
Viewpoints

On text reuse and the scholarship of science

Miguel Roig

Department of Psychology, St John's University, 300 Howard Avenue, Staten Island, New York 10301, USA; roigm@stjohns.edu

Abstract On the basis of existing guidance, it is proposed that reuse of a limited number of one's own or of others' phrases and perhaps even longer word strings should be permissible only when the material copied is composed of highly technical descriptions of complex processes or phenomena, which are most often found in methodology sections. Reuse of simpler non-technical text is not consistent with excellence in scholarship and should be strongly discouraged, particularly amongst native speakers of the language of publication.

Keywords Text reuse; plagiarism; self-plagiarism; authorship; publication ethics; scholarship; research misconduct.

Composing a scientific journal article can be an exceedingly demanding task even for those experienced researchers who are also native speakers of English – the *de facto* language of science. A manuscript's grammar and syntax must approach near perfection, and the language must be exceptionally clear and concise and devoid of any slang and unnecessary jargon. Moreover, authors must also follow basic principles of scholarship, which include, but are not limited to, the use of appropriate evidentiary support for each claim made and of footnotes or citations to identify the source of others' ideas or words. Regarding the reuse of others' words, the main concern of this paper, authors are expected to observe two general standards:

When paraphrasing others' work, an author must thoroughly modify the original text in a way that the paraphrased version acquires the author's own "voice" while conveying the exact same message as the original. The source from which the information is derived must be identified in accordance with the style of writing being used (Vancouver system, APA style, Chicago style, etc) by using footnotes or parenthetical or other forms of citation that clearly indicate the origin of the material used. In those cases in which paraphrasing the original would be so difficult that the paraphrase runs the risk of altering the meaning or fails to capture the unique elegance or essence of the original, we can use the original language in our paper by enclosing the verbatim text in quotation marks and providing a citation or footnote that clearly identifies its origin.

Unlike in the humanities, a number of constraints in the sciences make it difficult to apply and/or enforce these two basic rules. Scientific writing is often laden with unique terms and phrases for which there are no substitutes.¹ Consequently, sometimes it is very difficult, if not impossible, to thoroughly paraphrase certain segments of others' work. Consider the following short paragraph:

Mammalian histone lysine methyltransferase, suppressor of variegation 39H1 (SUV39H1), initiates silencing with selective methylation on Lys9 of histone H3, thus creating a high-affinity binding site for HP1. When an antibody to endogenous SUV39H1 was used for immunoprecipitation, MeCP2 was effectively coimmunoprecipitated; conversely, αHA antibodies to HA-tagged MeCP2 could immunoprecipitate SUV39H1.²

Even authors who have a good conceptual understanding of the general area of research from which the paragraph was drawn and who are also native speakers of English will find it difficult to paraphrase. For the growing segment of the scientific community who are non-native speakers of English, paraphrasing in general can be most challenging, let alone paraphrasing complex, scientific prose.

To further complicate matters, some non-native English speaking authors seem to lack the necessary training in relevant aspects of scholarly writing, such as the need for textual originality³ and transparency with respect to the sources from which they derive their own work.⁴ Some of these authors may have been taught a set of values related to using and to acknowledging others' work that is fundamentally different from the values associated with current English-language scientific scholarship.

Certain traditions in science writing play an important role in authors' ability to express themselves in a manner consistent with standard scholarly practices. For example, in scholarly work in the humanities one can frequently find portions of others' verbatim text appearing in quotation marks. This technique is used for a variety of purposes, such as highlighting or further elucidating the quoted material, or when an author wishes to emphasize a phrase, sentence, or paragraph that would help justify a particular position or counter-argument. Whether it is a matter of style or tradition, quoting others' text is an uncommon practice in the sciences⁵ that perhaps represents an expectation that authors should be able to articulate others' ideas in the authors' own words.

How much text reuse is acceptable?

It is not surprising, then, that so many cases of text plagiarism occur in the scientific literature.⁶ Putting language ability issues aside for the moment, the question of how much of others' verbatim text may be reused in one's publications and without attribution is one that sorely needs consensus. For example, some authors believe that copying a few sentences that do not embody an original idea is of "marginal importance"⁷ relative to the misappropriation of ideas. These authors also question whether both types of plagiarism should be treated in the same fashion.⁷⁻⁹

Others, who agree that distinctions should be made between the two types of plagiarism, take a more conservative approach and urge that the plagiarism label be used not only for the act of copying verbatim sentences from other sources, but also for the act of lightly paraphrasing them by changing only some of the words.¹⁰⁻¹¹ I note that at least some student writing guides in the sciences similarly caution readers about proper paraphrasing as a way of avoiding plagiarism.^{5,12}

The ideal, thorough paraphrase that we have come to expect is sometimes simply not possible – or even desirable – with the type of language often found in scientific journal articles. The question arises: when a thorough paraphrase is not feasible, how much text should authors be allowed to reuse? Few guidelines specifically address this important question. A guideline from the United States Office of Research Integrity (ORI)¹³ is helpful in this regard. ORI's working definition of plagiarism states: "Substantial unattributed textual copying of another's work means the unattributed verbatim or nearly verbatim copying of sentences and paragraphs which materially mislead the ordinary reader regarding the contributions of the author". The definition provides specific additional guidance on acceptable text reuse: "ORI generally does not pursue the *limited use of identical or nearly-identical phrases* which describe a commonly-used methodology or previous research because ORI does not consider such use as substantially misleading to the reader or of great significance" (emphasis mine).¹³ ORI allows for a limited amount of copying of phrases containing technical language, but not of sentences and paragraphs in a way that misleads the reader as to who the true author of the borrowed material really is.

Although ORI's definition is only applicable to instances of potential plagiarism in work that has been funded through the United States Public Health Service agencies, many US academic institutions, agencies, and professional organizations have adopted it in their research misconduct policies. The fact that a significant number of authors appear to stretch the concept of permissible text reuse from a limited number of phrases to entire sentences and paragraphs is perhaps the main reason for the explosive proliferation of articles on plagiarism.⁹

Perhaps the single most important dilemma for editors and authors when it comes to text reuse is the lack of an operationally defined guideline for when "limited use of identical or nearly-identical phrases" crosses the line from acceptable to unacceptable. At least one author has suggested a specific word count of 48 consecutive words for plagiarism,¹⁴ but this recommendation has never been enforced or even encouraged in any official capacity. Recently, Elizabeth Wager of the Committee on Publication Ethics (COPE) addressed this most difficult issue in a discussion paper that proposes a quantifiable distinction between major and minor plagiarism.¹⁵ While acknowledging the arbitrary basis for establishing any numerical threshold, she suggests that lapses that would qualify as major plagiarism include "verbatim copying of >100 words of original material in the absence of any

citation to the source material". For minor plagiarism, Wager suggests "verbatim copying of <100 words without indicating that these are a direct quotation from an original work (whether or not the source is cited), unless the text is accepted as widely used or standardized (eg, the description of a standard technique). Minor plagiarism would also include "close copying (not quite verbatim, but changed only slightly from the original) of significant sections (eg, >100 words) from another work (whether or not that work is cited)".

Although Wager's recommendations are an important and much needed step, they will probably need to be further elaborated because of the many clever forms in which inappropriate text reuse can occur that do not fit neatly into these preliminary categories. For example, with respect to the third guideline on minor plagiarism, consider an author who stitches together an entire article by misappropriating several paragraphs from one or more papers with one or two word substitutions per sentence and without attribution. In terms of the seriousness of the plagiarism, how would we compare such extensive copying to an author who has copied verbatim and without attribution two segments of 100-110 words in length? Again, the different ways in which text misappropriation can occur illustrates the difficulties with generating comprehensive, yet practical, guidance for authors and editors.

The problem of excessive text reuse from other sources is further compounded by the fact that some researchers, including native speakers of English, genuinely believe that as long as a citation is provided, segments of text from other sources may be reused with little or no modification.^{13,16} But, except perhaps for text segments that consist, to use ORI's wording, of "identical or nearly-identical phrases which describe a commonly-used methodology", such extensive reuse seems to me to fall short of scholarly excellence. I left out the phrase "or previous research", which is part of ORI's statement that was quoted earlier. In my view, the reuse of short amounts of others' verbatim text about previous research may, in some cases, be appropriate – but only when the language in question is technically challenging. Such language may be found throughout a paper, including the literature review, but it is most commonly found in the Methods section. Not all methods sections are challenging to rewrite, and even those that are highly complex may contain portions of text that are not difficult to paraphrase.

Most importantly, there seems to be an underlying assumption that just because a paper has been published, it cannot benefit from additional clarification or further elaboration.¹⁷ As most readers of the biomedical literature know full well, lapses in clarity, omission of key details, and other ambiguities in the writing are fairly common in journal articles. Rather than reusing the same written material, authors should, at the very least, view their new manuscript as a unique opportunity to possibly improve and further clarify what has already been written. Of course, such an approach is meaningful when the manuscript is being prepared by experienced authors who also have a full command of English.

Reusing portions of others' literature reviews is problematic for several reasons. Deception is associated with not having actually written the material, and the author may not even have read the articles cited in the misappropriated portions of the review. Perhaps with some exceptions, citing a particular work represents a declaration to the reader that we have read that work and that we are summarizing or distilling its relevant findings in our own words – unless, of course, we note otherwise by the use of a footnote or some other convention that informs the reader of a different situation. More importantly, literature reviews that are constructed, in part, as a patchwork of text from other sources could result in potentially serious misrepresentations of the scientific record.⁶

Reuse of own published text

Perhaps an even more contentious issue is the question of the extent to which authors may reuse their own previously published text in new publications. The practice appears to be relatively common, with at least one study showing that 60% of authors sampled reused at least 10% of their own text in subsequent publications.¹⁸ As with traditional forms of plagiarism, the practice of recycling one's own text probably ranges in scope from the reuse of a few stock phrases to the reuse of several paragraphs of a journal article.^{18,19}

Reusing certain key phrases or expressions is probably unavoidable and perhaps even desirable, as when the reuse occurs with portions of a previously published method section or some other highly complex description or process elsewhere in a paper. On the other hand, changing precise descriptions merely for reasons of appearing original risks altering their meaning in some subtle, but not inconsequential, way. However, the same cannot be said for literature reviews or discussion sections. Substantial text reuse in these usually non-technical sections not only falls short of scholarly excellence, it suggests intellectual laziness. Recycling our own text is not nearly as offensive as reusing others' text, but the practice conveys "poor scholarly etiquette" and may even trigger charges of misconduct if the reuse is sufficiently widespread.²⁰ Perhaps for these reasons limits for textual overlap had been suggested in the past, ranging from 10% to 30%. A recent model for determining whether self-plagiarism of text has taken place uses a 10% cut-off,²¹ and a consensus may be emerging in the medical editing community for a 10% maximum reuse.²²

While the reuse of one's own published text is far less serious than reuse of others' text, the increasingly multi-author nature of scientific publication can make matters more complicated for those who engage in this practice. Consider the researcher who was part of a research collaboration that published extensively but now joins a rival group working on the same type of research problem. Would this researcher be justified in reusing substantial portions of text that had been previously published with the old collaborative group, even if she or he had written the earlier material? Would the right to reuse previously co-authored text depend on his or her relative position in the lab (post doc, fellow, head of the lab)? And what if she or he didn't do any of the writing? As a coauthor of

previously published papers, is the researcher still entitled to reuse any portion of the writing from papers produced with the previous collaborators or from grant proposals, or other relevant documents? These are difficult questions and their resolution will likely depend on a detailed analysis of the many relevant variables present in each individual case beyond the crossing of a simple minimum percent threshold of text reuse.

Text reuse and English proficiency

An important concern is the question of how to handle text reuse committed by non-native English speakers. For example, should such authors be automatically held to more lenient standards of text reuse from their own publications? We often hear how limited English skills are used as a rationale for authors' tendency to misappropriate text.²³ As a group, these authors are probably at a linguistic, as well as at an economic, disadvantage relative to their native English counterparts. However, as I have argued elsewhere in this issue,²⁴ I believe that we commit a fundamental error when we assume that all non-native English speaking authors who plagiarize are, in fact, operating with the same low levels of linguistic and economic resources or that in all such cases the misappropriation is being committed for purposes of "scientific English."¹⁵ There are probably vast within-group differences in terms of ability to write in English, availability of resources to help them produce high quality manuscripts in English, and even motivation for and commitment to scientific truth. I urge flexibility in dealing with this group of authors but also caution that each case of a suspected ethical lapse should be judged on its own merits.¹⁷

Conclusion

Text reuse of the copy-paste type is not generally consistent with a "best practices" approach to scientific writing. While there is some practical justification for reusing limited amounts of technical text, the reuse of non-technical text, even if it is from our own published work, should be actively and unambiguously discouraged.

In those cases in which it is clear that the reuse was committed for purposes of "scientific English" and where the authors have very limited resources to address this problem, then perhaps a more lenient approach should be followed. If the offense is committed by an experienced native-speaker who should have known better, a less forgiving approach ought to be in order.

Why should we not allow non-native English speaking authors to "borrow part of a well-written phrase or even a whole sentence (with appropriate citation of course) from another published article to better express what they want to say?"⁹ The simple answer is that text reuse that is accompanied by a citation but not enclosed in quotation marks misleads the reader about the true authorship of that text. Given that accuracy and intellectual honesty are some of the hallmarks of scientific work, we should be encouraging those traits in every facet of the scientific process, including the dissemination of our findings. Bypassing long-established scholarly traditions and allowing authors to freely reuse their own or others' text,

even if only in small portions, is not consistent with a true scientific spirit. A more honest and transparent strategy other than paraphrasing, is to reuse text by enclosing it quotation marks and adding a citation to identify its origin. The biomedical editorial community should begin to consider encouraging the use of this option. If authors were more comfortable about using quotations, perhaps cases of text plagiarism would finally begin to decline.

Competing interests MR has written an online instructional resource, with support from the US Office Research Integrity, to help authors avoid plagiarism and other unethical writing practices.

References

- 1 Roig M. Some reflections on plagiarism: the problem of paraphrasing in the sciences. *European Science Editing* 2007;33(2):38–41.
- 2 Lunyak VV, Burgess R, Prefontaine GG, Nelson C, Sze SH, Chenoweth J, Schwartz P, Pevzner PA, Glass C, Mandel G, Rosenfeld MG. Corepressor-dependent silencing of chromosomal regions encoding neuronal genes. *Science* 2002;298(5599):1747–1752. doi:10.1126/science.1076469
- 3 Vasconcelos S, Leta J, Costa L, Pinto A, Sorenson MM. Discussing plagiarism in Latin American science. Brazilian researchers begin to address an ethical issue. *EMBO Reports* 2009;10(7):677–682. doi:10.1038/embor.2009.134
- 4 Heitman E, Litewka S. International perspectives on plagiarism and considerations for teaching international trainees. *Urologic Oncology: Seminars and Original Investigations* 2011;29(1):104–108. doi:10.1016/j.urolonc.2010.09.014
- 5 Knisely K. *A student handbook for writing in biology*. 3rd ed. Sunderland, MA: Sinauer and Associates, 2009.
- 6 Shafer SL. You will be caught. *Anesthesia and Analgesia* 2011;112(3):491–493. doi:10.1213/ANE.0b013e3182095c73
- 7 Bouville M. Plagiarism: words and ideas. *Science and Engineering Ethics* 2008;14:311–322. doi:10.1007/s11948-008-9057-6
- 8 Flowerdew J, Li Y. Language re-use among Chinese apprentice scientists writing for publication. *Applied Linguistics* 2007;28(3):440–465. doi:10.1093/applin/amm031
- 9 Habibzadeh F, Marcovitch H. Plagiarism: The emperor's new clothes. *European Science Editing* 2010;37(3):67–69.
- 10 Afifi M. Plagiarism is not fair play. *Lancet* 2007;369(9571):1428. doi:10.1016/S0140-6736(07)60662-X
- 11 Farrokhi F. Plagiarism: where unawareness makes a lame excuse. *Archives of Iranian Medicine* 2009;12(2):176–178.
- 12 Pechnick J A. *Short guide to writing about biology*. 7th ed. New York: Addison Wesley Longman, 2010.
- 13 ORI's working definition of plagiarism. *ORI Newsletter* 3(1):3. <http://ori.hhs.gov/policies/plagiarism.shtml> (accessed 29 May 2011).
- 14 Julliard K. Perceptions of plagiarism in the use of other authors' language. *Family Medicine* 1994;26(6):356–360.
- 15 Wager E. *How should editors respond to plagiarism? COPE discussion paper*. http://www.publicationethics.org/files/COPE_plagiarism_discussion_%20doc_26%20Apr%2011.pdf (accessed 28 May 2011).
- 16 Roig M. Plagiarism and paraphrasing criteria of college and university professors. *Ethics and Behavior* 2001;11(3):307–323.
- 17 Roig M. Plagiarism: consider the context. *Science* 2009;325(5942):813–814. doi:10.1126/science.325_813c
- 18 Bretag T, Carapiet S. A Preliminary study to identify the extent of self-plagiarism in Australian academic research. *Plagiary: Cross-Disciplinary Studies in Plagiarism, Fabrication, and Falsification* 2007;2:92–103. <http://quod.lib.umich.edu/p/plag/images/5240451.0002.010.pdf> (accessed 14 April 2011).
- 19 Roig M. Re-using text from one's own previously published papers: an exploratory study of potential self-plagiarism. *Psychological Reports* 2005;97:43–49
- 20 Scanlon P M. Song from myself: An anatomy of self-plagiarism. *Plagiary: Cross-Disciplinary Studies in Plagiarism, Fabrication, and Falsification* 2007;2:57–66. <http://quod.lib.umich.edu/p/plag/images/5240451.0002.007.pdf> (accessed 29 May 2011).
- 21 Bretag T, Mahmud S. Self-plagiarism or appropriate textual re-use? *Journal of Academic Ethics* 2009;7:193–205. doi:10.1007/s10805-009-9092-1.
- 22 Kravitz RL, Feldman MD. From the editors' desk: self-plagiarism and other editorial crimes and misdemeanors. *Journal of General Internal Medicine* 2011;26(1):1. doi:10.1007/s11606-010-1562-z
- 23 Yilmaz I. Plagiarism? No, we're just borrowing better English. *Nature* 2007;449(7163):658. doi:10.1038/449658a
- 24 Roig M. On plagiarism, laziness, and scholarship. *European Science Editing* 2011;37(3):69–70.

How I heard of EASE

I first heard of EASE in 1995 when I joined the Centre Technique pour la Cooperation Agricole et Rurale in Wageningen, NL, as an interim generalist consultant. One of my multifarious pick-up-the-pieces tasks after the sudden departure of a publications manager was to fast track the editing and publication of several conferences' proceedings. They were dallying unmanaged in a pipeline, and I was told to subcontract the editing to a Mr Brian Mills somewhere in southern Massif Central, as he knew the topics in question. A wily and wise editor, Brian happily submitted to my management, advising me all the time how to manage him and instructing me, a novice, in the editing process. Wonderful! He was an active member of EASE, and amongst his many counsels

was that I should sign up myself.

I demurred, never having held a blue pen in my life. Now, 16 years on, I survive the ends of my months as a translator and (sometimes scientific) editor, especially when the World Bank and several governments find no need for my mediating, writing, and filming skills. Now fully motivated to develop my editing skills, I joined Mediterrean Editors and Translators (Barcelona) on the advice of a client in Montpellier. Splendid advice. A few weeks ago, in MET, I learned of the Barcelona seminar on peer review, and that prompted me to put right my omission of years ago.

Paul Osborn
paul@osborn.nl