

Wissenschaftliche Literatur lesen, verstehen und präsentieren

HS Sprachtechnologie für eine bessere Welt (Winter term 2022/23)

Nils Reiter,
`nils.reiter@uni-koeln.de`

October 18, 2022

Sitzungsthemen und -zuordnungen

Datum	Thema	Person(en)
22.11.	Authorship obfuscation / attribution	
29.11.	Fake News Detection	
6.12.	Hate Speech Detection	
13.12.	Mental Health	

Was würden Sie Erstsemester:innen zum Umgang
mit wissenschaftlicher Literatur raten?

Section 1

Overview

Computational Linguistics Literature

- ▶ Computational Linguistics (CL): A young field
 - ▶ Compared to philosophy, physics, ...
- ▶ Interdisciplinary between computer science and linguistics
 - ▶ Pendular movement
 - ▶ Currently: Strongly in the CS field

Computational Linguistics Literature

- ▶ Computational Linguistics (CL): A young field
 - ▶ Compared to philosophy, physics, ...
- ▶ Interdisciplinary between computer science and linguistics
 - ▶ Pendular movement
 - ▶ Currently: Strongly in the CS field

Core Requirements for Scientific Literature

- ▶ Quality assurance

Computational Linguistics Literature

- ▶ Computational Linguistics (CL): A young field
 - ▶ Compared to philosophy, physics, ...
- ▶ Interdisciplinary between computer science and linguistics
 - ▶ Pendular movement
 - ▶ Currently: Strongly in the CS field

Core Requirements for Scientific Literature

- ▶ Quality assurance: Reviewing

Computational Linguistics Literature

- ▶ Computational Linguistics (CL): A young field
 - ▶ Compared to philosophy, physics, ...
- ▶ Interdisciplinary between computer science and linguistics
 - ▶ Pendular movement
 - ▶ Currently: Strongly in the CS field

Core Requirements for Scientific Literature

- ▶ Quality assurance: Reviewing
- ▶ Sustainability and (in principle) accessibility
 - ▶ It should be possible to access a work in the distant future

Computational Linguistics Literature

- ▶ Computational Linguistics (CL): A young field
 - ▶ Compared to philosophy, physics, ...
- ▶ Interdisciplinary between computer science and linguistics
 - ▶ Pendular movement
 - ▶ Currently: Strongly in the CS field

Core Requirements for Scientific Literature

- ▶ Quality assurance: Reviewing
- ▶ Sustainability and (in principle) accessibility
 - ▶ It should be possible to access a work in the distant future
- ▶ Publishing houses ensure both (in theory)
- ▶ “Scientific publishing” \neq making something available to others

Peer Review

- ▶ Scientific articles are reviewed by other researchers/scientists
- ▶ Blindness
 - ▶ Double blind: Reviewer and authors are anonymous
 - ▶ Single blind: Only reviewers are anonymous
 - ▶ Zero blind / “Open Review”: No one is anonymous
- ▶ Different fields have different preferences
 - ▶ and different people have different preferences
 - ▶ CL: Double-blind (recently reaffirmed)
 - ▶ But: Preprint servers are an important venue in machine learning!

Publication Venues

- ▶ Monographs (books): Except for theses, typically not reviewed
- ▶ Journal articles: Peer reviewed (details are journal-dependent)
- ▶ Conference articles: Peer reviewed (details are conference-dependent)
 - ▶ “Proceedings” = Collection of all conference articles

Publication Venues


- ▶ Monographs (books): Except for theses, typically not reviewed
- ▶ Journal articles: Peer reviewed (details are journal-dependent)
- ▶ Conference articles: Peer reviewed (details are conference-dependent)
 - ▶ “Proceedings” = Collection of all conference articles

Lengths and “Abstracts”

- ▶ Length varies
 - ▶ Conference articles < 10 pages
 - ▶ Journal articles ca. 10 – 50 pages
- ▶ “Abstract”
 - ▶ Literal meaning: A summary of an article
 - ▶ Conference abstracts (DHd/DH) \simeq short articles

Relevant Publication Venues for CL

Conferences

- ▶ ACL / NAACL / EACL / EMNLP: Conferences (double-blind)
 - ▶ Association for Computational Linguistics
 - ▶  ACL 2022: 604 long papers – ACL 2002: 65 papers

aclanthology.org

Relevant Publication Venues for CL

Conferences

- ▶ ACL / NAACL / EACL / EMNLP: Conferences (double-blind)
 - ▶ Association for Computational Linguistics
 - ▶ ⚠ ACL 2022: 604 long papers – ACL 2002: 65 papers
 - ▶ Co-located workshops with more specific focus
 - ▶ “Workshop” in CL: Mini conference
 - ▶ Workshops associated with *CL conferences also in anthology
- ▶ COLING, KONVENS: Smaller conferences

aclanthology.org

Relevant Publication Venues for CL

Journals

- ▶ CL: Uncommon
- ▶ Computational Linguistics direct.mit.edu/coli
 - ▶ Also in anthology: <https://aclanthology.org/venues/cl/>
 - ▶ Fully open access
- ▶ Digital Scholarship in the Humanities (Literary and Linguistic Computing) academic.oup.com/dsh
 - ▶ Partially open access via UB
- ▶ Journal of Computational Literary Studies jcls.io

Relevant Publication Venues for CL

Preprint-Servers

- ▶ Origin: Share preprints freely
- ▶ No review: Everyone can upload anything
- ▶ Popular for machine learning advances
- ▶ Many papers are later/also submitted to a conference

arxiv.org

Relevant Publication Venues for CL

Others

- ▶ DFG (funding agency): No reviewing → no worth
- ▶ Blogs – it depends on their authors
- ▶ Sammelbände / collections

Structure of a CL Paper

Common structure

- ▶ Introduction
- ▶ Background
 - ▶ Optional. What do we have to know about the phenomenon?
- ▶ Related Work
 - ▶ Work dealing with same or similar problem
- ▶ Approach (the core)
 - ▶ Description on conceptual level
 - ▶ Good: Point out assumptions the approach makes
- ▶ Data set(s) / Corpus
 - ▶ Inter-Annotator agreement
- ▶ Experiments
 - ▶ Baseline(s)
 - ▶ Evaluation Metric(s)
- ▶ Results
- ▶ Error Analysis
 - ▶ Types of errors the system makes
- ▶ Conclusions
 - ▶ Summary
 - ▶ Findings about concept(s)
 - ▶ Future work

Section 2

Reading CL Literature

How to Read?

- ▶ Reading scientific literature is work
- ▶ A work environment is important
- ▶ Reading multiple times is often necessary

How to Read?

- ▶ Reading scientific literature is work
- ▶ A work environment is important
- ▶ Reading multiple times is often necessary

References

- ▶ Scientific references consist in:
 - ▶ Markers in the text (e. g., “Doe (2015)” oder “[3]”)
 - ▶ Bibliographic details at the end
- ▶ Different styles
 - ▶ CL: author-year
- ▶ URLs or DOIs
 - ▶ <https://www.example.com>
 - ▶ 10.1515/9783110693973 ⇒ <https://doi.org/10.1515/9783110693973>

Guiding Questions

You should be able to answer (at least) these questions

- ▶ What was the task/the problem to be solved?
- ▶ What is the new aspect compared to previous research?
- ▶ How well did it work?
 - ⚠ Authors have an interest to highlight success and neglect failure
- ▶ Which experiments were made to measure it?
 - ▶ Which data and evaluation metrics were used?

Critical Reflection of Literature

- ▶ Was there an easier way to achieve similar performance?
- ▶ How many assumptions are incorporated (maybe implicit)?
 - ▶ What would be needed to redo it from scratch?
 - ▶ What would be needed to adapt it to another language/genre/domain?
- ▶ Why did the authors did it the way they did?
- ▶ Can the experiments actually show what the authors claim they show?
- ▶ Are the experiments “correctly” interpreted? Are there alternative interpretations that are just as reasonable?
- ▶ Is there evidence to generalize results to “the language”, “the text type X”, ...?

Section 3

Giving (Scientific) Talks

Group Exercise

1. What are the three most important recommendations you would give to a new student on talks in seminars?
2. What should they avoid at all costs?
3. Do you have a secret, game-changing tip?

Outlining the Topic I

- ▶ What do you want (and need) to say?
- ▶ **Focus**
 - ▶ The talk should have a clear focus. What's the context of the talk? What's the topic of the course? A paper may contain parts that are not relevant in the given context **and vice versa**.
- ▶ **Understanding**
 - ▶ The talk should be understandable. Explain and introduce as much as needed, but not more. In university seminars, imagine the others as knowing as much as you **before** starting to read.
- ▶ **Structure**
 - ▶ Divide the talk in parts and subparts. The structure of the paper is not necessarily a good structure for the talk (but can be).
- ▶ Write your outline down, but think of it as a draft!
- ▶ Do **not** start making slides now.

Outlining of the Topic II

- ▶ See your talk from the audience's perspective
 - ▶ What can you expect them to know? What did you need to look up?

Outlining of the Topic II

- ▶ See your talk from the audience's perspective
 - ▶ What can you expect them to know? What did you need to look up?
- ▶ Get (honest) feedback
 - ▶ But: Your talk, your decision, your responsibility
- ▶ Make necessary changes
- ▶ Repeat the process

Making Slides

- ▶ Use a presentation tool for making slides:
LaTeX+Beamer, MS PowerPoint, Apple Keynote, OpenOffice Presenter, ...
- ▶ Use one of the built-in themes
 - ▶ Preferably a simple one
- ▶ Make the structure visible to the audience
 - ▶ Head/footlines, section break slides, etc.
- ▶ Avoid animations, effects etc.
- ▶ No screenshots of tables and figures
 - ▶ Recreate them in the presentation program (for readability)
- ▶ Scientific talks have references and a bibliography at the end (Reiter, 2021)
 - ▶ But only show it when someone asks

Making Slides

- ▶ Use a presentation tool for making slides:
LaTeX+Beamer, MS PowerPoint, Apple Keynote, OpenOffice Presenter, ...
- ▶ Use one of the built-in themes
 - ▶ Preferably a simple one
- ▶ Make the structure visible to the audience
 - ▶ Head/footlines, section break slides, etc.
- ▶ Avoid animations, effects etc.
- ▶ No screenshots of tables and figures
 - ▶ Recreate them in the presentation program (for readability)
- ▶ Scientific talks have references and a bibliography at the end (Reiter, 2021)
 - ▶ But only show it when someone asks
- ▶ No running text on slides: The audience **either** reads **or** listens.

Preparing the Actual Talk

= Rehearsing

- ▶ Go through the slides
- ▶ Speak loudly what you want to say
- ▶ Note the points where you stumbled or had problems finding words
- ▶ Change the slides accordingly
- ▶ Write down what you want to say – at least in keywords
- ▶ Maybe: Script the first few sentences
- ▶ Pay attention to the time

Discussion Preparation

- ▶ Do not put supporting information in the main presentation
 - ▶ E.g., charts, tables, long examples or detailed numbers *that you do not talk about*
- ▶ Add slides to your presentation that are useful for the discussion
- ▶ The slides should be in the same file, but at the end

Discussion Preparation

- ▶ Do not put supporting information in the main presentation
 - ▶ E.g., charts, tables, long examples or detailed numbers *that you do not talk about*
- ▶ Add slides to your presentation that are useful for the discussion
- ▶ The slides should be in the same file, but at the end
- ▶ Discussion
 - ▶ What are the weak points that could come up as a question?

Giving the Talk

- ▶ Stage fright
 - ▶ Inability to breathe
 - ▶ Inability to stand up
 - ▶ Inability to operate brain
- ▶ That's normal – and to be expected



Giving the Talk

- ▶ Stage fright
 - ▶ Inability to breathe
 - ▶ Inability to stand up
 - ▶ Inability to operate brain
- ▶ That's normal – and to be expected



What to do about it

- ▶ Be prepared for it
- ▶ Avoid waiting in front of the audience
- ▶ Imagine the feeling afterwards
- ▶ Script the beginning
- ▶ Try out what works for you

Be Seen and Heard

- ▶ Don't talk to the wall, window or computer
- ▶ Choose someone in the back (ideally, a nodder) to talk to
- ▶ Make breaks for questions
- ▶ Finish on time!







References I