

EASE-Forum Digest: June to September 2012

You can join the forum by sending the one-line message "subscribe ease-forum" (without the quotation marks) to majordomo@helsinki.fi. Be sure to send messages in plain text format; the forum software does not recognize HTML-formatted messages. More information can be found on the EASE web site (www.ease.org.uk). When you first subscribe, you will be able to receive messages, but you won't be able to post messages until your address has been added manually to the file. This prevents spam being sent by outsiders, so please be patient.

Citing from the repository version of a paper

A colleague asked Reme Melero if there was any guidance for editors on how to deal with citations to the repository version of a paper. No one on the forum knew of any but the consensus was that it was preferable to cite the version of record. Angela Turner had seen repository versions of papers that differ in several respects from the final published version, when citing an earlier version would be misleading. However, in Tom Lang's experience authors usually posted the published pdf in a repository, often after an embargo so if it was openly accessible it could just as easily be cited as it would be identical to the final published version. Except, Liz Wager highlighted the final published version will be linked up, at least in theory, with any subsequent correction or retraction, whereas repositories may not have this linking facility. She also alerted the forum to CrossMark (<http://www.crossref.org/crossmark/>) which indicates the 'publisher-maintained' version of the paper.

In any event without a doi, Karen Shashok could see no way of being sure if a paper in a repository was identical to the published version, except to ask the author. Some corresponding authors she had asked provided the accepted manuscript rather than the final published version. Some had told her they had not been given the final pdf version by the publisher. She had noticed that some self-archived versions gave citation details of the final published version (including final page numbers and doi), which was helpful.

Mary Ellen Kerans pointed out that some publishers do not allow posting of final pdfs and no "best practice" advice for dealing with these circumstances has emerged but she advised authors who post a manuscript version to mark where page numbers had changed on the final journal-published version. That will aid authors who quote from the text when they cite (because quoting requires use of a page number).

Measuring the quality of written English

All Gina Vega wanted and asked the forum for was a method for evaluating the quality of written English. This request elicited 27 postings, indicating that while there would be great interest in the context, eg of managing and paying copyeditors, many factors need to be considered and no

practical index that could be used for such a purpose exists. The following is a summary of the quality factors discussed, with side discussions on the use of "this" and "that" at the beginning of sentences and cultural perceptions of short vs long sentences.

First Tom Lang established that quality (comprehension, recall, referenceability, and usability) in readability research is defined by the reader not by the text and, as Mary Ellen Kerans added, the target readership is a major variable (with differences in age, discipline, familiarity with research structure etc.). Both Tom and Mary Ellen considered readability formulas worthless; writing to the formula actually reduces comprehension.

Mary Ellen thought the only effective way to test text quality would be to ask a sample of readers for their holistic impression of whether a text was well written according to a scale like pain: 0 = no pain on reading, lovely experience; 10 = worst possible reading experience imaginable. A linguist would then need to analyse the worst-graded texts to identify their features. Iconic example texts of several readability levels could be created and validated to guide anyone who needs to assess readability. This approach is known as benchmarking. Mary Ellen suggested Gina might investigate the benchmark texts used by TOEFL (Test of English as a Foreign Language, www.ets.org/toefl/). But TOEFL is a writing test, not a readability index.

Joy Burrough listed some basics that could be used to measure text quality and their problems.

- No grammatical errors: but a succession of correct sentences does not make good writing and views differ on correct grammar.
- Appropriate sentence length: but this depends on cultural expectations and changes over time.
- Appropriate register (not too pompous, but not too informal): but a comparison of the style of articles of today with ones written 20 years ago shows that this changes too.
- Conciseness of language and precision of word choice: but applying a formula to this is particularly tricky.
- Text coherence (appropriate use of linking words, allusions to preceding or succeeding text, optimal positioning of key information in the sentence).
- No idioms or metaphors that are so culture-specific that they will confuse international readers. As a problem she gave the example of a Japanese reader who was confused by the biblical reference to Daniel in a *New Scientist* article.

Yet another challenge she saw for a writing quality test was the need to take account of non-native English problems. Words that are similar to English words but have different meanings in another language (eg "eventual" being used to mean "possible", "preservative" to mean "condom") have been identified as a major problem by on-going research which is seeking to automatically identify non-native-English errors (see eg <http://aclweb.org/anthology-new/W/>)

W12/W12-2028.pdf). Another problem is the absence or misuse of “a/an” and “the” by writers whose native language doesn’t have indefinite or definite articles, which is further complicated because indefinite and definite articles tend to be used less frequently in scientific than in normal English (I agree. I constantly edit them in). Non-native English speakers also misuse verb tenses, which an editor can only rectify by looking at the particular context in which the verb is being used.

Ed Hull proposed that poor quality of text frustrated readers and quality could be tested by analysing the causes of frustration. These causes include:

- Poor focus on the main messages: a feeling of “so what?”
- Lack of a “storyline”: the storyline ties the main messages together.
- Poor linking: every sentence should be linked to its previous sentence by either using a linking word or by repeating words in the previous sentence.
- Sentence construction including: subject/verb distance, parenthetical phrases, lack of end focus, etc.
- Wordiness: redundancy, repeating, wordy phrases etc.

David FitzSimmons particularly endorsed poor linking (Joy’s “text coherence”) as an indicator of poor quality and criticised the common use of “this” and “that” as demonstrative pronouns to link sentences without a substantive noun being referred to. The result is unclear and poor quality text. The number of demonstrative pronouns that are orphaned in this way might therefore be used as a text quality marker.

The ensuing debate on the forum centred on translators. It was prompted by Mary Ellen’s observation that non-native English speakers and “genre-naive” or “developing” translators use more demonstrative pronouns than scientific articles usually contain. These authors and translators may not know, for example, that “This includes the ...” (where “This” is a demonstrative pronoun) can be edited to “This process includes ...” (where “This” is a demonstrative adjective describing the noun “process”, ie the noun being referred to is clear). Kersti Wagstaff commented that the two are linked. Developing translators are unable to interpret what the authors have written and reproduce their lack of explicitness. Mary Ellen on the other hand said translation instructors have noted that inexperienced translators have a slight tendency to make the text explicit, ie, say more than the target reader needs them to.

Karen Shashok thought that before orphaned demonstrative pronouns are used as a quality marker we need to understand why they appear in translated text. Experienced translators may also perpetuate these orphans if they believe the source text needs to be followed or they have been instructed that the translated text must be “identical” to the source text. She agreed with Kersti that an inexperienced translator may not know enough about the subject to decide what the orphaned pronoun refers to, adding that the translator might not be allowed to work together with the author to ask for an explanation and improve clarity because of time or cost restrictions. Alternatively she suggested that because demonstrative pronouns are common (commonly

misused) in research articles authors might believe that they should begin sentences with such a pronoun. Translators might assume that the authors’ peers will understand the text without problems and directly translate it retaining the pronoun’s orphan status. Indeed, Mary Ellen pointed out that each text has to be assessed individually for whether an explanation of what “this” or “that” refers to, is or isn’t necessary. In any event Karen and Kersti agreed that a combination of poor original writing and poor translating results in miserable text.

David broached “translatability” as a potential criterion for evaluating quality of text. By this he meant the ease with which text could be translated into other languages. His idea was inspired by one of his colleagues at the WHO. She tested the quality of the original English text by using machine-assisted translation software to translate it into Spanish. She then recorded the time needed to revise the output into acceptable text in Spanish. Well-written text took 50% less time to revise than poorly written text.

Mary Ellen warned that using “translatability” as a quality marker could be problematic, because if the prose seems well written to a reader it could be because it’s actually a calque from that reader’s native language (see the “eventual” and “preservative” examples in the last paragraph of Joy’s list of basics).

According to Sylwia Ufnalska, David’s anecdote illustrated the common problems of a lack of cohesion and long and complicated sentences encountered when translating poorly written text. She explained that this was why the *EASE Guidelines to Authors and Translators* emphasise the need for logical ordering of information and using short and simple sentences.

But how short is short? Joy asked, referring to linguistic research which has shown that there is a cultural difference between perceptions of normal sentence length: Americans write shorter sentences than the British, and other nationals, eg the French, write even longer sentences. Also, like has to be compared with like. Sentence length in a French novel and in an American children’s book should not be compared. This is where “Winnie the Pooh” entered the fray. Joy noted that sentences in the first paragraph of this British children’s classic book written by a vicar in 1926, have an average length of over 15 words. Authors have individual as well as cultural preferences. She had also noticed that sometimes authors writing in a foreign language overcompensated by deliberately trying not to write in a “childish” style and ended up writing complicated sentences which fail to communicate effectively. Furthermore, Mary Ellen joined, sentence length can vary within one discipline, eg sentences in results sections tend to be longer than elsewhere in research articles but present no burden for the target readers, who often skim over them rather than read them like a story. The point she was making is that disciplinary expectations dictate what’s easy or not easy rather than length and if the grammar, punctuation and parallel structure are well done, very long sentences might be easy to read, especially if they are common in the particular literature.

Along the same lines, Tom thought sentence complexity rather than length affected comprehension. Shorter

sentences just have less chance of being complex. Mary Ellen concluded that a short sentence can sometimes be hard to understand and a long sentence can sometimes be easy to understand. This prompted Anna Sharman to post a reference to an article in *American Scientist* that has some wise things to say about sentence structure and length: <http://www.americanscientist.org/issues/pub/the-science-of-scientific-writing/>.

Tom gave an example of a 198-word sentence written by R. Buckminster Fuller which he thought could be well understood. Joy did not agree and quoted another example from *Winnie the Pooh*, this time of a 194-word sentence which she used to demonstrate understandable long sentences to her students. But Karen highlighted that Milne is telling a story reporting individual events in a chronological order and stories are easier to understand than the relationships between ideas and concepts which Buckminster Fuller was proposing when shorter sentences would be easier to understand. Within a research article the methods and results section report sequences of events. Sentences therefore do not have to be short to be comprehensible but sentences in the discussion which analyse, interpret and explain should be written more like narratives.

Quite right, agreed Ed Hull, a narrative (a story) is easier for readers to understand so why not write research articles as a story? He saw the structure of the standard fairy tale (Once upon a time...) as similar to that of the standard (IMRAD) article. He asks his students to write a storyline of 800 words containing 10 main messages which should be in every research article. The main messages must link together to form a "story" that is readable by the non-specialist. They form the "skeleton" which the author should then support by filling in the technical details of background, methods, results, discussion and conclusion. The resulting article

is readable at two levels: the non-specialist can skim over topic sentences of paragraphs for the main messages; and the specialist can read the details within the paragraphs to judge if they credibly support those main messages.

Katharine Timberlake felt that accuracy of thought was an important precursor for good quality English. She gave the example "AA did not contain X, similar to BB", in fact meaning "AA did not contain X, in contrast to BB [which did]". These examples show that the author was not aware of the difference between the two options. Sylwia regretted the paucity of thought diligence and clear thinking. During her session in Tallinn (<http://www.ease.org.uk/ease-events/triennial-conference/editing-digital-world-tallinn/tallinn-programme/parallel-session-c>), a delegate commented that she spends 70% of time on thinking and only 30% on actual writing of an article. This, Sylwia thought, should be a rule among scientists, but it isn't.

Katharine stressed that authors should however make sure that whatever they have written is accurate before it goes to a journal to avoid reviewers and copy editors being "faced with the massive challenge of spotting crazy infelicities wherever they may lurk." In the same vein, Mary Ellen felt that despite the difficulties of assessing the quality of English there needs to be some means of doing so before review, especially in modest journals that are nonetheless SCI indexed. In particular, the person reviewing the English needs to understand the science.

Elise Langdon-Neuner (compiler)
a.a.neuner@gmail.com

Discussion initiators

Reme Melero: melero@iata.csic.es
Gina Vega: gvega@salemstate.edu

This Site I Like

BioMed Central: all about open access publishing

www.biomedcentral.com

BioMed Central is the pioneer of the open access publishing model whereby all research is freely available on the Internet, without subscriptions or any other barrier to access. BioMed Central is the brainchild of Vitek Tracz, a visionary business entrepreneur who foresaw that the disruptive nature of the Internet would eventually challenge the traditional print subscription model for scholarly publications. The web allows, and makes inevitable, the emergence of a seamlessly interlinked research, and in 2000 Vitek started BioMed Central to facilitate open access publishing and to prove that the new model was financially sustainable.

Among the major publishing houses, Springer was the first one to recognise the benefits of open access and acquired BioMed Central in 2008. BioMed Central's systems and platforms are now used by a rapidly growing portfolio of over 230 journals in biology and medicine, as well as by more than 60 journals in the SpringerOpen programme

which uses BioMed Central technology to expand the open access offer into other research disciplines.

BioMed Central journals are widely indexed, including in PubMed. Within two working days of publication, new articles are deposited in PubMed Central. All BioMed Central articles are also searchable on Springerlink and interlinked with Springer's vast and prestigious journal and book programme.

The main customers for publishers of subscription-based journals are libraries; for open access publishers, the most important customer group are researchers as they decide where to submit their work for publication. Authors can therefore expect first-class services from BioMed Central.

BioMed Central's submissions system is easy to use and allows authors to upload their manuscripts and associated content and datasets, to propose and exclude reviewers and Editors, to select article types, and add keywords and required