

#SIMILARITY

In the Oxford English Dictionary (OED), the term #SIMILARITY initially describes an everyday meaning of likeness or resemblance and the ensuing details or aspects “in which a particular thing is similar to another”. Further uses and compounds largely focus on “systems or processes for indicating or determining the degree of similarity between particular things” with a strong reference to mathematics. *

From this, the mixed method projects derive the digital humanities uses of the term, in which common comparisons play a minor role. The focus seems to be on identifying degrees of similarity and relating them to each other or mapping them in absolute terms. “Similarity is a relation between two entities, which is characterized by their significant proximity to a tertium comparationis” (Digital Plato). The entities vary and may include for example numerical, semantic or stylistic properties. As such, the concept of similarity may shift into an entirely abstract digital realm accessed by #MODELLING and #HUMAN-IN-THE-LOOP practices: “Among the steps that lead from word frequencies to a quantification of stylometric similarity are lemmatization, filtering, occurrences and dispersion, normalization, and the use of different aggregation functions” (ReadingAtScale).

The findings are often interpreted and discussed by clustering and/or #VISUALIZATION, thus connecting back to the realms of humanities’ practices of comparison (ReadingAtScale, ANCI).

* “similarity, n.”. in: *Oxford English Dictionary* (OED), Third Edition, September 2019; most recently modified version published online March 2022, <https://www.oed.com> / [accessed: 20.05.2022].

Title: Reading at Scale

Team: Thomas Weitin, Simon Pöpcke, Katharina Herget, Anastasia Glawion, Ulrik Brandes

Corpus: Deutscher Novellenschatz (Heise & Kurz, Eds., 1871–1876)

Field of Study: German Literature, Statistics, Network Science

Institution: TU Darmstadt, ETH Zürich

Methods: Stylometry, Network Analysis, Text Analysis

Tools: Natural Language Processing, Statistics, Clustering, Questionnaire

Technology: R, visone, own implementations