



# Erkennung und Visualisierung attribuerter Phrasen in Poetiken

Andreas Müller (1)   Markus John (2)   Steffen Koch (2)  
Thomas Ertl (2)   Jonas Kuhn (1)

(1) Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart

(2) Institut für Visualisierung und Interaktive Systeme, Universität Stuttgart

DHd 2015

Von Daten zu Erkenntnissen: Digitale  
Geisteswissenschaften als Mittler zwischen Information  
und Interpretation



# Übersicht

- Projektkontext
- Erkennung attribuerter Phrasen
- Visualisierung attribuerter Phrasen
- Ausblick



## Projekt ePoetics

- Vom BMBF gefördertes Projekt
- Eines der Ziele: Literarische Analyse durch computerlinguistische Methoden unterstützen
- Hier vorgestellte Arbeit im Kontext von Erkennung von Äußerungen in Poetiken



## Äußerungen in Poetiken

- Semi-formale Zitate (Deutsche Rundschau 1886, S. 132)
- Direkte Rede ( "die Handlung", wie Lessing wohl sagen würde )
- Indirekte Rede (und eben daher sei der Lehrdichter kein Dichter)
- Zusammenfassungen und Paraphrasen (Irene Behrens hat gezeigt, dass sie erst am Ende des 18. Jahrhunderts aufgekommen ist.)
- attribuierte Phrasen ähnlich, aber nicht direkt Äußerungen



## Attribuierte Phrasen

- Beispiele: Schillers Poesie, Klopstocks Messias
- Referenzen auf Werke von Personen (Klopstocks Messias)
- Referenzen auf Werkarten von Personen (Schillers Poesie)
- Generelle Definition: Eine attribuierte Phrase ist eine Referenz auf ein einer Person zugeordnetes Konzept



## Anwendungen in literarischer Analyse

- Übersicht darüber wie in Poetiken über Personen gesprochen wird
  - Welche Schiller zugeordneten Konzepte werden in einer Poetik erwähnt?
- Vergleiche zwischen Dokumenten: Welche Konzepte A werden in mehreren Dokumenten Person B zugeordnet
  - Beispiel: Wenn mehrere Poetiken Klopstocks Messias erwähnen kann es interessant sein die zugehörigen Textstellen zu analysieren



## Linguistische Vorverarbeitung

- Tokenisierung und Satzerkennung (OpenNLP Tools)
- Part-of-speech tagging (mate tools) (Bohnet 2010)
- Nominalphrasenerkennung - MuNPEX noun phrase chunker
- Stanford named entity recognizer (Finkel et. al. 2010) mit den Modellen fürs Deutsche von Faruqui und Pado (2010)



## Suchmuster

- Person\_in\_genitive NP (zum Beispiel: Goethes Faust)
- NP von Person (zum Beispiel: die Poesie von Schiller)





# Evaluation

- Präzisionsbasierte Evaluation
- Phrasen in der Poetik von Staiger (1946) automatisch gefunden
- Jede attribuierte Phrase in eine von 4 Evaluationskategorien eingeteilt



## Evaluationskategorien

- Vier Evaluationskategorien
- Vollständig korrekt: Person und zugeordnete Phrase werden richtig erkannt
- Teilweise korrekt weniger: Teile von Person und/oder zugeordneter Phrase werden nicht erkannt
- Teilweise korrekt mehr: Mehr als die Person und die zugeordnete Phrase werden erkannt
- Inkorrekt: Weder Person noch Phrase werden richtig erkannt



## Evaluationsergebnisse

- Vollständig korrekt: 72
- Teilweise korrekt weniger: 25
- Teilweise korrekt mehr: 6
- Inkorrekt: 25



## Beispiele für Kategorie 2

- Goethes Forderung vs. Goethes Forderung an ein gutes Gedicht
- ein Gedicht Hebbels vs. ein Gedicht Hebbels, das "Lied" überschrieben ist



## Visualisierung attribuerter Phrasen - Anforderungen

- Schnelles Auffinden bestimmter Personen
- Übersicht wo in Poetiken Phrasen und Personennamen erwähnt werden
- Einfacher Zugang zu konkreten Textstellen
- Einfacher Vergleich konkreter Textstellen in mehreren Poetiken



# Visualisierung attribuerter Phrasen - Demo

- Demo



## Klassifizierung attribuerter Phrasen

- 4 Klassen:
  - Werke eines Autors (Beispiel: ein Gedicht Hebbels)
  - Teile von Werken eines Autors (Beispiel: vier Strophen von Mörikes «Verborgeneheit»)
  - Mengen von Werken eines Autors (Schillers Poesie)
  - Referenzen auf Äußerungen von Charakteren oder Autoren (Werthers Wort)
- Erste Ansätze zur Erkennung: Listen-basiert und orthographische Eigenschaften



## Klassifizierung attribuerter Phrasen - Werktitel

- Listen-basiert: Liste von über 86.000 Werken kann aus TextGrid extrahiert werden
- NP der attribuierten Phrase kann in Liste gesucht werden
- Vorteil: Wenn Werk in TextGrid kann Werk auch gleich verlinkt werden
- Orthographische Eigenschaften: Werktitel in Anführungszeichen





## Klassifizierung attribuerter Phrasen - Teile von Werken

- Manuell erstellte Liste von Teilen von Werken
- Wenn in Nominalphrase ein Teil eines Werkes vorkommt entsprechend klassifizieren
- Beispiel: vier Strophen von Mörikes «Verborgenheit»
- Gleicher Ansatz für Mengen von Werken



## Ausblick - Präpositionalphrasen als Indikatoren

- Präpositionalphrasen, Nominalphrasen und Verbalphrasen mit TreeTagger (Schmid 1995) finden
- "in den Zeilen", "aus dem Buch" => Indikatoren für Zitat
- Werkerkennung in Phrase: "im Faust" => Indikator für Zitat



## Phrase Clouds

- Phrasen als Überblick über Inhalt
- Beispiel: "Aristoteles Poetik" sehr hohe Frequenz in Scherer => gut sichtbar in der Phrase Cloud



## Sentiment Clouds

- Positive und negative Wörter in Phrasen
- Hinweise auf positive und negative Äußerungen des Autors
- Beispiel: "Er wäre verletzt"
- Beispiel für räumlich getrennte Phrase: Der Beifall ... verletzt uns in lyrischen Liedern.



## Zusammenfassung

- Erkennung attribuerter Phrasen
- Visualisierung attribuerter Phrasen
- Ausblick: Attribuierte Phrasen im Kontext von konkreten Anwendungen



- Danke für die Aufmerksamkeit! Fragen?



## References 1

- Bernd Bohnet. 2010. Very high accuracy and fast dependency parsing is not a contradiction. In Proceedings of the 23rd International Conference on Computational Linguistics (COLING '10). Association for Computational Linguistics, Stroudsburg, PA, USA, 89-97.
- M. Faruqui and S. Pado. Training and Evaluating a German Named Entity Recognizer with Semantic Generalization. Proceedings of Konvens 2010, Saarbrücken, Germany.
- Emil Staiger. 1946. Emil Staiger: Grundbegriffe der Poetik. Atlantis Verlag Zürich.



## References 2

- Jenny Rose Finkel, Trond Grenager, and Christopher Manning. 2005. Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling. Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL 2005), pp. 363-370.  
<http://nlp.stanford.edu/manning/papers/gibbscrf3.pdf>
- OpenNLP Tools: <https://opennlp.apache.org/>
- MuNPEX: <http://www.semanticsoftware.info/munpex>
- Helmut Schmid (1995): Improvements in Part-of-Speech Tagging with an Application to German. Proceedings of the ACL SIGDAT-Workshop. Dublin, Ireland.