Last Words

Festina Lente: A Farewell from the Editor

Paola Merlo
Linguistics Department
University of Geneva
Paola.Merlo@unige.ch

The unexamined paper is not worth publishing.

Dear reader of *Computational Linguistics*,

I am about to finish my tenure as editor of the journal. It is time to bid farewell to the readership of the journal, to take stock of its current state (of the journal, not the readership), muse on the debate that the field of computational linguistics, like others in science, is having these days about publication practices and venues in a growing, rapidly changing field, and point, despite all growing pains, to the bright future that lies ahead.

1. The Last Five Years of *Computational Linguistics*

In the last few years the field of computational linguistics has grown and has increased its scientific and technological relevance. With this growth come new scientific and scholarly challenges.

1.1 The Journal's Impact

Citation numbers and journal citation impact factors tell us that the field of computational linguistics is growing and its journal is growing with it: Although we do have to monitor the newer and upcoming publication venues, *Computational Linguistics* is doing quite well. The impact factor has been going up ever since the journal went open access.[1] *Computational Linguistics*' impact factor for 2016 is 2.528. This makes it number 39 out of 133 in Computer Science, Artificial Intelligence, and number 5 out of 182 in Linguistics. Let me clarify that these numbers, whether they go up or down, should be taken with a grain of salt, since the journal publishes very few papers each year. They are an average of a very small set of numbers and they can fluctuate substantially. Despite this reminder, we can be satisfied with our journal's continuing relevance. Another big

---

1 The impact factors in recent past years were as follows: 2011: 0.721; 2012: 0.940; 2013: 1.468; 2014: 1.224; 2015: 2.017; 2016: 2.528. The journal impact factor suddenly dipped when going open access, which was clearly unrelated to its quality.

indicator of quality is that we are indexed in the Web of Science, and we are considered a class A publication in many rankings.

1.2 Reviewing Practices and Reviewing Time

Journals allow authors to provide long rebuttals to reviews, establishing a dialogue between reviewers, editors, and authors that creates a level playing field. They publish science that is long-lasting, has a long citation life, has been verified in detail, selected, reviewed and re-reviewed, written and re-written. The cost of this process is time. Journals like *Computational Linguistics*, which publish long to very long articles, publish slow science. There is therefore a perception that this kind of publication is too slow. This is in part a misconception.[2] For an example, the first quarter 2017 *Computational Linguistics* report indicates "Average time to first decision including survey proposals is 33 days and average time to first decision excluding "reject (not suitable)" is 53 days." On a longer time-period, we have an average of 80–90 days to first decision. Considering that most of our papers are 30 to 40 pages long, and that writing and reading times and length of reviews are largely linear to the number of pages, we have the fastest per-page decision rate of all the ACL publication venues. Publishing long papers in a carefully reviewed journal might be worthwhile and pay off in the end.

1.3 Diversity in the Journal's Publications

The ACL community has been recently concerned with extending its diversity, both geographically and gender-wise.

Geographical Spread. We are the longest-standing journal for anything related to text processing and language, we do not charge publication fees, and we do not require participation in costly conferences. So we attract a geographically diverse set of submissions. For example, in 2017, we received submissions from 41 different countries, the most represented being China, Finland, Iraq, Ireland, Spain, and Portugal. I am very happy with this fact, as I think that one of the roles of a journal is precisely to provide a publication venue to all the interested scientists and not just to the well-endowed, well-connected, or mainstream.

Computational Linguistics Gender Statistics 2015. In 2015, we have been asked by the Linguistic Society of America to compile gender statistics, as part of a much larger effort concerning all linguistics journals. We were requested to collect the gender statistics for the editorial teams (one female permanent editor and two female guest editors of a special issue) and for the editorial boards in the last three years (editorial board 2015: 16 men, 8 women; editorial board 2014: 18 men, 6 women; editorial board 2013: 16 men, 8 women). We were also asked to provide gender statistics for authors of submitted papers. For co-authored papers, each author was tallied. These numbers include one special issue. These are the statistics we were able to gather in a reasonable amount of time for papers submitted in 2015 for which a decision had been reached. The number of papers submitted for review in total was 129 and the number of suitable papers was 53, for which we had 84 male authors, 36 female authors, and 18 unknown. Nineteen

---

2 We publish our publication rates every six months on the ACL wiki, so all the information is available for all five years.

suitable papers were accepted: 35 male authors, 17 female authors, and 4 unknown. The first author information for the accepted papers showed 13 male first authors, 5 female first authors, 1 unknown.

In compiling these statistics, I was happy to see that submissions by women, the under-represented gender, do not appear to fare worse than those submitted by men. I was, however, intrigued by the fact that most (suitable) submitted papers are single gender (either all men or all women). Here are the statistics for the papers submitted (including those for a special issue): female only 6, male only 26, mixed 15, other 5. This indicates, I think, that the field is somewhat segregated. They also show that women are more represented in the journal, where they are roughly one-third of the authors, compared with the community in general, where they are only 15%. It might be worthwhile asking if different modes of publishing affect genders differently.

1.4 Concerns for the Future and Some Attempted Solutions

The growth of the field and its rapid rate of development and change has also brought some concerns and difficulties of which we need to be aware, both from the scientific and scholarly point of view.

The Journal's Falling Presence Within the Field. There is a perception that *Computational Linguistics* does not publish the interesting, topical research in our field. The indicators of this decrease in attractiveness are citation counts. See Figure 1, which shows Google scholar citations on March 2, 2018. This negative conclusion is modulated by looking at other indicators, such as impact factors, and by keeping in mind that these numbers depend primarily on the total number of papers published in each of these venues. These citation counts, however, certainly indicate that the field has grown, and that most of the research is channelled toward dissemination venues that are possibly not reviewed, or that are reviewed by fast, one-shot procedures.

It is not the role of a journal that publishes long papers to compete with conference publications. This is not possible in a community like ours, where research moves fast, very fast (too fast?), and the practice to publish conference papers is entrenched. *Computational Linguistics* receives more publications that are the considerably extended version of conference and workshop papers and therefore describe research that is more established. This is common for all journals that publish long papers. The work we do, and the mandate we have, is to be the archival publication for a whole body of work, with reproducible results and sound scholarship.

The real concerns is, therefore, not that publications in *Computational Linguistics* are not forward-trend indicators, but that the community does not consider that the reference publications must be journal papers. This is worrying, as we are now a community that not only publishes much of its best work in internally produced conference proceedings that are not counted in citation indices, but one that is diverting publications from long papers in prestigious journals and conferences to unreviewed archival sites. This is either the dawn of a new era in scientific dissemination or collective scholarly suicide.

Special Issues. The editorial board members of the journal had a conversation last year on how to make the journal more interesting and attract more readership. One suggestion was to make more use of special issues. Special issues develop hot topics and attract attention to new developing areas in a coherent way. Our field has had some memorable special issues, like the one in 1993 on corpus-based and statistical methods, edited by

	Publication	h5-index	h5-median
1.	arXiv Computation and Language (cs.CL)	79	140
2.	Meeting of the Association for Computational Linguistics (ACL)	67	103
3.	Conference on Empirical Methods in Natural Language Processing (EMNLP)	58	108
4.	Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (HLT-NAACL)	47	79
5.	International Conference on Language Resources and Evaluation (LREC)	43	71
6.	International Workshop on Semantic Evaluation	37	62
7.	International Conference on Computational Linguistics (COLING)	35	43
8.	Transactions of the Association for Computational Linguistics	32	56
9.	Conference of the European Chapter of the Association for Computational Linguistics (EACL)	31	57
10.	Computational Linguistics	31	55
11.	Computer Speech & Language	28	42
12.	Language Resources and Evaluation	25	43
13.	IEEE Spoken Language Technology Workshop (SLT)	24	39
14.	Workshop on Statistical Machine Translation	23	36
15.	Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)	23	34
16.	Conference on Computational Natural Language Learning (CoNLL)	21	36
17.	International Conference on Computational Linguistics and Intelligent Text Processing	19	28
18.	Joint Conference on Lexical and Computational Semantics	18	43
19.	Workshop on Building Educational Applications Using NLP	18	24
20.	Workshop in Computational Approaches to Subjectivity and Sentiment Analysis	16	34

Dates and citation counts are estimated and are determined automatically by a computer program.

Figure 1
Google Scholar citation counts in the field of computational linguistics.

Susan Armstrong, that marked a turning point in our field. However, special issues are a double-edged sword, as they crystallize interest in a community and define the "must read" papers on a topic and everything that does not end up in the special issue is forgotten. Special issues also compete for the limited space in the journal, potentially squeezing out unsolicited papers. Unsolicited papers should always remain our primary form of publications, as the only kind of submission judged solely on its intrinsic quality. Despite the special issues' ambivalence, we have had the fortune of receiving several very interesting proposals for special issues, which have appeared or which have been accepted and will see the light in the near future.[3]

2. Let Us Proceed Slowly, For We Are in A Hurry

At the ACL 2017 business meeting there was a panel on publication practices in our field. Currently, it is very common to post papers on arXiv—a self-publication venue—assuming that comments from peers will make the papers better (see Figure 1). As the

---

3 Formal distributional semantics, editors A. Herbelot and G. Boleda, issue 42:4 in December 2016; The language of social media, editors M. Taboada, D. Inkpen, and F. Benamara, to be published in 2018; Computational approaches in historical linguistics, editors T. Rama, S. J. Greenhill, H. Hammarstrom, G. Jaeger, J. M. List, and 'adjunct editor' R. Sproat from the board. Call is out and will be closed July 15, 2018.

editor of the journal, I pointed out the potential problems with unreviewed papers. Based on the most frequent shortcomings of papers submitted to *Computational Linguistics*, we can anticipate the most common problems with self-publications: Flawed or inadequate scholarship and contextualization of work; overly positive assessment of novelty; methodological inaccuracies; poor, unclear writing. Although these problems might, in some cases, be corrected by peer replies and commentaries in arXiv, I think it is fair to expect that in many cases they will not. And even in the cases in which they will, it is not clear that quick dissemination of possibly flawed work followed by fast and frequent revisions is the best way to advance knowledge. I think then that we must ask ourselves the long-term questions of what creates progress in science and what progress we want for our scientific community.

- What is the status of a scientific community whose main means of scientific dissemination is not reviewed (or reviewed under the Publish Everything model)?

- What are fair publication practices that benefit everybody?

- What advances the scientific collective more: fast, small-stepped, possibly messy, progress or slow and methodical collective construction of correct results?

These are not idle or rhetorical questions, as there is real tension between publication practices that focus on the spearheading researchers and foster the driving forces of the scientific community—high-achieving, productive, innovative, well-supported researchers in prestigious, leading institutions—and publication practices that might be more accessible to individually brilliant scientists, under-represented or under-funded groups, more remote and geographically and topically more peripheral. There is also, more fundamentally, real tension between carefully reviewed work, which maximizes high scientific precision, and fast dissemination with frequent revisions, which maximises high scientific recall.

In this context—and in my opinion, please remember, this is an opinion piece—it is important to remind ourselves that fast dissemination suffers from what have been dubbed the Proteus phenomenon and the very closely-related winner's curse. Fast dissemination also often implies carving our research in smaller pieces that can be performed in less time and described in fewer pages, what has been called bite-size science.

Short Articles and Bite-size Science. In a 2012 study, Bertamini and Munafò discuss the effect of publishing shorter and shorter papers, and in particular the effect of smaller studies toward finding the true answers (Bertamini and Munafò 2012). Everybody agrees that larger studies will involve larger samples and therefore be more accurate, and smaller studies will have a wider spread around the mean. It is usually argued, though, that smaller, or even less accurate, but more numerous studies will have the same collective effect as larger studies. Both will, in the end, center on the true effect. Bertamini and Munafò argue, and show, that small studies are more biased in favor of positive results than large studies, while null or negative results will be dismissed. But small studies are also more likely than large studies to be false positives, if they are

statistically significant.[4] Bertamini and Munafò remind us, citing McManus (2002, 322), that false positives are a problem because "erroneous ideas much more easily enter the literature than leave it." These reflections do not bode well for the trend toward shorter publications and should make us ponder.

The Proteus Phenomenon and the Winner's Curse. The Proteus phenomenon—a term coined by Ioannidis and Trikalinos (2005)—describes the effect of rapidly alternating extreme research claims and extremely opposite refutations, particularly during the early accumulation of data. More precisely, Ioannidis and Trikalinos argue that early replication of results in a fast-moving scientific field are likely to refute the original findings, and conclude that the hotter a scientific field, the less likely the findings are to be true. The winner's curse describes the fact that the winner of an auction is the highest-bidder and, therefore, compared with the average, the one that most overestimates the value of the object. The winner's curse affects journals that only accept the most ground-breaking findings. Meta-research shows that first publication of results have a considerably higher chance of being inflated.[5]

The Publish Everything model—where papers are either not reviewed or rejected only if they have technical flaws—has been proposed as a remedy. But, in a simulation, de Winter and Happee (2013) compare the Publish Everything approach versus the Selective Publication approach, and show that the latter reaches the true effect faster and is therefore beneficial for the scientific collective. Their model simulates the situation where researchers worldwide are investigating the strength of an effect by means of identical experiments and that the observed effects appear in published articles. Formally, the observed effects are generated by independent random sampling of n subjects from a normal distribution. In the Publish Everything approach, observed effects are always published, irrespective of their magnitude or direction. In the Selective Publication approach, statistically significant findings are published and nonsignificant findings are not published. The model assumes that science is self-correcting: Past alternative hypotheses become null hypotheses to falsify. It also takes into account shortcomings in establishing the null hypothesis: "flawed scholarship" (which they approximate by ignoring the last three publications), overconfidence in one's theory, and ignoring the meta-truth.

They conclude that their simulation shows that "instead of publishing everything, it is worthwhile to be selective and publish only research findings that are statistically significant. After a number of publications, selective publishing yields a more accurate meta-analytic estimation of the true effect than publishing everything."

Dull is Good for Science. I would like to submit that these results alert us to the fact that we must find scholarly solutions that are not simply publishing more, faster and in smaller units. Instead, we need to hurry slowly, avoiding errors and repetitions and catching mistakes early in the process. In this context, it pays to remember the perhaps unfashionable words of the slow science manifesto.

---

4 They say: "A more subtle problem is that results in small studies are more likely to be false positives if they are statistically significant. This is because the false positive rate remains constant at 5%, whereas the true positive rate (the power) depends on sample size. Therefore, as power/sample size decreases, the ratio of true positives to false positives among those studies that achieve statistical significance also decreases."

5 "In studies involving many tests on one sample of the full population, the consequent stringent standards for significance make it likely that the first person to report a significant test (the winner) will also report an effect size much larger than is likely to be seen in subsequent replication studies." (Ioannidis 2008)

The Slow Science Manifesto

We are scientists. We don't blog. We don't twitter. We take our time.

Don't get us wrong–we do say yes to the accelerated science of the early 21st century. We say yes to the constant flow of peer-review journal publications and their impact; (...) All of us are in this game, too.

However, we maintain that this cannot be all. Science needs time to think. Science needs time to read, and time to fail. Science does not always know what it might be at right now. Science develops unsteadily, with jerky moves and unpredictable leaps forward—at the same time, however, it creeps about on a very slow time scale, for which there must be room and to which justice must be done.

(...)

—Bear with us, while we think.

3. Goodbye

Before going back to my pottery and my own research, I would like to thank all those that have made these five years possible: Robert Dale, for handing over the journal to me in such great state and for having established many good practices; the wonderful editorial assistants over the years, Tanja Samardzic, Sarah Ouwayda, Sharid Loaiciga, and Cristina Grisot, who have made my life so much easier, with unfailing patience; all the people at MIT Press, especially Levi Rubeck, for being such wonderful publishers; the squib editors and book review editors, Pierre Isabelle, Graeme Hirst, Mike White, and Hwee Tou Ng, for running their part of the journal so impeccably. I am very grateful to all the reviewers and members of the editorial board who over the years have devoted much of their precious time to reading the papers and providing careful, competent, rich reviews that have greatly improved all the submissions. Finally, to all the authors, for trusting that our journal would treat their work, often the effort of several years, with diligence, care, and respect, providing an influential platform to make their ideas known. And to the readers, for taking the time to read us.

It has been an interesting experience and a privilege to be the editor of this great journal and to support for five years its main form of publication: the long paper.

References

Bertamini, Marco and Marcus R. Munafò. 2012. Bite-size science and its undesired side effects. *Perspectives on Psychological Science*, 7:67–71.

Ioannidis, John. 2008. Why most discovered true associations are inflated. *Epidemiology*, 19(5):640–648.

Ioannidis, John and Thomas A. Trikalinos. 2005. Early extreme contradictory estimates may appear in published research: The Proteus phenomenon in molecular genetics research and randomised trials. *Journal of Clinical Epidemiology*, 58(6):543–549.

McManus, I. Chris. 2002. *Right Hand, Left Hand: The Origins of Asymmetry in Brains, Bodies, Atoms and Cultures*. Harvard University Press/Harvard University Press, London, UK/Cambridge, MA.

de Winter, Joost and Riender Happee. 2013. Why selective publication of statistically significant results can be effective. *PLOS ONE*, 8(6):1–7.