

Sentimentanalyse mit und ohne NN

für autobiographische Quellen von 1750-1850

Patrick D. Brookshire, M.A.

1) Hintergrund

Sentimentanalyse basiert in der Regel entweder auf speziellen **Lexika** oder **Neuronalen Netzen (NN)**, die auf gegenwartssprachlichen Korpora trainiert sind. Historische Daten unterscheiden sich jedoch u.a. durch Orthographie und/oder Sprachwandel von diesen.

→ „Wie gut eignen sich gängige Verfahren zur Analyse historischer Daten?“


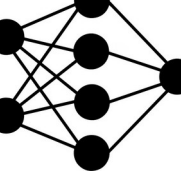
2) Korpus

Herrnhuter Lebensläufe bieten sich für erste Untersuchungen an, da **seit 1747** jeder aus der **pietistischen** Gruppierung einen verfassen musste (vgl. Böß 2016).

Aus dem Projekt *Moravian Lives* der Universitäten von Bucknell & Göteborg liegen **87 deutschsprachige** vor.

3) Coverage-Vergleich

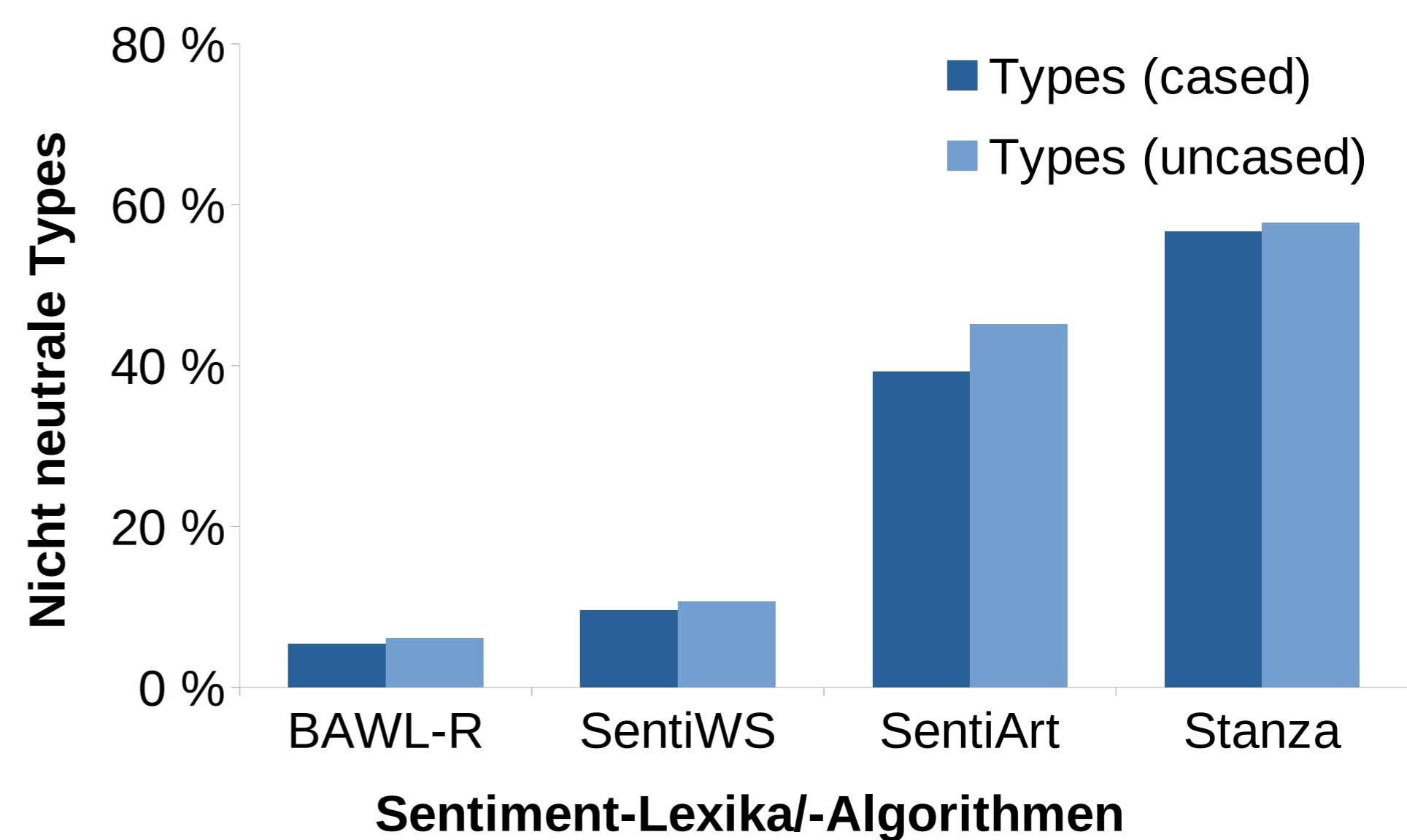
(angelehnt an Herrmann & Grisot 2022)

- gängige Lexika (BAWL-R/SentiWS) 
- mit NN vergrößertes Lexikon (SentiArt)
- NN-Classifer (Stanza) 

→ „Wieviele Types sind nicht neutral?“

4) Ergebnisse

- **Gängige Lexika** sind **ungeeignet**, da sie nur 5-10% der Types erkennen.
- Beide NN-basierten Ansätze decken mehr ab, ohne Domänen-spezifisch zu sein.
- Kleinschreibung spielt nur bei SentiArt eine gewisse Rolle.



5) Ausblick

Ein Annotationstask ist in Vorbereitung, um einen **Goldstandard** zu generieren, mit dem sich bestehende Algorithmen evaluieren bzw. neue NN (nach-)trainieren lassen.

da ich in meiner **Seelen-Noth** zum **Heiland** **schrie**, hat er mir meine **Sünden** **vergeben**.



Literatur

- Böß, S. (2016). Gottesacker-Geschichten als Gedächtnis. Eine Ethnographie zur Herrnhuter Erinnerungskultur am Beispiel von Neudietendorfer Lebensläufen. *Studien zur Volkskunde in Thüringen*, 6.
- Buechel, S., Hellrich, J. & Hahn, U. (2016). Feelings from the Past - Adapting Affective Lexicons for Historical Emotion Analysis. *LT4DH@COLING*, 54-61.
- Herrmann, J.B. & Grisot, G. (2022). Lieblingsgedenken, Fenster und Mauern. Zur emotionalen Enkodierung von Raum in Deutschschweizer Prosa zwischen 1850 und 1930. *Dhd*, 166-170.
- Jacobs, A.M. (2019). Sentiment Analysis for Words and Fiction Characters From the Perspective of Computational (Neuro-)Poetics. *Frontiers in Robotics and AI*, 6: 53.
- Qi, P., Zhang, Y., Zhang, Y. et al. (2020). Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. *ACL*, 101-108.
- Remus, R., Quasthoff, U. & Heyer, G. (2010). SentiWS - A Publicly Available German-language Resource for Sentiment Analysis. *LREC*, 1168-1171.
- Schmidt, T., Burghardt, M. & Wolff, C. (2018). Herausforderungen für Sentiment Analysis-Verfahren bei literarischen Texten. *INF-DH*: 16.
- Vö, M.L.H., Conrad, M., Kuchinke, L. et al. (2009). The Berlin Affective Word List Reloaded (BAWL-R). *Behavior Research Methods* 41, 534-538.