



# F-AG 7 Kurationsprojekt 3: Semantische Annotation für Digital Humanities



COMMON LANGUAGE RESOURCES AND TECHNOLOGY INFRASTRUCTURE

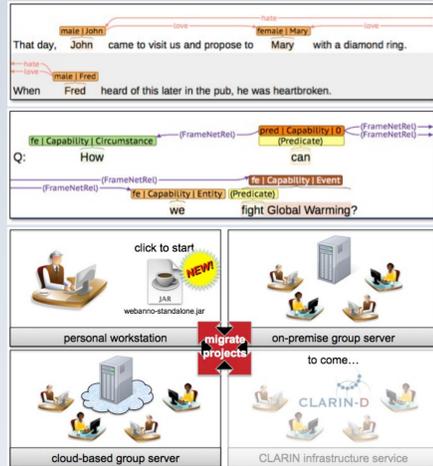
Die Kurationsprojekte der F-AG7 „Angewandte Sprachwissenschaften und Computerlinguistik“ haben zum Ziel, die Qualität von computerlinguistischen **Werkzeugen und Ressourcen für das Deutsche** zu verbessern. Im Zentrum stehen **Nicht-Standard-Varietäten des Deutschen**, sowie die Aufnahme spezieller **Bedürfnisse, Fragestellungen und Interessensgebiete aus den (Digital) Humanities** für den Einsatz computerlinguistischer Verfahren.

## Kurationsprojekte 1 & 2 → Kurationsprojekt 3

In der ersten Phase von CLARIN-D wurden die Kurationsprojekte **Implementierung einer webbasierten Annotationsplattform (WebAnno)** (KP1) und Entwicklung von **Guidelines und Best Practices für die Annotation von NSV des Deutschen** (KP2) durchgeführt. Ziel von KP3 ist es, die erfolgreiche Arbeit von KP1 & KP2 zu konsolidieren und in neuen Aspekten weiterzuentwickeln. Der Fokus von KP3 liegt auf **semantischer Annotation für DH**.

### WebAnno

- Intuitive Benutzeroberfläche
- Benutzerdefinierbare Annotationsebenen
- Arbeiten: alleine oder in Gruppen
- Annotation vergleichen und zusammenführen
- Automatisch erstellte Annotationen korrigieren
- CLARIN-kompatibel



#### Neu in WebAnno 3

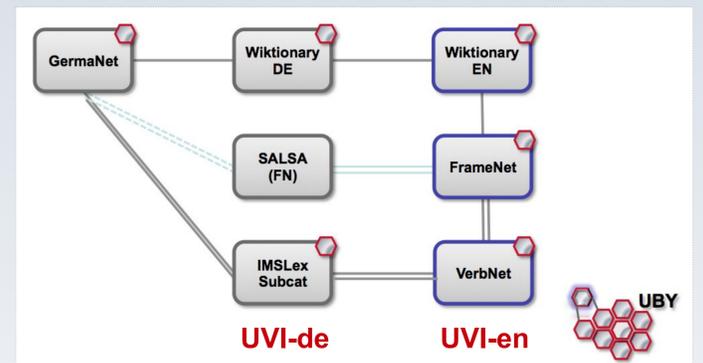
- Vorbereitung um WebAnno als CLARIN Service anbieten zu können
- Nutzerfeedback sammeln und zentrale Anforderungen umsetzen

### Unified Verb Index für Deutsch

- Kuration von Ressourcen für **Ereignisannotation: UVI-de**
- Projektionsbasierte **Integration** semantischer Lexika für das Deutsche
- **Qualitätsabschätzung** durch partielle manuelle Annotation



- UBY Web-API
- UBY LMF XML
- Linked Open Data (LOD)



### Semantische Annotation



```
# Semantic argument constraint
%Lemma.value = „shot“
label = „Action“

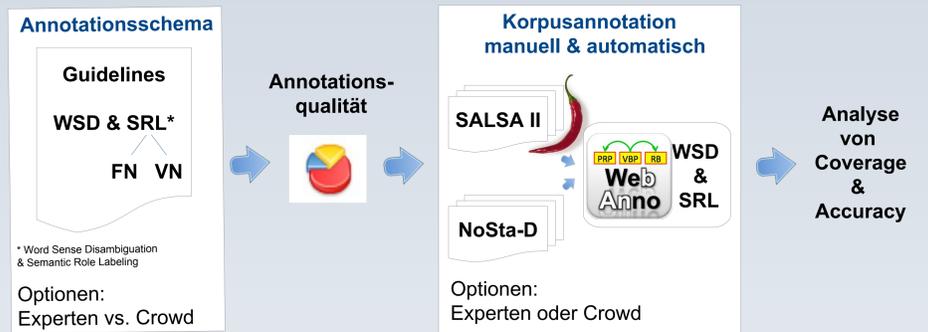
label = „Action“
arguments.role = „Actor“ |
arguments.role = „Target“

# Dependency relation constraint
governor.pos.posValue = „VB“ &
dependent.pos.posValue = „PRP“
depRel = „nsubj“
```

Zur besseren Unterstützung semantischer Annotationen und freier Annotationsarten wird WebAnno um neue Features erweitert:

- **Template-basierte Annotationen** – füllen von Slots, z.B. zur Annotation von Prädikat-Argument-Strukturen oder Ereignissen
- **Constraints** – kontextbasierte Einschränkung möglicher Annotationen

### WSD & SRL Annotation



#### Resultate

- Prototyp für verlinkte semantische Lexika im Deutschen (inkl. LOD)
- Robustes Annotationsschema für Prädikat-Argument-Strukturen

#### Perspektiven für DH: Ereignisannotation

### Shared Tasks und Annotation

„Lizenzierung“ in Koordination mit GSCL/DGfS-CL

- **NER für NSV** (aus KP1 & KP2): KONVENS 2014
- **Sentiment Tagging**: KONVENS 2014, gefördert durch GSCL
- **PoS-Tagging für IBK**: in Vorbereitung, gefördert durch GSCL

Optionen in KP3

- **Dependenzparsing für NSV** (mit Daten aus KP2)
- **Kompositanalyse**

### Referenzen

- Bonial, C., Stowe, K. & Palmer, M. (2013): Renewing and Revising SemLink. In: Proc. of LDL-2013: Representing and linking lexicons, terminologies and other language data, S. 9-17.
- Burchardt, A., Erk, K. & Frank, A. (2005): A WordNet Detour to FrameNet. In: Proc. of the GLDV 2005 GermaNet II Workshop, S. 408-421.
- Burchardt, A., Erk, K., Frank, A., Kowalski, A., Padó, S. & Pinkal, M. (2009): Using FrameNet for the Semantic Analysis of German: Annotation, Representation and Automation. Boas, H. C. (ed.), Multilingual FrameNets in Computational Lexicography - Methods and Applications, S. 209-244. Mouton de Gruyter.
- Burchardt, A., Padó, S., Spohr, D., Frank, A. & Heid, U. (2008): Constructing Integrated Corpus and Lexicon Models for Multi-Layer Annotations in OWL DL. Linguistic Issues in Language Technology, 1, S. 1-33.
- Chotikov, K., Eckle-Köhler, J. & Gurevych, I. (2014): Automated Verb Sense Labeling Based on Linked Lexical Resources. In: Proc. of EACL 2014, S. 68-77.
- Eckart de Castilho, R., Biemann, C., Gurevych, I. and Yimam, S.M. (2014): WebAnno: a flexible, web-based annotation tool for CLARIN. In Proceedings of the CLARIN Annual Conference (CAC) 2014, Soesterberg, Netherlands.
- Fürstenau, H. & Lapata, M. (2012): Semi-supervised Semantic Role Labeling via Structural Alignment. Computational Linguistics, 38(1): S. 135-171.
- Gurevych, I., Eckle-Köhler, J., Hartmann, S., Matuschek, M., Meyer, C. M. & Wirth, C. (2012): UBY - A Large-Scale Unified Lexical-Semantic Resource Based on LMF. In: Proc. of EACL 2012, S. 580-590.
- Hartmann, S. & Gurevych, I. (2013): FrameNet on the Way to Babel: Creating a Bilingual FrameNet Using Wiktionary as Interlingual Connection. In: Proc. of ACL 2013, S. 1363-1373.
- Palmer, M. 2009. Semlink: Linking PropBank, VerbNet and FrameNet. Proceedings of the Generative Lexicon Conference. : GenLex-09.
- Yimam, S.M., Eckart de Castilho, R., Gurevych, I., and Biemann C. (2014): Automatic Annotation Suggestions and Custom Annotation Layers in WebAnno In: Proceedings of ACL-2014, demo session, Baltimore, MD, USA
- Yimam, S.M., Gurevych, I., Eckart de Castilho, R., and Biemann C. (2013): WebAnno: A Flexible, Web-based and Visually Supported System for Distributed Annotations. In Proceedings of ACL-2013, demo session, Sofia, Bulgaria.



UBIQUITOUS  
KNOWLEDGE  
PROCESSING



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT



UNIVERSITÄT  
HEIDELBERG  
ZUKUNFT  
SEIT 1386



GEFÖRDERT VOM