

Lingua Franca and Beyond

Plagiarism: Is it a crime or just temptation?

How often when writing any text, you start by searching on the internet, using key words that are loosely linked to your vague idea of what to write and finally finding inspiration? Isn't it just one step to "borrowing" some words from what you found? Well, stop! If you do so, you will plagiarise, and act in an unethical and unfair way. But do you know what constitutes plagiarism and what does not? Andrew J. Davis explains in this article what plagiarism is, when a given text is considered plagiarised, and advises how to avoid plagiarism and, in particular, self-plagiarism. He also discusses the issue of acknowledgements and the complicated matter of copied images. It's

an interesting and relevant article for medical writers, and not only because of the main topic. I was fascinated to learn of the work of Elisabeth Bik: her outstanding skills for pattern recognition and her passion in tracing image manipulation in publications.¹ I also learned an abbreviation which I did not know before – *cf* – *confer* meaning compare with...

Finally, when searching for inspiration 😊 I found that Andrew is a biologist with an interest in environmental sciences and sustainability so maybe a good start for cooperation with our Sustainability Special Interest Group (<https://www.emwa.org/signs/sustainability-sig/>)?

SECTION EDITOR



Maria Kołtowska-Häggström, MD, PhD
maria.koltowska-haggstrom@
propermedicalwriting.com

Something to consider?

Last, but definitely not least, I would like to thank Andrew for contributing to our section, and to invite all of you to learn more about plagiarism.

References

1. https://en.wikipedia.org/wiki/Elisabeth_Bik. (Accessed 2023 April 8.)

Plagiarism, lack of acknowledgement, and image fraud

Andrew J. Davis

English Experience Language Services,
Göttingen, Germany
Institute of Zoology & Anthropology,
University of Göttingen, Göttingen, Germany
English.Experience@mayalex.u-net.com

doi: 10.56012/bqnf5251

Plagiarism

Plagiarism is the use of text written by other people but without indicating where the text comes from. This worries journals because they want to be sure of the link between the text of the work and the stated authors. As a consequence of this worry, many journals now automatically screen incoming manuscripts for evidence of plagiarism. The question is, though, how much reuse counts as plagiarism? There are some claims that journals have penalised authors for reusing very short pieces of text, or even for using organisational titles (e.g., World Health Organization). This is clearly ridiculous and should not happen. In such cases, authors are justified in protesting, and protesting strongly.

There are, however, several "official" definitions of how much reuse counts as

plagiarism. COPE (Committee on Publication Ethics¹) gives one definition. And individual anti-plagiarism applications (I use iThenticate) have their own recommendations. Taken together, these different definitions suggest that an overall identity of 15% or less is definitely not plagiarism and 25% or more definitely is. (And plagiarism of 15%–25% is an area where journal editors should exercise their discretion).

However, there is a problem with overall rates. This problem is that a text might have a low (<15%) overall rate of plagiarism but still have one or more paragraphs where the identity is much greater than 25% – even 100%.

As a journal editor, I therefore use this procedure. If the text has an overall identity of >20% I warn the authors. If the identity is >70% I reject the text on the grounds of plagiarism and inform the authors accordingly. Then, whatever the overall identity is, if any paragraph has identity >20% I warn the authors. Furthermore, if more than 20% of paragraphs have more than 20% similarity or any one paragraph identity

>70%, I reject on the grounds of plagiarism.

So, how might authors respond to claims from journals that their papers are plagiarised?

The answer depends on how much of a stand the authors want to make. If the overall level of identity is really low and the authors want their manuscript to be accepted, then they might as well make adjustments to remove that low degree

of plagiarism. The changes needed won't be very great.

However, if they want to take a stand, they should point out (politely!) that their overall level of similarity is well below what COPE and major software packages define as plagiarism and so their manuscript should not be rejected on those grounds. If, in contrast, the level of identity is high, the authors will have to come up with a convincing, but innocent, reason, or else, considerably

rewrite their manuscript so that there can be no suspicion of plagiarism.

Whatever thresholds are adopted, the COPE guidelines on this are worth following.

I ought here to mention the phenomenon of

If the text has an overall identity of >20% I warn the authors. If the identity is >70% I reject the text on the grounds of plagiarism and inform the authors accordingly.

“text recycling”. This differs from plagiarism only in that the text reused comes from other publications of the same author – it is, so to speak, self-plagiarism.² Authors often claim that they reuse their own text because, well, there’s just no better way of writing what they want to say. This claim is made particularly for Introduction or Methods sections. (I find it difficult to accept it could ever apply to the Results or the Discussion. If your Results and Discussion can be expressed with reused text then is your manuscript really a new manuscript?) But reuse of text is not required in the Introduction or the Methods anyway because you can refer to your previous papers. If, for example, you’re examining the same question, then you can write in the Introduction “The importance of this question is explained in detail in our previous paper (Yorick et al. 2021),” according to the normal rules of acknowledging ideas. And if the Methods are the same, you again cite your previous work. “We again analysed our samples with FT-MIR spectroscopy (cf Yorick et al. 2021).” There is thus no need to reuse blocks of text and risk raising suspicions of plagiarism.

Lack of acknowledgement

Reuse of ideas or content might be lack of acknowledgement, but it’s not plagiarism.

If, without using the exact words of another author, someone uses their ideas or content they should acknowledge the other author. They do that either by citing their published work, other easily available sources (such as preprints), or a personal communication. If they don’t acknowledge the other authors, and really did use their ideas, they are acting unethically.

It is extraordinarily difficult, however, to prove that someone is using your ideas or content. It is more than possible that different people independently have the same or similar ideas. For example, several people had evolutionary ideas before or contemporaneously with Charles Darwin. And Hooke, Schwann, and Schleiden all recognised the importance of cells. The idea or content has to be very special before you can prove that it’s specifically taken from you.

In general, as an editor, I would want ideas and content that have already been published to be acknowledged, but thereafter things get very complicated.

In general, as an editor, I would want ideas and content that have already been published to be acknowledged, but thereafter things get very complicated



Photo: Unsplash Ning Wang

on who is reusing, how much they’re reusing, why they’re reusing, who holds the copyright of the reused material, whether “fair use” applies, the specific policies of the publisher that first published the material, and also the specific policies of the publisher of the reused material. COPE says a great deal about these things but not in a single coherent place.

Copied images

Journals are not only worried about copied text but also about copied images. The reason for the worry is the same, journals want to be certain about the link between the creators of the material and the specified authors of a manuscript. Journals have a further worry about

images: this is that image reuse is very difficult to identify. Most known cases have only come to light because scientists have not been very clever and have, for example, used the same image in many different papers (in some cases more than

10 different papers!). And yet, image duplication is widespread.³ Therefore, to detect this problem, journals are implementing checks of all images in manuscript submissions. These checks use humans, computers, or both, as covered in a recent *Nature* article⁴ and elsewhere.⁵ Some humans are very good at identifying copied images (e.g., Elizabeth Bik). And, although automated identification of copied images is in its infancy, and as yet, none of the available systems are 100% effective, they are quickly becoming more powerful.

But again, it’s difficult to prove that an image is copied from a particular source unless there’s something very characteristic about the image (such as a specific arrangement of points). For example, if I produce an image of the cell lineages in *C. elegans* then it’s going to look similar to such an image produced by someone else. That is unavoidable because it’s showing the same phenomenon. To prove that the image was copied, and copied from a specific source, it would be necessary to show that the suspect image was very closely similar to the model and also, the suspected author had, in fact, seen the



Photo: Getty images

“Data is not copyrightable. Unless there is a creative element to data depiction that is being reproduced without alteration, fair use can be asserted, with attribution.”

model. Neither of these are easy tasks.

The problem of duplicated images is complicated by the enormous differences in publisher policies. For example, Princeton University Press⁶ says “Where a chart, graph, or table is being reproduced in a critical study of the work or to buttress an argument of the writer, no permission is needed. Data is not copyrightable. Unless there is a creative element to data depiction that is being reproduced without alteration, fair use can be asserted, with attribution.” However, and in contrast, Oxford University Press⁷ says “As a guide, you should always seek permission for: ...pictures (paintings, drawings, charts, engravings, photographs, cartoons, and so on); Figures and maps; Tables.”

However, to help authors decide on best policy, there is a broad brush set of guidelines at the STM site.⁸ (Learn more about STM at <https://www.stm-assoc.org/about-stm/>.) But finding guidelines for a particular case is very difficult indeed – there are just too many variables.

A wise course for authors, however, for text as

well as for images, is not to reuse any material without acknowledging its source and clearly attributing it to its original creators.

Disclosures and conflicts of interest

The author declares no conflicts of interest.

References

1. Committee on Publication Ethics. Available from: <https://publicationethics.org/>
2. Text recycling - COPE. Available from: https://publicationethics.org/files/Web_A29298_COPE_Text_Recycling.pdf
3. Bik EM, Casadevall A, Fang FC. The prevalence of inappropriate image duplication in biomedical research publications. *mBio*. 2016;7:e00809–e00816
4. Van Noorden R. Journals adopt AI to spot duplicated images in manuscripts. *Nature*. 2022;601:14–5. doi:10.1038/d41586-021-03807-6
5. Bucci EM. Automatic detection of image manipulations in the biomedical literature. *Cell Death Dis*. 2018;9:400. doi:10.1038/s41419-018-0430-3

Cell Death Dis. 2018;9:400.

doi:10.1038/s41419-018-0430-3

6. Permissions – Princeton University Press. Available from <https://press.princeton.edu/permissions>
7. Crediting copyrighted material – Oxford University Press. Available from <https://academic.oup.com/pages/authoring/books/preparing-your-manuscript/crediting-copyrighted-material?lang=en&cc=gb#lev2>
8. Permission Guidelines – STM Association. Available from https://www.stm-assoc.org/2022_01_27_STM_Permission_Guidelines_2022.pdf

Author information

Andrew Davis, PhD, has been a journal editor and reviewer for more than 30 years. He trained as a science text editor in the 1980s. Since then, he’s been a science editor and translator with English Experience Language Services. From 1994 onwards he has taught scientific writing at Universities in Germany and Japan.