An exploratory study of the Gender spectrum using N-Gram analysis to show its distribution in latest edition of DDC through literary warrant

Ankira Sanyal

College Librarian, Sarojini Nadu College for Women, 30, Jessore Road, Dumdum, Kolkata, West Bengal, India

ankira.sanyal@gmail.com

ABSTRACT

Gender spectrum is a term which describes all the gender present in our society. It cannot establish the Gender, Male and Female as polar opposite. In 6th September, 2018 Supreme Court declares that Gay marriage is legal and it's natural. People have no control over it. Being a library professional a curiosity was there to study the Gender spectrum and it's expand in different classification system through literary warrant. N-gram analysis is used to see the frequency of terms used in literature and world cat database is used to see the literary warrant of selective concepts.

KEYWORDS: Gay, Gender spectrum, Lesbian, Ngram study, N-gram value, worldcat database etc.

1. INTRODUCTION

Gender spectrum is an umbrella term which includes all the genders present in our society. The word LGBTQIA is an acronym stands for Lesbian, Gay, Bisexual, Transgender, Queer, Intersex and Asexual. History of LGBT in India and abroad was very old. The main example is the *Kamasutra* written by Vatsyayana is an ancient Indian text which dedicates a complete chapter on homosexual behavior. *Arthashastra* is an example where different sexual practices were described with man or woman. From Arthashastra we came to know that Homosexuality was treated as a very minor offence at that time. After that from the writings of Al-Biruni, we came to know that Medival Hindu society disapproved it while in Muslim society it was very common. In Mughal period Babur was a Mughal Emperor. He wrotes *Baburnama*. It was his autobiography where he says about his infatuation with a teenage boy. ("LGBT history in India", 2020). Literature of Mughal periods tells us that homosexuality was a common practice at that time. After some time in the year 1757, the battle of Plassey was happened and the East India Company expanded their power in India. British Raj criminalized homosexuality based on "Buggery act 1533". The Indian Penal Code (IPC) is a criminal code which compiles a significant amount of criminal law with in a particular jurisdiction. The draft was first prepared by T.B. Macaulay in the year 1834. Finally, it came into operation in 1st January, 1862

although Macaulay died on 1859. ("Section 377 of the Indian Penal Code", 2020.). The Buggery act 1533, was an act of Parliament of England which was passed during the time of Henry VIII. Buggery is an unnatural sexual act against the will of God and Man. Modeled on the Buggery act 1533, section 377 was introduced in 1861which declares that sexual activities against the order of nature is illegal. ("Buggery Act 1533", 2020). In the year 2009, an NGO named NAZ Foundation started a fight against section 377 to decriminalize it. After nine years, in 6th September 2018, The Supreme court of India declares that Gay sex is natural and people have no control over it. Now the study of Gender spectrum uses Ngram analysis to show the frequency of terms used in literature. Here a vocabulary (by using different authentic lists) of gender spectrum is prepared and world cat database is used to see the literary warrant of concepts.

2. STATEMENT OF THE WORK

The Problem of the proposed research focuses on the existence of the Gender related terms in published documents and their treatments in DDC. The problem of the proposed research also give its focus on increasing and decreasing frequencies of the use of the concepts in literature over time by using google book Ngram viewer. Now under Gender spectrum, Lesbian, Gay, Bisexual, Transgender, Queer, Intersex, Asexual all are present. To investigate the existence of these terms, with a focus on the development of the literary warrant, the emphasis is given on the form heading present in the latest edition of DDC.

3. OBJECTIVES

The purpose of this work is to discover the existence & growth pattern of the different kinds of Gender in last fifty years, to show their n-gram value and show their distribution in latest edition of DDC. Specific objectives are:

3.1 To trace the existence of the concepts under Gender spectrum and their availability in public domain.

3.2 To trace the existence of the concepts under Gender spectrum and their n-gram value.

3.3 To ascertain the evolution of literary warrant of concepts under Gender spectrum in last fifty years and to analyses their position in the latest editions of DDC

3.4 To examine whether all the concepts are directly given class number and form heading in DDC or not?

4. RESEARCH QUESTIONS

4.1 What are the concepts of Gender spectrum available in different published lists? How can they be identified?

4.2 Are all these concepts have n-gram value? Or not?

4.3 Whether all these concepts are increasing over time or not? How does it identified?

4.4 Whether all these concepts are present in latest edition of DDC or not?

4.4 How did the literary warrant on terms, all over the world, have evolved in course of time? How does this trend get its reflection in classification system? What methodology should be followed in this purpose?

5. METHODOLOGY

Content analysis (Document research), Key word analysis, and Survey of the world cat cataloguing database and survey of Google book Ngram viewer is to be done to achieve the stated objectives of the proposed research work.
In document research at first different authentic published lists of Gender spectrum is to be collected. By consulting all these lists, consolidated lists of terms have to be prepared alphabetically.

Now world cat cataloguing database is used to see the growth of literature over 50 years.

After that reflections of these terms can be meticulously studied in latest editions of DDC. In this research work focus should be given on the trend of the development of the subject because DDC, UDC gets enriched with new subjects only after the emergence of new literary warrant. World cat's advanced search option will be the best option for collecting literary warrant. Restrictions should be given on the year for last hundred years. Otherwise there would be no restrictions on audience, content, format and language.

At last the frequency of these concepts is being studied by using google book Ngram viewer in between 1945 to 2019. Actually after the 2nd world war, different movements related to LGBT people was started. So, 1945 may be taken and as google books N gram viewer is giving a result up to 2008 for the some corpus of literature, the year 2019 has to be taken as a cutoff date. Ngram viewer gives the graphical representation of concepts by analyzing the frequency of their use in literature of selected corpus. For this study, the corpus of "English 2012" may be selected which will represent "Books predominantly in the English language published in any country". ("Google Ngram viewer", 2013).

6. LITERATURE REVIEW

Library play an important role in dissemination of knowledge in the various sections of our society. Most of the library ignored to keep LGBT related documents as there was very low or less demand on LGBT related documents. There is a phobia related to LGBT people. We called it Homophobia or Transphobia etc. Some study reveals that overlook of Transgender people is common in our society. But these people also have information needs. As society have some phobia on them, they mainly use Internet as an Information resource. They felt shame to come to the Library for their information needs. Most of the library don't have current collections which can meet their information needs. They face difficulties in searching documents and using controlled vocabulary. They felt discomfort to approach library staff. Privacy issue is also there. (Thompson, 2012). 'The study of LGBT behavior' shows the importance of need of information for the LGBT people at their transitional phase of life. In this study, they have taken LGBT people belonging 18-30 years of age. This study reveals their information need which may protect them from suicide or HIV infections. This study also says that Prejudice or ignorance against LGBT people should be stopped and if society treat them sympathetically, it will be better for them. (Vazquez & Teruel). Some library like Oak Park Public Library (OPPL), Oakland, California's public library, The San Francisco Public Library, started collection development for the LGBT community and served them in an orderly manner. (Thompson, 2012). Classification and cataloguing are two major operational work in the library which runs the library in an orderly manner. Most of the classification scheme and subject heading lists are not up-to-date and does not contains all the terms related to Homosexual concepts. Previously in LCSH, Homosexuality was under Sexual deviance. "After agitation by the cataloging section of SRRT's Task Force on Gay Liberation, LC pulled 'Homosexuality' from the shadow of 'Sexual deviations' into the clear descriptive light of 'Sexual life' ". This study also says that LOC Classification and subject heading notify by the LOC are often wrong and should be corrected. There is a bias in library classification structure and subject heading lists. If terms are not appropriately kept in the schedule then it will leads problem in respect of queer theory and also LGBT related theory. Particular category always reflect a particular ideology. So bias is not expected there. (Drabinski, 2013). The study of "Google book Ngram viewer in socio cultural research" shows that how google book Ngram viewer can play an important role in research. Different words and group of words are tested here to see their frequencies, increasing and decreasing

frequencies with time. By reviewing all these literature it has been found that the frequency of use of Gender spectrum related concepts in literature by using google book Ngram viewer and their treatments in DDC against literary warrant is still untouched.

7. VOCABULARY ON GENDER SPECTRUM

According to Oxford Advanced Learner's Dictionary, vocabulary means "all the words that a person knows or uses". (Hornby, 2010, pp 1722). Here lists are taken from different authentic sources. Amnesty International USA is a nonprofit making organization published a list named "LGBTQI Glossary". UC DAVIS (The University of California, Davis) is a public research university published a list named "LGBTQIA resource centre glossary". LGBT resource centre of USC (University of Southern California) published a list named "LGBTQI Terminology". Stonewall is the Scottish Government Consultation Act for Trans equality in Scotland. They published a list named "Glossary of terms". Australian Institute of family studies, the Australian Government, published a list named "LGBTIQA+ communities Glossary of common terms". These lists are being consulted to make a consolidated list of terms. Here only ten terms are taken as these terms are directly related to the concept of Gender spectrum.

7.1 Consolidated list of concepts: After consultation of all the above lists, ten terms are to be selected as they are directly related to the Concept "Gender Spectrum". Terms are given below:

7.1.1 Allosexual: It is a "sexual orientation generally characterized by feeling sexual attraction or a desire for partnered sexuality". Both Men and women can fall into this category. ("LGBTQIA resource centre glossary", 2020).

7.1.2 Blaq/BlaQueer: Black African queer persons are called Blaq/BlaQueer.

7.1.3 Bisexual: A person who can attracted to either sex.

7.1.4 Cross Dresser: In our society we dress ourselves as per the social norms. There is distinguishing between male and female dress. When a person dresses himself/herself of another gender, then he/she is called cross dresser.

7.1.5 Gay: A sexual and affection orientation toward people of the same gender.

7.1.6 Homosexual: Homosexual means when a person attracts towards same sex person. It is somehow an out date of term now a day. Homosexual may be of two types like Gay or Lesbian. Gay refers to a male person, attracted to a male person and Lesbian refers to a female person attracted to a female person.

7.1.7 Lesbian & Lipstick lesbian: It is a form of homosexuality where a female person attracted mentally or sexually towards a female person. A lesbian became a lipstick lesbian when she shows a feminine approach like wearing makeup, dressing etc.

7.1.8 Queer: It is an umbrella term which encompasses sexual and gender minorities who are neither heterosexual nor cisgender. Lesbians, gays, bisexuals, Trans people, intersex persons are includes in this term.

7.1.9 Trans gender: These people have a gender identity which differs from their sex assigned at birth. Their sexual orientation does not depend on their gender identity.

7.1.10 Twospirit: They are neither male nor female. They have attributed of both the genders. (Sanyal & Bhattacharyya, 2020).

8. LITERARY WARRANT OF TERMS IN LAST 50 YEARS

The concept of literary warrant was first devised by E.Wyndham Hulme in the year 1911 in his book "Principles of book classification" in the Library Association record. He was a Librarian of the patent office, London. Literary

warrant means a subject is listed on the scheme when some literature has appeared on it. ("General theory of library classification", n.d.). World cat is a union catalogue of 17900 libraries collections which are distributed over 123 countries and territories. It is maintained by OCLC. Previously the full form of OCLC was Ohio computer Library centre. Now it is named as Online Computer Library Centre. ("World cat", 2020). Here world cat cataloguing advanced search option is being used to see the frequencies of these concepts as subject. The table is given below:

Selected Terms	Documents in World cat between 1971-1980	Documents in World cat between 1981-1990	Documents in World cat between 1991-2000	Documents in World cat between 2001-2010	Documents in World cat between 2011-2020
Allosexual	0	0	0	0	0
Blaq/Blaqueer	0	0	0	0	5
Bisexual	157	493	1978	2659	6710
Cross Dresser	409	479	767	560	505
Gay	8007	11595	20366	30975	64055
Homosexual	239	406	1074	1940	2295
Lesbian	955	1810	5293	8854	15968
Queer	343	506	2106	7664	21062
Trans gender	341	593	1870	3920	22331
Twospirit	0	0	0	0	1

Table1 : Literary warrant of concepts in subject approach

From table 1 it has to be found that for Allosexual there is no any literary warrant in last fifty years. For Blaq/Blaqueer and Twospirit it is only one and five in lat 10 years. The literary warrant of the concepts like Bisexual, Gay, Homosexual, Lesbian, Queer, Trans gender as a subject are increasing day by day whereas for Cross dresser it is more or less same for the entire period.

9. REFLECTION OF THESE TERMS IN LATEST EDITION OF DDC

Dewey Decimal Classification System is a very important classification tool in the library. The system was devised by Melvil Dewey, an American Librarian and educator, in the year 1873 and first published in 1876. (Dewey, 2011). Now the latest edition is 23rd edition contains 4 volume. Volume 1 contains Introduction, Glossary, Index to the introduction and Glossary, Manuals: a guide to use the DDC, and six tables which are: Table 1: Standard Subdivisions; Table 2: Geographic Areas, Historical Periods, Biography; Table 3: A-C Subdivisions for Arts, for Individual Literatures, for Specific Literary Forms; Table 4: Subdivisions of Individual Languages and Language Families; Table 5: Ethnic and National Groups; Table 6: Languages.

Volume 2 contains Schedules: The organization of knowledge from 000 to 599. 000 – Computer science, Information & General works, 100 – Philosophy & Psychology, 200 – Religion, 300 – Social sciences, 400 – Language, 500 – Pure Science.

Volume 3 contains Schedules: The organization of knowledge from 600 to 999. 600 – Technology, 700 – Arts & Recreation, 800 – Literature, 900 – History & Geography.

Volume 4 Relative Index: It is an alphabetical lists of subjects along with their class number. As the terms with class number is present (with page number is not present), it is called relative index.
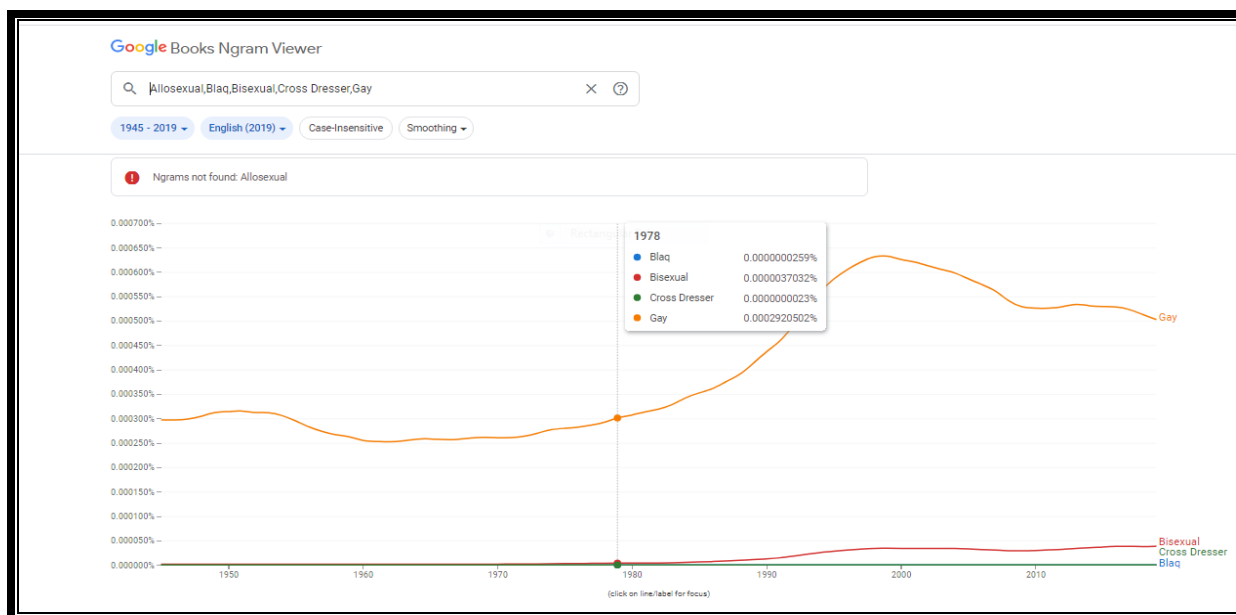
After analyzing all these terms it is to be found that Most of the concepts are belongs to the same class number and whereas some of the terms don't have class number although their literary warrant is present. The form heading of Queer should be included in DDC as there is huge literary warrant of the term Queer as a subject approach.

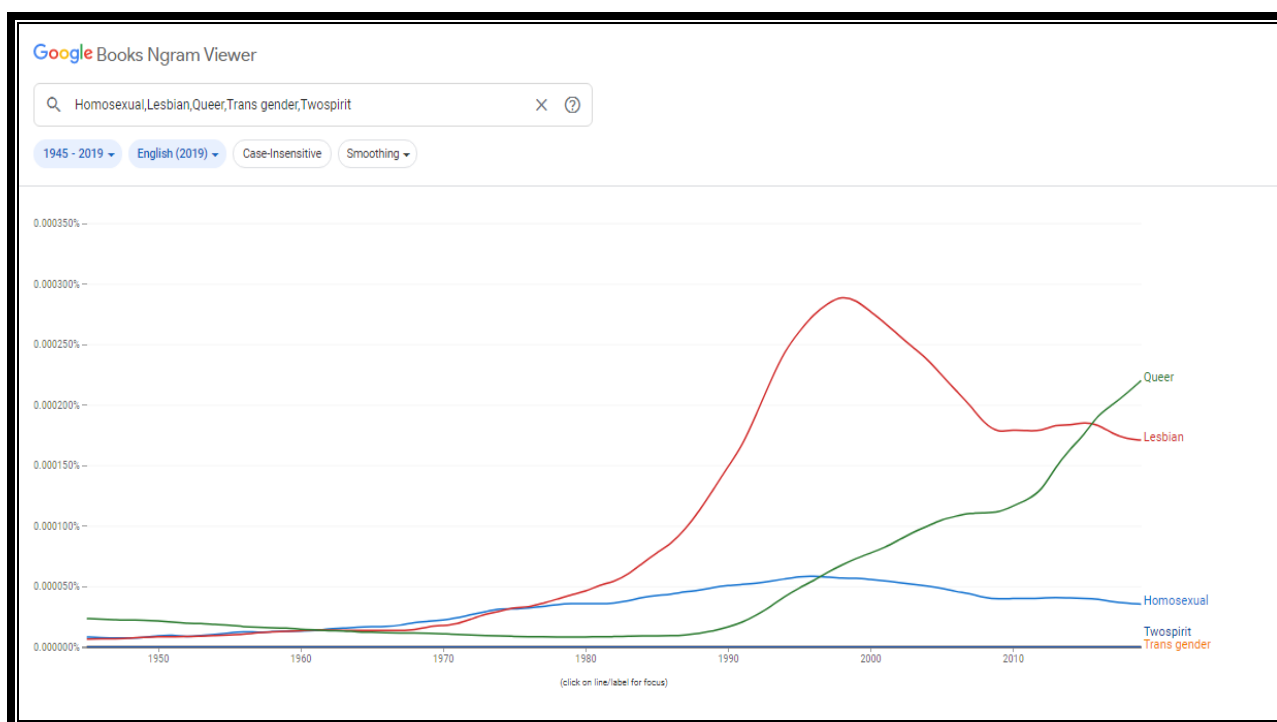| Selected Terms	Class number in DDC 23rd edition
Allosexual	Absent
Blaq/BlaQueer	Absent
Bisexual	306.765
Cross Dresser	306.7
Gay	306.7662
Homosexual	306.766
Lesbian	306.7663
Queer	Absent
Trans gender	306.76
Twospirit	Absent

10. NGRAM STUDY OF THESE CONCEPTS

Google books was launched in October 2004 as a service provided by Google Inc. They scanned a huge number of books and Magazines and store them in their digital database by using OCR technology. Optical character Recognition (OCR) is a technology which allows converting different types of documents captured by a digital camera into editable and searchable data. ("Google books", 2020). Google books Ngram viewer is an online search engine which depicts its result in graphical form. Here any one can search any set of comma-delimited search strings using a yearly count of grams found in sources printed between 1500 and 2019. Google text corpora are English, Chinese (simplified), French, German, Hebrew, Italian, Russian, or Spanish. There are also some specialized English corpora, such as American English, British English, English Fiction, and English One Million; and the 2009 version of most corpora is also available. ("Google Ngram viewer", 2020). In this study, the year was given from 1945 to 2019. 2019 is given as it is the Cutoff date in the dataset which is available now. In near future 2019 will be changed as the scanning of books is on ongoing process. Here 1945 is selected as because immediately after 2nd world war, a number of homosexual rights group from Britain, France, Germany, Netherland, across the western world including United States came into the light. They preferred the term Homophile rather than Homosexual. This movement was started in late 1940's and around 1945 to 1960s. The Gay liberation movement was started in the year 1969 which continues till 1974. In the year 1973 The American Psychiatric association removed homosexuality from the diagnostic manual of Mental illness. LGBT rights movement was started in the year 1972 and continues till date. But the origin of LGBT movement was present in Homophile movement which was started in the year 1945. ("LGBT social movements", 2020). After placing all these terms in Google book N gram viewer it is being found that these concepts are used in literature. Their frequencies are increasing and decreasing over time. Results are given below:

Pic 1: N gram analysis of Allosexual, Blaq, Bisexual, cross dresser and Gay.

From Picture 1 it has to be found that there is no ngram value for the term allosexual whereas there are ngram value for the terms like Blaq, Bisexual, cross dresser and Gay.

Pic 2 : N gram analysis of Homosexual, Lesbian, Queer, Trans gender and Twospirit.

From Picture 2 it has to be found that there is ngram value for the terms like Homosexual, Lesbian, Queer, Transgender and Twospirit. The ngram value of the queer is increasing with time.

CONCLUSION

There are many concepts on Gender Spectrum are available in different authentic published lists and literature in our society. There are some concepts present in DDC some are absent in DDC. If the literary warrant is compared with

the N gram study, then it has to be found that for those subject like Allosexual the literary warrant is zero on the other hand for the same subject N-gram value is not present. For the term Blaq/Blaqueer ngram value is very low and literary warrant is also very low. But for the term queer, Lesbian, homosexual, and Gay the literary warrant and Ngram value both are high. So as a subject approach these terms are present and demand to get proper value in DDC for organization of documents.

REFERENCES

[1] Buggery act 1533. (2020). In *wikipedia*. Retrieved January 11,2020, from

https://en.wikipedia.org/wiki/Buggery_Act_1533

[2] Dewey, M. (2011). Dewey decimal classification and relative index. USA :OCLC online computer library centre.

[3] Drabinski, E. (2013). Queering the catalog: Queer theory and the politics of

correction. *The Library quarterly*. 83, 2, 94-111. Retrieved on March 23, 2020 from

https://www.jstor.org/stable/10.1086/669547

[4] General theory of library classification. (n.d.). Retrieved March 20, 2020 from

http://egyankosh.ac.in/bitstream/123456789/33078/1/Unit-3.pdf

[5] Google books. (2020). *In wikipedia*. Retrieved March 26, 2020 from

https://en.wikipedia.org/wiki/Google_Books

[6] Google Ngram viewer. (2020). *In wikipedia*. Retrieved March 26, 2020 from

https://en.wikipedia.org/wiki/Google_Ngram_Viewer

[7] Hornby, A.S. (Ed). (2010). *Oxford Advanced Learner's Dictionary* (8th ed.). New york : Oxford university press

[8] LGBTIQA+ communities glossary of terms. (2020). Retrieved from

https://aifs.gov.au/cfca/publications/lgbtiq-communities

[9] LGBTQI Glossary. (n.d.). Retrieved from https://www.amnestyusa.org/pdfs/AIUSA_Pride2015Glossary.pdf

[10] LGBT history in India. (2020). Retrieved March 29, 2020 from

https://en.wikipedia.org/wiki/LGBT_history_in_India

[11] LGBT social movements. (2020). *In wikipedia*. Retrieved March 26, 2020from https://en.wikipedia.org/wiki/LGBT_social_movements

[12] LGBTQI terminology. (2003-2004). Retrieved from

https://lgbtrc.usc.edu/files/2015/05/LGBT-Terminology.pdf

[13] Section 377 of the Indian Penal Code. (2020.). In *wikipedia*. Retrieved January11,2020, from https://en.wikipedia.org/wiki/Section_377_of_the_Indian_Penal_Code

[14] Thompson, K. J. (2012). Where's the "T"? : Improving library service to community members who are Transgender identified. *B Sides*. Retrieved on 23rd January 2020 from

https://ir.uiowa.edu/cgi/viewcontent.cgi?article=1032&context=bsides

[15] Vazquez, T. A., Teruel, A. G. Information behavior of lesbians, gays,bisexuals and transsexuals (LGBT). Retrieved on March 22, 2020 from http://eprints.rclis.org/24616/1/Andres_GT.pdf

[16] Worldcat. (2020). In *Wikipedia*. Retrieved January 16, 2020, from https://en.wikipedia.org/wiki/WorldCat

_____