

VORWORT

Der Aufbau, die Annotation und die Auswertung digitaler Korpora und Sprachressourcen haben im letzten Jahrzehnt weltweit einen großen Aufschwung genommen. Es ist das Ziel des vorliegenden Bandes, einen repräsentativen Überblick über einschlägige Forschungen in Österreich zu geben. Diese beziehen sich auf synchrone, schriftliche und mündliche, standardsprachliche, medien sprachliche, historische, dialektale und kindersprachliche Korpora, die in Österreich aufgebaut werden.

Der Band befasst sich mit Forschungsinfrastrukturen, mit der Korpuserstellung unter Anwendung verbreiteter Standards und mit Möglichkeiten der Korpusanreicherung, etwa durch morphosyntaktische Annotationen und Kodierungen; weiters mit Korpusvergleichen, der Entwicklung und Evaluierung von Annotationswerkzeugen und der Beantwortung konkreter Forschungsfragen, die korpusbasiert effizient bearbeitet werden können.

Die hier versammelten Beiträge bieten eine Auswahl der Vorträge, die Ende 2013 im Rahmen der 40. Österreichischen Linguistiktagung im Workshop „Korpusbasierte Linguistik in Österreich“ an der Universität Salzburg gehalten worden sind:

Karlheinz Mörth gibt als Koordinator von CLARIAH-AT einen Überblick über Aufbau, Funktionen und Ziele von internationalen Forschungsinfrastrukturinitiativen in Österreich.

Jutta Ransmayr, *Matej Ďurčo* und *Karlheinz Mörth* stellen das Austrian Media Corpus (AMC) vor, das in exhaustiver Weise die Texte aller österreichischen Printmedien umfasst und bereits am Institut für Corpuslinguistik und Texttechnologie und am Austrian Centre for Digital Humanities an der Österreichischen Akademie der Wissenschaften beforscht wird.

Claudia Resch und *Ulrike Czeitschner* berichten über den Aufbau und die Annotation des Austrian Baroque Corpus (ABaC:us) und zeigen, wie sich stilistische Spezifika und musterhafte Regularitäten in Texten von Abraham a Sancta Clara identifizieren und auswerten lassen.

Branko Tošović beschreibt den Aufbau und die morphosyntaktische Annotation eines umfangreichen mehrsprachigen Grazer Korpus (Gralis).

Katharina Korecky-Kröll widmet sich dem Aufbau und der Beforschung eines Wiener longitudinalen Erstspracherwerbskorpus im Vergleich zur Erwachsenensprache und den Prinzipien des internationalen CHILDES-Projekts.

Peter Ernst lotet die Möglichkeiten der korpusbasierten Eigennamenforschung bezüglich Identifikation, Klassifikation und Annotation in verschiedenen Korpusinitiativen aus.

Martina Werner, Karlheinz Mörth und *Wolfgang U. Dressler* untersuchen die Entstehung und Entwicklung von Pluraldubletten im historischen Austrian Academy Corpus (AAC) und im Austrian Media Corpus (AMC).

Miroslava Hliničanová, Matej Ďurčo, Karlheinz Mörth und *Wolfgang U. Dressler* kontrastieren phonotaktische und morphonotaktische Konsonantengruppen in slowakischen und deutschen Korpora.

Friedrich Neubarth und *Harald Trost* informieren über die Methoden ihres Projekts zur statistisch-maschinellen Übersetzung aus dem Standarddeutschen in den Wiener Dialekt.

Die HerausgeberInnen und AutorInnen dieses Bandes sind am Institut für Corpuslinguistik und Texttechnologie und am Austrian Centre for Digital Humanities der Österreichischen Akademie der Wissenschaften tätig beziehungsweise an den Instituten für Sprachwissenschaft und Germanistik der Universität Wien, am Institut für Slawistik der Universität Graz und am Österreichischen Forschungszentrum für Artificial Intelligence sowie am Institut für Artificial Intelligence der Medizinischen Universität Wien.

Mit der Herausgabe dieses Bandes wollen wir in der Reihe „Linguistik und Kommunikationsforschung“ des Verlages der Österreichischen Akademie der Wissenschaften einen neuen Akzent setzen, zur intensiveren Nutzung digitaler Korpora anregen und zur Verbreitung digitaler Methoden in den Geisteswissenschaften beitragen.

Wien, im März 2016

Claudia Resch

Wolfgang U. Dressler