Case Report

Plagiarism through Paraphrasing Tools—The Story of One Plagiarized Text

Libor Ansorge 1,*, Klára Ansorgeová 2 and Mark Sixsmith 3

Abstract: This paper describes a unique case study wherein real plagiarism revealed in a scientific journal is compared with the original article. The plagiarized text contains many typical errors, such as inconsistent terminology, unclear meanings of sentence, missing tables and figures, and an incorrect literature list. The occurrence of similar errors in other manuscripts may serve as a warning against plagiarism. During the analysis of the plagiarized text, it was assumed that a paraphrasing tool was used for preparing this plagiarized text. To confirm this assumption, the chosen paraphrasing tool was used to create a paraphrased version of the article and this version was compared with the plagiarized text. The paraphrased version had far fewer changes from the plagiarized text than the plagiarized text had from the original article. Thus, it was confirmed that the plagiarized text was created using a paraphrasing tool. Information contained in this article can be used for detecting this type of plagiarism.

Keywords: plagiarism; paraphrasing; text spinning; context comparison; plagiarism detection

1. Introduction

The opening sentence in the article by Khan et al. [1] stated that “in the twenty-first century, the academic world is facing new threats to maintain integrity”. Singh [2] wrote that one of the greatest challenges for researchers is producing scientific papers, “be it the language checks or adhering to the strict journal guidelines and finally summing one’s scientific research into a prescribed word count”. He added that this requires “the best writing skills using the relevant words and correct flow in language to bring out the salient features highlighting every single step and methodology performed”. And he argues that the discussion in particular requires “creativity and originality from the author”. That is indeed true. Progress in language research and the use of modern IT tools, such as neural networks and language models, have led to the development of paraphrasing tools (also known as text rewriting or text spinning tools). These tools can then serve, and unfortunately do serve, activities that can be described as academic dishonesty. However, there is no clear and standardized definition of academic dishonesty; that the notion of scientific dishonesty is inexact makes the question of definition elusive [3] and several definitions have been published [4,5]. Jones [6] describes academic dishonesty as “‘cheating’, ‘fraud’, and ‘plagiarism’, the theft of ideas and other forms of intellectual property—whether they are published or not”. Academic dishonesty is a persistent problem for universities [7,8] since, as Ellahi et al. [9] put it, “dishonest behaviour of students becomes very severe when they exercise the same practice at their place of work”. Several authors [10–12] have described examples of students using online paraphrasing tools to transcribe someone else’s work and pass it off as their own. This article shows that not only students use these tools. Within the academic community, all dishonesty based on the use of a piece of work without appropriately acknowledging the
source falls under the umbrella of plagiarism [13]. Adam [14] identified three different perspectives of plagiarism: “as a moral issue, . . . a regulatory issue, and . . . a natural part of learning to write from sources”.

Perhaps the worst case is when the author takes over someone else’s entire work and makes it his or her own. These instances of academic dishonesty are often motivated by pressure on publishing activity by authors known as “publish or perish”. Eshchanov et al. [15] describe “publish or perish” as “the pressure in academic discipline to frequently publish research papers in scholarly journals and advance one’s career”. They add that “academics who fail to publish eventually perish by either not finding jobs or losing their existing positions”. Kun [16] argues that it is “the pressure to produce eye-catching results, which are publishable in prestigious journals, that undermines the integrity of science”. Although the pressure to publish may seem to be particularly recent, this is not the case. The need for “publish or perish” was already mentioned by Archibald Cary Coolidge in 1932 [17] and N. R Barrett wrote “to kill the dragon called ‘publish or perish’” in 1962 [18]. Despite the pressure to publish from academics’ rating and remuneration systems, publishing other people’s ideas or whole works under his/her own name is a sign of personal failure by the individual. In reaction to the importance of ethical issues linked to “publish or perish”, Warsy and Warsy [19] paraphrased this topic into the “Publish ethically or perish” and Kun [16] into “Publish and Who Should Perish: You or Science?”.

One of the roles of peer review is to detect this type of academic dishonesty. Reviewers and editors have to identify this academic dishonesty manually or can use plagiarism detection systems, such as Turnitin or iThenticate, which use some detection methods. Several such methods have been described in the literature in recent decades [20–22]. The presence of plagiarism can be demonstrated only by comparison with a previously published original source [23].

Automatic paraphrase detection has an important role in various tasks, including plagiarism detection [24]. Foltýnek et al. [25] described a quiz when the participants try to identify machine-paraphrased text. The average efficiency of participants was below 80%, but the authors [25] believe that “efficiency will be lower in a realistic scenario, in which readers do not pay special attention to spotting machine paraphrases”. On the other side, Weber-Wulff [26] highlighted that plagiarism detection systems “cannot determine plagiarism”, adding that they can help with identifying problems, “but not for discriminating between originality and plagiarism”. While using any plagiarism detection systems, the user must be aware of the limitations of automatically generated comparisons [27]. Also, such a system cannot catch what it cannot access, for example, if a plagiarist used an article from a journal that does not post materials online [28]. Davies and Howard [29] demonstrate limitations of automatic detection systems and highlighted pedagogies as “an important part of preventing online plagiarism”. Ultimately, a person must make the final decision about plagiarism.

This paper describes a unique case study when real plagiarism was revealed in a scientific journal and is compared with the original article. During the analysis of plagiarism, it was assumed that a paraphrasing tool was used for preparing this text. To confirm this assumption, the chosen paraphrasing tool was used to create a paraphrased version of the article and this version was compared with the plagiarized text. This article provides some recommendations for detecting this type of plagiarism.

The article continues with a description of materials and methods, which describe the analysed article, tools and methods used for analysis. In the ‘Results’ section, outputs from contextual and statistical analysis are described. In next section, several questions are discussed which were identified during the analysis. The last section concludes the results from the analysis.

2. Materials and Methods

This article describes a case where the full text of the original article has been reproduced, and therefore the plagiarized text is not quoted in this article and only the
original article is quoted. Quotation of plagiarism improves the statistics of these articles in indexing services such as Google Scholar and Research Gate.

2.1. An Analysed Article Description

Water literacy is a relatively new field of research. For example, in August 2021, the database Scopus contains only 36 records, which used “water literacy” in the title, abstract, or as a keyword. Scholar indexing services such as Google Scholar cover a wide range of scientific sources, making research results published in less prestigious journals or publications available to the scientific community. This service contains 1320 records with the term “water literacy” in August 2021. Google Scholar is an online search engine, which means, as Falagas et al. [30] point out, “there are no limits on the languages covered or list of covered journals, provided that an electronic edition exists for the latter”. Academic social networking sites such Research Gate or Academia.edu provide a similar service [31].

There is no standardised definition of “water literacy”. McCarroll and Hamann [32] analysed 26 different definitions of water literacy and highlighted that “current water literacy definitions, understandings, and applications vary substantially”. Fielding et al. [33], for example, define water literacy as “water-related knowledge”. McCarroll and Hamann [32] approach water literacy as a set of key knowledge from various perspectives and levels divided into eight unique and overarching themes.

In 2020, the conference paper “Encouraging Water Literacy” was found on Research Gate. This paper was published in special issue 236A of 'Research Journey’ International Multidisciplinary E- Research Journal (ISSN 2348-7143) from the conference on “Introspection, Prognosis and Strategy For Global Water Resources” in January 2020. In March 2021, the article “Advocating Water Literacy” by Otaki et al. [34] from 2015 was found using Google Scholar. These two papers have the same content, which is based on classification of water literacy and dividing of water literacy into three categories:

1. Practical water literacy;
2. Living water literacy;
3. Social water literacy.

According to plagiarism taxonomy [35], analyzed plagiarism can be characterized as intelligent plagiarism since paraphrasing for text manipulation was used.

The original article [34] is divided into these parts:

- Title;
- Abstract (1 paragraph);
- Keywords (4 keywords);
- Chapter 1. Introduction (2 paragraphs);
- Chapter 2. Definition of Water Literacy (8 paragraphs);
- Chapter 3. Classification of Water Literacy (2 paragraphs);
  - Subchapter Practical water literacy (1 paragraph);
  - Subchapter Living water literacy (2 paragraphs);
  - Subchapter Social water literacy (3 paragraphs);
- Chapter 4. Need for the Concept of Water Literacy (1 paragraph);
  - Subchapter 4.1 Reviewing the Historical Transition of Our Relationship with Water and Formation of the Modern Water System (2 paragraphs);
  - Subchapter 4.2 Merits and Demerits of Modern Water Systems (0 paragraph);
    - Subchapter 4.2.1 Invisible Water (2 paragraphs);
    - Subchapter 4.2.2 Mass Consumption and Its Effect (4 paragraphs);
    - Subchapter 4.2.3 Distrust for Water (2 paragraphs);
- Chapter 5. Effectiveness of the concept of Water Literacy (6 paragraphs).

In total, the original article has 36 text paragraphs (Abstract included). The plagiarist used 35 paragraphs, deleted one paragraph (4th paragraph in subchapter 4.2.2) and added four paragraphs as a new chapter entitled “Suggestion”. They also added the new heading
“Conclusion” between the third and fourth paragraphs in chapter 5. Effectiveness of the concept of Water Literacy. Comparison of structures and content between the original and plagiarized text is shown in the Supplementary Material.

2.2. Methods for Plagiarism Analysis

In the first step, Microsoft Word and the online tool DiffChecker (https://www.diffchecker.com/ accessed on 3 September 2021) were used for comparison of changes in the plagiarized text and original article and errors in the plagiarized text were identified. The Abstract and 34 paragraphs, which are in the original document and in the plagiarized text too, were used for statistical analysis.

In the second step, several online paraphrasing tools were used to create a paraphrased abstract. The paraphrasing tool (https://paraphrasing-tool.com/ accessed on 3 September 2021) that produced the most outputs similar to the plagiarized text was used for preparing the new paraphrased article.

In the third step, Microsoft Word and DiffChecker were used for comparison of changes in the plagiarized text and paraphrased article.

3. Results

3.1. Comparison of the Plagiarized Text and Original

The plagiarist made a few alterations to the text, but the structure of the article is the same. The main alterations are:

- Changed several keywords;
- Changed some words in each sentence of the body of the article (see Supplementary Material);
- Changed some words in several titles in the article (see Supplementary Material);
- Deleted 1 paragraph that describes a case study in Japan;
- Added a new title Conclusion;
- Added a new chapter Suggestion;
- Changed the list of literature.

One of the main motivations of a plagiarist is to make his/her job easier. Therefore, the text in question may also contain errors by which plagiarism could be identified; several were found in the analyzed plagiarized text.

3.1.1. Unclear or Confusing Terminology

The term “water education” is used many times for the term “water literacy”, although the original term water literacy is used in a different place for the same term. For example, water literacy was divided into three categories: practical, living, and social water literacy. These categories are used by the plagiarist too in chapter 3, but in the Abstract of the plagiarized text there is the sentence: “Water literacy was isolated into three classifications: down to earth, living, and social water education”. There are other, similarly confusing substitutions of terminology in the text. Another example of confusing terminology is “wellbeing of water”, which is a paraphrase of the original “safety of water”, or “creating nations” for “developing countries” and “created nations” for “developed countries”.

3.1.2. Quality of Language

Paraphrased texts often suffer from impaired language quality. This has been described by other researchers, for example by Kannangara [36]. A large number of language-poor sentences were also used in the studied plagiarized text, such as “Lopsided conveyance of water causes contrasts in water accessibility and expertise on the most proficient method to get enough water, just as customs in water use from area to area”. (original: “Uneven distribution of water causes differences in water availability and know-how on how to obtain enough water, as well as customs in water use from location to location”). Another example is the opening sentence of the abstract (even the first word) that immediately raises suspicions about the reliability and quality of the text: “Momentum water use in our day by day life is in no way, shape or
form feasible, and the ecological and social issues rise to the top”. (original: “Current water use in our daily life is by no means sustainable, and the environmental and social problems come to the surface”).

Language errors do not necessarily prevent the reader from understanding what an author is trying to say, but halfway through the Introduction we do run into the problem of not actually comprehending the message: “There are numerous spots where individuals are compelled to oversee in trouble”. (Original: “There are many places where people are forced to manage in difficulty”). Even using the context created by the previous part of the paragraph, it is almost impossible to see the meaning of this sentence.

Even if we do not yet suspect plagiarism or paraphrasing, we must surely ask how the following text (from the Introduction) managed to get through peer reviews and into a journal: “At present, more than one billion individuals out of six billion individuals of the total populace don’t approach safe water and beyond what three billion individuals can’t forestall water-related sicknesses due to the absence of sanitation”. (original: “Currently, more than one billion people out of six billion people of the world population do not have access to safe water and more than three billion people cannot prevent water-related diseases because of the lack of sanitation”).

The Conclusion also includes mystifying sentences: “The 21st century is an age wherein various societies acknowledge singular contrasts and focus on concurrence and manageable”. (original: “The 21st century is an age in which diverse cultures accept individual differences and aim at coexistence and sustainable”). Interestingly, the original contains a mistake (the adjective sustainable without a noun) which the plagiarized text copies.

Any doubts about the use of paraphrasing are dispelled by the following sentence: “Toward the start of the advanced period, the word [literacy] alluded to the capacity to impart successfully in a proficient society, and was esteemed as an apparatus for individual achievement”. The use of advanced period, impart, and proficient suggest that these words were randomly chosen rather than used intentionally. (Original: “At the beginning of the modern era, the word referred to the ability to communicate effectively in a literate society and was valued as a tool for personal success”).

In the following text, the use of ought to raises the suspicion of paraphrasing: “Along these lines, expanded innovation ought to be presented and different strategies ought to be advanced . . . ”. The original uses the correct modal verb of obligation (should expresses a subjective opinion, whereas ought to is for an objective truth): “Thus, diversified technology should be introduced and various policies should be promoted . . . ”. In fact, all four uses of the word should have been replaced by ought to in the plagiarized text.

3.1.3. Missing Table and Figure

Reference is made to both a table and an image in the text, yet no table or image is included in the plagiarized text.

3.1.4. Using the Wrong Personal Pronoun

Unlike the original, only one author is listed in the plagiarized text, yet the text includes the personal pronoun “we” several times. For example, “we advocate the idea of water education” or “we characterize water literacy as . . . ”.

3.1.5. Title in the Wrong Place

The plagiarist added a new title Conclusion. The paragraph before the title (in the original) starting with “Secondly, water literacy . . . ” (it was changed to “Besides, water literacy . . . ” in the plagiarized text) and the paragraph after starting with “Third, water literacy . . . ”.

3.1.6. Attached Unrelated Text

The plagiarist added a new part Suggestions. This new text describes the use of water quality testing packs by children in India. This newly inserted text has no connection to the previous text.
3.1.7. Wrong Quotation and List of Literature

Citation-based plagiarism detection [37,38] can be used for identification of plagiarism. Therefore, some plagiarists change the literature list. In the analyzed plagiarized text, the list of literature was changed too. That leads to a situation when the first and third item in the list of Literature is not quoted in the text of article. An amusing situation occurs in the case of the third item when author “King, R.” was paraphrased to “Ruler” in the sentence “Ruler called this unique literacy practical education” (original sentence is “King [2] called this original literacy functional literacy”).

3.2. Result of Statistical Analysis

DiffChecker identified 817 differences between the original article and plagiarized text in 35 paragraphs of the original article, which creates the Abstract and the body of the original article and are used in the plagiarized text. For the paraphrased text, it was 821 differences between the paraphrased manuscript and the original article, but only 426 differences between the paraphrased article and the plagiarized text.

There are 3328 words in 35 paragraphs of the original article which creates the Abstract and the body of the original article and are used in the plagiarized text. The plagiarist changed every sentence (100%) in this text. In total, 1583 words were changed. This represents 48% of the text; the proportion of changed words varies between 28% and 65% in individual paragraphs. The number of words increased by 95 to 3423 words in these 35 paragraphs of the plagiarized text; in individual paragraphs it varied between −5 to +13 words.

Comparison of the paraphrased version with the original article shows similar values. The paraphrasing tool changed every sentence (100%) in this text. In total, 1643 words were changed. This represents 49% of the text; the portion of changed words varies between 28% and 72% in individual paragraphs. The number of words increased by 101 to 3429 words in these 35 paragraphs of paraphrased text; in individual paragraphs it varied between −4 to +14 words.

The number of changes between the paraphrased version and the plagiarized text is significantly different from the number of changes between the plagiarized and original article. The paraphrasing tool created 13 sentence equal to the plagiarized text. In total, the paraphrased text differs from the plagiarized text only in 486 words. This represents 14% of the text; the portion of changed words varies between 5% and 42% in individual paragraphs. The number of words increased by 6 words in these 35 paragraphs of paraphrased text; in individual paragraphs it varied between −6 to +7 words.

4. Discussion

The many errors in the plagiarized text illustrate both the author’s maximum efforts to simplify his/her work and the significantly poor (or no) peer review prior to publication of the text. Scientific peer review is a process when other researchers or experts in the appropriate field (reviewers) evaluate research findings for competence, significance, and originality. This process is not perfect since some reviewers may be unqualified and others are biased due to personal or professional rivalry, etc. However, when peer review is carried out casually, journals publish research that is flawed [39,40]. Grainger [41] stated that the peer review process is as good as the participants and recommended facilitating professional training it. The lack of quality in peer reviewing is an indicator of the predatory behavior of some publishers or conference organizers [42–44]. Makvandi et al. [45] suggested establishing a system for evaluating the quality of scientific conferences. Like scientific journals, it is difficult to cover every global scientific conference with an appropriate evaluation system. Renowned indexing services, such as Web of Science or Scopus, therefore only cover selected journals or conferences and conduct their own regular evaluation. Indexing services, such as Google Scholar, or social networks, such as Research Gate or Academia.edu, do not have such evaluation mechanisms by the very nature of how these services operate.
The plagiarist in question is an academic researcher from Walchand College of Arts and Science, Solapur Maharashtra, India. Mohanty [46] described several cases of scientific dishonesty in India. Misra et al. [47] identified 46 cases of retracted articles from India in the 2010–2017 period due to plagiarism in PubMed Database. In the Scopus database, there were 491 retracted documents with authors affiliated to India in August 2021 (from 2,547,508 documents). This may not seem to be an important portion of documents, but there was a clearly increasing trend between 2007 and 2017, with stagnation after this year. Shahare and Roberts [48] analyzed the historic background of academic dishonesty in the Indian academic system. One potential explanation could be the minimum qualifications required by academic staff in India; according to the University Grants Commission, in terms of published work, assistant professors need at least two papers, associated professors need at least seven papers, and full professors must have produced at least ten [49].

In 2010, Satyanarayana [50] proposed a draft National Plan of Action to combat plagiarism and argued for the necessity of urgent and serious consideration. In 2015, Juyal et al. [51] drew attention to the dangers that the practice posed to the Indian academic system and mentioned that India lacked an effective statutory body to deal with research misconduct in academia. This was changed in 2018, when India’s University Grants Commission (UGC) published regulations on plagiarism [52]. In the case of this type of plagiarism, the culprit shall be denied two annual increments and he/her shall not be allowed to be a guide to any undergraduate, postgraduate, M Phil, or PhD scholar for a minimum of 3 years [53]. In the context of this regulation, it is hard to understand the motivation of someone from an academic institution in preparing a plagiarized text. It is to be expected that they knew the UGC regulation and yet created an easily detectable plagiarized text and, in addition, posted it on Research Gate.

The unanswered question is why there are so many easily detectable errors in the plagiarized text. Any reader who speaks English at a reasonable level should easily detect them. It seems as if the author did not read the plagiarized text before submission to the conference. The author was asked by email about the number of errors in the plagiarized text, but this query went unanswered. It can therefore be assumed that plagiarism errors are the result of plagiarism and similar mistakes will be found in other plagiarism cases. This is corroborated, for example, by Fotlýnek et al. [54], who describe a case of the master’s thesis of a former Czech Minister of Justice, where several entire pages were adopted from another work and included grammatical errors, formatting, and typos. Weber-Wulf concluded that observing small quirks and mistakes in the text may help spot plagiarism [23].

An important task in paraphrased plagiarism detection is the decision about paraphrasing. In the first step, some spin detectors could be used. In this study, the original article and plagiarized text were analysed with an online spin detector [25]. The original article was marked as “likely written by a human”, and the plagiarized text was marked as “likely paraphrased by a machine”. The next logical step in identifying plagiarism is to try to find the original. Google is a very strong indexing machine. In the plagiarized text, the example of Iran and Spain was used. When the keywords “water literacy”, “Iran”, and “Spain” were used for searching on www.google.com, the original article was in second place in the list of results. However, it is not always successful. The newly added chapter Suggestion is probably also paraphrased, but the original source of these paragraphs was not found.

Plagiarism can lead to reinforcement of negative attitudes toward plagiarists, or even researchers from the same university or country, in academic communities all around the world [55]. Some journals blacklisted an author who regularly submitted plagiarized manuscripts [56]. The general approach to the effects of plagiarism is often influenced by laws and traditions relating to theft or copyright, which often do not fulfill the characteristics of plagiarism [5]. The Indian UGC recommends strict action against any faculty, staff, or researcher found guilty of the practice [53].

In this case, the plagiarist was contacted and, currently, the text in question is not available on his/her Research Gate profile. In addition, the publisher of ‘Research Jour-
ney’ International Multidisciplinary E-Research Journal was contacted by email too. The editorial staff re-sent this email to the plagiarist, asking for an explanation and an offer to retract the article. The answer, by email, was very interesting: “I am apologize for a paper, Research paper was writing project PG stundents, i was not cheked for the same. Please remove that”. (sic). If this admission is true, then it means that the plagiarist did not read the article by his/her students (the errors in the article are evident) and published it as his/her own. Poor English in the email response may support the argument that this type of academic misconduct could be due to low skill levels when working in a non-native language [57]. On the one hand, prevention is better than witch (plagiarist)-hunting. On the other hand, the presented plagiarized text is wholly inadequate.

Although the paraphrased version of the article is different from the plagiarized text, the changes between the plagiarized and paraphrased versions are importantly lower than the changes between the original and the plagiarized text. This can be seen as evidence that one of the paraphrasing tools was used in the creation of the plagiarized text. Another possible way to create such a text is to use translation into and out of another language. In this case, however, it is unlikely as all the sentences in the text have been changed, including very short and simple sentences. The translator would have translated these sentences unchanged. As Prentice and Kinden [11] have shown, translators maintain the usually settled terminology and do not always try to change the sentence’s text.

5. Conclusions

The content of one plagiarized text published in a scientific journal was analysed. In this case, the plagiarist used almost all of the original article. Every sentence from the original article was changed. To avoid the possibility of detection, he/she probably used one of the available paraphrasing tools. To confirm this assumption, the process used by the plagiarist was repeated and the newly created paraphrased version was compared with the plagiarized text. Conformity of the paraphrased version with the plagiarized text was significantly higher than when comparing the plagiarized text with the original.

Readers can easily identify many mistakes in this plagiarized text. These are typical for outputs from the paraphrasing tool. These typical errors contain inconsistent terminology, unclear sentence meaning, missing tables and figures, and the wrong literature list. It can be assumed that these typical errors in the analyzed text indicate a plagiarist’s effort to make the work of creating an article as easy as possible. Description of these errors can help identify this type of academic dishonesty in the other works.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/publications9040048/s1. Text of original article, plagiarized text and paraphrased text.


Funding: This research received no external funding.


Informed Consent Statement: Not Applicable.

Data Availability Statement: Data analysed in this case report is contained within the supplementary material.

Acknowledgments: We would like to thank anonymous reviewers for their helpful comments.

Conflicts of Interest: The authors declare no conflict of interest.
References
15. Eshchanov, B.; Abduraimov, K.; Ibragimova, M.; Eshchanov, R. Efficiency of “Publish or Perish” Policy—Some Considerations Based on the Uzbekistan Experience. Publications 2021, 9, 33. [CrossRef]
32. McCrorall, M.; Hamann, H. What We Know about Water: A Water Literacy Review. Water 2020, 12, 2803. [CrossRef]
45. Makvandi, P.; Nodehi, A.; Tay, F.R. Conference Accreditation and Need of a Bibliometric Measure to Distinguish Predatory Conferences. Publications 2021, 9, 16. [CrossRef]