# Annotation in DH (annDH)

August 6-10, 2018, Sofia, Bulgaria

Welcome and Introduction

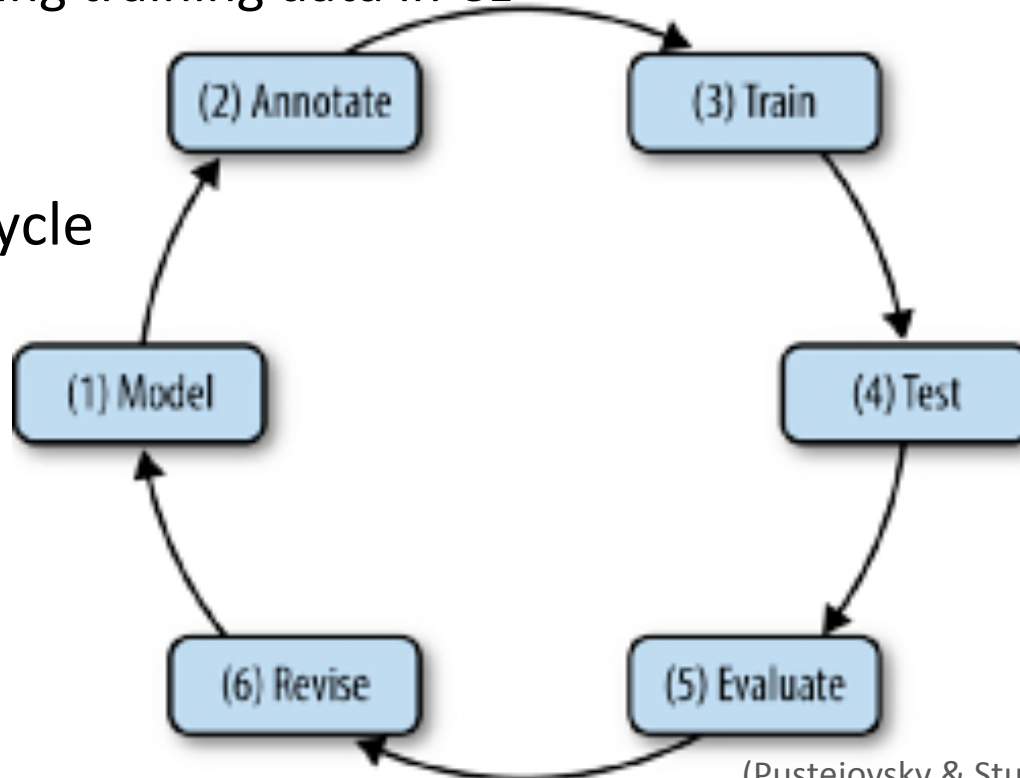Sandra Kübler & Heike Zinsmeister

Workshop at

# Motivation: Digital Humanities

- Exploring digital data and computational means in humanities' disciplines
  - Data: unstructured text
  - Annotation: going beyond the surface word forms
- Goals:
  - Discovering structures / tendencies / relations (by distant reading methods)
  - Identifying relevant data for further analysis
  - Creating training data for machine learning to explore 'big data'
- How can linguistics/computational linguistics (CL) help with annotation in DH?

# Motivation: Annotation in CL

- Long tradition
  - Of using annotated corpora in linguistics
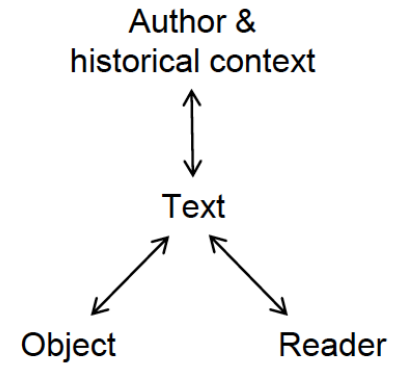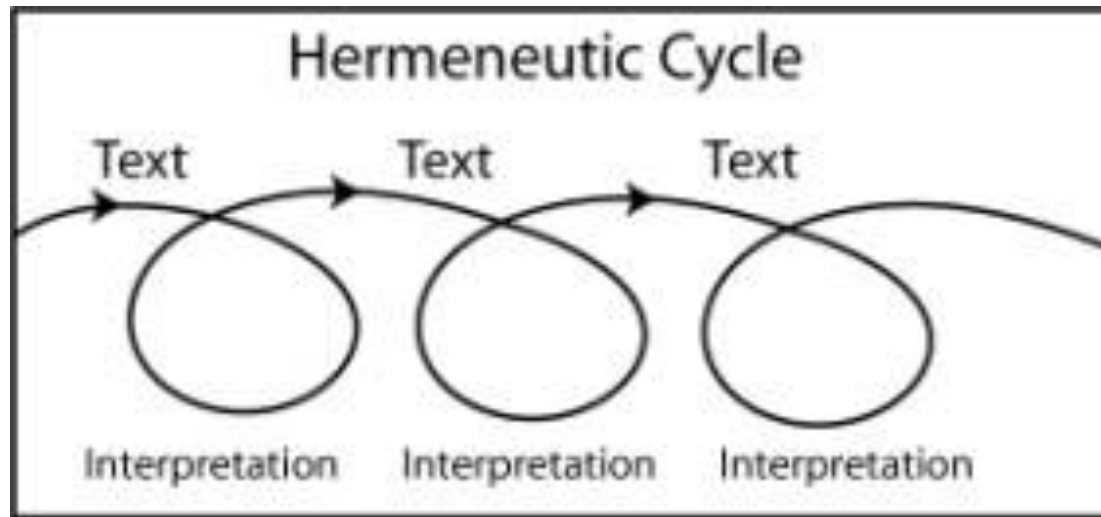  - Of creating training data in CL

Development cycle
MATTER



(Pustejovsky & Stubbs 2013: 23 ff.)

# In the Humanities

Author &
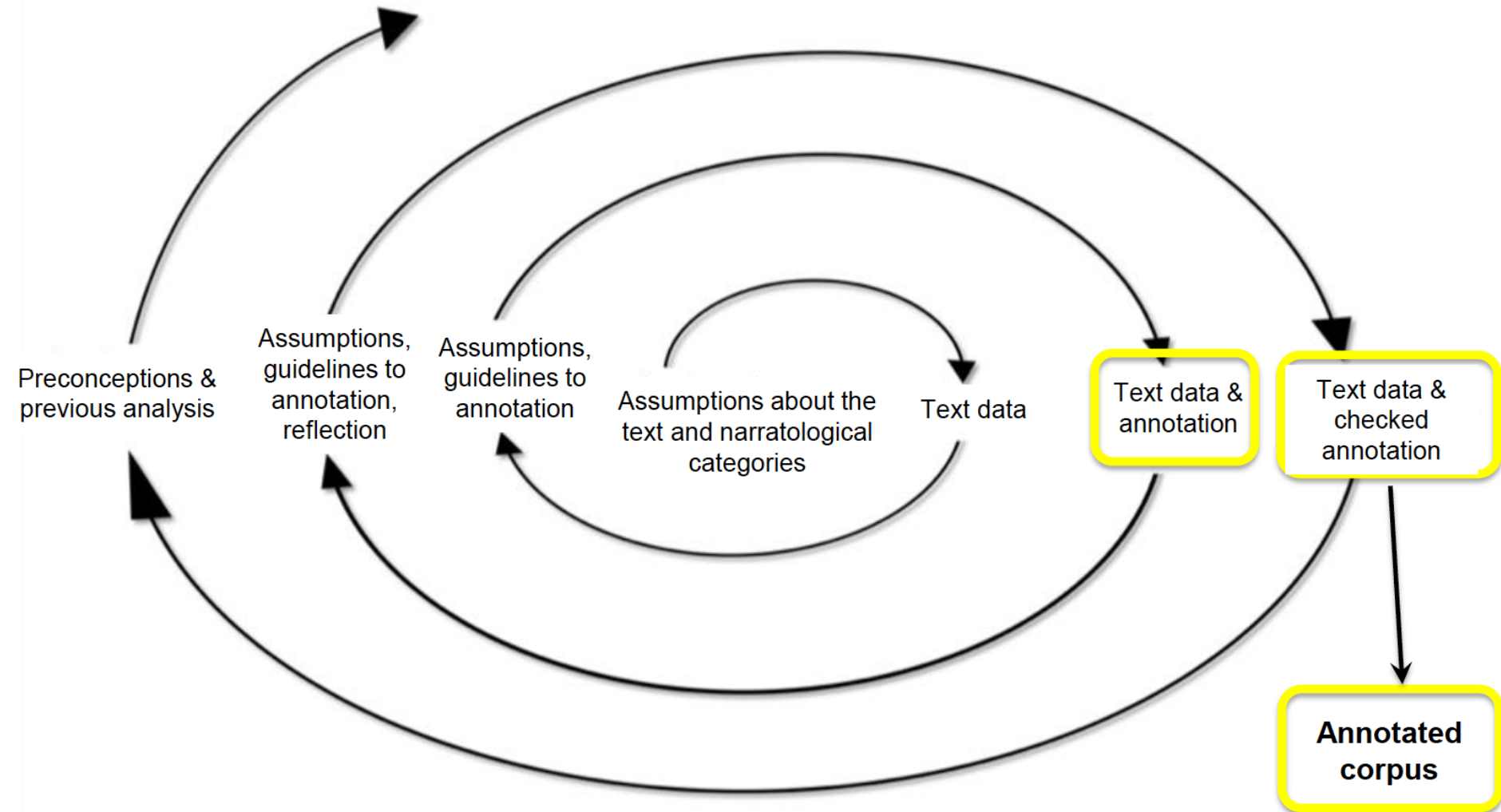historical context
↕
Text
↙ ↘
Object     Reader

## Exploring text meaning in the hermeneutic tradition

**Hermeneutic Cycle**

Text → Text → Text

Interpretation   Interpretation   Interpretation

(following Gadamer 1979, Image from Murphy 2003)

# An Extended Hermeneutic Cycle



Preconceptions & previous analysis

Assumptions, guidelines to annotation, reflection

Assumptions, guidelines to annotation

Assumptions about the text and narratological categories

Text data

Text data & annotation

Text data & checked annotation

**Annotated corpus**

Exemplary display of a hermeneutic circle extended by annotation (yellow boxes) for a narratological analysis (adapted from Bögel et al. 2015)
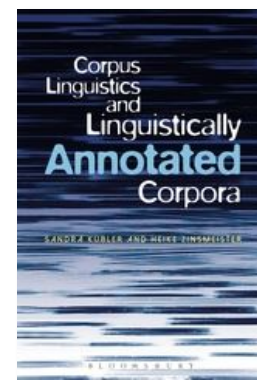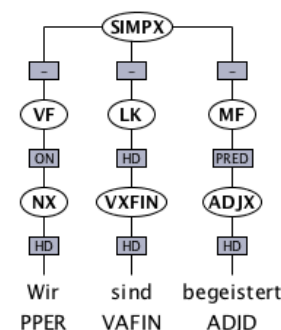
# Annotation and Familiar Concepts

# annDH People: Workshop chairs

- Sandra
  - computational linguistics and corpus linguistics
- Heike
  - (computational) linguistics and philosophy (some literary studies)
- Together
  - worked on German TüBa-D/Z treebank (~2005)
  - Book: Corpus Linguistics and Linguistically annotated corpora
  - Teach4DH workshop Berlin, 2017 (co-chairs Thierry Declerck and Peggy Bockwinkel)

# annDH People: Program Committee

| | | | |
|---|---|---|---|
| Melanie Andresen | University of Hamburg | Liliana Melgar-Estrada | University of Amsterdam |
| Fabian Barteld | University of Hamburg | Marco Passarotti | Catholic University of the Sacred Heart |
| Sabine Bartsch | University of Darmstadt | | |
| Peggy Bockwinkel | University of Stuttgart | Georg Rehm | DFKI Berlin |
| Peter Boot | Huygens Institute for Netherlands History | Nils Reiter | University of Stuttgart |
| | | Thomas Schmidt | IDS, Mannheim |
| Fritz Breithaupt | Indiana University | Ulrike Schneider | University of Mainz |
| Simon Clematide | Zurich University | Olga Scrivner | Indiana University |
| Thierry Declerck | DFKI Saarbrücken | Caroline Sporleder | University of Göttingen |
| Stefanie Dipper | Ruhr-Universität Bochum | Kenneth Steimel | Indiana University |
| Kim Gerdes | Sorbonne nouvelle Paris | Thorsten Trippel | University of Tübingen |
| Evelyn Gius | University of Hamburg | Mihaela Vela | University of the Saarland |
| Fotis Jannidis | University of Würzburg | Gabriel Viehhauser | University of Stuttgart |
| Hannah Kermes | University of the Saarland | Andreas Witt | University of Cologne |
| Lothar Lemnitzer | BBAW, Berlin | Amir Zeldes | Georgetown University |
| Harald Lüngen | IDS, Mannheim | Kalliopi Zervanou | Utrecht University |

# Special Issue



- Proposal for

- https://academic.oup.com/dsh
  - previously known as Literary and Linguistic Computing
  - drawback: Embargo before ‚Author Version' can be made publicly available on own webpage: 24 months
- Other suggestions?

# Proceedings

- CEUR Workshop Proceedings

http://ceur-ws.org/Vol-2155/

# Schedule

https://anndh18.github.io/program.html

**Friday 10th** August: **5.15 pm  - 6.45 pm**

- Beth disseration prize during student session
- FOLLI General Meeting

# Discussion:
## What Is DH? Interaction between CL and DH?

1. What is your disciplinary **background**?

2. What is the **aim** of your annotation effort?

3. What types of **annotation categories** do you use?

4. What kind of **data** do you work with?

How can linguistics/computational linguistics help with the annotation?

# Discussion:

What Is DH? Interaction between CL and DH?

1. Where do you see connections between CL and DH?

2. What types of support from CL would be useful?

   Annotation tools, semi-automatic annotation, visualization tools, support for consistency checking?

# References

- Bögel, T., Gertz, M., Gius, E., Jacke, J., Meister, J.C., Petris, M. d& Strötgen, J. 2015. Gleiche Textdaten, unterschiedliche Erkenntnisziele? Zum Potential vermeintlich widersprüchlicher Zugänge zu Textanalyse, Extended Abstract, Digital Humanities 2015 " Von Daten zu Erkenntnissen: Digitale Geisteswissenschaften als Mittler zwischen Information und Interpretation", Graz, Österreich.
    - See also: Gius, E. & J. Jacke. 2017. The Hermeneutic Profit of Annotation. On Preventing and Fostering Disagreement in Literary Analysis. International Journal of Humanities and Arts Computing 11: 2, Special Issue: Explanatory Annotation in the Context of the Digital Humanities (2017), 233–254.

- Gadamer, Hans-Georg. 1979: Truth and Method (W. Glen-Doepel, Trans. 2nd edition ed.). London: Sheed and Ward.

- Murphy, David C. 2003. "Mediated Pedagogical Design: The Cycles of Iteration Interface."  MA thesis. Simon Fraser University 1995. url: http://www.sfu.ca/media-lab/cycle/extras/cycle_design.htm.

- Pustejovsky, James & Amber Stubbs. 2013. Natural Language Annotation for Machine Learning. Sebastopol, CA: O'Reilly.