

Last Words

Are We Near the End of the Journal?

Nat Lang*

State University of Euphoria, Greece

In a *Last Words* piece at the end of 2009, the editor of this journal suggested that moving *Computational Linguistics* to open access and electronic-only availability was just the beginning of a sequence of changes we might see in how the journal operates. In fact, I think this step was, for the *Computational Linguistics* journal, the beginning of the end. But this isn't a bad thing: it's a positive thing that *CL* has started down this road before many of its competitors. We should be accelerating the pace towards the demise of the traditional journal.

In an age where most of us have our own Web pages, and many of us have our own blogs, it's possible to publish ideas without the inexcusable delays introduced by the reviewing process and the trappings of "professional publishing." The journal is dead as a mode of scholarly communication because it's a stultified and outmoded way of disseminating research findings. Here are some reasons why I think this is the case.

1. Forty-page Articles Are Dead

The traditional model of journal publication in our field involves writing forty-page articles. There are many fields that don't see the need for such a long-winded approach to reporting research results. Even our own field is somewhat ambivalent about this way of doing things: It's well recognized that the proceedings of the annual conferences of the Association for Computational Linguistics are just as much part of the archival literature as are journals. Many of these conference papers are heavily cited, so it's clearly not necessary to consume 40 pages to present a research contribution that other people can build on. A typical 8-page conference paper is probably something like 12 pages in the format adopted by this journal. It takes less time to write, and it takes less time to read, and in a fast-paced research environment, both of these are good things.

At the ACL's business meeting in Singapore in 2009, Ido Dagan, then ACL's Vice President, proposed a model whereby the papers that are currently published in the proceedings of the conference would instead be published as short journal articles. Some fears have been expressed that this might kill the long journal article: If it's possible to get a journal publication, and thus appease institutional bean counters who care only about journal publication, by writing a 12-page paper, why would anyone bother to write longer pieces? So, the argument goes, a major motivation for writing journal papers of the kind we see today will disappear, and in the absence of that incentive, long journal articles will die out. No great loss, I'd say. Learning the art of

* Centre for Rational Empiricism, State University of Euphoria, Themiscyra, Greece.
E-mail: natlang363@gmail.com.

saying what you want to say in fewer words is a skill that most of us (readers as well as writers!) would benefit from in any case.

2. Survey Articles Are Dead

It might be argued that journals are still a sensible home for longer pieces that serve as surveys or review articles. This journal has published a few survey articles in recent years, and there are entire journals dedicated to survey articles. Editors know that they are good for citations, and the more citations you get, the better your impact factor, so you can see why journals would be keen to have them.

But why would anyone want to read a survey article that's almost guaranteed to be out-of-date the day it is published? In almost all subfields of computational linguistics, progress is still ongoing, so static survey articles are at best snapshots of the state of the field at some arbitrary point in time. Surely something dynamic, updated as a field progresses, is much more appropriate for this kind of resource. This is one of the things that the Wikipedia model excels at, of course. Ah, but of course, authors don't get credit for Wikipedia articles, so they don't cater to the bean-counting incentives that drive both authors and journal editors. Google has made a stab at providing an alternative here: Knol lets individual authors write attributed pieces on topics on which they have expertise. So far, this hasn't been seen as a serious alternative to more traditional publication models, but it's just so much more sensible than a static survey article that it, or something similarly dynamic, is bound to take off.

3. Journal Issues Are Dead

Traditional paper journals are published as separately bound issues, typically four times a year. Of course, the research findings reported in the journal's pages are not produced according to such a rigid schedule; authors submit papers when their written-up results are ready. In the case of a quarterly journal, that means that an article which doesn't quite make the deadline for a given issue could easily wait for a further three months before it appears in published form. This might be justifiable in a paper-based age, where all sorts of production-oriented concerns would make it infeasible to publish each article as soon as it is ready. But in an electronic age, there can surely be no such justification. The notion of a journal issue just doesn't make sense when there is no printing requirement. Some electronic journals have seen the light here, but a surprisingly large number have not, including the one you are reading now. Any aspect of the publishing process that delays the publication of a research result one day longer than is necessary should be questioned.

4. Page Numbers Are Dead

Most journals, and many conferences, continue to insist that we provide page numbers for the papers we list in our reference lists. The argument is made that this detail is an important element in making it possible for the reader to locate the cited work. In an age of search engines, this just isn't very convincing. In fact, in many cases information about page numbers is unavailable because the version of the paper the author has at hand has been downloaded in a version where no page numbers are present; and the reader may well find the same version by typing the title into a search engine in

any case. Of course, if we really care about uniquely identifying cited works, what we should really be doing is providing Digital Object Identifiers (DOIs) for the items in our reference lists. It's particularly annoying that the present journal doesn't insist upon this, given that many of the references in *CL* articles are to papers that reside in the ACL's online Anthology, and every one of those has a DOI.

But there's another reason why page numbers are pointless. In an age when we may be reading on a phone, an iPad, a TV screen, or soon even on a tabletop surface or heads-up display, having fixed-format pages just doesn't make sense. Papers, like Web pages, should be automatically rendered optimally for the device on which they are viewed. Page numbers are just another hangover from the paper-bound 20th century.

Of course, page numbers allow us to be more precise in our citations, indicating just where in some paper it is that the author has made the claim we want to take issue with; but this function is just as easily achieved, and probably better served, by using numbered paragraphs instead. If it's good enough for lawyers (if anyone has a need for an effective citation mechanism, they do) and for Wittgenstein, then perhaps it should be good enough for us. Being forced to think about one's research in paragraph-sized citable pieces might also improve the quality of writing.

5. Copy Editing Is Dead

Older journals incorporate a copy-editing step in their publication process, whereby a disinterested third party combs the text for spelling, grammatical, and typographic errors, and ensures that the article is presented in a manner that is consistent with the house style of the journal that will host it. Some would go so far as to say that this step distinguishes a professional publication from an amateur one.

But some newer on-line journals, which survive on a minimal budget and volunteer labor, don't make use of copy editors. And why should they? With most journal articles never being cited, why waste time and money polishing text that perhaps no one will ever read? Copy editing adds a significant cost, and it also introduces an unnecessary delay in making the material available. A better model would be to optionally apply copy editing *after* publication, once the level of interest in the article in question can be assessed.*

6. Peer Reviewing Might Be Dying Too

A distinguishing feature of scholarly activity is the institution known as peer review. Before an author's research findings can be published, they must be first judged appropriate and correct by a committee of his or her peers.

It's often acknowledged that the process of peer review is far from perfect, but the claim is made that it is better than any of the alternatives. Is that really true? An April Fool's joke doing the rounds on the Internet in 2010 suggested that *Science* and *Nature* were going to jointly publish an open-access journal. Among various radical characteristics that this new journal would possess, it was proposed that articles would be reviewed by means of a voting mechanism on Facebook, thus replacing small committees of potentially self-interested reviewers with a "wisdom of the crowds" model. Although some of the other proposed features of the mythical journal were clearly daft,

* Editor's note: This article has been copy edited.

it's not so clear that this aspect is all that silly. After all, if we consider citation counts to be a valid means of assessing the impact or quality of a piece of work, why do we need reviewing at all?

7. The Article as a Unit of Publication Is Dead

Just as there's no need for a journal to collect articles together to make an issue, there's no real need for an author to collect together enough ideas to make a paper of a predefined length. The minimal publishable unit can be as small as a paragraph if that paragraph contains an important idea that impacts on others, so why shouldn't it be citable regardless of the larger document context in which it appears, and why shouldn't the author get credit for it?

The difficulty, of course, is that of assigning credit. But that's only because the currency we use at the moment doesn't have the most appropriate denominations: The only unit we have is the article or paper. What we need is the scholarly analog of micropayments. This fits comfortably with the earlier suggestion that we get into the habit of writing our papers as sequences of numbered paragraphs: This will make it easy to cite individual paragraphs. And if we can cite individual paragraphs, then there's nothing to stop us citing comments on blog posts, for example. Or even tweets.

An otherwise flawed paper can have one or two good ideas buried among the misconceptions and confusions; surely a citation mechanism that allows us to pinpoint and reward those ideas, without giving undue credibility to the shakier material that surrounds them, is better than the coarse-grained mechanism we use today.

8. A New Beginning for Scholarly Publishing?

If we follow these lines of thought to their logical conclusions, we inevitably converge on a mode of publication that is quite different from the one we are familiar with today, and one which is much faster and much slimmer. The slimmness comes from the potential for the removal of the redundancy that current practices encourage. I am surely not the only researcher to become impatient when reading several articles by the same author, all subtly different but clearly involving reuse of large slabs of material. A better model is one where any given individual's research on a particular topic is collected together, without redundancy, in one place. This might be a single Web page or a collection of pages, which for the lack of a better term I'll call a 'unipaper'. A unipaper is something that is added to and developed over time, with corrections and revisions incorporated; a single individual might produce only one unipaper in his or her lifetime, and few would create more than two or three, each dedicated to distinct research topics. No more tiny increments on last year's paper wrapped up as eight pages of text, with 50–75% being recycled from earlier publications: If this year's contribution can be summed up in a few (numbered!) paragraphs, then these individually citable elements get added to the unipaper.

This model does away with much of the publication process that we see today. That's not to say that there is no place for a future evolution of the notion of the journal. At the end of the day, a journal is just the view of a fairly random group of people (the editorial board) of what is worth reading. Amazon has those too; they're called Listmania lists. That function of journals can be retained without all the other baggage they carried in the last century and beyond; all you need is a Web page, with the DOIs

of the papers, or paragraphs within unipapers, that are worth reading. It doesn't even matter if several "journals" point to the same papers.

So let's abandon all the 20th-century baggage of traditional journals, and move to a more rational model for scholarly publication, with no copy editors, no reviewers, no redundancy, and no unnecessary delays. A concrete step would be to give each ACL member a DOI for a unipaper, and then ask them to non-redundantly populate this with a sequence, or a tree, of numbered paragraphs that consolidate all their work on a topic. Then, to get things moving, the present journal could insist that some proportion of citations be to paragraphs within these unipapers, with hyperlinks embedded right there in the citations. What are we waiting for?

