Annotation as a Didactic Means

Heike Zinsmeister, Melanie Andresen, Fabian Barteld,
Jiayin Feng, Johanna Flick, Renata Szczepaniak

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Institut für Germanistik

Motivation

The term annotation is a polysemous notion. In linguistic contexts, it denotes (i) the **process,** in which we enrich text by systematically adding interpretative information to it, (ii) the added information in terms of **conceptual** classes, hierarchies and other relations that **model** some linguistic phenomenon, and (iii) the **digital representation** of the added information (cf. Leech 1997, Kübler & Zinsmeister 2015).

In this poster, we explore annotation in all of its three dimensions from a didactic perspective.

The process of annotation is detailed as an **iterative procedure**, which enables students to conceptualize linguistic pheno-mena in a systematic way.

The development of conceptual classes in terms of an annotation tagset is done either **inductively** by inducing genera-lizations and classifications from the data, or **deductively**, based on a given annotation scheme or theory.

Even if annotation can be done with paper and pencil, **annotation tools** allow for an efficient realisation with a sustainable outcome in class.

The Annotation Cycle

Cf. the MAMA annotation development process in computational linguistics: model-annotate-evaluate-revise (Pustejovsky & Stubbs 2012) and the extended hermeneutic circle in the humanities: assumption-annotation-guidelines (Bögel et al. 2015).

LEA: Linguistic Exercises with Annotation Tools

Creating their own **digitally annotated data** can be a difficult task for linguistic beginners. This is especially true within the philologies where most students lack the computational expertise needed for corpus linguistics. Moreover, there is often not much time to teach the related methods.

With **LEA** we present a set of e-learning packages* that combine linguistic exercises with easy to use annotation tools, simplified annotation guidelines, and tools for correction and evaluation. Thus, while doing their homework students learn to create sustainably annotated data.

Advantages for students
- Helps to understand linguistic concepts via new ways of visualization
- Introduces students to annotation tools and guidelines

Advantages for teachers
- Tools to automatically analyze and evaluate students' answers
- Ready to use in introductory courses
- Own exercises can be easily integrated

*The development of the first LEA packages is funded by the „Innovationsfonds für Studium und Lehre" (Hamburg University, Faculty of Humanities). We thank the student assistants Christiane Höltmann and Alexandra Lindt.

Study: Inductive Annotation of Causality I

The goal of this task was for students to explore different types of causality.

(1) Die Rohre sind geplatzt, weil Frost herrscht.

'The pipes burst because it is freezing.'

(2) Es herrscht Frost. Weil, die Rohre sind geplatzt.

'It is freezing. [I know this] because the pipes burst.'

Scheme: RESULT, weil 'because' REASON

Task: Inductive development of an annotation scheme considering the syntax and semantics: the choice of connector, the source of coherence etc.

Causality describes a relation between two events: reason (a causing event) and result (a caused event). It is often indicated by causal markers, such as connectives (weil, deshalb...), prepo-sitions (wegen, aufgrund...) etc.

According to the source of coherence, we distinguish between a semantic and a pragmatic relation.

Semantic relation: two propositions are causally influenced; **Pragmatic** relation: the reason justifies the claim of the speaker or explains the speech act itself. (cf. Breindl & Walter 2009)
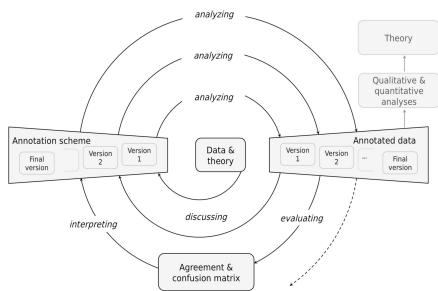
## Study: Inductive Annotation of Causality II

Setting
- 25 linguistics students (B.A. and M.A., including teacher training)
- 90 minutes: group discussion (4-6 students) and individual work
- Paper and pencil
- Prior knowledge from previous classes: different layers of annotation; discourse annotation with Rhetorical Structure Theory (RST)

Table suggested for the initial scheme:

Label	Definition	Cues	Comments

Steps

1. **Group**: Initial discussion of five examples that exemplified differ-ent readings; agreement on an initial annotation scheme.
2. **Individual:** Annotation of eight more sentences according to the initial annotation scheme.
3. **Group**: Comparison and evaluation of annotation results. Mismat-ches? Typical confusion catego-ries?
4. **Group**: Discussion of problematic cases. Is it possible to achieve agreement?
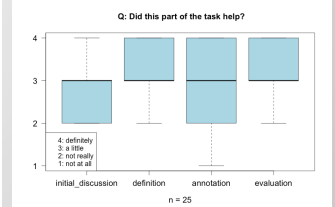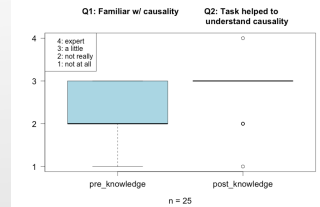
Discussion
- Mixed but overall positive evaluation results.
- Introduction was too short, goal of the task was not clear enough.
- Partly confusion about actual annotation step.
- One group got completely "lost".

Conclusions
- → Introducing "expert advisors".
- → Comparison between groups for more variety in annotations and discussions.
- → Relaxing time frame by reducing individual annotation load.

15 of 25 will be likely using annotation as a didactic means on their own.

Demo: Deductive Annotation of PoS with LEA

Every LEA package contains:
- The exercise
- Information how to obtain the annotation tool
- Annotation guidelines
- A manual for students and teachers
- The sample solution
- A tool for automatic correction and evaluation

The first LEA package deals with parts of speech. The exercise comes in a tab-separated-value file that can be edited with common spreadsheet applications like Libre Office Calc. The correction and evaluation tool is created in Java and outputs helpful statistics (e.g. mean number of errors, common mistakes, a confusion matrix) for the teacher (see figure on the right).

The second LEA package is still work in progress. Using Synpathy as annotation tool it will cover phrases and syntactic functions
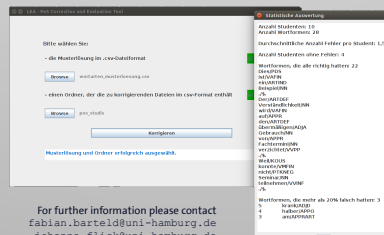
For further information please contact
fabian.barteld@uni-hamburg.de
johanna.flick@uni-hamburg.de

References

- Bögel, Thomas, Michael Gertz, Gius, Evelyn, Janina Jacke, Jan Christoph Meister, Marco Petris, and Jannik Strötgen. 2015. "Gleiche Textdaten, Unterschiedliche Erkenntnisziele? Zum Potential Vermeintlich widersprüchlicher Zugänge zu Textanalyse." In: Von Daten zu Erkenntnissen. Book of Abstracts - Vorträge, 119—26. Graz. http://gams.uni-graz.at/o:dhd2015.abstracts-vortraege.
- Breindl, Eva and Maik Walter. 2009. "Der Ausdruck von Kausalität Im Deutschen." Eine korpusbasierte Studie zum Zusammenspiel von Konnektoren, Kontextmerkmalen und Diskursrelationen. Mannheim: Institut für Deutsche Sprache.
- Kübler, Sandra and Heike Zinsmeister. 2015. Corpus Linguistics and Linguistically Annotated Corpora. Bloomsbury Publishing.
- Leech, Geoffrey. 1997. "Introducing Corpus Annotation." In Corpus Annotation. Linguistic Information from Computer Text Corpora, edited by Roger Garside, Geoffrey Leech and Tony McEnery, 1—18. London / New York: Longman.
- Lemnitzer, Lothar and Heike Zinsmeister. 2015. Korpuslinguistik. Eine Einführung. Tübingen: Narr.
- Pustejovsky, James and Amber Stubbs. 2012. Natural Language Annotation for Machine Learning. O'Reilly.
- Zinsmeister, Heike. 2011. "Exploiting the 'Annotation Cycle' for Teaching Linguistics." Slides presented at the Workshop Corpora in Teaching Linguistics and Linguistics, Humboldt-Universität zu Berlin. https://www.linguistik.hu-berlin.de/en/institut-en/professuren-en/korpuslinguistik/events-en/CTLL/ctll_abstracts/ctll_zinsmeister
- Synpathy: http://www.mpi.nl/tools/synpathy.html