\[
y_t = y_0 * b^t \quad \text{(Since } t=0, y_0 = a \text{= number of scientists in the beginning)}.
\]

Or \( y_t = y_{t-1} * b \)

Since \( b > 1 \), obviously the number of scientists during any period \( t \) is greater than those existing during any particular period in the past.

3.13.11 PRICE’S 80 X 20 LAW

It is observed in library and information field that:
Most of the documents are hardly circulated/used/cited and very few books are frequently circulated/used/cited.
Most of the authors publish very few articles and very few authors publish more frequently.

This phenomenon is explained fairly by Price in his 80-20 rule. That is 80 per cent of the documents contribute to 20 per cent of the total circulations/citations received. In a collection of publications, 20 per cent of the documents contribute to 80 per cent of the total circulations/citations received. The 20 per cent of such documents are called "core documents/core collection". Similarly it is observed that 20 per cent of the given authors (at any given time are called "Core authors") contribute to 80 per cent of the total literature output.

3.13.12 LIMITATIONS OF THE STUDY

Records for this study have been drawn exclusively from Indian Academy of Science which follows its own standards for the inclusion or omission of items or periodicals in its coverage. The contributions of authors with an address pertained to India or its component cities only have been counted into the study. Though Indian authors are there elsewhere outside the geographical boundaries of India and their contributions have not been included into the present study.
REFERENCES


