BHARATI INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH & DEVELOPMENT (BIJMRD)

(Open Access Peer-Reviewed International journal)

DOI Link :: https://doi-ds.org/doilink/06.2024-25472452/BIJMRD/Vol -2 / 4/A30/Kalipada Jana1 Bidya

Available Online: www.bijmrd.com|BIJMRD Volume: 2| Issue: 4| May 2024| e-ISSN: 2584-1890

Keyword Analysis of a Subject: A Study from a Linguistic Viewpoint

Kalipada Jana[1] & Bidyarthi Dutta[2]

1. Research Scholar in Vidyasagar University
2. Associate Professor, & Dept.of Library and Information Science, Librarian, Basanti Devi College, Vidyasagar University

Email:-librarian@basantidevicollege.edu.in & Email:- bidyarthi.bhaswati@gmail.com

Abstract:

Keywords are used in our everyday life. The link between user-generated content and what they are seeking for is created via keywords. In order to adequately describe a subject's content, keywords are crucial. An interdisciplinary field, linguistics draws from the study of languages, particularly English, as well as disciplines like psychology, sociology, cognitive science, and so forth. This study's primary goal is to ascertain the process by which keywords in a certain subject are generated. There are numerous methods in linguistics for coining new words. The classification of the linguistic analysis's keyword components is explained in this study. The study of onomastic words (OW), endocentric words (EW), exocentric words (XW), copulative words (PW), Subject-Specific Compound Words (SSW), and Domain-Specific Compound Words (DSW) is highlighted in this work from the linguistic as well as semantic viewpoint.

Keywords: *Linguistics; Onomastic Words (OW); Subject-Specific Compound Words (SSW); Domain-Specific Compound Words (DSW).*

Introduction:

LIS is an interdisciplinary or multidisciplinary area in nature. It deals with collecting, processing, organizing, preserving, and disseminating different types of information resources to maximize the utilization of those resources. Keywords are the bridge between what people are searching for and the content provided to fill that need.

Since 1971, the growth of online search systems for large databases has been a striking development in information retrieval. In the Information Retrieval system keywords play an important role. Keywords are terms extracted from documents or single sentences that generate sense when clustered in context. To describe the same document, different indexers assign different keywords from the controlled vocabulary, like LCSH, DDC, Thesaurus, etc. So the keyword is a subject descriptor. In different documents, the same keywords might be present, but the structure and position of keywords build different meanings, which highlights keyword importance (Pawar, Sanket S., et al., 2016). Keywords are daily used terms in two or more groups to represent information from phrases or short sentences and require concepts generated by keywords.

Related Works:

Linguistics is an interdisciplinary area that draws from the study of languages, including English, and fields such as psychology, sociology, cognitive science, computer science, and anthropology (Bolin, 2017). Library and Information Science (LIS) is also interdisciplinary and can be studied using techniques from the humanities, social science, and science. The many theories and methods of linguistic research can be useful and have significant explanatory power for LIS. Warner (1994) points to the problems for LIS if the meaning of words must be partially inferred from a socio-linguistic context. It is clear that simply matching query words to index words, no matter how sophisticated a partial match and ranking algorithm one has, will always have a low precision because the semantics are not equally well-defined. Overload, noise, lack of precision, and ignorance of recall are modern document retrieval problems (Søren, 2006). Palmer (1981), states, "Dictionaries appear to be concerned with stating the meaning of words and it is, therefore, reasonable to assume that the word is one of the basic units of semantics". He also stated that, since meaning is a part of language, semantics is a part of linguistics. Words are not independent entities but are mutually inter-related among themselves through meaning. So in linguistics, meaning is a core entity. In a study, the subject has been logically interpreted as a collection of well-defined and semantically related sets of words (Dutta &Dutta, 2013). Hartley (2003) described the usefulness of keywords in science journals. Voorbiz (1998) conducted a comparative study between title keywords and subject descriptors in the humanities and social sciences. Strader (2011) examined the overlap between author-assigned keywords and cataloger-assigned Library of Congress Subject Headings (LCSH). Hurt (2010) examined the differences between author keywords and automatically generated keywords for polymer science literature. O'Connor (2010) observed that many indexes had much lower rates of match between title keywords and subject headings. Gbur (1995) gave suitable guidelines for the selection of optimal keywords in the subject field of statistics. The correlation between derived keywords from titles in bibliographic records and LCSH terms was studied by Frost (1989). Frank (1999) explored the automatic keyword extraction methods in specific subject domains. Hurth (2003) discussed the automatic keyword or keyphrase extraction process from a linguistic point of view. Cleverdon (1967) showed that each indexing system was made up of a basic vocabulary system.

Objectives of the Study:

Keywords denote the subject area of any discipline. Research is an ever-growing process. New terms, i.e., keywords may be raised in any research area. These new keywords may not be included in the existing subject access tools like LCSH, DDC, and Thesaurus due to the lack of updating or extension of the said tools.

The objectives of the study are:

- To find out the keywords for a specific subject.
- To show the various components of keywords from a linguistic viewpoint.
- To interpret the compounding of Keywords in the area of study carried out.

Methodology and Limitation:

The main objective of this paper is to identify the significant keywords of a specific subject area and categorize those keywords from a linguistic viewpoint. To carry out the study, 2445 journal articles published from the years 2004 to 2021 have been retrieved from the Web of Science database by using the search term "Hawking Radiation". "Hawking Radiation," indicates a specific subject domain under the broad area of astrophysics. As the contents of the research are reflected through the assigned keywords, the keywords have been selected from the abstracts and contents of the retrieved articles. Then 16480 assigned

keywords were culled out from these 2445 articles. Only 7182 of the 16480 allocated keywords are unique due to the recurrence of the same phrases in a particular year (Table 1).

One or more words combine to make a keyword. These words may be called "**Components of the Keyword**". Say, "TUNNELING OF VECTOR PARTICLE". It is a keyword. Now it has 4 components i.e. "TUNNELING", "OF", "VECTOR", and "PARTICLE". For another example say," AdS BLACK HOLE". In this keyword, there are 3 components i.e. "AdS", "BLACK", and "HOLE". All the distinct keywords are separated here in this way.

From a linguistic viewpoint, "TUNNELING", "OF", "VECTOR", "PARTICLE", "BLACK", and "HOLE" are Linguistic Words (LW), and "AdS" is an Abbreviation (AA). Again, the Linguistic Words (LW) maybe categorized as Semantic words (SW), Onomastic Words (OW), and Form Words (FW). Again, the Semantic words (SW) may be categorized as Root Words (RW), Stem Words (TW), Compound Words (CW), and Loan Words (LW). The Compound Words (CW) are categorized into Endocentric Words (EW), Exocentric Words (XW), Copulative Words (PW), General Words (GW), Subject-Specific Words (SSW), and Domain-Specific Words (DSW). The components of keywords by different levels are shown in Figure 1.

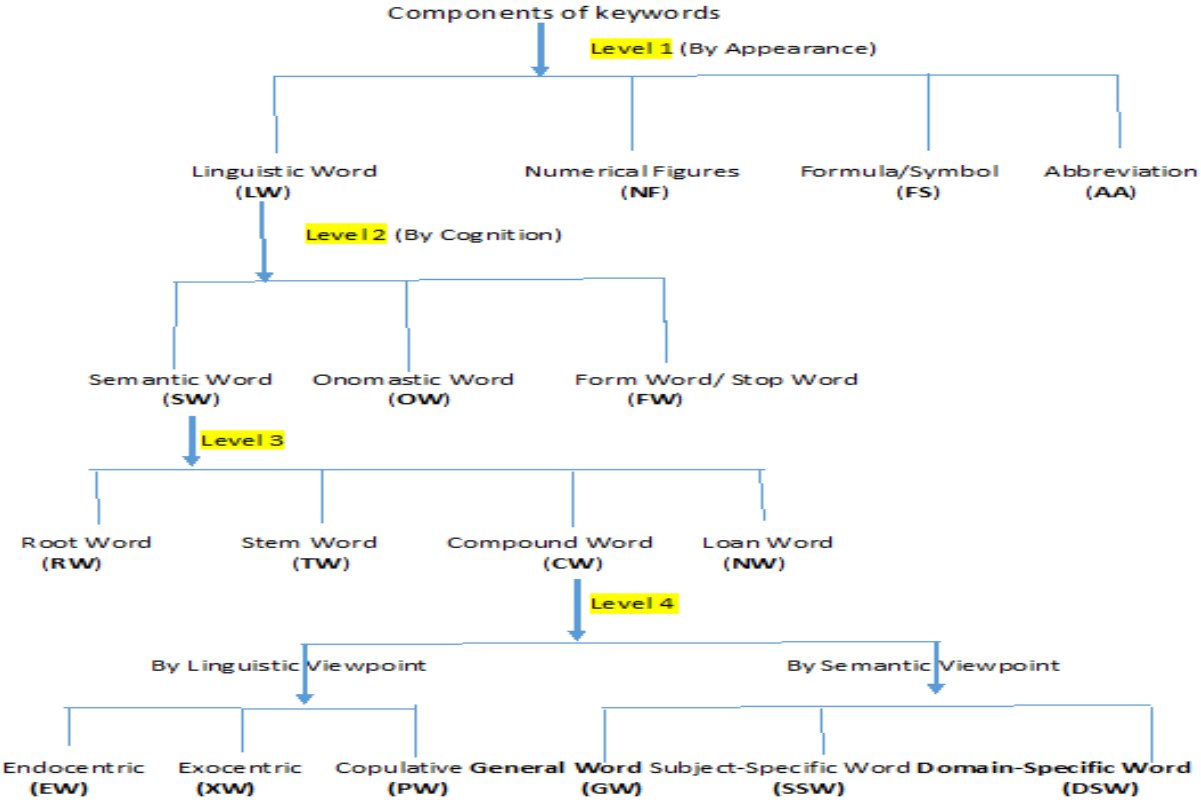

Figure- 1

Year-wise occurrences of the components of Keywords have been calculated. For example, components of keywords for 3 conjugative years' i.e., 2009, 2010, and 2011 have been shown in Table 2, Table 3, and Table 4 respectively along with diagrams 1, 2 and 3.

Again, from a linguistic viewpoint, Compound words have been categorized into three groups viz., Endocentric (EW), Exocentric (XW), and Copulative (PW).

Endocentric Compound (EW): If a compound contains a semantic head, it is called an endocentric compound.

Suppose, “AB” = “A” + B, where “A” and “B” are the two different words to make a new word “AB”. The meaning of ‘‘AB’’ depends on either ‘‘A’’ or “B’’.

Say, **Football=Foot+Ball**, here **Football** denotes a type of **Ball**, not **Foot.** Therefore**, Football** is an **Endocentric word.**

Exocentric Compound (XW): If a compound word contains neither its semantic heads, it is called an Exocentric compound.

Suppose, “XY”= “x” +”Y”, where “X” and “Y” are the two different words to make a new word “XY”. The meaning of “XY” depends on neither ‘X” nor “Y”. So “XY” is also a headless compound.

Say, **“Fat-Belly”**= Fat +Belly. “Fat-Belly” denotes a ‘person’, neither Fat nor Belly. Therefore “Fat-Belly” is an Exocentric compound word.

Copulative Compound (PW): In this compound, there are two semantic heads. Each head equally contributes to the meaning of the whole.

Suppose, the word is **“bittersweet”.**There is no semantic relation between “bitter” & “sweet”. Both of them are the head. They have a meaning when they are coordinated.

In this study, all the compound words have been categorized and shown in Table 6.

Again, Compounding has been categorized into three groups by their semantic viewpoint. They are **General Words (GW), Subject-Specific Words (SSW),** and **Domain-Specific Words (DSW)** (Table7)**.**

Result and Analysis:

From Table 2, It has been seen that the total number of components of keywords in 2009 is 427. Out of 427 (100%), the number of Compound Words (CW) is 39 (9.13%), the number of Root Words (RW) is 181 (42.39%), the number of Stem Words (TW) is 112 (26.23%), the number of Onomastic Words (OW) is 63 (14.75%), the number of Form Words (FW) is 5 (1.17%), the number of Abbreviations (AA) is 23 (5.39%) and the number of Numerical Figure (NF) is 4 (0.94%). It is clear that after the Root word and Stem Word, the Onomastic Word (OW) plays an important role in forming a keyword, especially for science subjects. The result is almost identical for the years 2010 and 2011 (Table 3 & Table 4 with the respective diagrams 2 &3). For the stipulated time i.e., from 2004 to 2021 the result is identical (Table 5 along with diagram 4).

Table 1: Sample of the study

Year	No of Articles(A)	No. of selected keywords (B)	Total no. of occurrences of all keywords (C)
2004	57	200	374
2005	62	195	356
2006	109	338	706
2007	145	370	934
2008	170	406	1145
2009	165	433	1252
2010	151	431	1182
2011	127	403	912

2012	124	381	911
2013	117	352	820
2014	147	398	960
2015	129	429	891
2016	143	421	1014
2017	146	543	1188
2018	154	503	1095
2019	139	395	747
2020	184	495	1024
2021	176	489	969
Total	**2445**	**7182**	**16480**

Table 2: Components of Keywords for the year, 2009

Word Type	Count	%
CW	39	9.13
RW	181	42.39
TW	112	26.23
OW	63	14.75
FW	5	1.17
AA	23	5.39
NF	4	0.94
Total	**427**	**100**

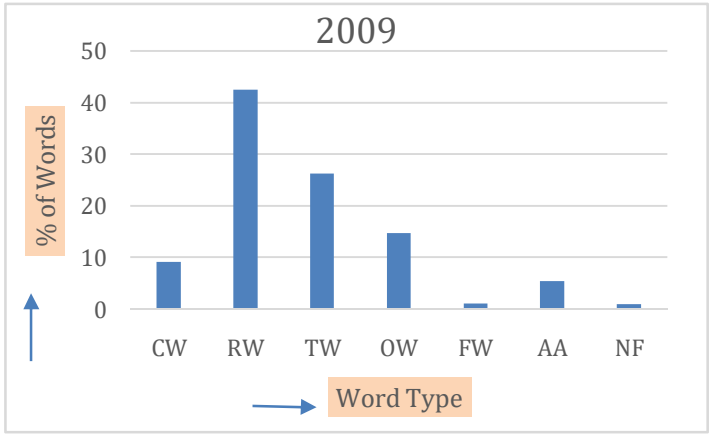

Diagram 1

Table 3: Components of Keywords for the year 2010

Word Type	Count	%
CW	34	8.11
RW	183	43.67
TW	113	26.97
OW	60	14.32
FW	5	1.2
AA	19	4.53
NF	5	1.2
Total	**419**	**100**

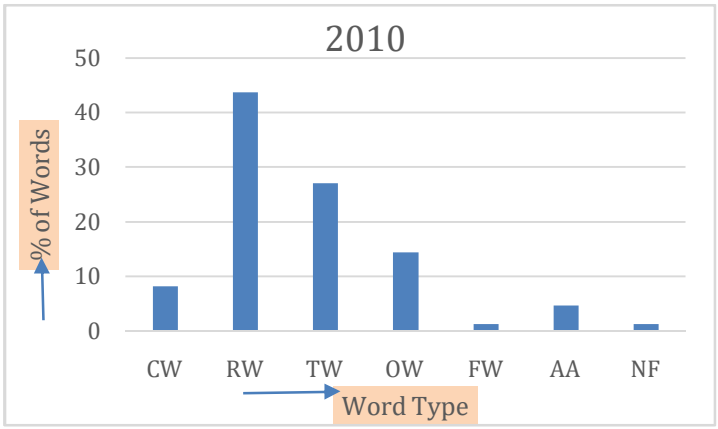

Diagram 2

Table 4: Components of Keywords for the year 2011

Word Type	Count	%
CW	33	8.05
RW	180	43.9
TW	115	28.05
OW	54	13.17
FW	4	0.98
AA	21	5.12
NF	3	0.73
Total	**410**	**100**

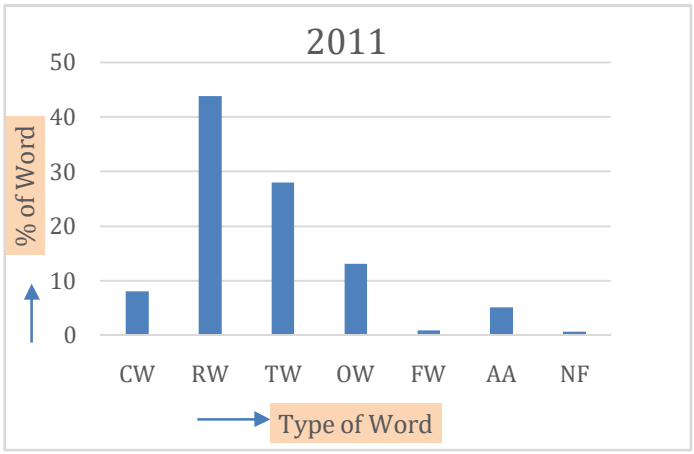

Diagram 3

Table 5: Components of Keywords for the period 2004-2021

Word Type	Count	%
CW	154	9.22
RW	596	35.69
TW	512	30.66
OW	276	16.52
FW	15	0.9
AA	99	5.93
NF	18	1.08
Total	**1670**	**100**

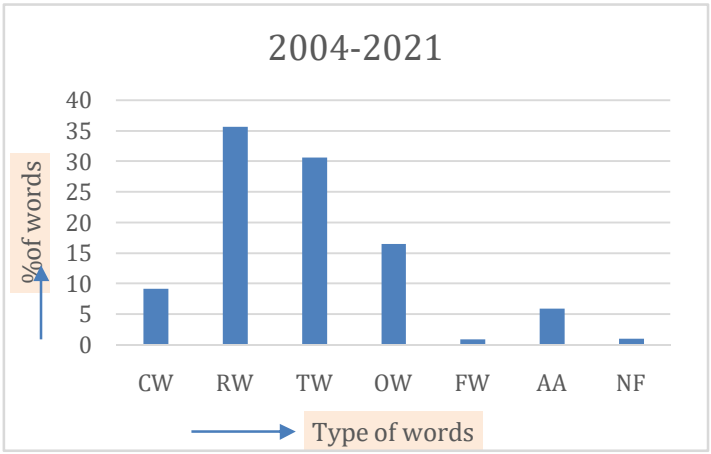

Diagram 4

Table 6 shows that total Compound Words (CW) is 154. Out of 154 Compound words, Endocentric Word(EW) is 70 (45.45%), Exocentric Word(XW) is 29 (18.43%) and Copulative word (PW) is 55(35.72%), From this study it is observed that the Endocentric Compound Words (EW) are highly used in formation of keywords.

Table 6: Number and Percentage of different types of Compound words (by Linguistic Viewpoint)

Type of Compound words	Number	Percentage (%)
EW	70	45.45
XW	29	18.83
PW	55	35.72
Total	**154**	**100**

Table 7 shows different type of Compound Words from the semantic viewpoint. Out of 154 Compound Word only 12 (7.80%) are General Words (GW), 70(45.45%) are Subject-Specific Word (SSW) and 72(46.75%) are domain-Specific word (DSW). For example, Background, Coefficient, Degenerate, Hierarchy, Holography are the General words (GW) where Microcavities, Quasispectrum, Superconductivity, Supernovae are the Subject-Specific Words (SSW) and Acoustoelectric, Agegraphic, Bran world, Antiferromagnets are the Domain Specific Words (DSW) under the study of "Hawking Radiation".

Table 7: Number and Percentage of different types of Compound words (by Semantic Viewpoint)

Type of Compound words	Number	Percentage (%)
GW	12	7.80
SSW	70	45.45
DSW	72	46.75
Total	**154**	100

Conclusion:

This study looks at keyword analysis related to the "Hawing Radiation" subject domain. Statistical analysis has been done on the different parts of the keywords that were allocated to the journal articles' titles and contents over the provided time period. The study focuses on the onomastic word in the formation of keywords. In Hawking Radiation, the onomastic word is incredibly dynamic to generate a keyword. Furthermore, noted are endocentric, exocentric, copulative, subject-specific, and domain-specific compound terms from a linguistic and semantic perspective. Gaining insight into the composition of keywords is made easier by this examination. Other science-related subject areas might also be included in this study.

References:

https://www.quattr.com/optimize-content/topics-vs-keywords

Pawar, Sanket S., et al. "Keyword search in information retrieval and relational database system: Two class view." *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*. IEEE, 2016. Bolin, Mary K. "Linguistics and LIS: a research agenda." (2017).

Palmer, Frank Robert. *Semantics*. Cambridge University Press, 1981.

Dutta, Bidyarthi, and ChaitaliDutta. "Concept of 'subject' in the context of library and information science from a new angle." *Annals of Library and Information Studies (ALIS)* 60.2 (2013): 78-87.

Hartley, James, and Ronald N. Kostoff. "How useful arekey words' in scientific journals?." *Journal of information science* 29.5 (2003): 433-438.

Strader, C. Rockelle. "Author-assigned keywords versus library of congress subject headings." *Library resources & technical services* 53.4 (2011): 243-250.

Hurt, Charlie D. "Automatically generated keywords: A comparison to author-generated keywords in the sciences." *Journal of Information and Organizational Sciences* 34.1 (2010): 81-88.

O'connor, John. "Correlation of indexing headings and title words in three medical indexing systems." *American Documentation* 15.2 (1964): 96-104.

Gbur Jr, Edward E., and Bruce E. Trumbo. "Key words and phrases—the key to scholarly visibility and efficiency in an information explosion." *The American Statistician* 49.1 (1995): 29-33.

Frost, Carolyn O. "Title words as entry vocabulary to LCSH: Correlation between assigned LCSH terms and derived terms from titles in bibliographic records with implications for subject access in online catalogs." *Cataloging& classification quarterly* 10.1-2 (1989): 165-179.

Frank, Eibe, et al. "Domain-Specific Keyphrase Extraction. JCAI'99: Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence, 668-673." (1999).

Hulth, Anette. "Improved automatic keyword extraction given more linguistic knowledge." *Proceedings of the 2003 conference on Empirical methods in natural language processing*. 2003.

Cleverdon, Cyril. "The Cranfield tests on index language devices." *Aslib proceedings*. Vol. 19. No. 6. MCB UP Ltd, 1967.

Warner, Julian. *From writing to computers*. Routledge, 1994.

Brier, Søren. "The foundation of LIS in information science and semiotics." (2006).

Ningsih, AyuWidia, and Rusdi Noor Rosa. "Types And Processes of Compound Words Used In Headline News Columns In "The Jakarta Post" Newspaper." *English Language and Literature* 1.2 (2013).

Booij, Geert. *The grammar of words: An introduction to linguistic morphology*. Oxford University Press, 2012.