TALLIP Perspectives: Editorial Commentary

The State of the Journal

This year, 2014–2015, has been my second year as Editor-in-Chief of the journal and, like last year at this same time, I am happy to report progress on a number of fronts.

First and foremost, *Transactions on Asian Language Information Processing* has now become *Transactions on Low-Resource and Asian Language Information Processing*, broadening its scope to cover not only the (already broad) set of languages of Asia, but also low-resource languages from most of the world.

To that end we have also added a new category of paper—*TALLIP Notes*—which is intended for relatively short articles that discuss work on low-resource languages. This was done in part to counter a perennial problem in the field, namely, the constant demand that a new article propose new methodology. Articles submitted to *TALLIP Notes* need not propose new methods, so long as they demonstrate a nontrivial application of some computational approach to a low-resource language. This is as things should be: The field doesn't always advance by inventing new methods. How many machine learning approaches have been proposed over the last couple of decades? Certainly, in many cases these have represented true advances affording improvements in at least some areas of application. But the fact that people tend to pounce on them in droves—so that a meeting of the Association for Computational Linguistics can have flocks of papers on, say, using CRFs for a range of problems (only to be replaced by something else the following year)—suggests that there is an aspect of faddishness to this. It is high time there were more outlets for work that may use familiar, even old, techniques, but which applies them to new linguistically interesting problems. *TALLIP Notes* is a small effort in this direction.

Submissions to the journal continue to increase. Between July 1, 2014 and June 30, 2015, there were 173 submissions, up 15% over the previous year. And we have added several Associate Editors in the past year to accommodate the broadening of scope as well as the increase in submissions.

That's the good news. Now for the bad news.

In addition to gaining a few editors, we have also lost a few editors who have requested to step down: For those who were on the board long enough to actually serve as editor for one or more articles, we thank them for their service.

The turnaround time for reviews continues to worsen, to an average of 65 days, up from 58 days. I am not sure what to blame this on: Perhaps the new ACM requirement that in order to accept an article, we must have *three* reviews that agree on “accept” has played a role here though I do not think that can explain all of the increase. It is possible that my encouraging reviewers and AE's to move the reviewing process along has had the opposite effect from the one intended. And the result is puzzling since I have increased the number of desk rejects both on the basis of obvious problems with the technical content, as well as another issue that we shall turn to immediately. The

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.

2015 Copyright held by the author/owner(s).

2375-4699/2015/10-ART19 \$15.00

DOI: <http://dx.doi.org/10.1145/2823512>

desk rejects should have helped to reduce the overall average turnaround time. In any case, whatever the cause, it is not an outcome I would have hoped for.

But the most serious issue that has arisen over the last year is plagiarism. Since the end of 2014, articles submitted to TALLIP have been routinely sent to the iThenticate's CrossCheckTM system. In most cases what the system detects as potential plagiarism are silly things like the boilerplate rights footnote found at the bottom of every article submitted to an ACM journal. Titles of articles in the bibliography were another source of false positives. Then there were cases of what is often incongruously termed "self-plagiarism". While the frequent reuse of the same material by an author is at best tedious, and at worst in poor taste, I find it hard to take it seriously as a case of academic misconduct: The whole concept of plagiarism revolves around the notion of appropriating the thoughts and language of *someone else* and making them appear to be one's own. "Self-plagiarism" is therefore an oxymoron, and as such I ignore it. After eliminating such cases, there were still sixteen articles where the authors had copied nontrivial amounts of material from other sources without clearly indicating that the material was lifted from those sources. In one case, the material was from the web page of a course that I had taught a few years previously! (In that instance, I suggested to the author that it was particularly ill advised to lift material from the work of the editor of the journal to which one is submitting.) Articles where this had occurred were rejected immediately with an explanation to the author of why their article was rejected.

A couple of the sixteen were repeats of the same article, where the author, upon receiving the first rejection, made some cosmetic changes, mostly changing a few words here and there, hoping that the system would not catch the now perforated sections of text. Unfortunately for them it did, and, in those cases, I of course, rejected the article again and put a block on further submissions from the author until they could demonstrate to me directly that they fully understood what "plagiarism" means.

There is, incidentally, a very apparent nonuniformity in the distribution of countries from which the offending articles are submitted. I am not going to reveal any more details here since I do not want to cause offense, and also because the sample is after all still rather small. Nonetheless it is striking.

Dealing with cases of plagiarism is my second least favorite aspect of running the journal, coming right after chasing after some of the reviewers and Associate Editors to get reviews and decisions in on time. I cannot simply accept the summary report from CrossCheck at face value. Perhaps the authors actually did make it clear that the offending passage was indeed a quote, and CrossCheck missed that. So I have to check both the submitted article and the putative source. This means finding the latter, which is often hidden behind a paywall. (CrossCheck does provide the text of the original source, but since this is plain text and garbles any kind of formatted material such as equations, it is effectively useless.) Sometimes I have access through my employer (but that requires being on the corporate network when I hunt down the paper), and sometimes I can get it via a major university library that I have had access to until recently. But, in some cases, I cannot get the original at all, and I have to look for another segment in the submitted article that appears to have been copied and hunt down *that* reference. Then once I have a couple of clear cases (I usually try to find two), I write a rejection letter to the author including the evidence and instructions on what they need to do if they plan to resubmit. All of this takes time: I estimate that I spend between a half hour to an hour whenever I have to deal with an article that involves plagiarism. And the worst part of it that, while I personally have little sympathy towards cases of plagiarism, it is quite likely that the authors involved simply do not know that this kind of behavior is inappropriate in academic settings. And this

leads, of course, to the question of what the educational system they went through was teaching them (or failing to teach them) about academic integrity.

If there is any justification for plagiarism, it might be that sometimes someone has said something so well that it would be hard to say it in a better way. I hardly think that the web page to my own course that I mentioned would fall into that category, but I am prepared to believe that perhaps some of the other lifted material does. If the original material is really that good, then, of course, the authors should use it: But should basic ethics not demand that the authors indicate that they are using someone else's material and give full credit to the source?

I can only assume that such cases of plagiarism must have happened and gone undetected in earlier issues of TALIP. Based on the numbers I have observed, a quick back-of-the-envelope calculation suggests that between five and ten percent of articles that have appeared on the pages of this journal in the past may have involved serious cases of plagiarism. For all of that, I do not suppose that TALLIP is worse off in this regard than other journals (though I have no data to support any claims about that). Unfortunately it seems that education on academic integrity is not uniform.

I can only hope that the message is getting out: that plagiarism is unacceptable and grounds for immediate rejection no matter what the technical merits of the work may be. If so, then one may hope that in my next State of the Journal contribution next year, I will be able to report that this problem has gone away. But in so saying, I am probably setting myself up to be offered a bridge over the East River.

Finally, my next State of the Journal contribution will also be my last: I have decided after much thought not to renew my editorship, so TALLIP will be looking for a new editor. Anyone who wishes to take on the job should feel free to let me know.

Richard Sproat
Editor-in-Chief