

Encouraging Editorial Flexibility in Cases of Textual Reuse

Miguel Roig

Department of Psychology, St. John's University,
Staten Island, NY, USA

Received: 10 January 2017

Accepted: 12 January 2017

Address for Correspondence:

Miguel Roig, PhD

Department of Psychology, St. John's University, 300 Howard
Avenue, Staten Island, NY 10301, USA
E-mail: roigm@stjohns.edu

Because many technical descriptions of scientific processes and phenomena are difficult to paraphrase and because an increasing proportion of contributors to the scientific literature are not sufficiently proficient at writing in English, it is proposed that journal editors re-examine their approaches toward instances of textual reuse (similarity). The plagiarism definition by the US Office of Research Integrity (ORI) is more suitable than other definitions for dealing with cases of ostensible plagiarism. Editors are strongly encouraged to examine cases of textual reuse in the context of both, the ORI guidance and the offending authors' proficiency in English. Editors should also reconsider making plagiarism determinations based exclusively on text similarity scores reported by plagiarism detection software.

Keywords: Plagiarism; Publication Ethics; Periodicals as Topic

INTRODUCTION

Plagiarism, a concept grounded in the humanities, refers to the misappropriation of others' work (e.g., ideas, text, images, design elements, structural properties, data, processes, musical notes) as one's own. A less serious and somewhat controversial concept, which is often discussed in the context of plagiarism, is self-plagiarism. It refers to the passing off of one's own previously disseminated work (ideas, text, images, etc.) as new content. These types of scholarly misbehaviors have become a subject of great concern in the sciences. Although most scholarly journals warn prospective authors against committing plagiarism and self-plagiarism, these are some of the most common transgressions observed by biomedical journal editors (1).

The situation has become so alarming that some editors have publicly complained about the large number of journal submissions with plagiarized or self-plagiarized materials (2-4). Cases of plagiarism have also been discovered in published articles, resulting in corrections or retractions (5,6). Plagiarism is now listed as one of the major forms of research misconduct in the national policies of countries whose scientists contribute a significant portion of the scientific literature (7).

CURRENT GUIDANCE

Regardless of whether the medium is an Introduction, Methods, Results and Discussion (IMRAD)-type journal article, a review of the literature, or a book chapter, there is a general expectation that when using the ideas and/or work of others authors will follow long-established rules of quotation and attribution. Specifically, when writing about the work of others or their own previously published work authors must either: 1) enclose in

quotation marks any verbatim (word-for-word) text or, depending on the length of the borrowed material, block-indent it, and add a proper citation to identify its source; or 2) thoroughly paraphrase and/or summarize the material and add a citation.

Paraphrasing and summarizing are two distinct writing strategies that may be used either separately or in combination. When we paraphrase someone else's work we restate in our words that author's ideas and typically do so using roughly the same amount of text. When we summarize, we refer to scientific facts, convey only key ideas, and do so in a more condensed fashion. For example, we might summarize the contents of an entire paper in several paragraphs, a single paragraph, or even in a single sentence, depending on how much detail we wish to provide about that paper.

The use of quotation marks as a means of indicating that the quoted text is not the author's own is common in the humanities but not in the natural sciences. Perhaps one reason for that tradition is that there is nothing so uniquely elegant about the descriptions of processes and/or scientific phenomena that researchers write about to merit their verbatim repetition. Instead, the emphasis in scientific journal articles lies in maintaining authors' objectivity and precision in their concise descriptions of observations and in striving for utmost clarity in their writing. Another possible reason for the infrequent appearance of quoted texts in scientific papers is because there has been a general expectation that scientists should be able to read, analyze, and convey others' work using their own words and writing style. Although the latter rationale may have been reasonable decades ago, it has become less so in recent years given the increasing numbers of non-native English-speaking researchers with limited ability to produce polished English.

WRITING PRACTICES THAT MAY BE DEEMED AS PLAGIARISM

Evidence suggests that some authors, regardless of their discipline and language proficiency, engage in inappropriate writing practices that could be construed as plagiarism (8,9). For example, there are those who believe that it is acceptable to simply reuse others' text in their papers and provide a correct reference. Steven Shafer has coined such a practice as 'technical plagiarism' (3) and others deem it unacceptable as well (10).

Another instance is when an author substantially paraphrases others' work but fails to provide a citation (2,11) or, arguably the most common practice, an author superficially paraphrases others' text by making only minor superficial modifications (e.g., changing a couple of words) and adding a citation (12). All these practices are problematic and, depending on the specific circumstances such as the extent of the questionable content, may be classified as plagiarism. It should be stressed, however, that paraphrasing with minor modifications may be quite acceptable for some forms of highly technical text, for example, in methods section.

LIMITATIONS OF CURRENT DEFINITIONS OF PLAGIARISM

Concern about the incidence of plagiarism is evident from numerous papers on publication ethics published recently (13,14). A review of 63 editorials that included discussion of plagiarism, published in 2008–2012 revealed a fairly uniformed message: plagiarism and self-plagiarism are major problems in science; editors are alarmed by these and related matters of research integrity; publishers have tools to catch offenders; and authors' misdeeds will result in serious negative consequences (15).

Despite these cautionary messages, little agreement exists about key aspects of plagiarism. With respect to paraphrasing, how extensively should the original text be modified to pass the 'plagiarism test?' Specifically, how many consecutive words in a sentence can an author reuse from other sources without incurring a charge of plagiarism? Unfortunately, there has been little discussion, let alone universal agreement, about this fundamental question. Similarly, the question of how much reused text (i.e., percent similarity score) should a manuscript have to merit a charge of plagiarism varies across journals and can range from 5% to 25% (16,17). And such determinations may depend on whether the percent of similar text is derived from a single source as opposed to many sources (18). Some of these same questions apply to so-called text recycling (i.e., self-plagiarism): how much of their own previously disseminated text may authors be free to covertly recycle (i.e., without a citation and quotation marks) in a new paper? Also, should it matter whether the recycled text comes from one or more papers by the same

author? One informal poll (19) revealed that acceptable similarity scores for text recycling range from 10% to 30% with some editors allowing text reuse from several sources as long as citations are provided (20,21). At least one editor, however, insists that any such recycling must be enclosed in quotation marks (22).

CURRENT DILEMMA

Journal articles have long been the means by which scientists communicate their findings to their peers and although various languages have played a key role at different periods, English is now the de-facto language of scientific communications. Even scholarly journals from non-English-speaking countries are now publishing exclusively in English. Yet, a significant portion of the scientific literature is generated by scientists from non-English-speaking nations where education and cultural traditions do not emphasize relevant concepts of intellectual property as much as in English-speaking western nations. Perhaps not surprisingly, the former group is likely responsible for a large proportion of submissions containing plagiarism (23,24).

The fact is that scholarly writing is not easy even for some native English-speaking authors. In addition to building up a good language vocabulary and familiarizing themselves with the many nuances of English grammar and syntax, scholarly authors must also learn the vocabulary of their specialty, have a solid conceptual understanding of pertinent issues, and be able to express themselves in the unique writing style commonly used in the sciences. It can take many years for native-English speakers to develop the proper skills to produce good papers. The types of difficulties experienced by non-English scientists, including limited access to necessary resources (e.g., effective and affordable translation services), together with the shortcomings of definitions of plagiarism and the lack of uniformity with which editors view these issues (15) justify continuing discussion of possible reforms.

Vessal and Habibzadeh (25) proposed that plagiarism of text be reappraised in the context of scientific publishing. Others have similarly called for greater emphasis on the 'science' reported rather than on the 'language' with which the science is being reported (26,27). These proposals are understandable given that the strict application of scholarly rules—rules that were conceived from the perspective of the humanities—may be too constraining when applied to technical language. In fact, evidence suggests that when faced with difficult-to-paraphrase text, students and university professors seem to naturally appropriate longer text strings in their paraphrases of others' work (9,28). This same tendency, which has also been observed in non-English writers (29) may, for obvious reasons, occur with greater frequency in this particular group. Such writing strategies may be unavoidable in situations where textual material is

made up of unique terms that have no appropriate equivalents. For example, consider the following excerpt: “Mammalian histone lysine methyltransferase, suppressor of variegation 39H1 (SUV39H1), initiates silencing with selective methylation on Lys⁹ of histone H3” or “When an antibody to endogenous SUV39H1 was used for immunoprecipitation, MeCP2 was effectively coimmunoprecipitated; conversely, α HA antibodies to HA-tagged MeCP2 could immunoprecipitate SUV39H1” (30). Certainly, this type of material can be paraphrased, but not with the same degree of ease as one is able to paraphrase the less technical text (31), and the ability to do so correctly will surely depend on the author’s writing skills.

SIMILARITY SCORES AND INAPPROPRIATE PARAPHRASING

Given the above considerations, perhaps it is time for editors to reconsider the use of arbitrary percent similarity scores when determining whether an author has plagiarized. A rethinking of this approach is especially needed at a time when textual infractions, namely, inappropriate paraphrasing, are made by increasing numbers of well-meaning scientists whose primary intentions are to contribute to science, but who lack the necessary writing skills in English. To that end, editors should be encouraged to take a close look at the definition of plagiarism offered by the US Office of Research Integrity (ORI) (32). That definition states (underlined sections represent my emphasis):

“As a general working definition, ORI considers plagiarism to include both the theft or misappropriation of intellectual property and the substantial unattributed textual copying of another’s work. 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.”

ORI’s definition uses the word ‘substantial,’ which means ample, significant, considerable, a large amount or quantity. A question does arise as to what proportion of textual copying corresponds to the word ‘substantial.’ Be that as it may, however, the important point of the definition is that large amounts of verbatim or nearly verbatim copying of sentences which mislead the reader about the true contributions of the author, represent plagiarism. This means that taking verbatim sentences from another source and adding a citation is considered plagiarism as is the inappropriate paraphrasing of sentences by merely changing 1 or 2 words, thus making them ‘near verbatim.’ Consistent with traditional definitions, ORI’s definition of plagiarism assumes that paraphrasing others’ text must be accomplished by substantially modifying the original material and adding a citation to clearly indicate the source of the content.

There is, however, one element of ORI’s definition that appears to recognize the need for flexibility when paraphrasing techni-

cal text (32): “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.”

Those who wrote ORI’s definition of plagiarism seem to have understood that there are segments of text that are extremely difficult, if not impossible, to paraphrase according to traditional rules of scholarship without running a risk of altering the intended meaning of the original. Thus, the need to convey accuracy that can potentially result in a near verbatim paraphrase likely to be flagged as plagiarism must be carefully weighed. By the same token, editors may have to sacrifice their high standards of scholarship for lightly paraphrased material that contains too many “identical or nearly-identical phrases” but “which describe a commonly-used methodology or previous research.”

As per ORI’s definition, it is important to emphasize that phrases are not sentences, that any reuse even at the phrase level should be kept to a minimum, and that minor modifications are only acceptable with text that describes a complex technical method or previous research that similarly involves highly technical language. Again, a question arises as to what proportion of lightly paraphrased material should be allowed before the questionable corpus crosses the threshold of unacceptability. Here is where editors must avoid making a decision based on some arbitrarily-set, percent of text similarity threshold. These situations call for a thoughtful assessment as to what types of textual material are being flagged by plagiarism-detection software (33) and do so with special consideration for the language proficiencies of the authors.

CONCLUSION

Even though the above suggestions are quite modest, there may be some editors unwilling to sacrifice their high scholarly standards of zero tolerance for textual reuse. On the other hand, editors who wish to ‘do the right thing’ may also be unable to do so because there is no agreed-to standard for determining authors’ actual level of English language proficiency. In spite of these and other shortcomings, but in view of the difficulties experienced by a growing segment of our non-native-English peers, it is hoped that the community of science editors will carefully consider the above proposal.

NOTE

This opinion piece is based on presentation titled: “Scientific vs. academic plagiarism: Is it time for editors to make a distinction?” given on November 18, 2016 at the IV Brazilian Meeting on Research Integrity, Science and Publication Ethics (BRISPE) conference held at The Federal University of Goiás, Goiás, Brazil

(<http://brispe2016.org/>).

DISCLOSURE

The author, whose native language is Spanish, is now dominant in English. He is also the author of the instructional module on plagiarism hosted by the US Office of Research Integrity (ORI) titled: "Avoiding plagiarism, self-plagiarism, and other unethical writing practices: A guide to ethical writing." The opinions expressed in this paper are his own and not those of ORI.

ORCID

Miguel Roig <http://orcid.org/0000-0001-5311-5651>

REFERENCES

1. Wager E, Fiack S, Graf C, Robinson A, Rowlands I. Science journal editors' views on publication ethics: results of an international survey. *J Med Ethics* 2009; 35: 348-53.
2. Hausmann L, Murphy SP; Publication Committee of the International Society for Neurochemistry (ISN). The challenges for scientific publishing, 60 years on. *J Neurochem* 2016; 139 Suppl 2: 280-7.
3. Shafer SL. You will be caught. *Anesth Analg* 2011; 112: 491-3.
4. Zhang Y. Chinese journal finds 31% of submissions plagiarized. *Nature* 2010; 467: 153.
5. Almeida RM, de Albuquerque Rocha K, Catelani F, Fontes-Pereira AJ, Vasconcelos SM. Plagiarism allegations account for most retractions in major Latin American/Caribbean databases. *Sci Eng Ethics* 2016; 22: 1447-56.
6. Fang FC, Steen RG, Casadevall A. Misconduct accounts for the majority of retracted scientific publications. *Proc Natl Acad Sci U S A* 2012; 109: 17028-33.
7. Resnik DB, Rasmussen LM, Kissling GE. An international study of research misconduct policies. *Account Res* 2015; 22: 249-66.
8. Roig M. Plagiarism and paraphrasing criteria of college and university professors. *Ethics Behav* 2001; 11: 307-23.
9. Sun YC, Yang FY. Uncovering published authors' text-borrowing practices: paraphrasing strategies, sources, and self-plagiarism. *J Engl Acad Purp* 2015; 20: 224-36.
10. The insider's guide to plagiarism. *Nat Med* 2009; 15: 707.
11. Jacobs H. From and to a very grey area. *EMBO Rep* 2011; 12: 479.
12. Foster RL. Avoiding unintentional plagiarism. *J Spec Pediatr Nurs* 2007; 12: 1-2.
13. Habibzadeh F, Marcovitch H. Plagiarism: the emperor's new clothes. *Eur Sci Ed* 2011; 37: 67-9.
14. Weems M. Plagiarism in Review. *Pediatr Rev* 2017; 38: 3-5.
15. Roig M. Journal editorials on plagiarism: what is the message? *Eur Sci Ed* 2014; 40: 58-9.
16. Peh WC, Arokiasamy J. Plagiarism: a joint statement from the Singapore Medical Journal and the Medical Journal of Malaysia. *Med J Malaysia* 2008; 63: 354-5.
17. Swaan PW. Publication ethics--a guide for submitting manuscripts to pharmaceutical research. *Pharm Res* 2010; 27: 1757-8.
18. Zhang Y, Jia X. A survey on the use of CrossCheck for detecting plagiarism in journal articles. *Learn Publ* 2012; 25: 292-307.
19. Kravitz RL, Feldman MD. From the editors' desk: self-plagiarism and other editorial crimes and misdemeanors. *J Gen Intern Med* 2010; 26: 1.
20. Drummond GB. Reporting ethical matters in The Journal of Physiology: standards and advice. *J Physiol* 2009; 587: 713-9.
21. Kohler CS; American Diabetes Association. Updates to policies and procedures related to potential scientific and academic misconduct in the Journals of the American Diabetes Association. *Diabetes Care* 2012; 35: 189-90.
22. Bonnell DA, Hafner JH, Hersam MC, Kotov NA, Buriak JM, Hammond PT, Javey A, Nordlander P, Parak WJ, Schaak RE, et al. Recycling is not always good: the dangers of self-plagiarism. *ACS Nano* 2012; 6: 1-4.
23. Bohannon J. Study of massive preprint archive hints at the geography of plagiarism [Internet]. Available at <http://news.sciencemag.org/scientific-community/2014/12/study-massive-preprint-archive-hints-geography-plagiarism> [accessed on 9 January 2017].
24. Citron DT, Ginsparg P. Patterns of text reuse in a scientific corpus. *Proc Natl Acad Sci U S A* 2015; 112: 25-30.
25. Vessal K, Habibzadeh F. Rules of the game of scientific writing: fair play and plagiarism. *Lancet* 2007; 369: 641.
26. Flowerdew J, Li Y. Language re-use among Chinese apprentice scientists writing for publication. *Appl Linguist* 2007; 28: 440-65.
27. Habibzadeh F, Shashok K. Plagiarism in scientific writing: words or ideas? *Croat Med J* 2011; 52: 576-7.
28. Roig M. When college students' attempts at paraphrasing become instances of potential plagiarism. *Psychol Rep* 1999; 84: 973-82.
29. Sun YC. Does text readability matter? A study of paraphrasing and plagiarism in English as a foreign language writing context. *Asia Pac Educ Res* 2012; 21: 296-306.
30. Lunyak VV, Burgess R, Prefontaine GG, Nelson C, Sze SH, Chenoweth J, Schwartz P, Pevzner PA, Glass C, Mandel G, et al. Corepressor-dependent silencing of chromosomal regions encoding neuronal genes. *Science* 2002; 298: 1747-52.
31. Roig M. Avoiding plagiarism, self-plagiarism, and other questionable writing practices: a guide to ethical writing [Internet]. Available at <https://ori.hhs.gov/sites/default/files/plagiarism.pdf> [accessed on 9 January 2017].
32. Office of Research Integrity Newsletter (US). ORI provides working definition of plagiarism [Internet]. Available at https://ori.hhs.gov/images/dbblock/vol3_no1.pdf [accessed on 9 January 2017].
33. Lykkesfeldt J. Strategies for using plagiarism software in the screening of incoming journal manuscripts: recommendations based on a recent literature survey. *Basic Clin Pharmacol Toxicol* 2016; 119: 161-4.