

Proseminar Linguistische Annotation

Ines Rehbein und Josef Ruppenhofer

SS 2010



Seminarplan

- I. Linguistische Annotation - Überblick
 - ▶ Welche Arten von linguistischer Annotation gibt es?
 - ▶ Wozu sind sie gut?
- II. Der Annotationsprozess
 - ▶ Wie wird annotiert?
 - ▶ Welche Probleme treten dabei auf?
 - ▶ Welche Faktoren können die Annotation beeinflussen?
- III. Wie gut sind unsere Annotationen?
- IV. Wie bekomme ich größere Mengen an annotierten Daten?

- III. Evaluation - Wie gut sind unsere Annotationen?
 - ▶ Evaluation gegen einen manuell annotierten Goldstandard
 - ▶ Inter-Annotator Agreement
 - ▶ Einsatz der Daten als Trainingsset für Systeme der automatischen Sprachverarbeitung (Task-based evaluation)
- Thema: Methodik der Evaluation (Artstein & Poesio, 2008)
 - ▶ Evaluationsmaße
 - ▶ Umgebung mit mehreren Annotator/innen, Inter-Annotator Agreement
 - ▶ Beispiele aus der Praxis

Seminarplan

- **II. Der Annotationsprozess**
 - ▶ Wie wird annotiert?
 - ★ Erstellung von Annotationsrichtlinien
 - ★ Training
 - ★ Annotationsprozess
 - ★ Qualitätssicherung
 - ▶ Welche Probleme treten dabei auf?
 - ▶ Welche Faktoren können die Annotation beeinflussen?
 - ★ Annotations-Tools
 - ★ Richtlinien
 - ★ persönliche Eignung und Neigung der Annotator/innen
- **Thema:** Verlässlichkeit der Annotation, Inter-Annotator Agreement, Wie entwickle ich ein gutes Annotationsschema? (Maxwell, 1977; Bayers et al., 2003; Duffield et al., 2007)
- **Thema:** Lernen mit fehlerbehafteten Daten (Beigman and Beigman Klebanov, 2009)
- **Thema:** Sentiment Analysis: Unterschiede in der Annotation, spezifische Probleme (Wiebe et al., 2005)

- IV. Wie bekomme ich große Mengen an annotierten Daten?
- **Thema:** Einfluss von automatischer Vorannotation, Annotationsumgebung (Tools) und Erfahrung der Annotator/innen (Dandapat et al., 2009)
- **Thema:** Semi-automatische Annotation (8, 9, 10)
- **Thema:** Bootstrapping (16, 17, 18)
- **Thema:** Active Learning (11, 12, 13, 14, 15)

Seminarplan

- IV. Wie bekomme ich große Mengen an annotierten Daten?
- Thema: Semi-automatische Annotation (8, 9, 10)
 - ▶ 8. Chou et al., 2006. A semi-automatic method for annotating a biomedical proposition bank.
 - ▶ 9. Xue et al., 2002. Building a large-scale annotated Chinese corpus.
 - ▶ 10. Kuzman et al., 2007. Semi-Automated Named Entity Annotation.
- Thema: Bootstrapping (16, 17, 18)
 - ▶ 16. Hwa et al., 2005. Bootstrapping parsers via syntactic projection across parallel texts.
 - ▶ 17. Riloff et al., 2003. Learning subjective nouns using extraction pattern bootstrapping.
 - ▶ 18. Xu et al., 2002. A Domain Adaptive Approach to Automatic Acquisition of Domain Relevant Terms and their Relations with Bootstrapping.

- IV. Wie bekomme ich große Mengen an annotierten Daten?
- Thema: Active Learning (11, 12, 13, 14, 15)
 - ▶ 11. Tomanek and Olsson, 2009. A Web Survey on the Use of Active Learning to support Annotation of Text Data.
 - ▶ 12. Reichart et al., 2008. Multi-task active learning for linguistic annotations.
 - ▶ 13. Zhu and Hovy. Active Learning for Word Sense Disambiguation with Methods for Addressing the Class Imbalance Problem.
 - ▶ 14. Ringger et al., 2007. Active Learning for Part-of-Speech Tagging: Accelerating Corpus Annotation.
 - ▶ 15. Baldrige and Palmer, 2009. How well does active learning actually work? Time-based evaluation of cost-reduction strategies for language documentation.