

Donnerstag
09.03.2017
10:30 – 11:15
B4 1, Foyer

Semi-automatische TEI Repräsentation der Diskursstruktur eines deutschsprachigen Blogkorpus

Holger Grumt Suárez
*Justus-Liebig-Universität
Gießen*

Suarez@germanistik.uni-
giessen.de

Natali Karlova-
Bourbonus
*Justus-Liebig-Universität
Gießen*

Natali.Karlova-
Bourbonus@germanistik.uni-
giessen.de

Henning Lobin
*Justus-Liebig-Universität
Gießen*

Henning.Lobin@uni-
giessen.de

Das Poster „Semi-automatische TEI Repräsentation der Diskursstruktur eines deutschsprachigen Blogkorpus“ informiert über das Vorgehen sowie die ersten Forschungsergebnisse und die weiteren Ziele der Kompilierung und Annotation eines deutschsprachigen Blogkorpus.

Bislang existiert kein Standard für das Repräsentieren von sogenannten Computer-Mediated Communication-Daten (kurz CMC), allerdings arbeitet die TEI CMC Special Interest Group (vgl. Beißwenger 2016) seit 2013 an einem Schema für die Repräsentation von CMC-Genres. Unser Forschungsvorhaben leistet einen Beitrag zur Standardisierung der Text Encoding Initiative (TEI) für die CMC. Das Hauptziel des Vorhabens umfasst die semi-automatische Kompilierung sowie die Repräsentation der Blogdiskursstruktur. Dabei sollen die Relationen zwischen den textuellen und multimodalen Elementen (Blogbeiträge, Kommentare, Hyperlinks, Bilder und Töne) und den verschiedenen Textproduzenten (Blogger, Kommentatoren) abgebildet werden. Die Grundlage des Korpus bildet das Wissenschaftsblogportal SciLogs – Tagebücher der Wissenschaft (SciLogs 2016). Zusammenfassend soll das Poster nicht nur unser Vorhaben vorstellen, sondern auch einen Einblick in unser grundsätzliches Vorgehen bei der Erstellung eines CMC-Korpus geben.

References: • Abendroth-Timmer, D. et al. (2014). *Corpus d'apprentissage INFRAL (Interculturel Franco-Allemand en Ligne)*. Banque de corpus CoMeRe. Ortolang.fr: Nancy. <https://hdl.handle.net/11403/comere/cmr-infral>. • Beißwenger, M. (2016). *SIG:Computer-Mediated Communication*. http://wiki.teic.org/index.php/SIG:Computer-Mediated_Communication. • SciLogs (2016). *SciLogs. Tagebücher der Wissenschaft*. Spektrum der Wissenschaft. • WebCorp (2013). *Birmingham Blog Corpus*. WebCorp: Linguist's Search Engine. Birmingham City University.