

Carolin Odebrecht, Thomas Krause, Anke Lüdeling
Humboldt-Universität zu Berlin

Austausch von historischen Texten verschiedener Sprachen über das LAUDATIO-Repository

LAUDATIO ist ein offenes Repository zur Speicherung historischer Texte und ihrer Annotationen. Das Ziel ist es, dass einmal erhobene Daten in neuen Projekten wiederverwendet werden können. So können Daten, die beispielsweise für geschichtswissenschaftliche Projekte erhoben wurden, unter anderen Fragestellungen auch für linguistische Forschung interessant sein. Um die Forschungsergebnisse, die auf diesen Daten basieren, nachvollziehen zu können, ist der langfristige Zugang zu der Dokumentation und zu den Quell- und Metadaten des Korpus notwendig. Dazu wurde ein umfangreiches Forschungsdatenmodell und Metadatenschema basierend auf TEI (Odebrecht 2014) entwickelt, das als Grundlage für die Weboberfläche des Repositoriums dient. Bisher ist LAUDATIO nur auf die Anforderungen zur Speicherung von annotierten Korpora, die aus historischen deutschen Texten von Althochdeutsch bis zum modernen Deutsch bestehen, angepasst. Korpusprojekte aus verschiedenen Disziplinen stellen schon jetzt ihre Korpora mit insgesamt 2000 Texten und mit etwa zwei Millionen Wortformen über LAUDATIO mit freien Lizenzen¹ zur Verfügung. Sofern die Korpora in Formaten vorliegen, die mit den Konvertierungstool Pepper (Zipser & Romary 2010) konvertiert werden können, können sie auch auf unserer öffentlichen Installation der generischen Korpussuchmaschine ANNIS (Krause & Zeldes erscheint) bereitgestellt werden. Damit ist ohne den Aufbau einer eigenen Infrastruktur nicht nur die langfristige Speicherung der Daten eines Projektes sichergestellt, sondern auch deren Analyse durch ANNIS.

Unser Ziel ist es, den Nutzerkreis von LAUDATIO zu erweitern und auch historische Texte anderer Sprachen und Zeitstufen zu unterstützen. Die Schwierigkeit besteht darin, das Forschungsdatenmodell und die Weboberfläche so weiterzuentwickeln, dass sie einerseits flexibel und generisch sind, aber andererseits auch die unterschiedlichen Texte und Annotationen umfassend und detailliert abbilden können. Durch solche Anpassungen können nun auch Korpora wie das „Historische Syntax des Jiddischen“²-Korpus oder die koptischen Korpora des Projektes „Scriptorium“ (Zeldes & Schroeder 2014) in LAUDATIO integriert werden. In Zukunft wird LAUDATIO neben neuen Sprachen und Zeitstufen verstärkt auch Texte weiterer geisteswissenschaftliche Disziplinen aufnehmen.

Referenzen

Zeldes, Amir/ **Schroeder**, Caroline T. (2014) Multi-layered Annotation for Multi-disciplinary Research on Digitized Coptic Literature. Digital Humanities 2014 Conference. 8.7.-12.7.2014, Lausanne. URL: <http://dharchive.org/paper/DH2014/Paper-849.xml>

Krause, Thomas/ **Zeldes**, Amir (erscheint) ANNIS3: A New Architecture for Generic Corpus Query and Visualization Literary and Linguistic Computing.

Odebrecht, Carolin (2014) Modeling Linguistic Research Data for a Repository for Historical Corpora. Digital Humanities 2014 Conference. 8.7.-12.7.2014, Lausanne. URL:<http://dharchive.org/paper/DH2014/Paper-737.xml>

Zipser, Florian/ **Romary**, Laurent (2010) A model oriented approach to the mapping of annotation formats using standards In Proceedings of the Workshop on Language Resource and Language Technology Standards, LREC 2010. Malta. URL: <http://hal.archives-ouvertes.fr/inria-00527799/en/>

¹ Creative Commons <http://de.creativecommons.org/>

² Ressource <http://hdl.handle.net/11022/0000-0000-24F9-F> und Projekt <http://www.indogermanistik.uni-jena.de/Web/Projekte/HSJ.htm>