How to Write a Paper*

MeMentor Session

Julia Kreutzer

24 January

*Disclaimer: still learning…

Slides: shorturl.at/joV56

Today

After this session, you should *ideally*:

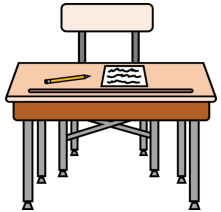
- Know a few concrete **exercises and tricks** to help you get writing.
- Have a better idea about the **world around [NLP] papers**.
- Find the whole process **less frightening** than before.
- Know a few **friends** who are on the same journey as you.

This session is not about finding topics or problems to write about, but rather about the *process* from an experiment/insight to a paper, specifically for NLP.

Introductions

Terminology

Double-blind



Desk-reject

Reviewer 2

Rebuttal / author response

Program Chair / Area Chair

SoftConf / EasyChair / OpenReview

Dual submission

Non-archival

Preprint / ArXiv

Proceedings

Camera-ready

LaTex templates

Appendix

Bibliography

Agenda

1. **AfricaNLP Workshop**
2. **The Writing Process**
    a. The Story
    b. Selling the Paper
    c. Writing the Paper
3. **Technicalities**
    a. LaTeX, Overleaf, SoftConf

Please leave questions in the chat or raise your hand.

The AfricaNLP Workshop

AfricaNLP Workshop

Join the slack workspace! Link on website

When? EACL April 19/20

Where? Online

What? See website, 2020's papers for inspiration

Important dates:

February 5: Paper submissions due

March 1: Notification of acceptance

https://sites.google.com/view/africanlp-workshop

AfricaNLP Workshop

Paper format:

- Length: long (up to 8 pages) and short (up to 4 pages)
- Types: research papers, resource papers, position papers, tutorial/overview papers, and system design papers

Good to know:

- Extra focus on reproducibility of empirical results and accessibility of resources
- Unlimited appendix and space for the author names
- We encourage shorter submissions
- Optional: include an abstract in your home language
- Conference sponsorship for authors

AfricaNLP Workshop

Submission:

- Non-archival
- Dual submissions are allowed as well but must be indicated
- Anonymized
- Submit through the softconf platform
- In PDF following the official EACL LaTeX template (zip) or ACL LaTeX/Word templates
- Will be reviewed by 2-3 reviewers
- Will be presented in virtual format if accepted

The Paper Writing Process

Life of a Paper

1. Initial idea / incentive / question
2. Research: the hard work
3. Writing: even harder (?)
4. Review
5. Optionally restart from 1, 2, 3, 4 until accepted
6. Online upload
7. Conference presentation
8. Follow-up: citations, implementations, Twitter hype ...

"The Paper"

Life of a Paper

1. Initial idea / incentive / question
2. Research: the hard work
3. **Writing: even harder (?)**
4. Review
5. Optionally restart from 1, 2, 3, 4 [illegible]cepted

---

6. Online upload
7. Conference presentation
8. Follow-up: citations, implementations, Twitter hype ...

Necessary evil but underestimated impact!

"The Paper"

Example

Related Work

Mine

<My Paper>

?

I'm not a writer.
I've done the main job already.
It doesn't matter.
No one will like/read it.

Which title?
Which tables?
Which plots?
How to introduce?

The Story

Exercise: The Story

The most significant feature of my sculpture is that ...

What I learned when building it was ...

It distinguishes itself from others by ...

Having done this, the next step would be to ...

It serves the purpose of ...

Exercise: The Story

The most significant feature of my sculpture is that ...

What I learned when building it was ...

It distinguishes itself from others by ...

Having done this, the next step would be to ...

It serves the purpose of ...

Selling the Paper

Example: Contribution

The most significant feature of my so[illegible]e is that ...

New feature

New data

What I learned when building it was

New method

New insight

It distinguishes itself from others by ...

Novel perspective

It serves the purpo[illegible]

New application

Having done this, the next step would be to

New avenues

Example: Contribution

The most significant feature of my solution is that ...

New feature

New data

Focus on the strongest one(s), reviewers will be looking for it!

What I learned when building it was ...

New method

New insight

It distinguishes itself from others by ...

Novel perspective

It serves the purpose ...

New application

Having done this, the next step would be to ...

New avenues

Empirical Results

	Metric 1	***Metric 2***
Baseline	30.1	**28.9**
Mine	**30.4***	7.1*
- w/o feature #1	30.3*	27.9*
- w/o feature #2	30.1	27.1*

Core result

*Significant difference to baseline according to test X

Empirical Results

	Metric 1	***Metric 2***
Baseline	30.1	**28.9**
Mine	**30.4***	**29.5***
- w/o feature #1	30.3*	27.9*
- w/o feature #2	30.1	27.1*

Comparison against prior or simpler work

*Significant difference to baseline according to test X

Empirical Results

	Metric 1	***Metric 2***
Baseline	30.1	**28.9**
Mine	**30.4***	
- w/o feature #1	30.3*	
- w/o feature #2	30.1	27.1*

A second metric that measures something different

*Significant difference to baseline according to test X

Empirical Results

	Metric 1	***Metric 2***
Baseline	30.1	**28.9**
Mine	**30.4***	27.1*
- w/o feature #1	30.3*	27.9*
- w/o feature #2	30.1	27.1*

The reality is often more complex than one would like.
Try to explain it and stay true to the data.
Candidate for future work or analysis.

*Significant difference to baseline according to test X

Empirical Results

	Metric 1	*Metric 2*
Baseline	30.1	**28.9**
Mine	**30.4***	27.1*
- w/o feature #1	30.3*	27.9*
- w/o feature #2	30.1	27.1*

Significance tests.

Plus points for multiple repeated runs and mean and variance reporting.

*Significant difference to baseline according to test X

Empirical Results

	Metric 1	***Metric 2***
Baseline	30.1	**28.9**
Mine	**30.4***	27.1*
- w/o feature #1	30.3*	27.9*
- w/o feature #2	30.1	27.1*

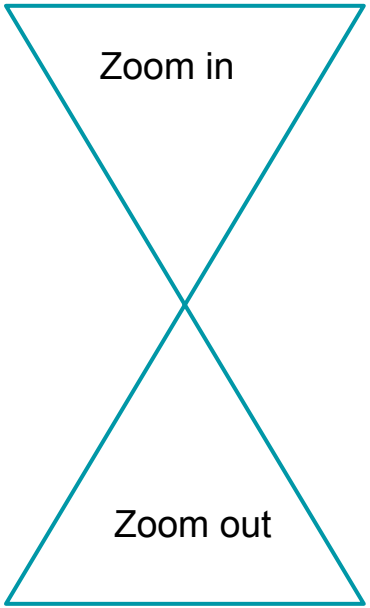
Ablation for your new model: how important is each component?

*Significant difference to baseline according to test X

Writing the Paper

Suggested Order(s)

1. Story
2. Structure
3. Skeleton
4. Core tables & plots [might also start here]
5. Descriptions and interpretation
6. Methods [might also start here]
7. Related
8. Introduction
9. Conclusion
10. Abstract
11. Adapt length (appendix!), reorganize, refine, ...



Order Suggestion(s)

1. Story
2. Structure
3. Skeleton
4. Core tables & plots [might also start here]
5. Descriptions and interpretation
6. Methods [might also start here]
7. Related
8. Introduction
9. Conclusion
10. Abstract
11. Adapt length (appendix!), reorganize, refine, ...

Reviewers will be looking for:

- Clarity
- Relevance
- Novelty
- Substance
- Reproducibility
- Contributions
- Claims and Support

There’s not only one way

RQ1 Is a *local* view of individual gaze trace beneficial for syntactic sequence labeling in comparison to an aggregate *global* view, where information is traced via i) the central tendency (mean) or ii) the variability (variance) of the gaze behavior?

RQ2 How well does learning from decontextualized gaze data represented at the *type*-level (as dictionary) perform in comparison to learning from contextualized gaze data, represented at the *token*-level (via multi-task learning)?

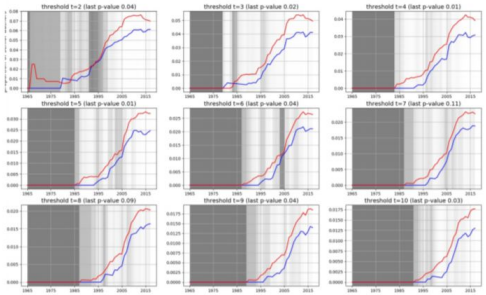


Figure 1: Proportion of male (red line) and female (blue line) mentors from 1966 to 2017, for different thresholds of “mentor seniority”. A disparity in these proportions has been increasing since the 1990s, which follows the general field of Computer Science. The whiter the background the more significant the difference in proportions for the corresponding year, with the p-value for 2017 in the title.

$$\lim_{\alpha \to 0} \mathrm{J}_\alpha(q \,||\, p_\theta) = \mathrm{KL}(p_\theta \,||\, q) \quad (10)$$

$$\lim_{\alpha \to 1} \mathrm{J}_\alpha(q \,||\, p_\theta) = \mathrm{KL}(q \,||\, p_\theta) \quad (11)$$

$$\mathrm{J}_{1/2}(q \,||\, p_\theta) = 4 \cdot \mathrm{JS}(q \,||\, p_\theta) \quad (12)$$

We prove these relationships in App. A and App. B. For ease, we define $\mathrm{J}_1 := lim_{\alpha \to 1} \mathrm{J}_\alpha$ and $\mathrm{J}_0 := \lim_{\alpha \to 0} \mathrm{J}_\alpha$. We note the following two equivalences for these special cases.

Proposition 1. $\nabla_\theta \mathrm{J}_1(u \,||\, p_\theta) = \nabla_\theta \mathrm{H}(q, p_\theta)$. *In words, the gradient of the loss with GER as $\alpha \to 1$ is equivalent to the gradient of the loss augmented with* ***label smoothing****.*

Proposition 2. $\nabla_\theta \mathrm{J}_0(u \,||\, p_\theta) = \nabla_\theta \mathrm{H}(p_\theta)$. *In words, the gradient of the loss with GER as $\alpha \to 0$ is equivalent to the gradient of the loss augmented with the* ***confidence penalty****.*

the learned distribution.

Magnitude. Figure 1 shows the different divergence measures between u and p_θ. We see that $\mathrm{J}_1 = \mathrm{KL}(u \,||\, p_\theta)$ **(label smoothing)** is much larger than $\mathrm{J}_0 = \mathrm{KL}(p_\theta \,||\, u)$ **(confidence penalty)** at values of p_θ farther from u. This indicates that J_1 would be a stronger regularizer than $\mathrm{J}_{<1}$, i.e. penalize values of p_θ far from u more heavily, given the same strength coefficient β. Note that it is not always the case that $\mathrm{J}_{<1}(u \,||\, p) \leq \mathrm{J}_1(u \,||\, p)$ for fixed p. We can, however, bound J_α from above and below by other quantities.

Proposition 3. *The divergence $\mathrm{J}_\alpha(u \,||\, p)$ is not a monotonic function of α for all distributions p.*

A proof by counter example is shown in Figure 2.

Proposition 4. *For fixed p, J_α has bounds:* $0 \leq \mathrm{J}_\alpha(u \,||\, p) \leq \mathrm{KL}(u \,||\, p) + \mathrm{KL}(p \,||\, u)$.

See App. E for a proof.

Input: the s-th source sentence in the training data $\mathbf{x}^{(s)}$, the s-th target sentence in the training data $\mathbf{y}^{(s)}$, the set of model parameters $\boldsymbol{\theta}$, the limit on the length of a candidate translation l, and the limit on the size of sampled space k.
Output: sampled space $\mathcal{S}(\mathbf{x}^{(s)})$.

```
S(x^(s)) ← {y^(s)}; // the gold-standard translation is included
i ← 1;
while i ≤ k do
    y ← ∅; // an empty candidate translation
    n ← 1;
    while n ≤ l do
        y_n ~ P(y_n | x^(s), y_<n; θ); // sample the n-th target word
        y ← y ∪ {y_n};
        if y_n = EOS then
            break; // terminate if reach the end of sentence
        end
        n ← n + 1;
    end
    S(x^(s)) ← S(x^(s)) ∪ {y};
    i ← i + 1;
end
```

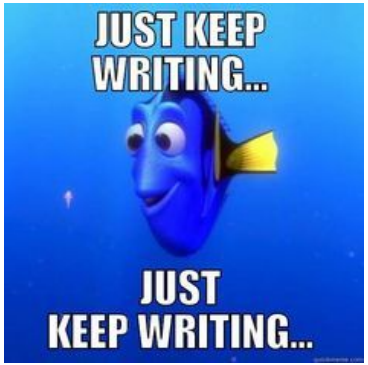
Algorithm 1: Sampling the full search space.

Explore the ACL Anthology
https://www.aclweb.org/anthology/

OK, WRITING SHOULDN'T BE SO HARD.
EVERYTHING I WANT TO SAY IS IN MY HEAD...
I JUST HAVE TO TRANSFER IT FROM MY BRAIN TO THIS COMPUTER.
24 in
61 cm
WRITING: THE MOST IMPOSSIBLE SHORT DISTANCE IN THE HISTORY OF HUMANITY.
c'mon!
JORGE CHAM © 2014
WWW.PHDCOMICS.COM

Practical Exercises

1. **“Why?”**
2. What does the paper not try to do? Set boundaries (profs won’t do it for you)
3. Pitch to Senior
4. **Explain to non-expert**
5. Explain findings to your past self
6. **Write the first word: one sentence per section**
7. Collect questions
8. Write down all doubts
9. Read paper aloud, for stylistic improvements



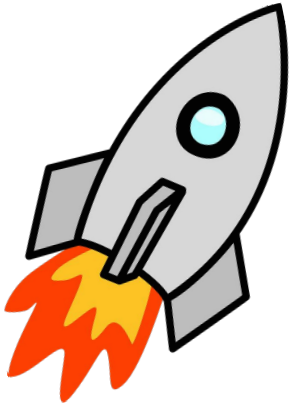
2 General Tips

1. **Guide the reader.**
   - Anticipate questions and doubts
   - Simulate their reading process: what do they (not) know?
   - Make it “interesting”: everyone likes to learn.
   - Structure should reflect logic.

2. **Welcome feedback, but treat it with care.**
   - Reviews can be harsh.
   - Don’t be afraid to “kill your darlings”.
   - If something is perceived as “unclear”: focus on this point for revisions.
   - Spellcheckers rarely hurt.

Tips on Collaborative Writing

- Align on the mission and story
- Commit on a time plan
- Identify everybody’s expertise and strengths
- Divide and conquer where possible
- Respectful edits / comments / suggestions
- Feedback rounds
- Support each other
- Tackle hard parts together
- Celebrate



Self-Doubt & Impostor Syndrome

1. **Trust in the process**
    - Small steps: take apart the problem
    - Track progress
2. **Reflection**
    - Think about your past self
    - *“What’s the best that can happen?”*
    - Don’t make assumptions about others
3. **Be your own friend**

Inspiration: CVs of failure

Technicalities

Anonymization

Goal: reviewers should not be biased by knowing the author identity.

- No author names
- No URLs that reveal authors (e.g. personal GitHub, hosting of corpus) unless it is about previous work
- Talk about previous work as “they”
- Mask tool/company/group name if relevant (“internal”, “in-house”, XXX)
- Anonymization period

In practice there are watermarks, e.g. “2049 hidden units”, table formatting, baseline names, number of TPUs, preprint versions ...

The Archive: arXiv.org

- ‘Preprint’: not peer reviewed, revisions possible
- Endorsement for new authors needed (happy to help!)
- Status field: add conference
- Research fields: e.g. cs.CL, cs.AI, cs.LG
- Compilation: from LaTeX source, .bbl file instead of .bib
- Sanity Preserver: http://www.arxiv-sanity.com/

When citing papers from arXiv, check if they’ve been published at a peer-reviewed venue and use that version.

LaTeX

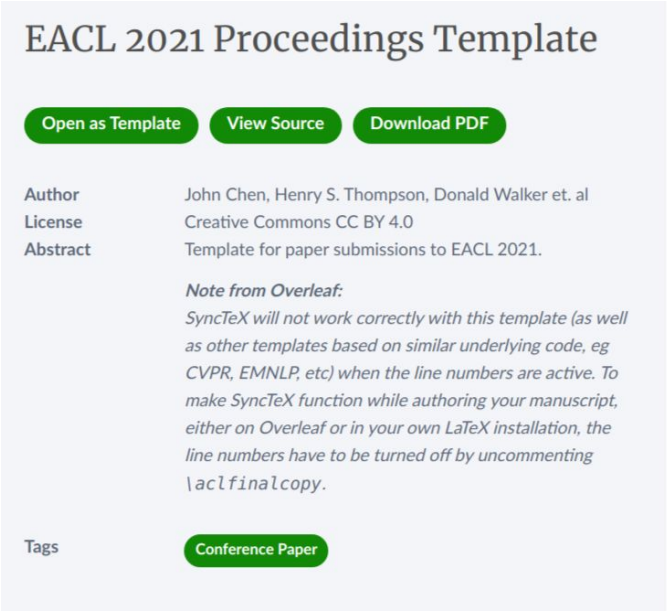
- Like a programming language → needs compilation
- Commands for formatting, e.g. \textbf{...} instead of Ctrl+B
- Lots of online help, documentation, templates
- Online and offline editors
- For theses, presentations, CVs, books, …
- Math typesetting

https://www.latex-project.org/about/

*“LaTeX is not a word processor! Instead, LaTeX encourages authors **not to worry too much about the appearance** of their documents but to concentrate on getting the right content.”*

Make sure you only use line breaks after blocks of text
LATEX
why?
LATEX
LaTeX Error: There's no line here to end.
3
LATEX

Overleaf Example

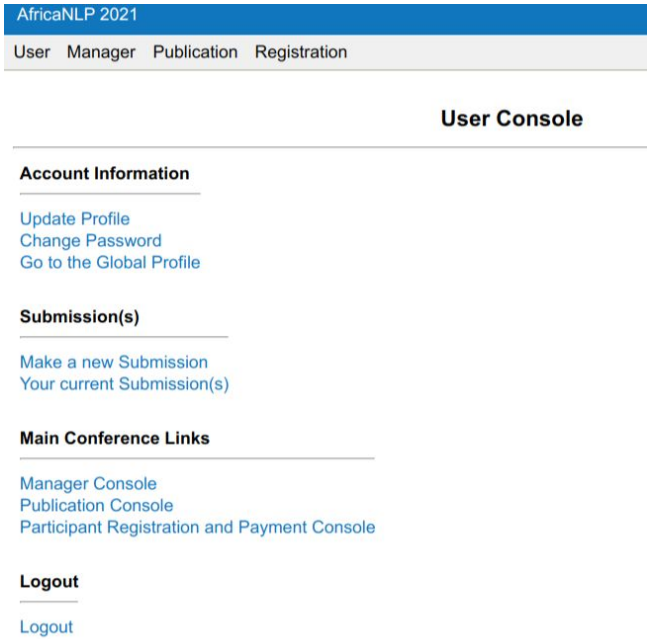


https://www.overleaf.com/latex/templates/eacl-2021-proceedings-template/jprrhhtnbrrm

SoftConf

https://www.softconf.com/eacl2021/AfricaNLP2021/

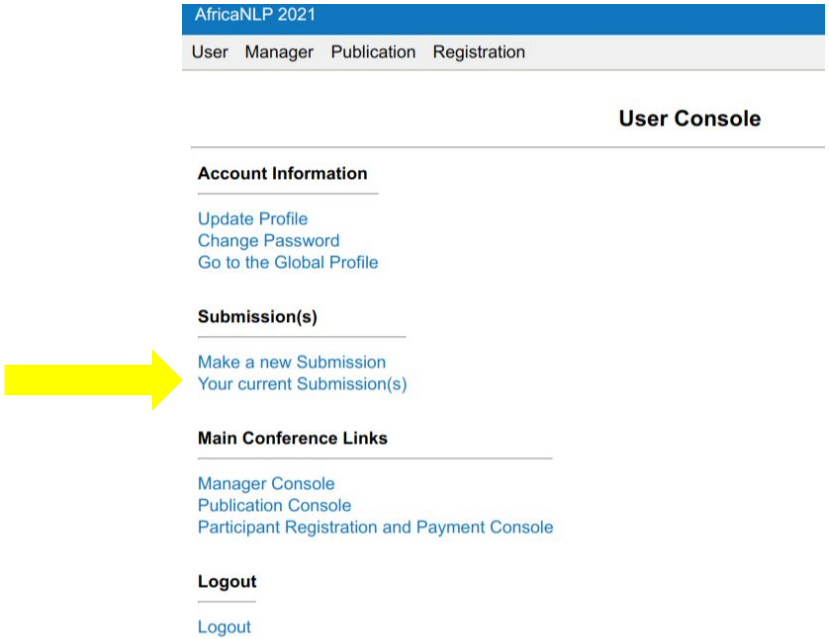
1. Register
2. Fill out “Global Author Profile”
3. Remember your username
4. Make a new submission



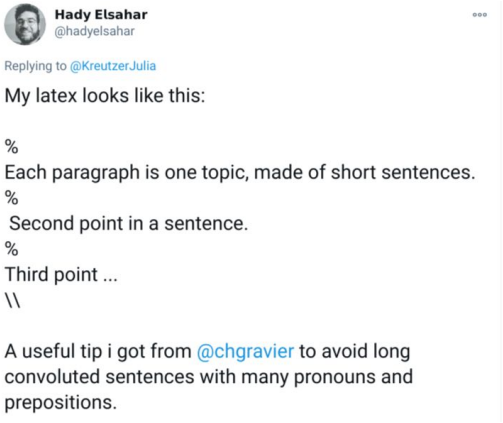
SoftConf

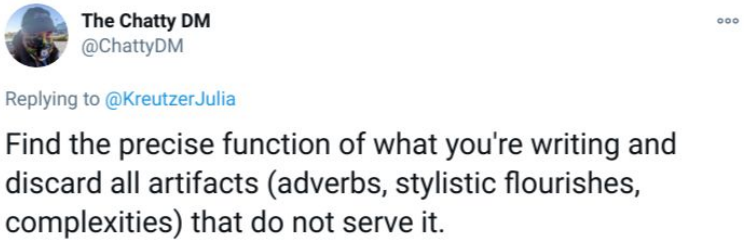
https://www.softconf.com/eacl2021/AfricaNLP2021/

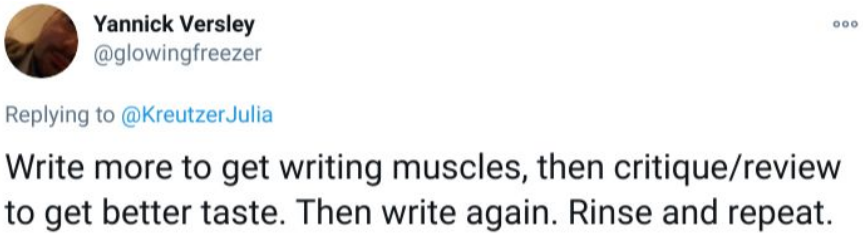
1. Register / Login
2. Fill out “Global Author Profile”
3. Remember your username
4. Make a new submission
    a. Title
    b. Abstract(s)
    c. List of authors
    d. Upload PDFs
        i. Main paper: page limit!
        ii. Appendix: optional, unlimited

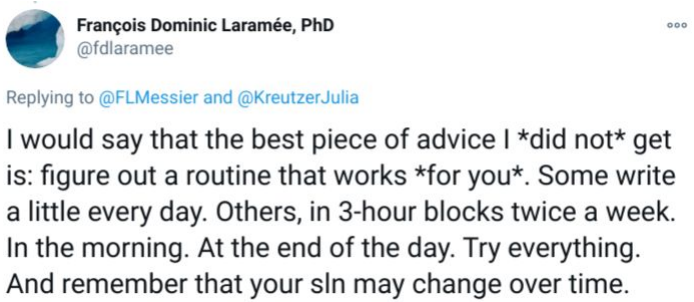


More Advice









Jason Eisner: [“Write the paper first.”]

Questions or Paper Pitches?

jkreutzer@google.com

Or on AfricaNLP Slack or Twitter (@KreutzerJulia)