

## NEUES VOM HEUTIGEN DEUTSCH

Empirisch - methodisch - theoretisch

MITTWOCH, 14. MÄRZ 2018, 10.00 UHR

### „WÖRTER IM SPIEGEL DER ZEIT“: WORTSCHATZWANDEL QUANTITATIV UNTERSUCHT ANHAND ALLER AUSGABEN DES ‚SPIEGEL‘ UND DER ‚ZEIT‘

Alexander Koplenig / Carolin Müller-Spitzer / Sascha Wolfer (IDS)

Im Zentrum unseres Vortrags steht eine vollständige systematische Analyse sämtlicher Ausgaben des Nachrichtenmagazins ‚Der Spiegel‘ und der Wochenzeitung ‚Die Zeit‘, welche seit kurzer Zeit am IDS als Korpusquellen zu Verfügung stehen. Dazu gliedert sich der Vortrag in drei Teile. Zunächst möchten wir im ersten Teil zeigen, warum die Analyse von natürlichsprachlichen Daten unter Zuhilfenahme klassisch-quantitativer Verfahren der Statistik aufgrund der stark ungleichen („Zipf“-)Verteilung von Worthäufigkeiten und des Umstandes, dass sich Interdependenzen auf Textebene ergeben (Montemurro & Pury 2002), schwieriger ist, als man es zunächst vermuten würde (vgl. u.a. Koplenig 2017). Aus diesem Grund benötigt man zum einen für die Analyse von natürlichsprachlichen Daten andere, besser geeignete Verfahren und zum anderen meist Zugang zu den Rohdaten (also das rohe, nicht-aggregierte Textmaterial in seiner ursprünglichen Abfolge), auch wenn dies wiederum aufgrund der großen Datenmengen und möglicher Urheberrechtsbeschränkungen eine Herausforderung darstellen kann. Entsprechend dieses Zwischenfazit wenden wir im zweiten Teil unseres Vortrags ein relativ neues statistisches Verfahren aus der Informationstheorie an, welches auf den so genannten generalisierten („Tsallis“-)Entropien beruht und mit dem jüngst vielversprechende Ergebnisse in der Analyse von lexikalischen Daten erzielt wurden (Altmann et al. 2017; Gerlach et al. 2016). Dieses Analyseverfahren erweitern wir um ein neues methodisches Prüfverfahren, welches den speziellen Anforderungen Rechnung trägt, die mit der quantitativen Analyse von Textdaten einhergehen. Basierend auf diesem Ansatz widmen wir uns im letzten Teil unseres Vortrags der eigentlichen Frage, wie sich der Wortschatz in den beiden untersuchten Printmedien über die Zeit hinweg entwickelt hat.

#### Literatur:

- Altmann, Eduardo G., Laércio Dias & Martin Gerlach. 2017. Generalized entropies and the similarity of texts. *Journal of Statistical Mechanics: Theory and Experiment* 2017(1). 014002.
- Gerlach, Martin, Francesc Font-Clos & Eduardo G. Altmann. 2016. Similarity of Symbol Frequency Distributions with Heavy Tails. *Physical Review X* 6(2). doi:10.1103/PhysRevX.6.021009. <https://link.aps.org/doi/10.1103/PhysRevX.6.021009> (8 January, 2018).
- Koplenig, Alexander. 2017. Against statistical significance testing in corpus linguistics. *Corpus Linguistics and Linguistic Theory* 0(0). doi:10.1515/cllt-2016-0036. <https://www.degruyter.com/view/j/cllt.ahead-of-print/cllt-2016-0036/cllt-2016-0036.xml?format=INT> (8 January, 2018).
- Montemurro, Marcelo A. & Pedro A. Pury. 2002. Long-range fractal correlations in literary corpora. *Fractals* 10(04). 451-461.