Plagiarism: An Essay in Terminology

M.P. Satija* and Daniel Martínez-Ávila

#Guru Nanak Dev University, Amritsar - 143005, India
#São Paulo State University (UNESP) Marilia, Brazil
+E-mail: satija_mp@yahoo.com

ABSTRACT

The terminology on plagiarism is not hard and fast. It is fluid, a bit ambiguous, and still emerging. It may take some time to settle the terms more clearly, concretely and exhaustively. This paper aims to provide a terminological discussion of some important and current concepts related to plagiarism. It discusses key terms/concepts such as copyright, citation cartels, citing vs. quoting, compulsive thief, cryptomnesia, data fakery, ignorance of laws and codes of ethics, information literacy, lack of training, misattribution, fair use clause, paraphrasing, plagiarism, plagiarism detection software, publish or perish syndrome, PubPeer, retraction, retraction vs. correction, retraction watch, salami publication, similarity score, Society for Scientific Values, and source attribution. The explanation and definition of these terms/concepts can be useful for LIS scholars and professionals in their efforts to fight plagiarism. We expect this terminology can be referred in future discussions on the topic and also used to improve the communications between the actors involved.

Keywords: Copyright; Cryptomnesia; Information literacy; Plagiarism; Research frauds; Retractions; Terminology management

1. INTRODUCTION

Concept theory is a fundamental part of Knowledge Organisation and Library and Information Science. While systems should ideally be able to communicate and retrieve concepts instead of terms (“Find what I mean, not what I say”), in practice, people and systems of any kind are dependent on the verbal forms that designate those concepts. In Wüster’s general theory of terminology and terminological lexicography, concepts were the starting point for the terminological work. In this sense, according to Francelin and Kobashi, the objective of terminology would be to establish clear-cut limit between concepts. In words of Ingetratu Dahlberg, “It is of course necessary for communication purposes to synthesize the concepts elements into one expression or a short word combination, in order to deal with it.”

Terminology is an essential part of communication between scientists, scholars, and professionals. Library and information science (LIS) scholars deal with terminology not only in the visible parts of the knowledge organisation processes and systems but also in the foundational discussions that permeate the social mediation between users and information. S.R. Ranganathan considered that scientific terminology is of dual importance to librarians: first, librarians and information scientists have to understand the terminology of their own discipline in order to discuss technical matters with colleagues for purposes of research and academic development in their discipline and profession; second, it is important to understand the technical terminology of other disciplines in order to organise knowledge and to communicate effectively with the experts in that field who come to seek some information in the library. As librarians know well, some of the main problems that affect communication and the organisation of knowledge include ambiguity and imprecision. In the topic of plagiarism, terminology must also be precise and well-defined. In this domain, not only good communication between scholars and between educators and students is at stake, but also the scientific foundations of the legal aspects.

In this paper we aim to provide a terminological discussion of some important concepts related to plagiarism. Many of these concepts must be clarified for Library and Information Science scholars and practitioners as the terms must be used by librarians and educators in course syllabi, library websites, and other documents listing the responsibilities of librarians in relation to plagiarism. We hope that this terminological essay on plagiarism helps to clarify these concepts and serves well for future discussions in the domain.

2. WHAT IS PLAGIARISM?

The definition of plagiarism is not obvious due to the wide and varied use of the term. Broadly speaking, plagiarism is a copyright infringement, yet not every copyright infringement is plagiarism. Copyright is a limited time monopolistic control given by law of the land to the creator of any artistic, intellectual or scientific entity. But the rights given to a copyright holder are not absolute as there is an inbuilt provision for its fair use for the progress of knowledge. Though the automatic and legally permitted use is neither copyright infringement, nor plagiarism.

Received: 10 December 2018, Revised: 30 January 2019
Accepted: 05 February 2019, Online published: 11 March 2019
Though the term “fair use” (or alternatively “fair dealing”) is detailed to some extent, yet it is far from unambiguous. Devil lies in its interpretation and implementation. Plagiarism indeed is in the infamous company of unfair means such as piracy, data-fabrication, research falsification, evidence cooking, ghost-writing, proxy writing, research-recycling, literary-thefts, copying graphics, industrial designs and trademarks, and committing other frauds in research, writings, and artistic and industrial creations. Strictly speaking, it is any lack of giving proper, formal and adequate credit to the original source of ideas or their expression in any writing and all other forms and modes of creative expressions. Lack of proper attribution to the original source is plagiarism and a serious research misconduct. Although the definition of plagiarism is not easy, a common aspect that can be found in the literature on the concept is someone using ideas or words by others as if they were his/her own.

3. WHAT CONSTITUTES PLAGIARISM?

Plagiarism is using someone’s intellectual or artistic creation without permission, acknowledgement, and credit. The term has its roots in the classical Latin word Plagirus, a person who abducted the child or slave of someone. A modern copycat plagiarist would be a kidnapper who, mercifully, does not ask for any ransom. Nevertheless, the act is illegal, unethical, and even criminal. Published works and other creative expressions are indeed for use of the public and benefit of the society at large. Lafollette, aptly writes: “scientists and scholars want their ideas to be used and their words to be quoted”. But using them in a way what constitutes plagiarism is simply their abuse and an act of dishonesty. It is immoral, it is an academic or research fraud, and a criminal offence in business and industry. Based on various dictionaries a synthetic definition of plagiarism would be “the “wrongful appropriation” and “stealing and publication” of another author’s “language, thoughts, ideas, or expressions” and the representation of them as one’s own original work”. Admitting that due to wide variations in its perception, practice and norms among nations, groups and disciplines, Lafollette identifies four elements present in most definitions:

- The unattributed use of words, text, ideas, or illustrations created by someone other than the author listed;
- The failure to credit the original (“real”) author in a manner appropriate to the communication;
- The implication (or statement) that the plagiarist is the original author; and
- The failure to obtain the original author’s consent”.

Any of these elements individually or collectively constitute plagiarism.

4. WHY IT HAPPENS?

There are several reasons why plagiarism may happen. Below we review some of the concepts and terms related to following reasons.

4.1 Publish or Perish Syndrome

Among academics, the maxim “publish or perish” (i.e., author more and more research papers or suffer stagnation in your career) is a threatening reminder of how publications are linked to promotions and job stability in higher education and research institutions. Many have criticized it as a barbarian threat which lays emphasis on quantity instead of quality and instigates academicians to resort to less than honest means to swell their needed number of publications. The publish or perish syndrome may lead to salami publishing to superficially enhance the number of publications and worse take to practices that cross the ethical line of research. Apart from mental stress on the researcher, this pressure is also widely believed as a cause of plagiarism and low-quality research.

4.2 Salami Publication

Segmented publication, also called “salami publication” is one of the forms of self-plagiarism, is a distinct form of duplicated publication which is usually characterized by publishing the same paper in different media. It could also be making more papers from an earlier published paper having much in common. Such deliberate acts present a serious threat to publication ethics which constitute the ethical problem of self-plagiarism and ways to handle such cases. “Salami publication can be roughly defined as a publication of two or more articles derived from a single study. Articles of such type report on data collected from a single study split into several segments just large enough to gain reasonable results and conclusions, also known as “minimal publishable unit”.

4.3 Ignorance of Laws and Codes of Ethics

This is a form of information illiteracy. Many authors and especially students are not aware of the legal and ethical implications of not properly attributing and citing sources. Deliberate infringement of copyright and ethical guidelines is not always the norm as many are ignorant of these matters. A further problem related to ignorance and plagiarism is the thin line between common knowledge and authorship. While some authors might omit citations believing something is common knowledge, other unexperienced authors might misattribute something to a low-quality paper that originally did not cite the source. Aspects related to information literacy skills, scientific argumentation, and academic etiquette are not always known to everybody and should be part of librarians and educators’ responsibilities.

4.4 Source Attribution

Within the context of creative writing, attribution means relating used or borrowed ideas or expressions to its real and ultimate source and appropriately and adequately giving credit to the creator or author. Even the most permissive of the copyleft licenses never relinquish the moral right of attribution. Non-attribution of sourced ideas leads to plagiarism.

4.5 Misattribution

Like misinformation, misattribution is attributing texts and ideas even events to something with which they really have no connection or association. It is making an incorrect attribution due to ignorance or lack of information skills.
4.6 Cryptomnesia

An illusion of the mind related with partial amnesia results in a biased memory. It is a selective memory loss. Under cryptomnesia one remembers the facts, or ideas but forgets their source. For stressed mind ideas, not their source, have the priority of retention. Thus, with time a cryptomnesiac person unconsciously starts attributing those ideas to self. This leads to unintended plagiarism.

4.7 Compulsive Thief

While some cases of plagiarism might be unintentional, many of them are intentional due to necessity or habit. Some authors are compulsive thieves that got used to get their way in a quick and easy way, many times at the expense of others in the academia. Other offenders get into the bandwagon of plagiarism as they perceive others are doing it with ease and impunity. The idea of possessiveness of ideas in which attribution is perceived as a devalue instead of a reinforcement of an argument is a problem too. Plagiarists tend to think that ideas and not arguments are the main focus of academic writings. They disregard the humble saying “standing on the shoulders of giants” as they would like to be portrayed as the giants, forgetting that “giants” is a metaphor for assiduously learning the work of many others and not only of those who author the latest paper.

4.8 Lack of Training

While plagiarism is linked to pressure to publish and quick and easy ways to deal with this productivism, lack of training can be one of the main barriers to the quality of research. There are several ways in which authors can be trained to avoid plagiarism and improve the quality of their research, and librarians can play a big role on that. Students and scientists need to acquire training on evaluation and access of sources, citation practices and styles, use of antiplagiarism software as a preventive measure, and the obligations and exceptions to copyright, among others. These are some of the aspects that can be included in information literacy programs.

4.9 Citing vs Quoting

Quoting is reproducing somebody’s expressions with full attribution in exactly the same words set apart usually by quotations marks or other typographical devices such as setting off in a new paragraph or in italicized typeface; Citing is referring to someone’s ideas, thoughts or expressions in one’s own words in a summarized or paraphrased way. Full attribution is necessary in both the cases. By citing or quoting others the author gets endorsement or discusses his/her ideas.

4.10 Paraphrasing

Restating or summarizing someone’s words into one’s own words and style for brevity of words and thoughts, or for more contextual clarity, or even to present someone’s ideas or expressions in an improved language and better literary style. Indeed, it is a highly skilled literary art. Unattributed paraphrasing amounts to plagiarism. Paraphrased texts need only citation, not quotation.

4.11 Misuse of Fair use Clause

No copyright law gives absolute rights to the author of the work, it has inbuilt provision for some taken for granted rights of the users. In the context of copyright, fair use refers to some granted and given rights of any and all users to make some genuine but limited and cautiously fair use of the work under the clause “fair use” (note: a similar clause exists in the United Kingdom under the denomination “fair dealing”). The spirit of fair use is to allow excerpts from the work without permission to review, criticize, quote or use for teaching and research purpose. Fair use is not infringement of copyright. In some countries though these parameters of the fair use are mentioned, yet limits of fair use remain open to debate and interpretation of legal experts. As there is not a clear line between fair use and copyright infringement, some authors may commit copyright infringement under the excuse of fair use.

5. THE ILL EFFECTS OF PLAGIARISM

Plagiarism has extremely negative effects for science and the academia, nay for the entire society. It is a theft and if not curbed will result in lack of faith in law and order. The credibility of the scientific output, researchers, and institutions becomes compromised once a plagiarism scandal happens and is made public. As Rothschild, plagiarism is “a threat to democracy” and closely related to corruption. Plagiarism has social, academic, and economic effects for nations. It discourages creativity and promotes imitation instead of innovation. Financial losses to commerce and industry are unaccountable when the culture of plagiarism is installed.

5.1 Data Fakery

Data fakery is also known as data fabrication or data cooking. In many occasions, this data fakery is caused by academic pressure and competitiveness. A very famous example of data fakery in the last decade was Woo Suk Hwang’s scandal involving stem-cell. Reported reasons that contributed to this data fakery included poor national mechanisms for accountability, competition, transparency, and ethical oversight. Check and Cyranoski evaluated the case stating that this scandal of data fakery would damage not only the public perceptions of stem-cell research, but also the image of science as a whole.

5.2 Citation Cartels

Citation cartels is when different individuals or groups mutually agree underhand to cite one another or manipulate the bibliometrics to increase the quantity of their citations. Indeed, it is inappropriately reciprocal back-scratching. It is a research misconduct which tends to reward quantity of metrics. The term was coined by Phil Davis who detected the existing of unholy alliances of scholars or editors to cite one another on reciprocal basis mostly in farfetched context. It also means to cite those who could benefit you, say the members of the editorial boards.
who could influence the publication of the submitted paper, or your senior/boss who could help you in the advancement of the professional career. Citation cartels and bibliometric tricks to boost the prestige of journals and authors also have very negative effects on science and society.

5.3 Piracy

Piracy is an euphemism for the act of making cheap, unauthorized, usually an underground and clandestine activity, copies of literary, artistic works, whether printed or digital, which are still in copyright, usually for commercial gains. Pirated documents are counterfeited inexpensive replicas of the originals. It is not plagiarism as pirated documents still have the full attribution to the original author(s) and publishers. This is a serious and very deliberate copyright infringement causing financial losses and many other commercial damages to the publisher. Advances in printing and digital technologies have given a boost to piracy A counterfeiter can be awarded financial punishment, or imprisonment, or both. On the other hand, authors such as Richard Stallman have discouraged the use of the term “piracy” to refer to copying (https://www.gnu.org/philosophy/words-to-avoid.en.html#Piracy). In this sense, a United States judge stated that terms such as “piracy,” “theft,” and “stealing” are smear words. Neutral terms such as “unauthorized copying” (or “prohibited copying” when it is illegal) or even a positive term such as “sharing information with your neighbor” are proposed as alternatives.

6. PUNISHMENT

Plagiarism has severe consequences for offenders. These consequences range from loss of reputation to economic fines and ruined careers. Students are expelled from their schools, and faculty fired in several cases of plagiarism. Doctoral degrees can be revoked and plagiarizing publications are retracted and cursed.

6.1 Retraction

Budd22, expresses the concept of retraction as follows: “Retraction is a serious action, taken when egregious misconduct is admitted to or shown to be the case. Editors may make statements of the retraction of a published work and, sometimes, third parties (including attorneys) may retract a paper. The most common (and, some say, the most serious) cause for retraction is the fabrication or falsification of data. Some researchers may create data from whole cloth, without the collection of legitimate data for legitimate analysis. Less common than fabrication of data, but still serious, is blatant plagiarism of substantial content from a previously published article. Similar to plagiarism is the duplication of publication—an author or set of authors publishing essentially the same paper in more than one journal. Retraction is not exactly a common occurrence; retractions represent a small portion of the published body of work” Retractions can be a positive thing that polices and discourages plagiarism in the academia. According to Oransky and Marcus33 retractions after genuine mistakes are considered an act of honesty that do not necessarily hurt someone’s career.

6.2 Retraction Watch

Launched by science writers Oransky and Marcus in August 2010, retraction watch <https://retractionwatch.com> is a blog that reports on retractions of scientific papers and on related topics. The blog has demonstrated that retractions are more common than commonly assumed as in many cases the reasons for retraction are not reported publicly by the editors. As a paradoxical law, more popular a journal, higher the rate of its retractions. One reason for this phenomenon is that authors want to get published in a popular journal by hook and crook, even by using devious means—and some of them later get caught.

6.3 Retraction vs Correction

The term retraction carries stronger connotation than the term correction. An alteration that changes the main point of the original statement is generally referred to as a retraction while an alteration that leaves the main point of a statement intact is usually referred to simply as a correction. A lesser withdrawal of content than a full retraction may be labeled a correction. Depending on the circumstances, either a retraction or correction is the appropriate remedy.

7. HOW CAN IT BE DETECTED?

In the analog days, plagiarism had to be manually detected by experts on the topic who had read the original text and remembered the source. The search for similarities in printed sources was a tedious and laborious work. The development of technologies though has facilitated most cases of plagiarism yet has also aided the detection and prevention of plagiarism.

7.1 Plagiarism Detection Software

It is any computer software, both proprietary and free software, which helps to know the source of borrowed or stolen textual expressions in the new document, if any, and also counts the percentage of similarity between the suspected piece and the huge database of the anti-plagiarism software. A plagiarism software does not check for plagiarism in a piece of work. Instead, it only checks a work against the huge but limited database held by the machine, and if there are instances where the new (suspected) writing is similar to or matches against any of the sources in the database, it gives data on overlapping with indication of the original sources. A database usually includes tremendously large number of web pages: both current and archived from the internet and a collection of documents. Though its database comprises thousands of journals, ETDs, e-books, and web pages, databases, and publications, yet it may not hold everything on the subject. Therefore, some plagiarized texts safely go undetected.

7.2 Similarity Score

It must be remembered that these are text matching systems and machines. In the process of detecting plagiarism through machines some segments of a later work may match with some of the text in database held by the machine. The similarity score simply makes us aware of any overlapping areas in a later paper. It can then be used as data for a larger process, to determine if the match is or is not acceptable. Machines by
comparison provides data only. Similarity reports provide a quantified summary of matching or highly similar text found in a submitted paper. Interpretation should be based on human judgment and on institutional policies. The UGC, India in its late 2018 regulations has prescribed varying punishments for varying similarity scores in a work—forgiving ten percent similarity score in any academic work! In a plagiarized work similarity may be of two types: semantic similarity, and syntactic similarity. The former is difficult to detect and measure. A paragraphed text may have full similarity with its original but the machine may be hoodwinked in taking it as original—as its syntactic similarity is poor.

8. AVOIDANCE?

Several legal experts, academicians and educators have suggested measures and strategies that authors, scholars, and librarians could follow to avoid plagiarism. Some of these measures include sensitizing students and researchers, and authors about the dangers of plagiarism. Many of these strategies, within the LIS field, relate to the area of information literacy.

8.1 Information Literacy

It is a trained capacity of an individual to feel the need of information and a set of skills to locate, validate sources, evaluate, and use effectively and ethically the needed information. Projected as a human right, the information literacy skills help a person to live and work in the society to the full potential. Information literacy is an essential skill for students, researchers and authors to be taught at the school level for lifelong learning. An information literate person is a well-informed citizen to participate effectively, intelligently, and actively in the society for promotion of democracy and its human progress. An information literate person is unlikely to commit unintended plagiarism—perhaps intended too.

Further solutions need to involve the activities of national and international institutions that watch and regulate the conditions and collaborations in the fight against plagiarism. Some of these institutions include PupPeer in the United States and the Society for Scientific Values in India.

8.2 PubPeer

Founded in 2012 by Stell and Barbour, Pubpeer <https://pubpeer.com/> is an online platform for post-publication peer review. It is a service run by the PubPeer, a California based not-for-profit foundation, for the benefit of its readers and commenters, who create its content. The overarching goal of the Foundation is to improve the quality of scientific research by enabling innovative approaches for community interaction. This post-publication has led to revelation of many academic misconducts and subsequent retraction of papers. Earlier the revelations of frauds were anonymous, now the foundation has changed the policy to name the whistle blowers on demand.

8.3 Society for Scientific Values, India

Keeping the growing academic frauds in mind, an Indian group of distinguished scientists with high international and national credentials, led by Prof. Avtar Paintal, FRS, set up the Society for Scientific Values (SSV) <www.scientificvalues.org/society.html> in 1986 at New Delhi. The society has obviously no legal or administrative powers, but it enjoys high moral credibility as watchdog. It has taken up cases from time to time, where values intrinsic to science, “scientific ethics”, have been compromised. The society, from time to time, organises meetings with a specific purpose. It posts on the page the cases where it has come to a specific conclusion. The SSV works on issues related to scientific ethics and misconduct. Though it has no legal popover but as a watchdog keeps an eye on misconduct, fraud, and scientific integrity. It also publishes SSV News and Views.

9. CONCLUSIONS

In this paper we have identified and explained some plagiarism related terms/concepts. This terminology is still amorphous, fluid, and not yet concrete. While many of these terms are recent and dependent on the new changes that the technological advances cause in the academic research and publications. Admittedly, the terminology listed in this pioneering paper is far from exhaustive. At the same time new terms are being thrown open as social, psychological, academic, and technological causes of plagiarism and its control thereof are emerging.

We hope this essay on terminology can help in the development of more stable and clear concept systems for the concerned community. The explanation and definition of these terms/concepts can be useful for LIS scholars and professionals in their efforts to fight plagiarism. As effective communications between educators and students, editors and authors, and scientists is essential to understand and avoid this plague. This endeavor can be referred to and may hopefully be applied in future discussions and publications on the topic of plagiarism.

REFERENCES


**CONTRIBUTORS**

**Dr. Mohinder Partap Satija** retired as Professor and UGC Emeritus Fellow from the Department of Library and Information Science, Guru Nanak Dev University, Amritsar, India. He has written more than twenty books, about 150 papers and numerous book reviews and conference papers published in India and abroad. Contribution in the current study is to take part in the planning and supervision of the work and preparation of the manuscript.

**Dr Daniel Martínez-Ávila** is an Assistant Professor at the Department of Information Science, São Paulo State University, Marília, Brazil. He is a member of the ISKO Scientific Advisory Council. He also collaborates with the Satija Research Foundation for Library and Information Science, India, and the Institute for Gender Studies (IEG) at University Carlos III of Madrid, Spain.
Table 1. Alphabetical list of the terms discussed along with the section numbers

<table>
<thead>
<tr>
<th>Terms/Concepts</th>
<th>Section/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citation cartels</td>
<td>5.2</td>
</tr>
<tr>
<td>Citing</td>
<td>4.9</td>
</tr>
<tr>
<td>Compulsive thief</td>
<td>4.7</td>
</tr>
<tr>
<td>Copyright</td>
<td>2</td>
</tr>
<tr>
<td>Correction</td>
<td>6.3</td>
</tr>
<tr>
<td>Cryptomnesia</td>
<td>4.6</td>
</tr>
<tr>
<td>Data fakery</td>
<td>5.1.6</td>
</tr>
<tr>
<td>Fair use clause</td>
<td>4.11</td>
</tr>
<tr>
<td>Ignorance of laws and codes of ethics</td>
<td>4.3</td>
</tr>
<tr>
<td>Information literacy</td>
<td>4.7, 7, 8.1</td>
</tr>
<tr>
<td>Lack of training</td>
<td>4.8</td>
</tr>
<tr>
<td>Law of retraction</td>
<td>6.2</td>
</tr>
<tr>
<td>Misattribution</td>
<td>4.5</td>
</tr>
<tr>
<td>Misuse of fair use clause</td>
<td>4.11</td>
</tr>
<tr>
<td>Paraphrasing</td>
<td>4.10,7.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Terms/Concepts</th>
<th>Section/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Piracy</td>
<td>5</td>
</tr>
<tr>
<td>Plagiarism</td>
<td>2, 3</td>
</tr>
<tr>
<td>Plagiarism detection software</td>
<td>7.1</td>
</tr>
<tr>
<td>Publish or perish syndrome</td>
<td>4.1</td>
</tr>
<tr>
<td>PubPeer</td>
<td>8.2</td>
</tr>
<tr>
<td>Quoting</td>
<td>4.9</td>
</tr>
<tr>
<td>Retraction</td>
<td>6.1</td>
</tr>
<tr>
<td>Retraction vs. Correction</td>
<td>6.3</td>
</tr>
<tr>
<td>Retraction Watch</td>
<td>6.2</td>
</tr>
<tr>
<td>Salami publication</td>
<td>4.2, 6</td>
</tr>
<tr>
<td>Self-plagiarism</td>
<td>4.2</td>
</tr>
<tr>
<td>Similarity score</td>
<td>7.2</td>
</tr>
<tr>
<td>Society for Scientific Values</td>
<td>8.3</td>
</tr>
<tr>
<td>Source attribution</td>
<td>4.4</td>
</tr>
<tr>
<td>Terminology</td>
<td>1</td>
</tr>
</tbody>
</table>