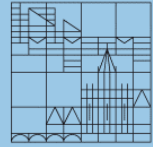


Exploiting the 'Annotation Cycle' for Teaching Linguistics

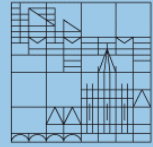
Heike Zinsmeister

Workshop on Corpora in Teaching
Languages and Linguistics, HU Berlin

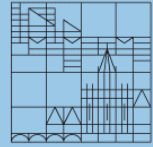
January 7th, 2011



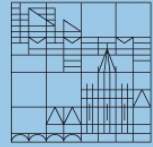
- Linguistics – the study of language form and meaning
- Teaching linguistics
 - making the students aware of linguistic forms and meanings (entities, generalization, ...)
→ objective: concepts
 - enabling the students to name, generalize, and explain the form and/or the meaning of particular language samples
→ objective: terminology, analyses and methods
 - making the students acquainted with linguistic research questions
→ objective: motivation for the undertaking



- Status quo (in many German linguistic programmes)
 - classroom lecture with homework exercises (that are discussed in tutorials held by student assistants)
- Linguistics is perceived as ‘dull’
 - partly familiar from school
 - ‘terminology catalogue’
 - analyses capture only ‘core grammar’ not every-day language
- Suggestion
 - let students be actively involved in class
 - apply analyses to every-day language
 - lecturer as moderator / expert consultant

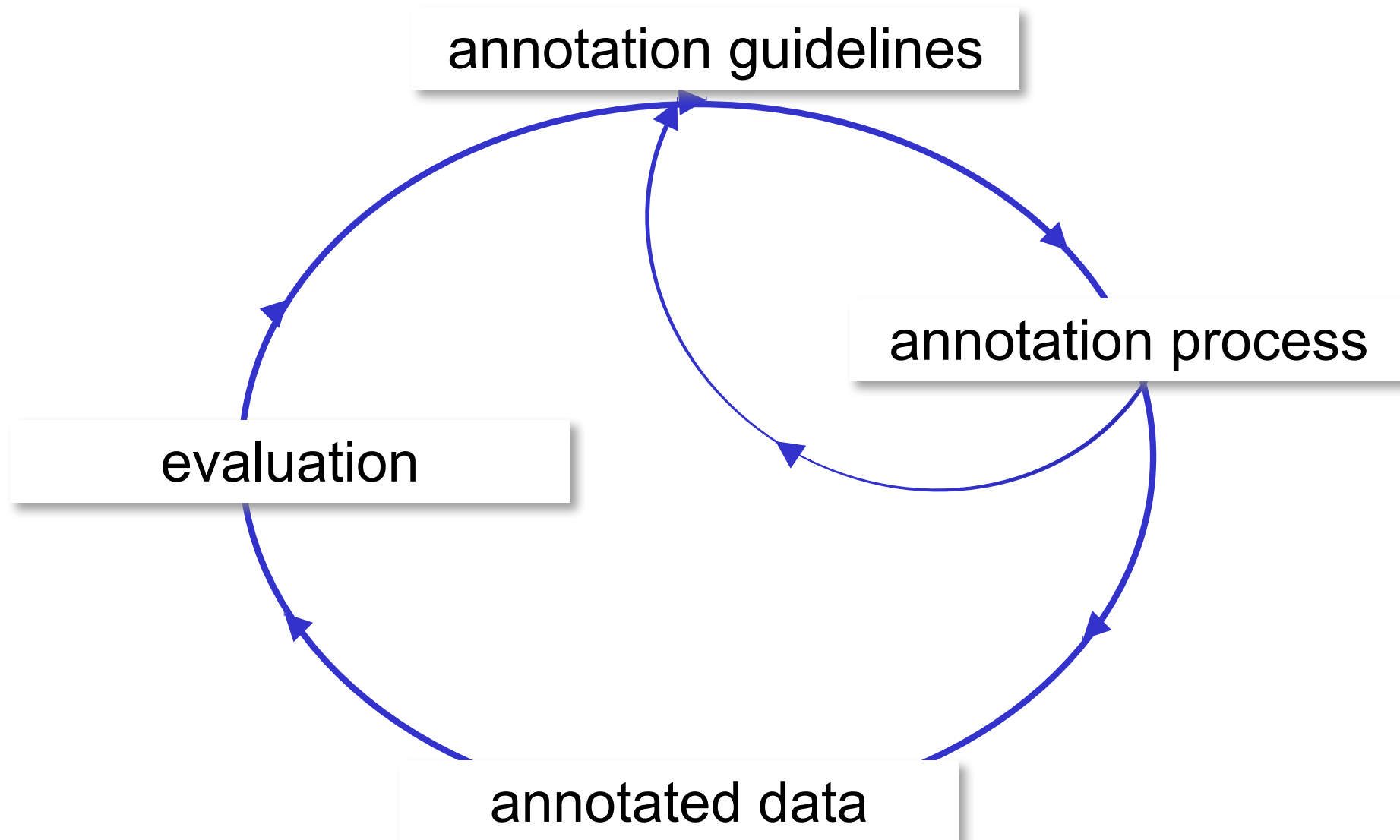
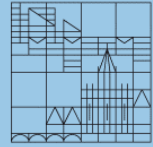


- General teaching premises
 - competences are acquired by ‘doing’
 - fulfilment of tasks is easier if sufficient resources are provided
 - the students’ performance has to be evaluated
- Consequences for the teacher
 - statement of explicit problems / work assignments
 - providing of examples, methods, expert knowledge, trainings
 - grading criteria are to be set in a comprehensible way

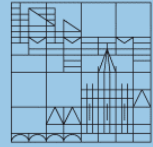


1. Motivation
- 2. The annotation cycle**
3. Annotation as a didactic means
4. Discussion of variables in the cycle
5. Sketch of two settings
6. Summary & discussion

The annotation cycle

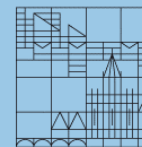


Example case



- Meaning of the third person pronoun 'it'
- Training data for a classifier
 - preprocessor for coreference resolution
- Coreference resolution is relevant for
 - information extraction
 - summarization
 - machine translation
 - etc.

Examples for English 'it'



Nominal anaphor

(1) The baby is lying in the cradle. **It** is sleeping calmly.

Abstract anaphor

(2) Gasoline prizes are rising again and I do not like **it**.

Placeholder for dislocated clausal argument

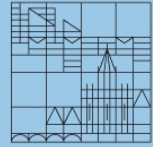
(3) She also made **it** clear that Conductive Education is not the only method.

Weather verbs / predicates involving time, place etc.

(4) **It** was raining / **It** was about midnight.

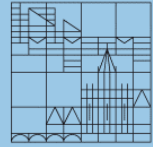
Cleft construction

(5) **It** was the military district commander who stepped in to avoid bloodshed.



- Definitions
 - classifications
 - labels
 - markables
- Linguistic tests
 - operationalized criteria for labelling a markable with a specific label
- Examples
 - uncontroversial cases
 - problematic cases
 - discussions
 - rules of thumb

Decision tree of linguistic tests



Does 'it' refer to some referent that is mentioned in the text elsewhere? test: *namely* paraphrase

yes= **referential pronoun**

Is the antecedent of the pronoun realised as an NP?
test: NP criteria

yes

nominal anaphor

no

abstract anaphor

no = **expletive**

Does it function as a placeholder for a dislocated clausal argument (including non-finite phrases)?
test: replacement; clausal criteria

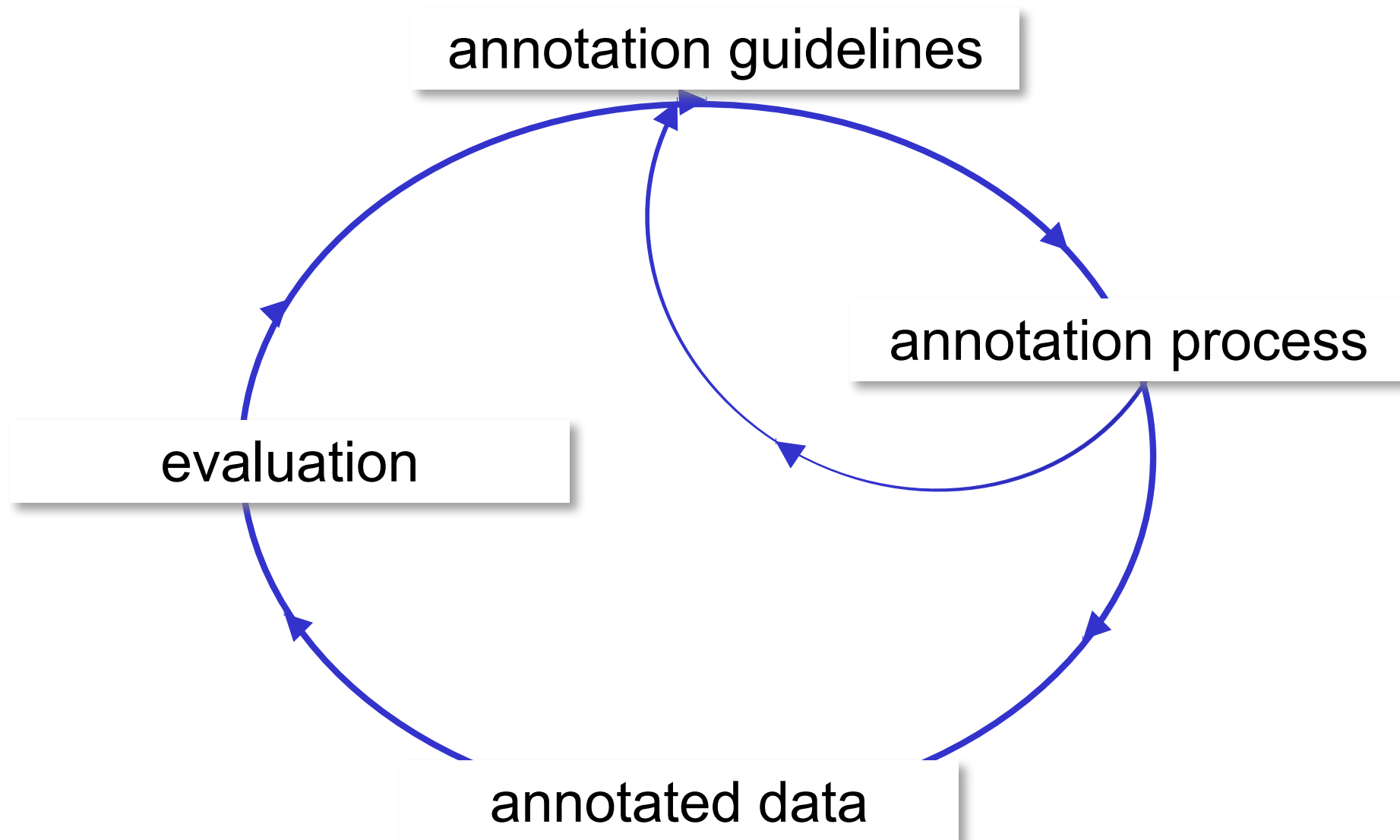
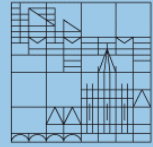
yes

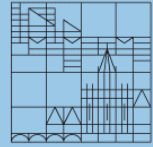
clausal placeholder

no

rest

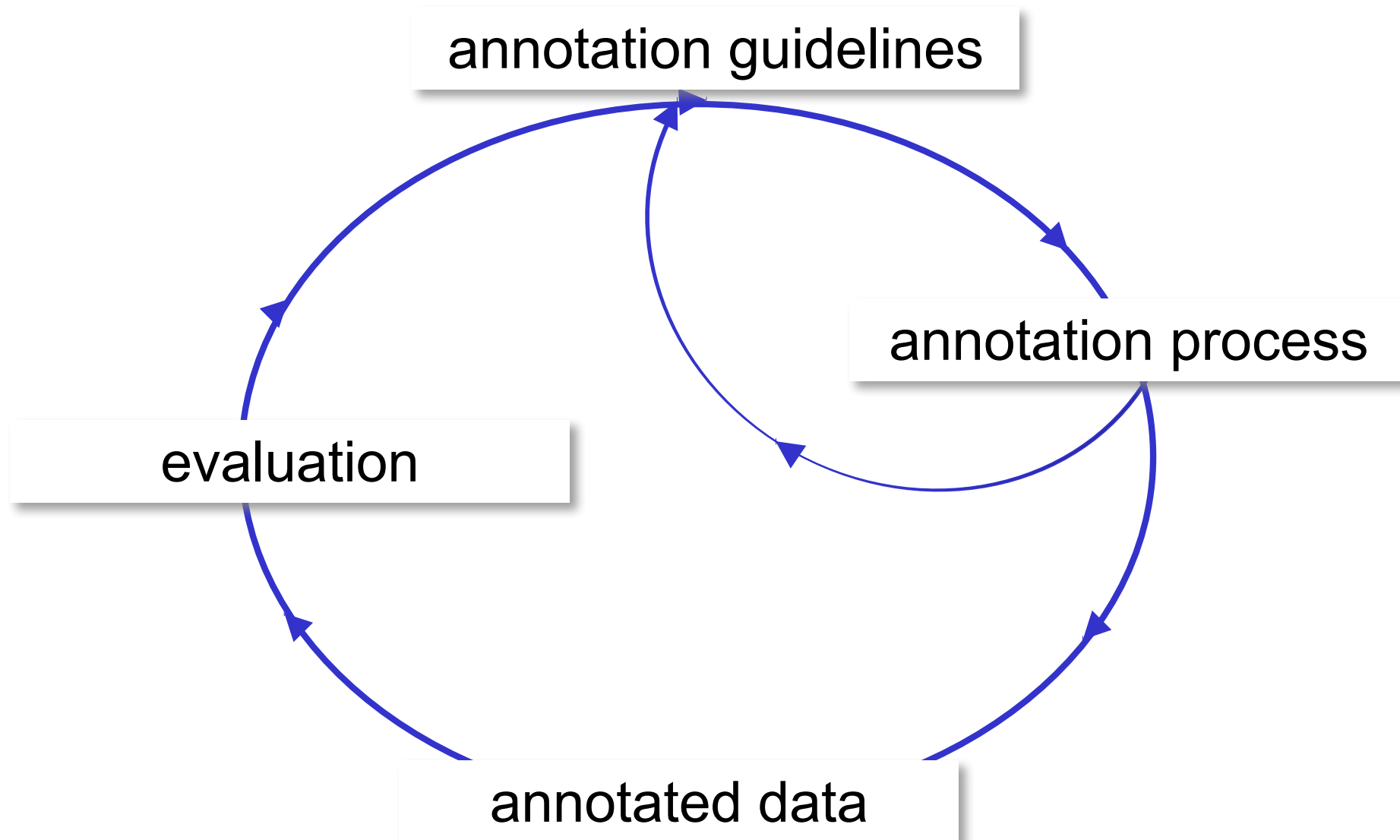
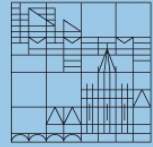
The annotation cycle

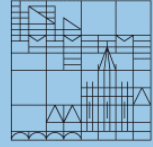




- Means
 - operational know-how
 - consistency of the annotation
 - sustainability of the annotated data
- Annotators
 - number
 - proficiency / training
- Data
 - data collection
 - encountering examples that do not fit the guidelines
- Time and Place

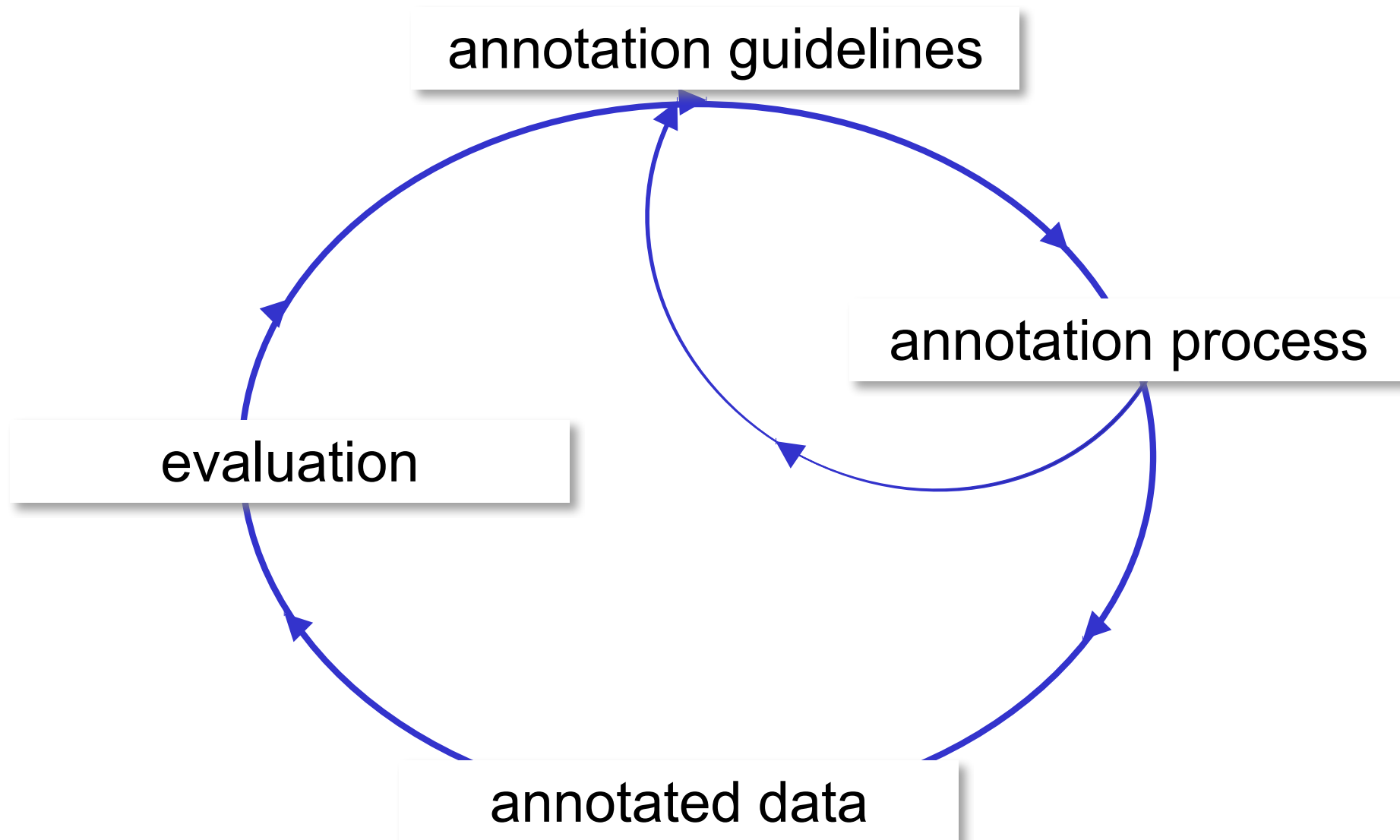
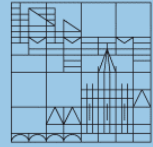
The annotation cycle

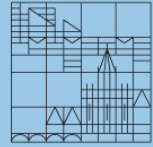




- Discussion of individual examples
- Comparison with a gold standard
- Inter-annotator agreement
 - abstracting away from chance agreement (see e.g. Artstein & Poesio 2008)
 - agreement on markables / labels
- Confusion matrices
 - labels that are easily confused
- Consistency of the annotation
 - reliability of the annotators
 - quality of the guidelines
 - difficulty of the phenomenon

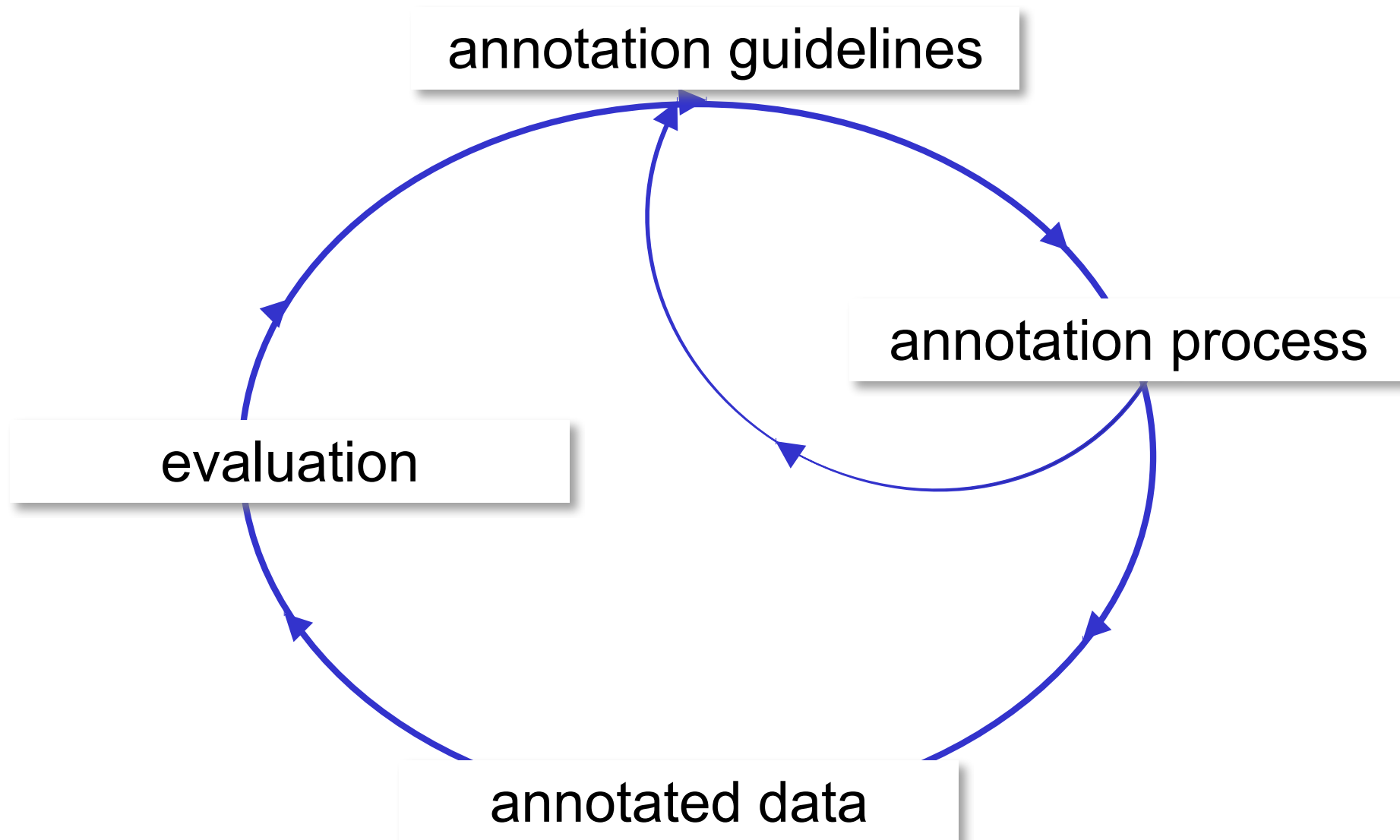
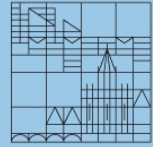
The annotation cycle



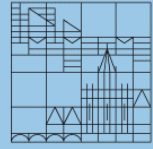


1. Motivation
2. The annotation cycle
- 3. Annotation as a didactic means**
4. Discussion of variables in the cycle
5. Sketch of two settings
6. Summary & discussion

The annotation cycle

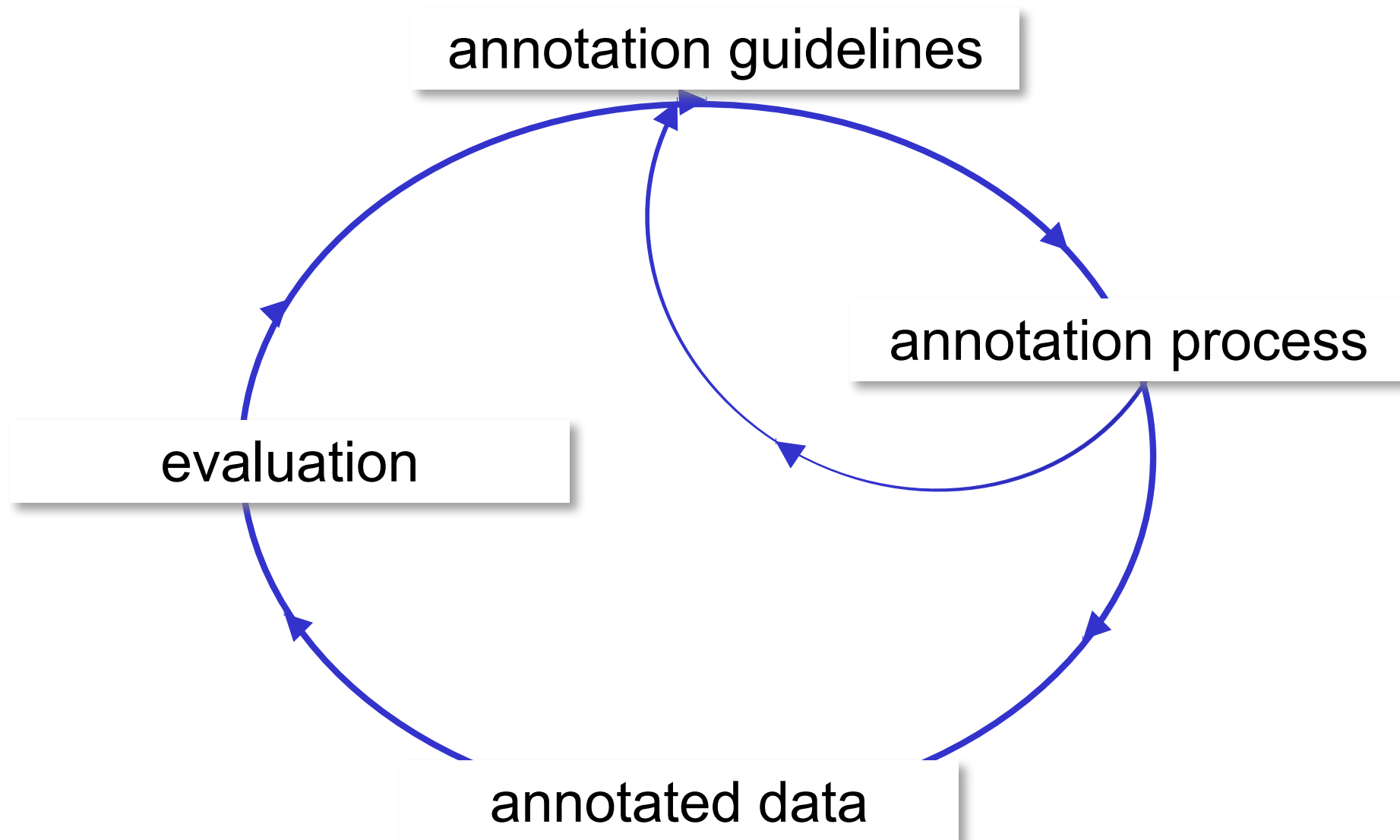
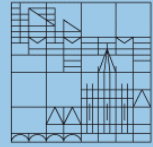


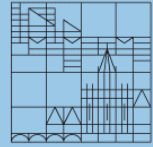
Creation of their own guidelines



- The students study grammars and linguistic articles with a purpose
 - operationalization of linguistic concepts
- Summarization / copy of definitions
 - “‘It’ is a referential pronoun, if it refers to an entity in the discourse model.”
- Collection / invention of examples
- Derivation of linguistic tests
 - “How do I know that an occurrence of ‘it’ is a referential pronoun?”

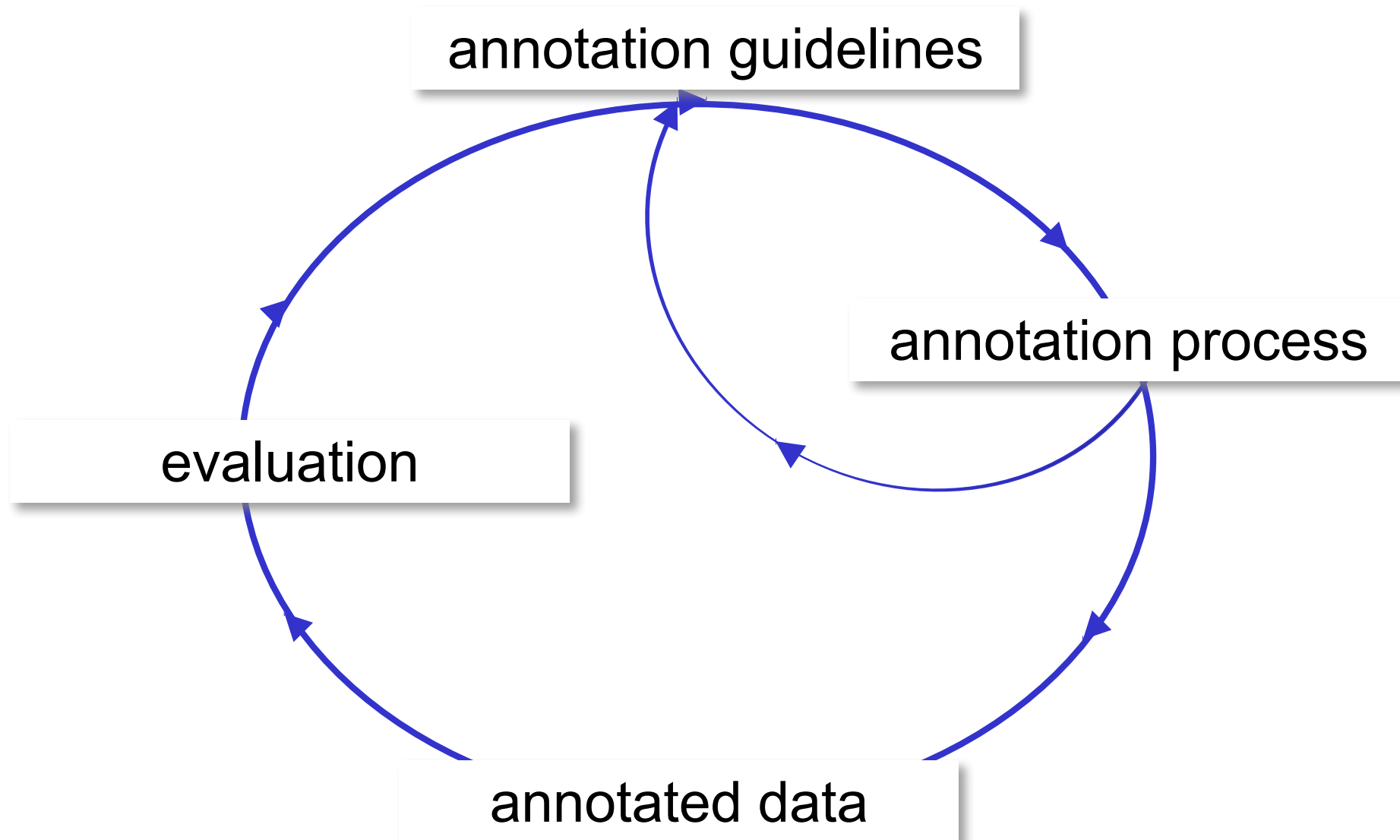
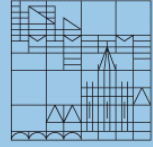
The annotation cycle

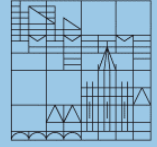




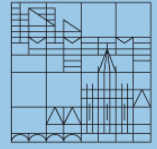
- The students put the analyses to use
 - learning by ‘doing’
 - challenge / fun (?)
- Uncontroversial examples help the students
 - to understand the classification / analysis
 - to memorize them
- Difficult examples encourage the students
 - to discuss the tests / the definitions / the classifications
 - use linguistic argumentation
 - understand that different analyses are competing models but not the ‘truth’ (i.e. the ‘language’, the language faculty)

The annotation cycle



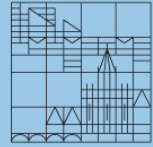


- Point them to difficulties and misunderstandings
 - ambiguous definitions
 - wrong definitions
 - missing exceptions / subclasses
 - vague phenomena
 - lack of understanding

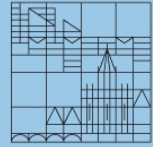


1. Motivation
2. The annotation cycle
3. Annotation as a didactic means
- 4. Discussion of variables in the cycle**
5. Sketch of two settings
6. Summary & discussion

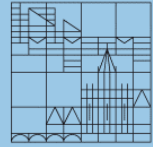
Variable: the guidelines



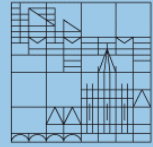
- Extreme values
 - to be created from scratch vs. completely given
 - all annotators work on all phenomena vs. expert groups
 - specific phenomenon vs. comprehensive sentence / text analysis
- Steps
 - study of grammars and analyses
 - classification / definition of labels
 - development of linguistic tests
- Pros and cons
 - learning effect
 - coverage of the linguistic facts
 - time needed for the task



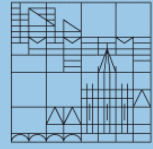
- Extreme values
 - pre-selected examples vs. free text
 - the same data for all annotators vs. individual examples for different annotators
 - pre-processed markables vs. free choice of string
- Pros and cons
 - coverage of the phenomena / diversity of the markables
 - learning effect
 - preparation time
 - time needed for the task
 - naturalness of the data
 - subsequent use of the data



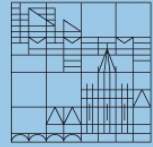
- Extreme values
 - paper and pencil vs. tools
 - graphical annotation / tag selection
 - evaluation scripts
 - visualization / search of the annotated data
- Pros and cons
 - technical requirements
 - operational know-how: training time of the instructor / the students
 - consistency of the annotation
 - reusability / replicability
 - learning effect
 - time needed for the task



- The temporal and local setting
- Extreme values
 - in a single session vs. accompanying a whole course
 - take-home task vs. lab
- Pros and cons
 - technical requirements: different platforms
 - time for exploration and discussion



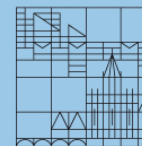
1. Motivation
2. The annotation cycle
3. Annotation as a didactic means
4. Discussion of variables in the cycle
- 5. Sketch of two settings**
6. Summary & discussion



- In-class introduction to coreference, anaphora and expletives
- Guidelines
 - predefined: general coreference / coherence annotation
- Data
 - Europarl (Koehn 2006) turns (w/ occurrences of 'es' (it))
 - 3 common, pre-selected examples and 3 random examples per student
- Paper& pencil vs. MMAX2 (90 min in class)
 - discussion groups but individual annotation
- No real cycle

Tool: MMAX2

Universität
Konstanz



The screenshot shows the MMAX2 software interface. At the top, there's a title bar with the text "[es]". Below it, a menu bar contains "One-click annotation", "Panel", and "Settings". A sub-menu bar has "ANA", "MARK", and "SENT" buttons. The main window is divided into two panes. The top pane contains annotation controls: a "< >" button, a "CandidateAnaphor" label, and several radio buttons: "na", "nominal_anaphor" (selected), "discourse_anaphor", "correlate", "unpersonal", and "vorfeld_es". Below these are fields for "Antecedent" (containing "621113") and "Comments" (an empty text box). At the bottom of this pane are "AnnotationCompleted" radio buttons: "na" and "yes" (selected). The bottom pane displays a text document with German text. The text includes several instances of "[es]" in yellow boxes. A red line is drawn across the text, starting from the first "[es]" and ending at the word "dieser". The text in the document is as follows:

Herr Präsident !

Ich denke , wir sollten unser Augenmerk nicht immer nur auf die Schiffskatastrophen und ihre Folgen richten , sondern sollten vielmehr etwas gegen die -- wie ich einmal sagen muss -- tägliche Schweinerei auf unseren Meeren unternehmen .

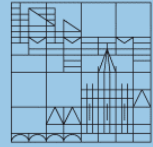
Deswegen an Sie , Frau Kommissarin , meinen herzlichen Glückwunsch zu [dieser Vorlage](#) .

Ich halte [\[es\]](#) für einen guten Vorschlag und möchte auch den Kollegen Pex , der leider abwesend ist , zu seinem guten Bericht beglückwünschen .

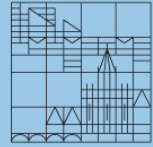
Wir müssen hier unbedingt etwas tun .

Allein im Jahr 2001 gab [\[es\]](#) in der Nord- und Ostsee zusammen ungefähr tausend illegale Einleitungen , und [\[es\]](#) darf nicht länger zulassen werden , [dass das als Kavaliersdelikt gilt und behandelt wird wie Falschparken oder eine Geschwindigkeitsübertretung](#) , sondern das muss endlich als kriminelles Delikt geahndet und strafrechtlich verfolgt werden .

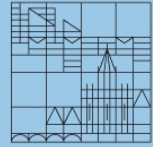
Und strafrechtliche Verfolgung , bedeutet harte Strafen , die spürbar sind , Geldstrafen , Gefängnis , aber eben auch -- wie der Ausschuss [\[es\]](#) beantragt hat -- [die Beschlagnahme von Schiffen](#) .



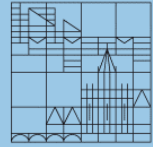
- Paper and pencil
 - easier to grasp the linguistic idea
 - training time limited to linguistic task
 - problem in comparing / evaluating the results (time-consuming)
 - results hardly reusable (even by the annotators themselves)
- MMAX2 (Müller & Strube 2006)
 - long training time due to operational issues
 - lab PCs (Macs), mouse-handling in tool
 - underlying file format hard to understand (XML standoff)
 - reusable results



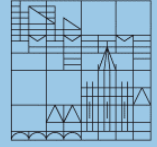
- Paper and pencil ideal for pilot study
- Using MMAX2 was more confusing but also more fun
- Future improvements
 - instructions about operational issues have to be absolutely clear
 - students need to be familiar with the hardware
 - support by student assistants
- General
 - annotation task is ideal for teamworking



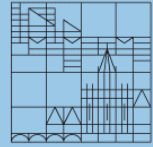
1. Motivation
2. The annotation cycle
3. Annotation as a didactic means
4. Discussion of variables in the cycle
5. Sketch of two settings
- 6. Summary & discussion**



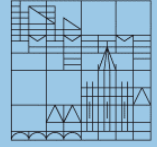
- Linguistic phenomenon
 - To what extent can it be subclassified / operationalized
- Requirements
 - availability of hardware, software
 - operational know-how of teachers (and students)
- Time needed for
 - the preparation / evaluation
 - the training in class
 - the task in class / at home



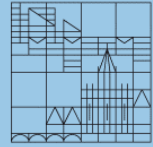
- Learning effect
- Quality of the annotation / analysis
- Reusability of the data
- Evaluation (grading) of both the documentation of the process and the outcome



- Learning effect
 - kind of problem-based learning
- Quality of the annotation / analysis
 - evaluation of individual annotators in multi-annotator settings, e.g. needed for ‘crowdsourcing’ data collection (cf. Carpenter 2008).
- Reusability of the data
 - representation of multiple / controversial annotation layers in XML-standoff formats (e.g. PAULA, Dipper 2005)
 - visualization and search of multi-layered data (e.g. ANNIS2, Zeldes et al. 2009)



Your own experiences with
annotation tasks in class?



- Artstein, Ron & Massimo Poesio. 2008. Inter-coder agreement for computational linguistics. *Computational Linguistics* 34, 555–596.
- Boyd, Adriane, Whitney Gegg-Harrison & Donna Byron. 2005. Identifying non-referential it: a machine learning approach incorporating linguistically motivated features. In *Proceedings of the ACL-05 Workshop on Feature Engineering for Machine Learning in Natural Language Processing*, 40–47.
- Carpenter, Bob. 2008. Multilevel Bayesian Models of Categorical Data Annotation. Technical Report. Alias-i. <http://lingpipe-blog.com/>
- Dipper, Stefanie. 2005. XML-based stand-off representation and exploitation of multi-level linguistic annotation schema. In *Proceedings of Berliner XML Tage 2005 (BXML 2005)*, 2005, 39–50.
- Müller, Christoph & Michael Strube. 2006. Multi-level annotation of linguistic data with MMAX2. In S. Braun, K. Kohn & Mukherjee, J. (ed.) *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*. Frankfurt a.o.: Peter Lang: 197–214.
- Zeldes, Amir, Julia Ritz, Anke Lüdeling & Christian Chiarcos. 2009. ANNIS: A search tool for multi-layer annotated corpora. In *Proceedings of Corpus Linguistics 2009*, Liverpool, July 20–23, 2009.