Last Words

What's the Future for *Computational Linguistics*?

Robert Dale*
Macquarie University, Sydney

You are reading the last issue of *Computational Linguistics* that will appear in printed hardcopy form. The beginning of 2009 heralds a new era for the journal in at least two major respects: As of the first issue of volume 35, *Computational Linguistics* will be published only electronically, and it will be open access. As editor, I'd like to take this opportunity to use these last words that will tumble from the presses to provide some explanation of the changes afoot at the journal, and to offer some thoughts on where this might take us in the future.

1. Why Open Access?

There are a number of definitions of the term 'open access' in circulation, but almost all share the key principle that scientific literature should be freely available for all to read, download, copy, distribute, and use (with appropriate attribution) without restriction. At the time of writing, the vast bulk of scholarly literature is not open access: Either you pay for access directly as an individual subscriber (for example, in the case of *Computational Linguistics*, via your annual membership subscription to the Association for Computational Linguistics), or you gain access via an institutional subscription (typically, your library's annual subscription to MIT Press).

There is an increasingly widely held view that this is just not right. Given that almost all the research published in scholarly journals is paid for by the taxpayer, it's reasonable to ask what justification there can be for restricting public access to this research by requiring that a further payment be made to read about it. And 'toll access', as it is sometimes called, is not only bad for the reading public; it has been frequently argued that it is bad for authors too, because any barriers to access may decrease the likelihood of citation and the general impact of the reported work.

So, as announced earlier this year at the annual meeting of the Association for Computational Linguistics in Columbus, Ohio, *Computational Linguistics* will become an open access journal as of the first issue of 2009. In fact, for some time the journal has had a rather unusual access status, whereby its content becomes freely available as part of the online ACL Anthology one year after publication.[1] As of the first issue of Volume 35, even this delay will be removed, with the journal's contents being freely available as soon as they are published.

---

* Centre for Language Technology, Macquarie University, Sydney, New South Wales 2109, Australia. E-mail: rdale@ics.mq.edu.au.

1 See `http://aclweb.org/anthology-new/`. In practice, for some material this one-year embargo on free access is closer to two years, as a consequence of the batching-up of uploads to the Anthology on an annual cycle, and the inevitable administrative delays.

2. Why Electronic-only?

That the journal is going electronic-only and open access at the same time are not unrelated events. Print publication is a major part of the cost of producing the journal, and this is a cost that is very considerably subsidized by the revenues generated by institutional subscriptions. Going open access means that institutions no longer pay these subscriptions, and so we need a sustainable financial model that does not depend on this income. Ceasing the production of the printed version of the journal is an important step in that direction. In an environmentally conscious age, it also has the additional benefit of removing the need to fly, four times a year, small plastic-wrapped wads of paper to two thousand ACL members around the world.

At the same time, there is no *necessary* connection between going open access and going electronic-only. In principle a journal could be open access and yet still be distributed in printed form, although then an alternative funding source has to be found for print production. One approach is that adopted by those publishers (such as Elsevier and Cambridge University Press) who offer authors the option of their work being published open access in exchange for a fee, which can be anywhere between $500 and $2,000 per article. This 'dissemination fee' model has become accepted practice in some disciplines, but not—so far, at least—in ours. There are other business models one could adopt to sustain print production. For example, free local newspapers and many trade magazines are also effectively open access, with their production costs being paid for by advertisers; but it seems unlikely that advertising revenue could sustain a scholarly journal like *CL*. Another alternative would be to seek sponsorship from corporations like Google and Microsoft. But in the Internet age, it just seems perverse to bend over backwards to find ways to kill more trees.

This is not to say that going electronic-only makes the journal free of production costs. Although the journal's editor, editorial board, and external reviewers provide their labor free of charge, it still costs real money to produce the journal: In particular, *CL* will continue to use the professional services of the MIT Press in managing the high standards of copyediting and typesetting to which our readers have become accustomed. For the foreseeable future at least, these costs will continue to be met by the ACL, as owner of the journal.

3. What Else is Changing?

The shift to electronic publication means that publishing an article in *Computational Linguistics* should be considerably faster than before, for a number of reasons. Under the old model, it can take up to six months from the time that the editorial office dispatches the author's final copy of an article to the publisher, to the time the article appears in print. Around half of this time is taken up by the print production process. The delay is further exacerbated by the fact that the first article that is ready for an issue has to wait until the remainder of the issue is ready before it can be published. By moving to electronic-only publication, articles will now appear online as soon as they have gone through the copy-editing process, on an article-by-article basis. You'll be able to access articles even before they have completed the copy-editing cycle. This, in fact, is not a new thing: for the last year or so, we have been publishing what are referred to as the 'first revises' of articles—these being the proofs that are sent to the authors for checking before final edits are made—via the MIT Press Web site under the label 'Early Access'. But the unavailability of an easy way to let individual subscribers see this toll-access

content meant that very few people were aware of its existence. With the move to open access, the Early Access articles will also be freely available to all.

Another consequence of being electronic-only is that there will be no artificial backlog resulting from the page limits imposed by physical print production. Previously, the journal's cost model meant that we could publish 160 or so pages each quarter; the removal of print production provides us with more flexibility, and the likelihood that the journal will publish a greater number of articles (although, as indicated previously, there will still be costs associated with this: Copyeditors and typesetters charge per page). Being freed from paper bindings does not mean that processing backlogs will completely disappear, because there are other places in the publication process where delays can occur. For example, an unpredictable copyediting workload may lead to delays when our demand for copyediting services outstrips the available resources.

We hope that readers who keenly look forward to receiving an issue of the journal in their pigeonhole every three months will not suffer too much from a sense of loss. Members of the ACL will instead be able to subscribe to a number of electronic alerting options which either direct them to a download location for the latest journal content, or provide the content in the form of email attachments. The full details of our electronic delivery mechanisms are still being worked out at the time of writing, but ACL members will already have been notified of them by the time this issue is published.

We are taking advantage of the shift to electronic-only publication to institute a number of other changes. Most visible for authors will be our move to an online journal manuscript management system, which should introduce efficiencies that will shave a few more days off the time our reviewing process takes. For readers, each article published in the journal will be available in both PDF format and in XML format, the latter allowing easy reading on screen.

4. What Does the Longer-term Future Look Like?

Over the next year or two we intend to introduce a number of further changes that become more straightforward as a consequence of our primary existence being electronic. Possibilities we are exploring include making available for download supplementary materials such as the code and data associated with articles, and providing each article with an online discussion forum where its content can be discussed. The editor welcomes suggestions and ideas for how we can improve the value of the journal to its readership along these lines.

There is a longer-term question, though, about the nature of journals and whether they are a sustainable means of scholarly dissemination in the Internet age. The publication landscape has changed significantly in the last 10 years. One challenge to the position of journals and existing forms of scholarly publication more generally comes in the shape of blogs and other online discussion fora. Obviously these are not rigorously reviewed in the way that journals are, but they are open to other means of quality assessment, via machinery like page rank and reader voting. For some purposes, the low cost and democratic openness these offer may outweigh the deficiencies that arise from the lack of a conventional peer-review process. Ultimately, how these alternative fora will fare as a means of scholarly publication may be determined by whether we can develop a credit assignment mechanism that gives due recognition for great ideas that first surface in these more informal spaces. This would have to be more fine-grained than our current model of journal impact factors and bald citation counts; but more traditional forms of publication would surely benefit from improvements here too.

Another challenge comes from online shared knowledge repositories such as Wikipedia and, more recently, Google's Knols project. Who needs a static survey article in a journal in an era of Web resources that are constantly kept up-to-date? The art of writing survey articles for journals will no doubt continue to be practiced for some time, but not because it's the best way to provide this kind of material; conventional survey articles will persist for only as long as there is no alternative means of gaining academic credit for work of this kind. It's entirely possible that, within a few years, Knols will frequently be cited in the pages of journals like this one.

These are future challenges for journals in any discipline. In our field, journals face a challenge from another direction, closer to home. It has often been noted that, in computer science generally, conference publication counts as much as journal publication. In fact, a paper in a good conference can be worth much more than an article in a mid-ranking journal. Given the substantial additional effort that is usually required to produce a journal article, and the longer reviewing process that is subsequently involved, it's perhaps a surprise that journals receive any submissions at all. So why don't we just let journals die, and allow conferences to become the primary means of scientific communication in the field?

Conferences play an important role in many disciplines, but they do not fulfil all the requirements for the ongoing development of a healthy scientific base. Over the last few years, there has been increasing disquiet about the way our premier conferences are evolving. A common complaint is that only 'safe' work delivering incremental results is accepted, which leads to a vicious cycle where this is the only type of work that gets submitted, and maybe even the only type of work that gets done. One also hears complaints that acceptance rates are too low, and that the events have become, well, just plain dull. Other disciplines adopt a different model, where journals remain the primary repositories for well-considered and rigorously reviewed research that is considered fit to enter the archives of knowledge in the field; conferences are first and foremost meeting places, where acceptance rates are high, and much of the work presented is work-in-progress, the initial outings of ideas that are yet to be further developed. Some would argue that such events are more exciting and engaging than conferences in our field, where the assumed high calibre of the work presented can be thrown into question by occasionally random and hasty reviewing processes.

The defining property of a scholarly journal, and one that persists independent of the medium used for content delivery, is the adherence to a robust and thorough quality control process, all the way from the use of carefully chosen expert reviewers and feedback–response cycles to detailed copyediting that improves the wording and makes sure bibliographic references are complete. Stripped of its covers in an electronic age, a journal is a brand; a journal name serves as an indicator of quality. The stamp of quality a journal provides is even more important in an age of abundant publication; with so much to read, it becomes imperative that journals short-circuit the quality assessment processes that readers would otherwise have to undertake themselves.

It is inevitable that journals will change in unforeseen ways in the coming decades, but they will continue to serve their function as trusted gateways to trustable resources. Here at *Computational Linguistics*, we aim to take on board ideas that arise from newer forms of media and find ways of adopting and adapting them; but we also aim to continue to be seen as a source of the highest quality research in the field. I hope you'll continue reading, and that you'll consider *Computational Linguistics* the publication forum of choice in our discipline.