JKMS

Check for updates

Plagiarism: A Bird's Eye View

Farrokh Habibzadeh (ID)

Past President, *World Association of Medical Editors* (WAME)
Editorial Consultant, *The Lancet*
Associate Editor, *Frontiers in Epidemiology*

OPEN ACCESS

ORCID iD
Farrokh Habibzadeh (ID)
https://orcid.org/0000-0001-5360-2900

ABSTRACT

Plagiarism is among the prevalent misconducts reported in scientific writing and common causes of article retraction in scholarly journals. Plagiarism of idea is not acceptable by any means. However, plagiarism of text is a matter of debate from culture to culture. Herein, I wish to reflect on a bird's eye view of plagiarism, particularly plagiarism of text, in scientific writing. Text similarity score as a signal of text plagiarism is not an appropriate index and an expert should examine the similarity with enough scrutiny. Text recycling in certain instances might be acceptable in scientific writing provided that the authors could correctly construe the text piece they borrowed. With introduction of artificial intelligence-based units, which help authors to write their manuscripts, the incidence of text plagiarism might increase. However, after a while, when a universal artificial unit takes over, no one will need to worry about text plagiarism as the incentive to commit plagiarism will be abolished, I believe.

Keywords: Plagiarism; Ethics; Scientific Misconduct; Medical Writing; Publishing

INTRODUCTION

Plagiarism is among the prevalent misconducts in scientific writing commonly identified by science editors and is a common cause of article retraction from scholarly journals.[1] The United States Office of Research Integrity (ORI) defines plagiarism as "the theft or misappropriation of intellectual property and the substantial unattributed textual copying of another's work."[2] Several types of plagiarism have so far been proposed,[3] however, from a pragmatic point of view, it can be divided into two broad categories — plagiarism of idea and plagiarism of text, also referred to as "verbatim."

PLAGIARISM

Plagiarism of idea is not acceptable by any means.[4] In the eyes of almost all scholars, it is tantamount to theft. Depending on the study field, wording of the text and thus plagiarism of text may be of paramount importance. For instance, in humanities and literacy, where the essence of novelty and originality of a work genuinely depend on the choice of the words and

the way they are collocated, verbatim is considered a blatant fault in almost all cultures.[4] In certain cultures, nevertheless, degrees of verbatim are not only tolerable but also encouraged! For example, there are techniques in Persian literature that use verbatim to supposedly give more beauty and eloquence to the text or a piece of poem. One of these techniques is to borrow a piece of text, usually from an eminent erudite scholar, and use it exactly — as it is (technically text plagiarism) — directly in your own text without clearly distinguishing the borrowed text from the rest of the text (e.g., by inserting it in double quotes or indenting it) or even exclusively mentioning the original source or that who the original author is. It is presumed that the borrowed text is such famous that [hopefully most] readers will figure out where the borrowed piece is taken from. This has been a common practice done even by world-renown great Iranian poets such as *Sædī* and *Hāfez*. This has not been considered a misconduct. Nor does it come with a shame. In Persian literacy, it has long been considered an artful skill. This practice is not confined to Persian literacy.

THE FLIP SIDE OF THE COIN

Plagiarism has not always been detrimental. *Schola Medica Salernitana*, the first modern medical school established in Western Europe, was flourished for the existence of a few textbooks. Authored by a Tunisian doctor, Constantinus Africanus (1020–1087), *Liber Pantegni* was among one of these influential books. However, based on an allegation made by Stephen of Antioch (first half of the 12th century), *Liber Pantegni* is a translation of *Liber Regalis* (*Kitāb al-mālikī*), a book written by an Iranian physician, Hāly Abbās (Alī ibn Abbās Al Majoussi).[5] By today's standards, what Constantinus Africanus did was a blatant misconduct, nonetheless, had he not plagiarized the Hāly Abbās' book, we would probably not have had *Schola Medica Salernitana* and today's modern universities around the globe!

NOBLES OBLIGE

Nowadays, when English is *de facto* the *lingua franca* of science, if we are going to publish, we need to follow the rules set; *nobles oblige*.[6] Although some cultures may still tolerate or even encourage recycling text and what is constitute a misconduct by other cultures, the modern western culture which significantly affects the whole science publishing enterprise round the globe, seriously banned such a practice and set penalties, sometimes very tough, for those committing every instance of plagiarism.[2]

PLAGIARISM IN SCIENTIFIC WRITING

I believe that while plagiarism of text in literature and the humanities is not acceptable by any means (for the reasons mentioned above), degrees of verbatim in scientific writing, where wordings are just a means to convey the scientific idea and not the foundation of the novelty and originality of a scientific work, might be tolerable, as long as we can be sure that the authors could correctly construe the text piece they borrowed.[4,7] Science content is itself complex enough that authors have degrees of difficulty in expressing their ideas. Presenting such complex things in a language other than our mother tongue makes things more complicated. I believe that recycling words in describing a well-known methodology by an author who is not an English native speaker and who just used the text for want of linguistic

JKMS

expertise should not be considered a deadly sin. It is better to just ignore such a *faux pas*, because I am not sure that those who consider it a serious misconduct would fare any better if they were to write a similar passage in a language which is not their own mother tongue. Not long ago, French and Arabic used to be the science *lingua franca*.[4]

Even if we could solve the problem some authors have with finding the appropriate words and expressing themselves in English, for example by helping them by employing authors' editors, such as what we had in the AuthorAID initiative,[8] there remain still some problems. Given the limitations existing in the number of words and acceptable legitimate phrases (from the syntactical and semantical points of view), there is only a limited number of ways we can describe a method, say measuring the blood pressure in a patient, in a certain language (e.g., English). This inevitable text recycling is clearly apparent in the case of using computer programming languages, such as *FORTRAN* or *C*, where the number of keywords is much less than the number of words available in a typical human language, such as English or Persian.[7,9]

WHERE VERBATIM MIGHT BE TOLERABLE

While recycling words in the methodology section of a manuscript would be tolerable, it is not acceptable in the results section; after all, that is supposedly the results of your own study and should not duplicate the findings of other studies. Text similarity in the results section may be a sign of salami publication.[10] Text similarity might be acceptable in part in the introduction, but it should be avoided or used rarely in the discussion. The reason is that the discussion is the only place where you can present what you think, express your views, and explain the observed findings, and the editor and reviewers should be confident that you really did understand and mean what you wrote and if you borrowed a piece of text from another source, they can never be sure about that. I believe, in most instances, it is just enough to write a comprehensible text. If the message of the study to be conveyed is clear enough to make sense so that the referees and editors clearly understand the message and appreciate the merits of the study, the editorial office will [hopefully on most occasions] take care of correcting the linguistic mistakes, if any, and present the final version of the manuscript in an acceptable standard way. At least, this is how mainstream journals operate, to the best of my knowledge.

WHY AUTHORS COMMIT PLAGIARISM

Many authors commit text plagiarism unintentionally because they are simply not aware that this is considered a misconduct and prohibited. This is particularly more likely in those who come from an Eastern culture where text recycling of erudite scholars is generally not discouraged.[11] In such instances, the authors are commonly young researchers with no formal training in Western institutes; in most cases, the source where the recycled text was copied from can be readily seen among the references.[3] On the other hand, there are those who commit text plagiarism intentionally just for academic laziness.[6,12,13] This is the most common reason for recycling text among English native speakers.[14] Most of these authors commit plagiarism intentionally to deceive the readers (including the editors and reviewers) and pretend that the text is their own. In such instances, the source where the recycled text was copied from is not cited. Some authors whose mother tongue is not English, although

being aware of the fact that plagiarism is not acceptable, do commit it intentionally not to deceive anyone but for want of linguistic proficiency and unwillingness to sacrifice accuracy and quality of the original English statement they copied.[4,6]

HOW TO DEAL WITH PLAGIARISM

Editors should manage plagiarism individually based on its context. As an editor, I will not be very strict with a junior author who has had no formal training in a Western institution, particularly if I see the source of the text copied is in the references list. I will explain for them that this kind of practice is not acceptable, that they should paraphrase the text, that they need to insert the borrowed text within double quotes and cite the original source, etc. The story will nonetheless be totally different if I face a senior author with years training in European or American institutes and who has already published many articles in international journals, particularly if I cannot find the source among the references list, which increases the likelihood of intentionally committing plagiarism — I will follow the recommendations of the Committee On Publication Ethics (COPE: https://publicationethics. org/guidance/Flowcharts?t=Plagiarism&sort=score); I will reject the submitted manuscript on the fly, ask the author(s) to clearly explain why they did commit plagiarism, and report their misconduct to the dean of their institute and ask the dean to take appropriate actions. But, before taking any actions, we need to detect text plagiarism.[14]

DETECTION OF TEXT PLAGIARISM

Detection of text plagiarism is mostly based on detection of text similarity.[3] There are numerous online platforms that do check text similarity for free. Examples are eTBLAST (http://etest.vbi.vt.edu/etblast3) and turnitin (https://www.turnitin.com/). However, the most powerful platform with its largest database for comparison of texts is the Crossref iThenticate (https://www.ithenticate.com/). Some journals check every submission and if the text similarity score exceeds a certain level, say 20%, the manuscript will be desk rejected. In a previous article, I have shown that text similarity, even as much as 30%, does not necessarily imply text plagiarism and that there is really no cut-off for text similarity to imply text plagiarism.[9] An editor with an in-depth understanding of plagiarism must examine the results of text similarity reports with scrutiny and decide if the similar text is a real instance of verbatim or not. There are many common statements that are mistakenly taken by text similarity software programs as verbatim.

WHAT THE FUTURE HOLDS

With recent introduction of artificial intelligence (AI)-based units and their help with scientific writing, the incidence of plagiarism may increase. AI-based units use a large-language model to generate text.[15] This is mostly done based on a probabilistic model derived from the frequency distribution of different words in a very large database. Although the size of the database is huge, given its limitations, the generated texts might not be as diverse as it would be if it was written by a human (for lack of novelty in the current AI units). AI units may also be (mis)used by reviewers in various ways, for instance, with writing their report under a short deadline.[16] This increase in the incidence warrants a pressing need for increasing

the awareness of the stakeholders on how to correctly use the AI units and knowing their strengths and limitations.[17] As the use of AI units in scientific writing, such as ChatGPT, has been increasing, editors need to have means to identify the machine-generated text pieces.[18] GPTZero (https://gptzero.me/) is among the first AI-based web-based software programs that has claimed to be able to differentiate machine-generated from human-written texts. However, although its specificity is acceptable, its sensitivity is low.[19] We need better tools.

Although the increase in the incidence of plagiarism attributed to the emergence of AI units and their help with writing will supposedly increase, the increase would be transient. As a matter of fact, the problem of plagiarism will be resolved completely after the rise of the Universal AI (UniAI), I believe.[20]

At the heart of any misconduct is the "intention to deceive." In plagiarism, the authors intend to deceive the readers [along with editors and reviewers] that the text in the manuscript is theirs. When the UniAI takes over, there will be no incentive to deceive others[7,20]; all scientific articles will be written by the UniAI with no intention to deceive us. Language, then, is only a medium for communicating science.

REFERENCES

1. Gupta L, Tariq J, Yessirkepov M, Zimba O, Misra DP, Agarwal V, et al. Plagiarism in Non-anglophone countries: a cross-sectional survey of researchers and journal editors. *J Korean Med Sci* 2021;36(39):e247.
   PUBMED | CROSSREF

2. US Office of Research Integrity. ORI policy on plagiarism. https://ori.hhs.gov/ori-policy-plagiarism. Accessed May 30, 2023.

3. Zimba O, Gasparyan AY. Plagiarism detection and prevention: a primer for researchers. *Reumatologia* 2021;59(3):132-7.
   PUBMED | CROSSREF

4. Vessal K, Habibzadeh F. Rules of the game of scientific writing: fair play and plagiarism. *Lancet* 2007;369(9562):641.
   PUBMED | CROSSREF

5. Glaze FE. Galen refashioned: Gariopontus in the later middle ages and renaissance. In: Furdell EL, editor. *Textual Healing: Essays on Medieval and Early Modern Medicine*. Leiden, The Netherlands: Brill Academic Publishers; 2005, 53-75.

6. Habibzadeh F, Marcovitch H. Plagiarism: the emperor's new clothes. *Eur Sci Ed* 2011;37(3):67-9.

7. Habibzadeh F. Plagiarism: what does the future hold for science writing? *Eur Sci Ed* 2014;40(4):91-3.

8. Shashok K. How AuthorAID in the Eastern Mediterranean helps researchers become authors. *Write Stuff* 2010;19(1):43-6.

9. Habibzadeh F. The acceptable text similarity level in manuscripts submitted to scientific journals. *J Korean Med Sci* 2023;38(31):e240.
   PUBMED | CROSSREF

10. Habibzadeh FA, Winker M. Duplicate publication and plagiarism: causes and cures. *Notf Rettmed* 2009;12(6):415-8.
    CROSSREF

11. Rokni MB, Bizhani N, Habibzadeh F, Farhud DD, Mohammadi N, Alizadeh A, et al. Comprehensive survey of plagiarism in Iran. *Pak J Med Sci* 2020;36(7):1441-8.
    PUBMED | CROSSREF

12. Kleinert S. Checking for plagiarism, duplicate publication, and text recycling. *Lancet* 2011;377(9762):281-2.
    CROSSREF

13. Habibzadeh F, Shashok K. Plagiarism in scientific writing: words or ideas? *Croat Med J* 2011;52(4):576-7.
    PUBMED | CROSSREF

14. Mehta P, Mukherjee S. Plagiarism and Its repercussions: a primer on responsible scientific writing. *Central Asian Journal of Medical Hypotheses and Ethics* 2022;3(1):52-62.
    CROSSREF

15. Park SH. Use of generative artificial intelligence, including large language models such as ChatGPT, in scientific publications: policies of *KJR* and prominent authorities. *Korean J Radiol* 2023;24(8):715-8.
PUBMED | **CROSSREF**

16. Donker T. The dangers of using large language models for peer review. *Lancet Infect Dis* 2023;23(7):781.
PUBMED | **CROSSREF**

17. Koga S. The integration of large language models such as ChatGPT in scientific writing: harnessing potential and addressing pitfalls. *Korean J Radiol* 2023;24(9):924-5.
PUBMED | **CROSSREF**

18. Zielinski C, Winker MA, Aggarwal R, Ferris LE, Heinemann M, Lapeña JF, et al. Chatbots, generative AI, and scholarly manuscripts. WAME recommendations on chatbots and generative artificial intelligence in relation to scholarly publications. https://wame.org/page3.php?id=106. Updated May 31, 2023. Accessed Jun 2, 2023.

19. Habibzadeh F. GPTZero performance in identifying artificial intelligence-generated medical texts: a preliminary study. *J Korean Med Sci* 2023;38(38):e319.
PUBMED | **CROSSREF**

20. Habibzadeh F. The future of scientific journals: the rise of UniAI. *Learn Publ* 2023;36(2):326-30.
CROSSREF