

# Austausch von historischen Texten

## verschiedener Sprachen über das LAUDATIO-Repository

Carolin Odebrecht, Thomas Krause, Anke Lüdeling | Korpuslinguistik und Morphologie, Humboldt-Universität zu Berlin

LAUDATIO-Projekt

37. Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft,  
4.-6.März 2015 Universität Leipzig



### Zielstellung für das Repository

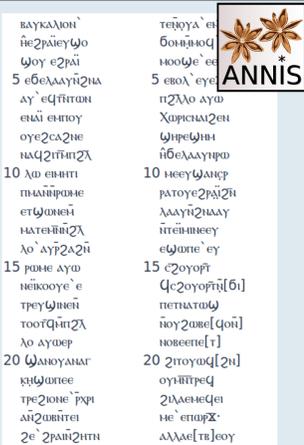
- **interdisziplinärer Austausch von historischen Texten und deren Annotationen für korpusgestütztes Forschen**
- einheitliche Beschreibung für Korpora unterschiedlicher Disziplinen und den damit verbundenen Registern, Sprachen, Annotationen und Formaten
- eigenständige Nutzung durch die Forscher

### Umsetzung

- Open Access Zugang zu wissenschaftlichen Forschungsdaten durch Creative Commons Lizenzen<sup>8</sup>
- Offenheit gegenüber allen Typen von Korpora durch ein generisches Datenmodell<sup>1</sup>
- Suche nach Korpora über ein technisch abstraktes Metadatenmodell<sup>2</sup> mit **allgemeinen Kriterien wie Sprachen, Orte, Daten, Textsorte**
- Suche und Analyse der Korpora durch eine Schnittstelle zu ANNIS<sup>3</sup>

### Shenoute (CC-BY 3.0)<sup>6</sup> Caroline T. Schroeder (University of the Pacific), Amir Zeldes (Georgetown University)

- Sprache: Sahidisch (Koptisch)
- Zeit: um 1000 v.Chr.
- Register: Brief (religiös)
- Ort: Atripe
- Aufbereitung: Transkriptionen, Normalisierung, Morpho-syntaktische Annotationen
- 2232 Token



### Historische Syntax des Jiddischen (CC-BY 3.0)<sup>4</sup>

Rosemarie Lühr (Friedrich-Schiller-Universität Jena)

- Sprache: Jiddisch
- Zeit: 15-18. Jahrhundert
- Register: Schöngestig-literarische Texte
- Ort: u.a. Basel, Prag, Worms, Krakau
- Aufbereitung: Transkriptionen und syntaktische, morphologische und lexikalische Annotationen
- 41878 Token

Alexander	Mukdun	der,die,das	sein	ein	gewaltig	König	
jidd.	jidd.	jidd.	jidd.	jidd.	jidd.	jidd.	
NE	NE	\$.	DD	VA	DI	ADJ	NA
NE	NE	\$.	DDS	VAFIN	DIA	ADJA	NA
		Masc.Sg.Nom.	3.Sg.Past.Ind.	Masc.Sg.Nom.	Pos.Masc.Sg.Nom.Mix.	Masc.Sg.Nom.	
subj		subj	præmpl	subj			
CE		CF_U_M					
v0		v2					
1C		1D					
narr		narr					
81v-2							

### Märchenkorpus (CC-BY SA 3.0)<sup>5</sup> Maik Walter (HU Berlin)

- Sprache: Deutsch
- Zeit: 19. Jahrhundert
- Register: Märchentexte
- Ort: u.a. Basel, Prag, Worms, Krakau
- Aufbereitung: Transkriptionen und Lemmatisierung, Wortarten
- 295880 Token

Rapunzel	ward	das	schönste	Kind	unter	der	Sonne	.
Rapunzel	werd	die	schön	Kind	unter	die	Sonne	.
NN	VFIN	ART	ADJA	NN	APPR	ART	NN	\$.

### Suche und Speicherung im Repository unabhängig von Sprache, Register und Umfang

für Linguistik, Philologien, Religionswissenschaften, Geschichtswissenschaften, Kulturwissenschaften, etc.

### Ausblick

- disziplinspezifische Anpassung des Interfaces/der Weboberfläche
- Schnellansicht Korpus/Dokumentation
- Einbindung in existierende Infrastrukturen wie DARIAH<sup>7</sup> und Anbindung an weitere Analysemöglichkeiten oder Datenquellen
- graphischer Editor für Metadaten

### Referenzen

<sup>1</sup> Zielke, D., Kindling, M. (2014) *Open Access Research Data Repository for Corpus Linguistic Data – A Modular Approach*. Open Repositories 2014 Conference. June 9.-13. Helsinki, Finland. | <sup>2</sup> Odebrecht, C. (2015) *Interdisziplinäre Nutzung von Forschungsdaten mithilfe einer technisch-abstrakten Modellierung*. DHd Jahrestagung. Von Daten zu Erkenntnissen. 25.02.2015. | <sup>3</sup> Krause, Thomas & Zeldes, Amir (2014): ANNIS3: A new architecture for generic corpus query and visualization. in: Digital Scholarship in the Humanities 2014 <http://dsh.oxfordjournals.org/cgi/content/abstract/fqu057?ijkey=GJBr0LhNfKW1g8i&keytype=ref> | <sup>4</sup> HSJ: <http://hdl.handle.net/11022/0000-0000-24F9-F> | <sup>5</sup> Märchenkorpus: <http://hdl.handle.net/11022/0000-0000-1F5B-9> | <sup>6</sup> Shenoute: <http://hdl.handle.net/11022/0000-0000-4680-0> | <sup>7</sup> <https://de.dariah.eu/> | <sup>8</sup> <https://creativecommons.org/>

