RESEARCH ARTICLE

Author's version vs. publisher's version: an analysis of the copy-editing function

Edward WATES and Robert CAMPBELL
Blackwell Publishing Ltd

© Edward Wates and Robert Campbell 2007

ABSTRACT: This report describes an informal study carried out by Blackwell Publishing to assess whether the copy-editing and proof-correction process alone results in a significant difference between the author's version and the publisher's version of an article accepted for publication. One hundred and eighty-nine articles were reviewed from 23 journals. The results indicate that a substantial number of changes are made. It is suggested that copy-editing has an equal role to play in both the printed and online environments, and that in the latter it contributes substantially to the accuracy of the electronic version. Copy-editing is therefore an important function within the publisher's overall responsibility towards the integrity of the article of record.

Edward Wates

Robert Campbell

Introduction

As the concept of open access (OA) has developed, two different ways of achieving it have emerged. The author-pays or pay-to-publish model was almost synonymous with OA initially, but now archiving articles on institutional or subject repositories (self-archiving) for free access over the Internet seems to be favoured by many research-funding bodies.

Proponents of this second model argue that self-archiving and journal publishing can coexist peacefully,[1] but many publishers feel that widespread availability of articles on institutional repositories could eventually lead to libraries cancelling subscriptions. A study by Ware[2] suggested this could happen when the majority of articles are being self-archived, and this has been confirmed by a recently published and more comprehensive study by Beckett and Inger.[3]

Another interesting conclusion from Beckett and Inger was that, although librarians appreciate the importance of peer review, they see little distinction between the different post-peer-review versions of an article: the 'author's version' and the 'publisher's version'. This finding is highly significant because until now publishers have been uncertain of the potential damage from the widespread availability of the author's version, with some publishers believing that librarians will continue to support the final published version. The publishers that have allowed authors to self-archive the author's version of the accepted article on publication may now review that policy.

It should be no surprise to publishers that their contribution to scholarly communication is undervalued. At the height of the Napoleonic wars the poet and co-founder of University College London, Thomas Campbell, successfully proposed a toast to

Napoleon at a literary dinner on the grounds that he had just been informed that Napoleon had executed a publisher. In a survey carried out by Rowlands, Nicholas and Huntington[4] one of the conclusions of this group (now at University College London) was that authors of research papers do not fully appreciate what publishers do.

Publishers have not been particularly good at promoting themselves, although they have made more of an effort recently in response to the OA lobby. Part of the problem is the lack of analysis of the value added in the various steps in the publishing process. As a small step we decided to look at the copy-editing function, to examine whether some librarians are right – as indicated by Beckett and Inger[3] – to disregard the value of the published version over the author's version of the accepted (i.e. post-peer review) article. Obviously, copy-editing is only one part of the production process, which includes a range of additional functions such as tagging, linking, image processing and general quality control (both in print and online).

Aim

The study set out to assess whether the copy-editing and proof-correction process results in a significant difference between the author's version and the publisher's version of an article. Although one aspect of copy-editing is the imposition of journal style (standardization of units and abbreviations, formatting of reference citations, styling of heading hierarchies, etc.), it was decided to exclude the application of such journal-specific conventions from the study. It can be argued that the traditional requirements to implement editorial and typographic consistency are part of the conventions that have grown up around printed journals and are less significant within the online environment. While not insignificant, the question of legibility and ease of understanding of on-screen material is more appropriately the subject of other studies.

This study therefore focuses on those changes that are made as a result of copy-editing and could materially distinguish the author's version of the article from that of the publisher.

Definition of terms

In this study, the 'author's version' is understood to mean the accepted version of the article that has been forwarded to the publisher by the editorial office. It therefore includes all those changes that have been made as a result of the peer-review process.

The 'publisher's version' is understood to mean the final version of the article as published within the journal. Only those articles that had been published in issue format with full bibliographic details (volume, issue, page number) were included in this study, although potentially the study could also have included those papers that had been published as individual articles online ahead of print (but which have not yet been allocated final page numbers, volume or issue details).

The copy-editing process involves an experienced copy-editor reading through the author's version of the article and applying journal-specific style as well as checking for sense, grammar and internal consistency. Once the article has been copy-edited, it is sent to the typesetter for conversion to XML and subsequent print pagination. The process of XML creation establishes the tags which both enable internal links to be made within the article and facilitate linking to external databases such as CrossRef and PubMed. Often queries from the copy-editing and proof-correction process concern the integrity of references, which, if not corrected, would hinder this linking to external databases.

Further details regarding Blackwell's copy-editing conventions are contained within the *Blackwell House Style Guide* (available at http://www.blackwellpublishing.com/housestyle/).

the study set out to assess whether the copy-editing and proof-correction process results in a significant difference between the author's version and the publisher's version of an article

Table 1 Number of articles and journals included in the study

	No. of journals	No. of articles
STM	17	145
HSS	6	44
Total	23	189

Table 2 Research questions

Manuscript	Proofs				
No. of author queries raised by copy-editor	No. of typographic changes to proof	No. of author changes to copy-editing	No. of alterations to the text	No. of changes made as a result of copy-editor's queries	No. of copy-editor's queries ignored

Methodology

The study was carried out during August and October 2006. A total of 189 articles was studied based on a random selection from 23 journals. Of these 23 journals, 17 were from science, technical and medical (STM) disciplines while six were from the humanities and social sciences (HSS). The split is illustrated in Table 1. The copy-editors did not know that the study was being carried out.

The researcher analysed the following material:

- author's original MS;
- edited version of the MS prior to type-setting;
- typesetter's PDF proof;
- author's corrected proof.

The researcher was then required to complete an Excel spreadsheet according to the categories shown in Table 2.

Number of author queries raised by copy-editor

An author query sheet is generated automatically as part of the pagination process and is included together with the PDF proof that the author receives. A sample author query form is shown in Figure 1. The researcher counted the number of author queries raised within each proof.

Number of typographic changes to proof

This category included all typographic corrections made to a proof, but not those that

---

Journal:

Article:

Dear Author,

    During the copy-editing of your paper, the following queries arose. Please respond to these by marking up your proofs with the necessary changes/additions. Please write your answers on the query sheet if there is insufficient space on the page proofs. Please write clearly and follow the conventions shown on the attached corrections sheet. If returning the proof by fax do not write too close to the paper's edge. Please remember that illegible mark-ups may delay publication. Many thanks for your assistance.

| Query No.	Query	Remark
1	Au: Dubbel et al. 1985—have changed spelling of Dubbel to match list = OK?	
2	Au: Is the text OK: Both sexes preferred black to white unbaited traps; however, females were repelled by white traps when they were also baited with a NHV, and there was a highly significant effect of colour for both sexes overall, with no interaction	
3	Au: Campbell, S.A. & Borden, J.H. (2006) Close-range, in-flight integration of olfactory and visual information by a host- seeking bark beetle. *Entomologia Experimentalis et Applicata*, in press.—any more details yet?	
4	Au: Please provide accepted date.	

Figure 1 Example author query form.
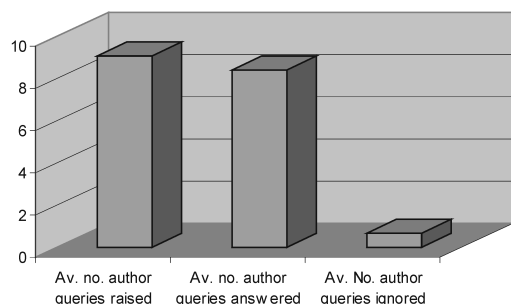
Figure 2 Mean number of queries per article responded to and ignored in relation to the total number raised.

Figure 3 Mean number of typographical changes and other major changes to proofs, together with the mean number of changes to the copy-editing.

substantively altered the text (this category is covered below). Since Blackwell uses the author's electronic file in the pagination process, it is unlikely that the traditional category of 'printer's error', resulting from the outdated process of rekeying, is the cause of these typographic corrections. An example from one of the articles included in the study indicates the types of error being corrected within this category:

- addition of digital object identifier (DOI);
- elimination of full points after author names (missed by copy-editor);
- introduction of white space between groups of data in a table;
- alteration to the presentation of data in a table (e.g. '$n = 8$' instead of '8 replicated');
- contraction written out in full (e.g. Douglas-fir beetle instead of DFB);
- initial capitals introduced into figure legends;
- minor changes to the visual arrangement of figures and tables to improve intelligibility;
- changes as a result of copy-editor queries (see below);
- addition of 'accepted' date.

Number of author changes to copy-editing

This category covers those cases where an author disagrees with any changes that the copy-editor has made within the text of the article. This should be distinguished from those occasions where an author either decides not to respond to a copy-editor
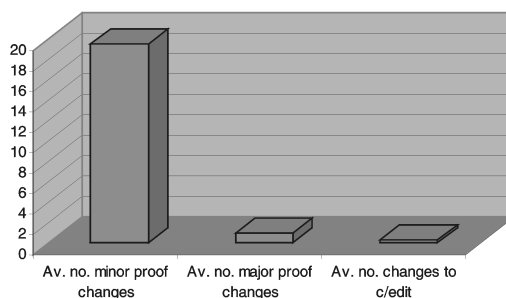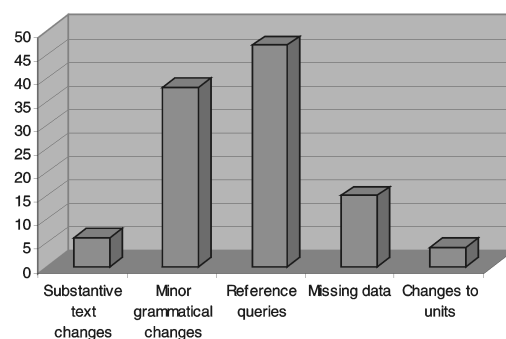

Figure 4 Analysis of the types of change being made by authors in response to 110 copy-editor queries.

query or decides that no action is required (see below).

Number of alterations to the text

This category includes those cases where an author makes a substantive change to the text on his or her own initiative and not as a result of a suggestion made by a copy-editor. An example of this type is the alteration of insect lengths in a particular study from the original 10–12 mm to 13–16.5 mm, or the addition of the word 'field' at proof stage in the following sentence: 'The relevance of these findings for designing *field* studies . . .'.

Number of changes made as a result of copy-editor's queries

These data were collected directly from each author's response to the queries raised in the author query sheet (see above).

authors are in general happy with the copy-editing that is carried out on their manuscripts

Number of copy-editor's queries ignored

This category includes those cases where an author either decided that the copy-editor's questions were inappropriate or else failed to answer the question.

Results

Author responses to copy-editor's queries

A total of 1,708 queries were raised by the copy-editor in the sample of 189 articles (mean = 9.0 queries per article). Authors responded to 1,586 queries (mean = 8.4 per article), which left 111 queries (mean = 0.6 per article) either ignored or unanswered. This is illustrated in Figure 2.

Alterations to proofs

a significant number of changes are made, and the largest single category relates to the accuracy of references

A total of 3,689 minor typographical alterations was made by authors during the process of correcting their proofs (mean = 19.5 per article). The number of unprompted changes to proofs resulting in significant textual alterations was 198 (mean = 1.0 per article). This category of correction includes those cases where an author decided to change the text during the process of reviewing the page proofs. However, the number of alterations made as a result of authors disagreeing with the changes imposed by the copy-editor was relatively trivial at only 48 (mean = 0.3 per article). This indicates that authors are in general happy with the copy-editing that is carried out on their manuscripts.

These data are illustrated in Figure 3.

What are authors changing?

A further analysis of the types of query being raised by copy-editors was carried out on a random selection of 110 author queries. The results are shown in Figure 4.

The majority of queries (n=47; 42.7%) related to references – either as a result of inconsistency between text and bibliography or else missing/incomplete references.

The second major category (n=38; 34.5%) included requests by the copy-editor to the author to check that the copy-editing was acceptable. This covered relatively trivial points of grammar, correction of spelling errors, etc.

There were also 15 (n=15; 13.6%) requests to supply missing data (identity of manufacturers of equipment listed, correspondence addresses, dates of acceptance) and 4 (n=4; 3.6%) requests to alter units in line with journal-specific conventions.

There were also six instances (n = 6; 5.5%) where as a result of the copy-editor's queries the author made an alteration that materially altered the sense of the text. These are listed in the Appendix.

Discussion

The aim of this study was to ascertain whether, as a result of the copy-editing and proof-correction process, substantive differences were made to the published version of the article compared to the author's version. The quantitative data indicate that a significant number of changes are made, and that the largest single category (42.7%) relates to the accuracy of references. A secondary category (34.5%) covers minor syntactical or grammatical changes that are approved of by authors, with a small number of changes (5.5%) correcting errors that might otherwise have led to misunderstanding or misinterpretation.

It is worth remembering that copy-editing was originally a printers' invention (printing and publishing being almost indistinguishable activities in the early days of the trade) but was gradually taken over by publishers from about the middle of the 20th century onwards. The primary aims of the copy-editor are to 'remove any obstacles between the reader and what the author wishes to convey, and also to save time and money by finding and solving any problems before the book [or journal] is typeset, so that production can go ahead without interruption'.[5]

While typesetting remains important for the majority of journals which continue to appear in print, within the online environment it can be argued that it is actually the process of XML creation (leading to HTML conversion) that now takes precedence. However, while automated software can carry out much of this task, manual intervention in the shape of copy-editing is still

required to oversee the syntactical accuracy of that tagging.

The traditional function of the copy-editor therefore remains within the online environment. It is now as much to do with the accuracy of tagging as with attention to grammar and journal style. Corrupt or missing references can be a source of minor irritation or major inconvenience – misquoted references increase the probability that a citation index such as Web of Science will not be able to link the citations to the source article. In today's metric-driven world, not receiving credit, in terms of citations, for the work that one has published can actually make a difference in terms of promotions, tenure, grant- funding, etc.

Minor errors in grammar or syntax (especially where the author does not have the benefit of writing in his/her first language) can also impede the reader's comprehension, or in the worst cases may lead to misunderstanding. Of equal importance is the correct tagging, for example, of authors' names – is this author's family name Changjian or Jiang? Not all software will notice the difference, while inconsistent attribution of authors' names could affect the electronic publication record of the author concerned.

The challenge for publishers is to establish what level of copy-editing is appropriate for a journal, at what cost level, and within what timeframe. Speed of publication is a decisive factor in this matrix, as is cost. Some would argue that the process is wholly dispensable, but this would be to ignore the contribution that copy-editing makes to the overall accuracy of the article of record. The value given to such accuracy may depend on the discipline. Obviously it is of some importance in clinical medicine, for example, where a dosage level might be critical; it may be considered less critical in fast-moving disciplines where much of the published material consists of the authors' own datasets.

Conclusion

In a sample of 189 articles the process of copy-editing resulted in a significant number of changes. None of these materially altered the conclusions of an article, which is more

the purview of the peer-review process, but they did produce a more consistent and accurate article of record. This is particularly important in the electronic environment, where accuracy of linking, for example, could be critical in establishing correct citation data.

References

1. Berners-Lee, T., De Roure, D., Harnad, S., and Shadbolt, N. 2005. Journal publishing and author self-archiving: Peaceful co-existence and fruitful collaboration. URL: http://eprints.ecs.soton.ac.uk/11160/
2. Ware, M. *ALPSP Survey of Librarians on Factors in Journal Cancellation.* Worthing, Association of Learned and Professional Society Publishers, 2006. URL: http://www.alpsp.org/ngen_public/article.asp?id=200&did=47&aid=157&st=&oaid=-1
3. Beckett, C. and Inger, S. *Self-archiving and Journal Subscriptions: Co-existence or Competition? An international Survey of Librarians' Preferences.* London, Publishing Research Consortium, 2006. URL: http://www.publishingresearch.org.uk
4. Rowlands, I., Nicholas D., and Huntington P. *Scholarly Communication in the Digital Environment: What Do Authors Want?* London, CIBER, City University, 2004. URL: http://www.ucl.ac.uk/ciber/ciber-pa-report. pdf
5. Butcher, J. *Copy-Editing: The Cambridge Handbook* (3rd edition), Cambridge, Cambridge University Press, 1992, p. 2.

Appendix: Changes to final articles as a result of queries by the copy-editor

Example 1

Author's MS: Both sexes preferred black to white unbaited traps; however, females were repelled by white traps when they were also baited with a NHV, and there was a highly significant effect of colour for both sexes overall, with no reaction.

Final version: In Experiment 3, there were insignificant reductions of male and female DFBs in white vs black baited traps. However, both sexes preferred black to white unbaited traps, females were repelled by white traps when they were also baited with a NHV, and there was a highly significant effect of colour for both sexes overall, with no interaction.

the process of copy-editing resulted in a significant number of changes

Example 2

Author's MS: To examine the number of visits each flower visitor group made within plots in response to variation in blossom density and species richness, linear multiple regression using visitation data from all periods, the flower visitor groups actually entered the plots, was used.

Final version: To examine the number of visits each flower visitor group made within plots in response to variation in blossom density and species richness, linear multiple regression using visitation data from all periods that the flower visitor groups actually entered the plots was used.

Example 3

Author's MS: A simultaneously operated vertical-pointing searchlight trap for sampling migrating insects from high altitude (up to 500 m above the ground) (Feng *et al.*, 2003), was placed about 450 m west away from the radar, on top of a house which was about 8 m above sea level (ASL).

Final version: A simultaneously operated vertical-pointing searchlight trap for sampling migrating insects from high altitude (up to 500 m above the ground) (Feng *et al.*, 2003) was placed about 450 m west of the radar, on top of a house that was about 8 m above sea level (ASL).

Example 4

Edited MS: and within which populations of a species are able to maintain a long-term average net reproductive rate that is $= 1$.

Final version: and within which populations

of a species are able to maintain a long-term average net reproductive rate that is ≥ 1.

Example 5

Author's MS: Within each plot, 100 haphazardly selected terminals of sea oxeye daisy with short (0.1–0.5 m) and tall (>1 m) stems.

Final version: Within each plot, 100 haphazardly selected terminals of sea oxeye daisy with short (0.1–0.5 m) and tall (>1 m) stems were sampled.

Example 6

Author's MS: Species numbers were counted from the category based on above information. For species which no subspecies differentiation, it was counted one species; for one species containing different subspecies (including nominated subspecies), the number subspecies was counted according to their occurrence in QTH.

Final version: Species numbers were counted from the category based on the above information. For species with no subspecies differentiation, it was counted as one species; for one species containing different subspecies (including nominated subspecies), the number of subspecies was counted according to their occurrence in the QTP–Himalayas.

Edward Wates and Robert Campbell
Blackwell Publishing Ltd
9600 Garsington Road
Oxford OX4 2DQ, UK
Email:
edward.wates@oxon.blackwellpublishing.com
robert.campbell@oxon.blackwellpublishing.com