



Language Technology: Research and Development

Review of Scientific Articles

Sara Stymne

Uppsala University
Department of Linguistics and Philology
sara.stymne@lingfil.uu.se

Partially based on slides from Joakim Nivre



Review of Scientific Articles

- ▶ Why?
 - ▶ Maintain standards of quality for scientific publications
 - ▶ Improve quality by giving feedback to authors
 - ▶ Evaluate quality of work for selection or promotion
- ▶ By whom?
 - ▶ Experts of the scientific community
 - ▶ At least the same level of seniority



Peer Review

- ▶ In general:
 - ▶ Evaluation of work by one or more people of similar competence to the producers of the work
- ▶ Editorial peer review
 - ▶ Reviewing a paper to determine its suitability for publication
 - ▶ *Philosophical Transactions of the Royal Society* (1665)
 - ▶ Firmly established in the scientific community
- ▶ Justification
 - ▶ Expert reviewers guarantee scientific quality (and novelty)
 - ▶ Multiple views guarantee diversity of opinion
 - ▶ Reviewers reduce workload for editors



Procedure

- ▶ The journal review cycle:
 1. Author submits article
 2. Editor appoints reviewers
 3. Reviewers submit reviews
 4. Editor makes decision:
 - ▶ Accept for publication (possibly with minor revisions)
 - ▶ Revise and resubmit (repeat cycle)
 - ▶ Reject

- ▶ Anonymity:
 - ▶ Double-blind: authors and reviewers are anonymous
 - ▶ Single-blind: reviewers (but not authors) are anonymous
 - ▶ Open: neither authors nor reviewers are anonymous



Language Technology Journals

- ▶ Computational Linguistics
 - ▶ Published by MIT Press on behalf of the ACL
 - ▶ Open access since 2009
 - ▶ Editor in chief elected for 5 years (Hwee Tou Ng since 2018)
 - ▶ Around 20 editorial board members (elected for 3 years)
- ▶ Other journals (selected):
 - ▶ Natural Language Engineering
 - ▶ Computer Speech and Language
 - ▶ Language Resources and Evaluation
 - ▶ Machine Translation
 - ▶ Traitement Automatique des Langues
 - ▶ Northern European Journal of Language Technology
- ▶ See also: <http://aclweb.org/aclwiki/index.php?title=Journals>



Conference Reviewing

1. Author submits article
2. Chairs assign reviewers
3. Reviewers submit reviews
4. (Author response + reviewers may update reviews)
5. Chairs make decision:
 - ▶ Accept
 - ▶ Reject
6. Author submits final version, taking reviews into account



Conference Reviewing

1. Author submits article
2. Chairs assign reviewers
3. Reviewers submit reviews
4. (Author response + reviewers may update reviews)
5. Chairs make decision:
 - ▶ Accept
 - ▶ Reject
6. Author submits final version, taking reviews into account
 - ▶ Responsibility of chairs typically divided between program chairs, senior area chairs, and area chairs



Language Technology Conferences

- ▶ Conferences organized by ACL and its SIGs:
 - ▶ Annual Meeting of the ACL
 - ▶ Conference of the European Chapter (EACL)
 - ▶ Conference of the North American Chapter (NAACL)
 - ▶ Conference of the Asia-Pacific Chapter (AACL)
 - ▶ Conference on Empirical Methods in NLP (EMNLP)
 - ▶ Conf. on Computational Natural Language Learning (CoNLL)
- ▶ All these conferences typically have a high number of associated workshops



Language Technology Conferences (2)

- ▶ Other conferences (selected):
 - ▶ Int. Conf. on Computational Linguistics (COLING)
 - ▶ Int. Conf. on Language Resources and Evaluation (LREC)
 - ▶ International Joint Conference on NLP (IJCNLP)
 - ▶ Nordic Conference on Computational Linguistics (NODALIDA)
 - ▶ Swedish Language Technology Conference (SLTC)



Language Technology Conferences (2)

- ▶ Other conferences (selected):
 - ▶ Int. Conf. on Computational Linguistics (COLING)
 - ▶ Int. Conf. on Language Resources and Evaluation (LREC)
 - ▶ International Joint Conference on NLP (IJCNLP)
 - ▶ Nordic Conference on Computational Linguistics (NODALIDA)
 - ▶ Swedish Language Technology Conference (SLTC)
- ▶ See also: <https://www.aclweb.org/anthology/>



Transactions of the ACL

- ▶ Started by ACL in 2012 for two purposes:
 - ▶ Improve review procedure for ACL conferences
 - ▶ Give the best conference papers journal status
- ▶ Hybrid model:
 - ▶ Conference-length papers (7–10 pages + references)
 - ▶ Journal style reviewing (but fast turnaround)
 - ▶ Publisher papers get a presentation slot at an ACL conference
- ▶ Editorial organization:
 - ▶ 2–4 editors in chief (Ani Nenkova, Brian Roark)
 - ▶ Around 80 action editors (appointed for 3 years)
 - ▶ A pool of standing reviewers (250–300)



Submission Policy

- ▶ Submissions must describe substantial, original, completed and unpublished work.
- ▶ Submissions will be judged on correctness, originality, technical strength, significance, and relevance to computational linguistics and natural language processing.
- ▶ We invite papers in the following four broad categories: theoretical computational linguistics, empirical/data-driven approaches, resources/evaluation, and applications/tools.



Review Form

- ▶ Numerical scores:
 1. Clarity
 2. Innovativeness
 3. Soundness/Correctness
 4. Related work
 5. Substance
 6. Impact of ideas or results
 7. Replicability
 8. Impact of accompanying software
 9. Impact of accompanying data set
 10. TACL-worthy as is?
- ▶ Comments:
 1. Detailed comments for authors
 2. Confidential comments for the editor
 3. Reviewer confidence



Clarity

For the reasonably well-prepared reader, is it clear what was done and why? Is the paper well-written and well-structured?

1. Much of the paper is confusing.
2. Important questions were hard to resolve even with effort.
3. Mostly understandable to me with some effort.
4. Understandable by most readers.
5. Very clear.



Originality/Innovativeness

How original is the approach? Does this paper break new ground in topic, methodology, or content? How exciting and innovative is the research it describes?

Note that a paper could score high for originality even if the results do not show a convincing benefit.

1. Significant portions have actually been done before or done better.
2. Pedestrian: Obvious, or a minor improvement on familiar techniques.
3. Respectable: A nice research contribution that represents a notable extension of prior approaches or methodologies.
4. Creative: An intriguing problem, technique, or approach that is substantially different from previous research.
5. Seminal: Significant new problem, technique, methodology, or insight – no prior research has attempted something similar.



Soundness/Correctness

First, is the technical approach sound and well-chosen? Second, can one trust the claims of the paper – are they supported by proper experiments and are the results of the experiments correctly interpreted?

1. Fatally flawed.
2. Troublesome. There are some ideas worth salvaging here, but the work should really have been done or evaluated differently.
3. Fairly reasonable work. The approach is not bad, and at least the main claims are probably correct, but I am not entirely ready to accept them (based on the material in the paper).
4. Generally solid work, although there are some aspects of the approach or evaluation I am not sure about.
5. The approach is very apt, and the claims are convincingly supported.



Related work

Does the submission make clear where the presented system sits with respect to existing literature? Are the references adequate? Note that the existing literature includes preprints, but in the case of preprints: Authors should be informed of but not penalized for missing very recent and/or not widely known work. If a refereed version exists, authors should cite it in addition to or instead of the preprint.

1. Little awareness of related work, or insufficient justification of benefits and discussion of limitations.
2. Only partial awareness and understanding of related work, or a flawed comparison or deficient comparison with other work.
3. Bibliography and comparison are somewhat helpful, but it could be hard for a reader to determine exactly how this work relates to previous work or what its benefits and limitations are.
4. Mostly solid bibliography and comparison, but there are a few additional references that should be included. Discussion of benefits and limitations is acceptable but not enlightening.



Substance

Does this paper have enough substance (in terms of the amount of work), or would it benefit from more ideas or analysis? Note that papers or preprints appearing less than three months before a paper is submitted to TACL are considered contemporaneous with the submission. This relieves authors from the obligation to make detailed comparisons that require additional experiments and/or in-depth analysis, although authors should still cite and discuss contemporaneous work to the degree feasible.

1. Seems thin. Not enough ideas here for a full-length paper.
2. Work in progress. There are enough good ideas, but perhaps not enough results yet.
3. Leaves open one or two natural questions that should have been pursued within the paper.
4. Represents an appropriate amount of work for a publication in this journal. (most submissions)
5. Contains more ideas or analysis than most publications in this journal; goes the extra mile.



Impact of Ideas or Results

How significant is the work described? If the ideas are novel, will they also be useful or inspirational? If the results are sound, are they also important? Does the paper bring new insights into the nature of the problem?

1. Will have no impact on the field.
2. Marginally interesting. May or may not be cited.
3. Interesting but not too influential. The work will be cited, but mainly for comparison or as a source of minor contributions.
4. Some of the ideas or results will substantially help other people's ongoing research.
5. Will affect the field by altering other people's choice of research topics or basic approach.



Replicability

Will members of the ACL community be able to reproduce or verify the results in this paper?

Members of the ACL community

1. could not reproduce the results here no matter how hard they tried.
2. would be hard pressed to reproduce the results. The contribution depends on data that are simply not available outside the author's institution or consortium; not enough details are provided.
3. could reproduce the results with some difficulty. The settings of parameters are underspecified or subjectively determined; the training/evaluation data are not widely available.
4. could mostly reproduce the results, but there may be some variation because of sample variance or minor variations in their interpretation of the protocol or method.
5. could easily reproduce the results.



Impact of Accompanying Software

If the authors state (in anonymous fashion) that their software will be available, what is the expected impact of the software package?

1. No usable software released.
2. Documentary: The new software useful to study or replicate the reported research, although for other purposes they may have limited interest or limited usability. (Still a positive rating)
3. Potentially useful: Someone might find the new software useful for their work.
4. Useful: I would recommend the new software to other researchers or developers for their ongoing work.
5. Enabling: The newly released software should affect other people's choice of research or development projects to undertake.



Impact of Accompanying Data Set

If the authors state (in anonymous fashion) that datasets will be released, how valuable will they be to others?:

1. No usable data sets submitted.
2. Documentary: The new data sets are useful to study or replicate the reported research, although for other purposes they may have limited interest or limited usability. (Still a positive rating)
3. Potentially useful: Someone might find the new datasets useful for their work.
4. Useful: I would recommend the new datasets to other researchers or developers for their ongoing work.
5. Enabling: The newly released datasets should affect other people's choice of research or development projects to undertake.



TACL-worthy as is?

In answering, think over all your scores above. If a paper has some weaknesses, but you really got a lot out of it, feel free to recommend it. If a paper is solid but you could live without it, let us know that you're ambivalent.

1. Poor: I'd fight to have it rejected.
2. Leaning against: I'd rather not see it appear in TACL.
3. Ambivalent: OK but does not seem up to the standards of TACL.
4. Worthy: A good paper that is worthy of being published in TACL.
5. Strong: I'd like to see it accepted; it will be one of the better papers in TACL.
6. Exciting: I'd fight to get it accepted; probably would be one of the best papers in TACL this year.



Reviewer Confidence

1. Not my area, or paper is very hard to understand. My evaluation is just an educated guess.
2. Willing to defend my evaluation, but it is fairly likely that I missed some details, didn't understand some central points, or can't be sure about the novelty of the work.
3. Pretty sure, but there's a chance I missed something. Although I have a good feel for this area in general, I did not carefully check the paper's details, e.g., the math, experimental design, or novelty.
4. Quite sure. I tried to check the important points carefully. It's unlikely, though conceivable, that I missed something that should affect my ratings.
5. Positive that my evaluation is correct. I read the paper very carefully and am familiar with related work.



Detailed Comments

Several purposes:

- ▶ Back up numerical scores by detailed arguments
- ▶ Help editor make a decision
- ▶ Help author improve the next version

Standard outline:

1. **Summary:** What has been done? What is the point?
2. **Major points:** What are the strengths and weaknesses?
3. **Conclusion:** Is it worth publishing? What can be improved?
4. **Minor points:** Presentational issues, typos, etc.



Conference reviewing

- ▶ Many NLP conferences has recently moved away from having several numerical scales
- ▶ Focus on textual motivations:
 - ▶ Strengths/weaknesses
 - ▶ Reasons to accept/reject



Conference reviewing: EMNLP 2020

1. In-Depth Review:

- ▶ The core review (text)
- ▶ Reasons to accept (text)
- ▶ Reasons to reject (text)
- ▶ Reproducibility (numerical)
- ▶ Overall recommendation (numerical)
- ▶ Reviewer confidence (numerical)
- ▶ Author response (yes/no)

2. Questions and Additional Feedback for the Authors (text)

3. Confidential Information (text)



Shift in reviewing model(?)

1. 2021: ACL rolling review
 - ▶ Monthly submission (not to a specific venue)
 - ▶ Quick turnaround
 - ▶ Possibility to revise and resubmit
 - ▶ Once reviews received, can commit the paper to a conference or workshop
 - ▶ Possible to publish an anonymous preprint at submission time
2. Used as a complement by some conferences and workshops in 2021
3. Adopted by ACL and NAACL in 2022



Conference reviewing: ARR form

1. Textual fields:
 - ▶ Paper summary
 - ▶ Summary of strengths
 - ▶ Summary of weaknesses
 - ▶ Comments, suggestions and typos
2. Numerical fields
 - ▶ Overall assessment
 - ▶ Confidence
 - ▶ Replicability
 - ▶ Best paper
 - ▶ software; datasets
 - ▶ author identity guess



Reviewing Advice

- ▶ Start with a summary of the paper – show the authors and editors that you have understood the main points
- ▶ Be specific – vague comments are unhelpful
- ▶ Be kind – write comments that you would like to receive
- ▶ Be honest – hiding weaknesses is not helpful (nor kind)
- ▶ Be organized – make clear what are required changes (as opposed to minor improvements)
- ▶ Don't leave reviewing to the last minute!

<https://acl2017.wordpress.com/2017/02/23/last-minute-reviewing-advice/>

A lot of good advice on recent conference sites



Reviewing for the Course

No numerical scores, only detailed comments

- ▶ But use the criteria for the numerical scores as a checklist

Try to follow the standard outline:

1. **Summary:** What is the paper about? Use your own words.
2. **Major points:** What are the strengths and weaknesses?
3. **Conclusion:** Summarize your review. Give specific suggestions (for improvements).
4. **Minor points:** Presentational issues, typos, etc.

Single-blind reviewing



Practicalities



EasyChair

- ▶ We will use a real conference system for first submission and reviewing: [EasyChair](#)
- ▶ You will need to:
 1. Get an EasyChair account:
<https://easychair.org/conferences/?conf=ltrd22>
 2. Send me an email with the following line:
FirstName LastName <EmailAddress>
Examples:
Sara Stymne <sara.stymne@lingfil.uu.se>
"Magdalena Carmen Frida" "Kahlo y Calderón"
<made.up@p.mx>
 3. Accept the invitation from EasyChair, which you should receive by email (could take a few days)



EasyChair

- ▶ We will use a real conference system for first submission and reviewing: [EasyChair](#)
- ▶ You will need to:
 1. Get an EasyChair account:
<https://easychair.org/conferences/?conf=ltrd22>
 2. Send me an email with the following line:
FirstName LastName <EmailAddress>
Examples:
Sara Stymne <sara.stymne@lingfil.uu.se>
"Magdalena Carmen Frida" "Kahlo y Calderón"
<made.up@p.mx>
 3. Accept the invitation from EasyChair, which you should receive by email (could take a few days)
 4. If you get no invitation within a few days: (1) check your spam folder (2) contact Sara again!



Term papers

- ▶ The paper submitted on December 13 should be a **complete** paper
 - ▶ Use the Latex templates available for TACL
 - ▶ Do not anonymize
- ▶ You may extend and revise it until the final deadline
- ▶ The final version should take review comments into account
 - ▶ For comments that you agree with, try to comply with them, or at least discuss the issues raised
 - ▶ If you do not agree with a comment, you do not need to follow it (But please consider all comments carefully)



Final seminar

- ▶ January 13
- ▶ Full day workshop on Campus
- ▶ Full and half class mix
 - ▶ Type of presentation based on topics
 - ▶ If you feel strongly for presenting to all or in a smaller group, let me know by email (We might not be able to accomodate all wishes, though)
- ▶ Social event after the workshop (if restrictions permit)
- ▶ Presentation schedule: January



Final seminar

- ▶ January 13
- ▶ Full day workshop on Campus
- ▶ Full and half class mix
 - ▶ Type of presentation based on topics
 - ▶ If you feel strongly for presenting to all or in a smaller group, let me know by email (We might not be able to accommodate all wishes, though)
- ▶ Social event after the workshop (if restrictions permit)
- ▶ Presentation schedule: January
- ▶ If you cannot present on Campus, email Sara as soon as you know, with a motivation of why



Deadlines

	Main	Backup (avoid!)
First version	Dec 13	Jan 14
Final version	Jan 14	Feb 18
Reviewing	Dec 22	Feb 18



In case you are late

- ▶ Note that you may not start the thesis course until you have passed R&D
- ▶ For students keeping to the original deadlines we make a quick pass assessment in connection to the deadline (before we finalize grades), in order for you to start on time
- ▶ For students missing the December 13 deadline
 - ▶ You may review (let Sara know if you want to do so by email)
 - ▶ In case you are sufficiently done with your project, you may present at the workshop (let Sara know by email)
 - ▶ You will probably not get any student reviews for your first version of the term paper
 - ▶ You will be delayed in starting the thesis course