

# Chapter G:I

## I. Scientific Toolbox

- ☐ Literature Research
- ☐ Oral Presentations
- ☐ Scientific Writing

# Scientific Writing

## Content of a Paper

- ❑ Most of the above hints on talks still hold

- Science is storytelling

Seminar: No scientific break-through expected,  
rather summarize and discuss.

- Science needs to be understood

- ❑ Papers are more complete

- Tell the whole story, avoid gaps in argumentation

- But: Include only relevant content

Don't expect too much prior knowledge.

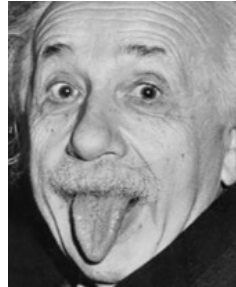
- But: No details on knowledge that can be presupposed

- ❑ Papers should be sound

- Need to be precise more than in talks

- Use logical arguments, from broad context to deep details

- Formalize concepts if needed / helpful



“Don’t make  
me think.”



Steve Krug

# Scientific Writing

## Structure of a Paper

### □ High-level structure

- Title, author information, abstract
- Introduction
- Usually 2–5 sections
- Conclusion
- References

Related work, approach, experiments, etc.

... and sometimes appendices

### □ Section structure

- Often numbered subsections (2.1, 2.2, ...)
- If any, subsubsections unnumbered
- Always have text introducing (sub)sections

### □ Section headings

- Stick to the standard: “Introduction” is first, “Conclusion” is last, etc.
- Short misleading headings worse than longer clear ones

#### The Impact of Modeling Overall Argumentation with Tree Kernels

Henning Wachsmuth  
Bauhaus-Universität Weimar  
Faculty of Media, Webis Group  
henning.wachsmuth@uni-weimar.de

Giovanni Da San Martino  
Qatar Computing Research Institute  
Arabic Language Technologies  
gmartino@hbku.edu.qa

Dora Kiesel  
Bauhaus-Universität Weimar  
Faculty of Media, VR Group  
dora.kiesel@uni-weimar.de

Benno Stein  
Bauhaus-Universität Weimar  
Faculty of Media, Webis Group  
benno.stein@uni-weimar.de

#### Abstract

Several approaches have been proposed to model either the explicit sequential structure of an argumentative text or its implicit hierarchical structure. So far, the adequacy of these models of overall argumentation remains unclear. This paper asks what type of structure is actually important to tackle downstream tasks in computational argumentation. We analyze patterns in the overall argumentation of texts from three corpora. Then, we adapt the idea of positional tree kernels in order to capture sequential and hierarchical argumentative structure together for the first time. In systematic experiments for three text classification tasks, we find strong evidence for the impact of both types of structure. Our results suggest that either of them is necessary while their combination may be beneficial.

#### (a) monological argumentative text

...The death penalty is a legal means that as such is not practicable in Germany...if for one thing, inevitable human dignity is anchored in our constitution...and furthermore no one may have the right to adjudge upon the death of another human being...Even if many people think that a murderer has already decided on the life or death of another person...this is precisely the crime that we should not repeat with the same.

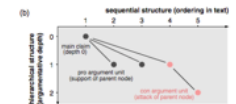


Figure 1: (a) Example text with five argument units, taken from the *Arg-Microtexts* corpus introduced in Section 3. (b) Graph visualization of the sequential and hierarchical overall argumentation of the text.

Figure 1(b) illustrates the interplay of the two types of overall structure in form of a tree-like graph.

#### 1 Introduction

Argumentation theory has established a number of major argument models focusing on different aspects, such as the roles of an argument's units (Toulmin, 1958), the inference scheme of an argument (Walton et al., 2008), or the support and attack relations between arguments (Freeman, 2011). The common ground of these models is that they conceptualize an argument as a conclusion (in terms of a claim) inferred from a set of pro and con premises (reasons), which in turn may be the conclusions of other arguments. For the overall argumentation of a monological argumentative text such as the one in Figure 1(a), this results in an implicit hierarchical structure with the text's main claim at the lowest depth. In addition, the text has an explicit linguistic structure that can be seen as a regulated sequence of speech acts (van Eemeren and Grootendorst, 2004).

Natural language processing research has largely adopted the outlined hierarchical models for mining arguments from text (Shay and Gurevych, 2014; Habernal and Gurevych, 2015; Peldszus and Stede, 2016). However, the adequacy of the resulting overall structure for downstream analysis tasks of computational argumentation has rarely been evaluated (see Section 2 for details). In fact, a computational approach that can capture patterns in hierarchical overall argumentation is missing so far. Even more, our previous work indicates that a sequential model of overall structure is preferable for analysis tasks such as stance classification or quality assessment (Wachsmuth and Stein, 2017).

In this paper, we ask and investigate what model of (monological) overall argumentation is important to tackle argumentation-related analysis tasks. To this end, we consider three corpora with fully

# Scientific Writing

## Abstract

- ❑ A concise high-level summary of the paper
- ❑ Usually 5–10 sentences
- ❑ One “approach”
  - Motivation and context (1 sentence)
  - Problem and why not solved (1–2 sentences)
  - Question addressed in the paper (1 sentence)
  - Approach with some details (2–3 sentences)
  - Evaluation, results, conclusion (1–3 sentences)
- ❑ Or in other words
  - What is the problem? Why is it a problem?
  - What is the solution? Why is it a solution to the problem?

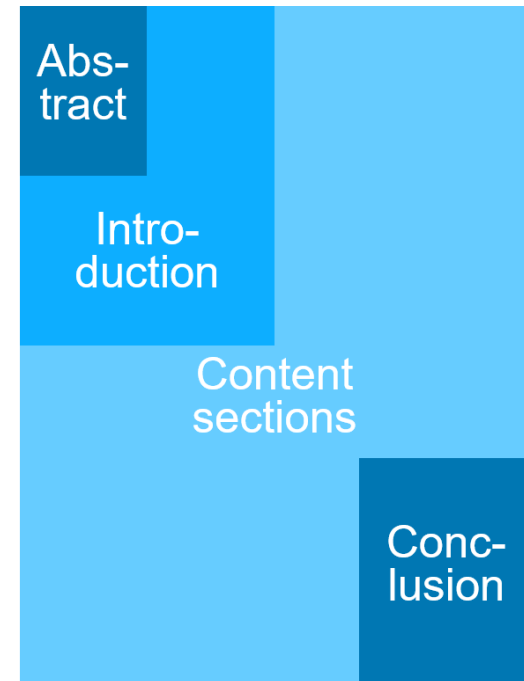
### Abstract

Several approaches have been proposed to model either the explicit sequential structure of an argumentative text or its implicit hierarchical structure. So far, the adequacy of these models of overall argumentation remains unclear. This paper asks what type of structure is actually important to tackle downstream tasks in computational argumentation. We analyze patterns in the overall argumentation of texts from three corpora. Then, we adapt the idea of positional tree kernels in order to capture sequential and hierarchical argumentative structure together for the first time. In systematic experiments for three text classification tasks, we find strong evidence for the impact of both types of structure. Our results suggest that either of them is necessary while their combination may be beneficial.

# Scientific Writing

## Sections

- ❑ Introduction
  - The abstract in more detail
  - Tell the whole story, from context to conclusion
  - Still high-level
  - Understandable for computer scientists
- ❑ Content sections
  - The introduction in more detail
  - Elaborate on related work, concepts, models, data, approaches, experiments, and results
  - More technical, for researchers from the area
- ❑ Conclusion
  - The introduction in less detail
  - Summarize story in retrospective, give outlook
  - Semi-technical



# Scientific Writing

## Style

- ❑ Write clearly, unambiguously, and concise
- ❑ Don't make things complex  
(common misunderstanding)
- ❑ Some guidelines:
  - Use impersonal or “we” form
  - Avoid pronouns with unclear references
  - Use explicit discourse markers, such as “because”
  - Blurring is non-scientific, such as “It could be . . .”
  - English sentences are short, one statement per sentence
  - Again: Avoid grammar and spelling errors
- ❑ Recommended reports from experienced researchers:
  - Justin Zobel: Writing for Computer Science
  - David Maxwell: [Writing up a PhD thesis](#)
  - George D. Gopen and Judith A. Swan: [The Science of Scientific Writing](#)



# Scientific Writing

## Style (continued)

- ❑ Hints from wordvice.com:
  - avoid nominalizations
  - eliminate prepositions
  - avoid fillers

# Scientific Writing

## Tables, Figures, Terms, and Footnotes

### □ Tables and figures

- In papers, just number increasingly

Figure 1, 2, ... Table 1, 2, ... (NOT: Figure 2.1, 2.2, ...)

- Tables: Horizontal lines suffice
- No included font larger than article font
- Explain in text and in caption

### □ Technical terms

- Introduce where needed, don't overformalize
- Use well-defined terms, AIA & AUA

Always introduce acronyms & avoid unnecessary acronyms.

- Don't use synonyms for terms

Reader is misled to check whether intentional differences exist.

### □ Footnotes

- Only for secondary information
- Reduced readability, should be an exception
- Don't cite literature using footnotes

	AAE-v2	Arg-Microtexts	Web Discourse
Argument units	6089	576	1149
Avg. units/text	15.1	5.1	3.4
Min. units/text	7	3	0
Max. units/text	28	10	16
Arguments	5687	443	560
Avg. depth	2.8	2.0	0.6
Min. depth	2	1	0
Max. depth	5	4	1
Texts	402	112	340

Table  
ments

#### (a) monological argumentative text

[1] The death penalty is a legal means that as such is not practicable in Germany. [2] For one thing, inviolable human dignity is anchored in our constitution, [3] and furthermore no one may have the right to adjudicate upon the death of another human being. [4] Even if many people think that a murderer has already decided on the life or death of another person, [5] this is precisely the crime that we should not repay with the same.

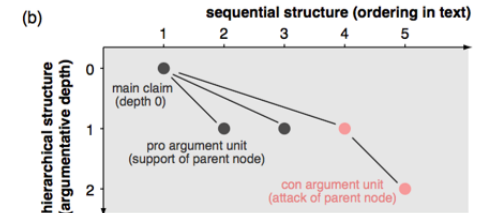


Figure 1: (a) Example text with five argument units, taken from the *Arg-Microtexts* corpus introduced in Section 3. (b) Graph visualization of the sequential and hierarchical overall argumentation of the text.

alternatives by modeling the stance of each unit towards its parent in the associated tree. This stance can be derived in all corpora.<sup>3</sup> All other unit and relation types from the specific models are ignored, since there is no clear mapping between them.

<sup>3</sup>Alternatively, the stance towards the main claim could be modeled. We decided against this alternative to avoid possibly wrong reinterpretations, e.g., it is unclear whether a unit that attacks its parent always supports a unit attacked by the parent.



# Scientific Writing

## Citations

### □ Citation

- In-text reference to a bibliographic source
- Different styles

Acronyms [ACW17], ACL style (Ajjour et al., 2017), numbers [42], . . .

mentation, namely, to classify the myside bias and stance of texts. For myside bias, [Stab and Gurevych \(2016\)](#) use features derived from discourse structure, whereas [Faulkner \(2014\)](#) and [Sobhani et al. \(2015\)](#) model arguments to classify stance. [Ong et al. \(2014\)](#) and we ourselves ([Wachsmuth et al., 2016](#)) do similar to assess the quality of persuasive essays and [Reisman, Klebanov et al. \(2016\)](#)

### □ What to cite

- Any reuse, paraphrase, summary, or translation of content
- Rule of thumb: Always clarify what is from you and what from others

Also have to cite yourself if you use your own sources.

- Better one citation too much than too few

### □ How to cite

- Direct reuse. Put in quotes (shorten with [. . .]), give source

Unit segmentation is “[. . .] the splitting of a text into argumentative segments” [ACW17].

- Other citations. Give source close-by

As Ajjour et al. point out, segmentation is the first task of an argument mining pipeline [ACW17].

- Large text portions. Give source once in the beginning

In the following paragraph, we summarize the segmentation approach of Ajjour et al. [ACW17].

# Scientific Writing

## References

- ❑ Bibliographical information at the end of the paper
  - ❑ Exactly those references cited in the text
  - ❑ Information should be complete and homogenous
  
  - ❑ Needed meta-information
    - All literature. Author, year, title
    - Conferences/Workshops. Proceedings, pages
    - Journals. Journal name, issue, number, pages
    - Books. Edition if any, publisher
    - Only online. Give URL with access date
  
  - ❑ Bibtex
    - LaTeX handles references automatically using bibtex
- See part on organizing literature above.

Aristotle. 2007. *On Rhetoric: A Theory of Civic Discourse* (George A. Kennedy, translator). Clarendon Aristotle series. Oxford University Press.

Beata Beigman Klebanov, Christian Stab, Jill Burstein, Yi Song, Binod Gyawali, and Iryna Gurevych. 2016. *Argumentation: Content, structure, and relationship with essay quality*. In *Proceedings of the Third Workshop on Argument Mining (ArgMining2016)*, pages 70–75. Association for Computational Linguistics.

Stefanie Brüninghaus and Kevin D. Ashley. 2003. *Predicting outcomes of case based legal arguments*. In *Proceedings of the 9th International Conference on Artificial Intelligence and Law*, pages 233–242.

Chih-Chung Chang and Chih-Jen Lin. 2011. *LIBSVM: A library for support vector machines*. *ACM Transactions on Intelligent Systems and Technology*, 2(3):27:1–27:27.

# Scientific Writing

## Plagiarism

- ❑ To sell another's ideas or expressions as one's own

See [en.wikipedia.org/wiki/Plagiarism](https://en.wikipedia.org/wiki/Plagiarism)

- ❑ On purpose or due to lack of giving sources

- ❑ Plagiarism is not(!) a trivial offense

In some countries considered as crime.

- ❑ Proper citing avoids all plagiarism issues

- ❑ Consequences

- Major cases lead to the denial of being published, graded, or worse
- Minor cases can still negatively affect a grade or review outcomes

- ❑ Webis Group [[www.webis.de](http://www.webis.de)]

- We do research on text reuse detection
- See publications, shared tasks, or the tool picapica [[www.picapica.org](http://www.picapica.org)]

