

55. Jahrestagung des Instituts für Deutsche Sprache vom 12. bis 14. März 2019

## DEUTSCH IN SOZIALEN MEDIEN

Interaktiv, multimodal, vielfältig

### KALEIDOSKOP

Mittwoch, 13. März 2019, 15:00 Uhr

#### DAS DIDI-KORPUS: INTERNETBASIERTE KOMMUNIKATION AUS SÜDTIROL, ITALIEN

Aivars Glaznieks und Jennifer-Carmen Frey

Das DiDi-Korpus ist eine Sammlung von soziolinguistisch und computerlinguistisch aufbereiteten Facebook-Statusmeldungen, -Kommentaren und -Chatnachrichten. Alle Texte stammen von privaten Profilen von 136 Schreiberinnen und Schreibern aus Südtirol und stellen deren gesamte Facebook-Kommunikation des Jahres 2013 dar. Die Schreiberinnen und Schreiber geben überwiegend Deutsch als ihre Erstsprache an (108/136) und verteilen sich auf unterschiedliche Altersgruppen. Neben der breiten Altersspanne der Schreiberinnen und Schreiber (14-74 Jahre) zeichnet sich das Korpus vor allem durch regionale Eigenheiten wie seinem hohen Anteil dialektaler und mehrsprachiger Texte (besonders Deutsch, Italienisch und Englisch) aus. Das Korpus ist frei zugänglich und kann über eine ANNIS-Suchoberfläche durchsucht oder als XML bzw. JSON heruntergeladen werden (<https://commul.eurac.edu/annis/didi>).

Auf der Methodenmesse werden wir das Korpus und daran durchgeführte Fallstudien vorstellen und Einblick in vorhandene Annotationen sowie in die Suchfunktionen der Korpusoberfläche geben.

Frey, Jennifer-Carmen, Aivars Glaznieks & Egon W. Stemle (2016): The DiDi Corpus of South Tyrolean CMC Data: A multilingual corpus of Facebook texts. In *Proceedings of the Third Italian Conference on Computational Linguistics CLiC-it 2016. 5-6 December 2016, Napoli*, Anna Corazza, Simone Montemagni, & Giovanni Semeraro, (eds). Torino: Academia University Press, 157-161.

Frey, Jennifer & Aivars Glaznieks (2018): The Myth of the Digital Native: Analysing language use of different generations on Facebook. *Proceedings of the 6th Conference on Computer-mediated Communication (CMC) and Social Media Corpora (CMC-corpora 2018)*. 17-18 September 2018, University of Antwerp, Reinhild Vandekerckhove, Darja Fišer and Lisa Hilte (eds.), Antwerp, 41-44. [<https://www.uantwerpen.be/en/conferences/cmc-social-media-2018/proceedings/>]

Glaznieks, Aivars & Alexander Glück (in press, 2019): From the Valleys to the World Wide Web: Non-Standard Spellings on Social Network Sites. Egon W. Stemle & Ciara R. Wigham (eds.): *Building computer-mediated communication corpora for socio-linguistic analysis*. Clermont Auvergne: Clermont Auvergne University Publishing House.