



UNIVERSITÄT
ZU KÖLN

Wissenschaftlicher Umgang mit Literatur

HS In Context Learning (ICL) (Summer term 2024)

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

April 18, 2024

Was würden Sie Erstsemester:innen zum Umgang mit Literatur raten?
Was sollten die über wissenschaftliche Literatur wissen?

Section 1

Overview

Scientific Literature

Two core requirements

- 1 Quality assurance – reviewing
- 2 Long-term availability – archiving

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, LREC: Smaller conferences

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

aclanthology.org

▶ COLING, KONVENS, LREC: Smaller conferences

▶ Journals: Uncommon

▶ Computational Linguistics

- ▶ Also in anthology: <https://aclanthology.org/venues/cl/>
- ▶ Fully open access

direct.mit.edu/coli

Relevant Publication Venues for DH

- ▶ New field, not yet fully established
- ▶ Venues from original H discipline (e.g., Journal of Literary Theory)
- ▶ DH, DHd
- ▶ 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
- ▶ DFG (funding agency): No reviewing → no worth
- ▶ Blogs – it depends on their authors
- ▶ Sammelbände / collections

Preprint-Servers

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

arxiv.org

Non-Scientific Literature

- ▶ Categories
 - ▶ All media for the general public (including newspapers and special interest journals)
 - ▶ E.g., Die Zeit, Segeln, GEO, ...
 - ▶ Blogs, YouTube channels and social media postings
 - ▶ E.g., spreeblick.com
 - ▶ Companies, lobby groups
 - ▶ Government publications

Non-Scientific Sources

- ▶ Is it ok to use non-scientific sources?
 - ▶ It depends
- ▶ When is it ok?
 - ▶ When we are explicitly dealing with public opinion or reception
 - ▶ When we are looking into other disciplines
 - ▶ When a scientific discourse on a topic does not exist
 - ▶ When the topic in question is not a scientific question
- ▶ What are better sources than others?
 - ▶ Sources with references
 - ▶ Sources with scientific references
 - ▶ Sources with (scientific) references that are correctly reproduced

Original/Primary Sources

- ▶ Scientific papers are usually written by the people who came up with the content
- ▶ Non-scientific sources are often mediated
- ▶ Original sources are better sources
 - ▶ Mediations may misrepresent the content substantially
- ▶ Examples
 - ▶ Press releases by universities about studies their researchers conducted
 - ▶ News articles about laws or decrees
 - ▶ Summaries of interviews

Finding Literature

- ▶ Specialised repositories
 - ▶ Computational Linguistics aclanthology.org
 - ▶ Digital Humanities [DH Index](#)
 - ▶ Generic preprints arxiv.org
- ▶ References of other papers
- ▶ Your library [USB Köln](#)
 - ▶ Don't underestimate the ebook collection!
- ▶ Search engines [Google Scholar](#) [Semantic Scholar](#)
 - ⚠ Google finds a lot of non-scientific literature
- ▶ Wikipedia pages have often very good references

Section 2

Reading Literature Scientifically

How to Read?

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

How to Read?

- ▶ Reading 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/DH: author-year
- ▶ URLs or DOIs
 - ▶ <https://www.example.com>
 - ▶ 10.1515/9783110693973 ⇒ <https://doi.org/10.1515/9783110693973>

Guiding Questions for CL/technical DH Papers

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 Technical 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”, ...?

Reading Non-Scientific Literature

- ▶ Make a local copy, ideally with all the meta data
 - ▶ Noone guarantees that it is still there tomorrow
- ▶ Who wrote it and why?
- ▶ Do they know what they are writing about?
- ▶ What's the track record of the author/venue?
- ▶ Are author/venue involved in any way?

Section 3

Giving (Scientific) Talks

Group Exercise

You're giving advice to younger students

1. What are the three most important recommendations on giving talks?
2. What should they avoid at all costs?
3. Do you have a game-changing tip?

Three Recommendations

To Avoid

A light gray grid background covers the entire page, intended for a table with two columns: 'To Avoid' and 'To Do'.

To Do

Outlining the Topic

- ▶ 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.

Outlining the Topic

- ▶ 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.

Outlining the Topic

- ▶ 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 a paper is not necessarily a good structure for the talk (but can be).
- ▶ Do **not** start making slides now.

Outlining of the Topic

Changing perspectives

- ▶ Write your outline down, but think of it as a draft!
- ▶ 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

Once you know what to say ...

- ▶ Use a presentation tool: LaTeX+Beamer, PowerPoint, Keynote, OpenOffice, ...
- ▶ Use a simple built-in theme
- ▶ 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
 - ▶ But only show it when someone asks

(Reiter, 2021)

Making Slides

Once you know what to say ...

- ▶ Use a presentation tool: LaTeX+Beamer, PowerPoint, Keynote, OpenOffice, ...
- ▶ Use a simple built-in theme
- ▶ 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
- ▶ But only show it when someone asks
- ▶ No running text on slides: The audience **either** reads **or** listens.

(Reiter, 2021)

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