Disruption and Societal Change: Exploring Knowledge Systems and Practices

René Dutschke[1] (rene.dutschke@tu-dresden.de) **// Jan Langenhorst**[1] (jan.langenhorst@tu-dresden.de) **// Ulrike Peifer**[1] (ulrike_marie.pfeifer1@tu-dresden.de) **// Sophia Seemann**[1] (sophia_marie.seemann@tu-dresden.de)
Authors listed in alphabetical order

[1]TUD Dresden University of Technology, Germany

Introduction

- "DiaDisK": interdisciplinary research project exploring knowledge systems and practices through a combination of linguistics and psychology, special interest: digital turn as a 'disruption' within central institutions of the so-called knowledge society
- dependence of language on cognition regarding epistemic concepts as representations of knowledge systems and practices, investigation of linguistic units as pairings of form and meaning (= cognitive units) → indirect study of cognition/behavior through language data
- comprehensive understanding of knowledge systems and practices through a combination of corpus-based evidence (L) and experimental data (P), exploration of 'knowledge' on an individual as well as on a collective level
- question of the cognitive/psychological reality of corpus data and analysis: "Corpora cannot stand in for experimental work. […] and corpus linguistics does not attempt to stand in for experimental work" (Arppe et al. 2010: 8-9)
- How can different types of evidence (L/P) be integrated and/or triangulated in order to gain a comprehensive understanding of knowledge systems and practices?

How can data from psychological experiments and word embedding models be compared to gain insights into knowledge systems and beliefs? What are the limitations of such a comparison?

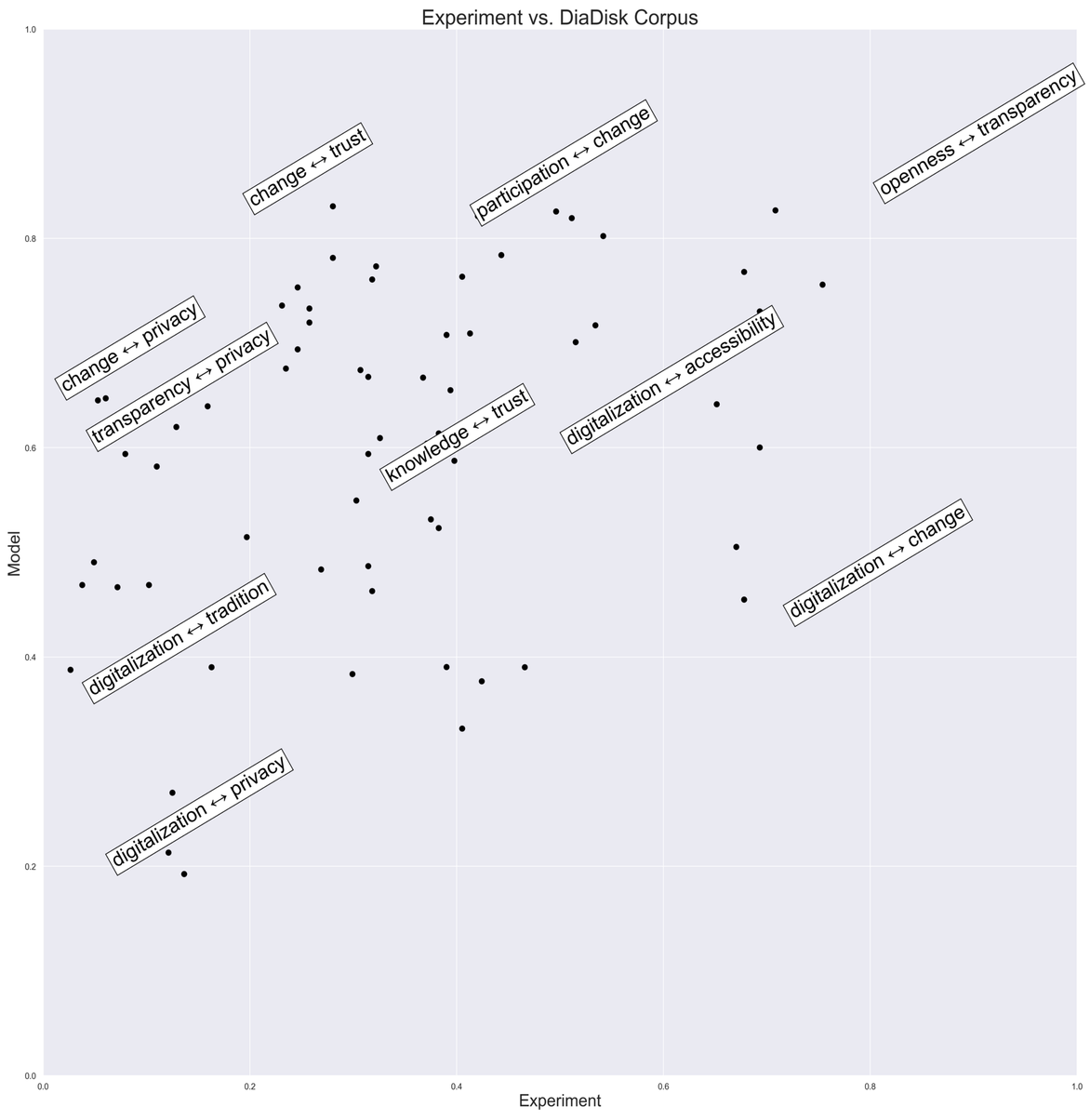
Methods & Data

- collection of written/spoken/digital texts from selected institutions (Saxon State and University Library Dresden & University School Dresden) under 'digital disruption', corpus of approx. 2 million tokens
- word embeddings: investigation of semantic relationships within text data (Almeida & Xexéo 2023), paradigmatic similarity of word pairs as indication of semantic similarity, calculation of cosine similarity values with range -1 to 1
- Trained model using the library *gensim* (Řehůřek & Sojka 2010) which implements the *Word2Vec* algorithm (Mikolov et al. 2013)

L

P

- experimental design based on theory of personal constructs/constructivism (Kelly 1955): individuals construct and interpret the world around them → focus on subjective reality
- personal constructs as a set of reference axes used by a person to make sense of the world; each relevant situation, person, and concept (= *event*) is evaluated with regards to the construct system of an individual
- triad test: judgement of similarity/difference for combinations of three events (*Which two are similar and/or different to the third?)* as a proxy for the structure of an underlying construct system → basis for calculation of individual similarity scores
- sample: N=24 (pilot study), criteria for participants: students and researchers in psychology, who are familiar with Open Science methods



Analysis

- comparison of similarity scores from the psychological experiment based on individual ratings (x-axis) and cosine similarity scores obtained from word embeddings model based on word occurrences in the corpus (y-axis)
- analysis of 78 pairs of words representing epistemic concepts related to knowledge systems/practices and digital disruption
- weak correlation of similarity scores from word embeddings model with the psychological experiment ($r = 0.36$, $p = 0.00125$, 95% CI [0.15, 0.54])
- plot (left) shows selected comparisons of word similarity

Discussion

- potential of visualizations: facilitating accessibility and exploration of data and hence interdisciplinary discussion → reflexion of commonalities/differences, generating working hypothesis
- prospect for further investigations: in-depth semantic analysis of word pairs, that differ significantly with respect to their similarity scores (including collocation profiles, other types of evidence and statistical measures)
- limitations: size of data samples, unequal distribution of data (institutions, text type, time), number/specific background of participants, etc.
- "[D]ifferent methods and types of evidence rarely yield exactly the same results. Nevertheless, even these differences help us gain a more accurate understanding of the linguistic phenomena studied" (Arppe et al. 2010: 6)

References

Almeida, Felipe & Geraldo Xexéo. 2023. *Word Embeddings: A Survey*. arXiv. http://arxiv.org/abs/1901.09069
Arppe, Antti, Gaëtanelle Gilquin, Dylan Glynn, Martin Hilpert & Arne Zeschel. 2010. Cognitive Corpus Linguistics: Five points of debate on current theory and methodology. *Corpora* 5(1). 1–27. https://doi.org/10.3366/cor.2010.0001
Kelly, George Alexander. 1955. *Personal construct psychology*. New York: Norton.
Mikolov, Tomas, Kai Chen, Greg Corrado & Jeffrey Dean. 2013. *Efficient Estimation of Word Representations in Vector Space*. arXiv. http://arxiv.org/abs/1301.3781.
Řehůřek, Radim & Petr Sojka. 2010. *Software Framework for Topic Modelling with Large Corpora.* https://doi.org/10.13140/2.1.2393.1847

Mitglied im Netzwerk von: DRESDEN concept SCIENCE AND INNOVATION CAMPUS

GEFÖRDERT VOM Bundesministerium für Bildung und Forschung

SACHSEN Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des vom Sächsischen Landtag beschlossenen Haushaltes.

TUD:SC Disruption and Societal Change