

Müssen Texte statistisch anders ausgewertet werden als Menschen?

Stella Neumann

Soweit in der Sprachwissenschaft statistische Methoden für die Auswertung empirischer Daten eingesetzt werden, sind diese den Humanwissenschaften entliehen. Dies ist unproblematisch, wenn die Daten aus Tests mit Versuchspersonen gewonnen wurden. In diesem Fall gibt es in der Tat keinen Unterschied zum Einsatz von Statistik in Psychologie und Sozialwissenschaft. Wenn jedoch nicht der Mensch einen Fall darstellt, sondern ein Text, dann ist zu fragen, ob noch die gleichen Tests eingesetzt werden können. Dabei ist zunächst zu fragen, ob ein Text überhaupt einen Fall darstellt. Möglicherweise stellt eine ganze Textsammlung einen einzigen Fall dar, oder, umgekehrt, jeder Satz oder gar jedes Wort ist separat als Fall in die Auswertung aufzunehmen. Aus dieser grundlegenden Frage ergeben sich auch weitere Fragen: Welches Skalenniveau kann für Häufigkeitsauszählungen zugrunde gelegt werden? Können verschiedene Annotationen an den gleichen Texten als wiederholte Messungen bezeichnet werden?

Der Beitrag setzt sich weniger zum Ziel, diese Fragen zu beantworten, sondern ist eher als Diskussionsbeitrag zu sehen. Die verschiedenen Fragen sollen in ihrer Bandbreite erläutert, an Beispielen aus dem CroCo-Projekt veranschaulicht und zur Diskussion gestellt werden.