

Was heißt eigentlich ‚offen‘?

Eine korpuslinguistische Untersuchung am Beispiel des bibliothekarischen Diskurses der SLUB Dresden

Simon Meier-Vieracker¹ (simon.meier-vieracker@tu-dresden.de) // Lucie Weigelt (lucie.weigelt@tu-dresden.de) // Ulrike Pfeifer¹ (ulrike_marie.pfeifer1@tu-dresden.de) // René Dutschke¹ (rene.dutschke@tu-dresden.de) // Alexander Lasch¹ (alexander.lasch@tu-dresden.de) // Stefan Scherbaum¹ (stefan.scherbaum@tu-dresden.de) // Sophia Seemann¹ (sophia_marie.seemann@tu-dresden.de)

¹ TU Dresden, Deutschland

Hintergrund

- Ausgangsfrage: Disruptive Auswirkungen der Digitalisierung für zentrale Institutionen der Wissensgesellschaft und damit einhergehender Paradigmenwechsel hin zur Offenheit von Wissenschaft
- Open Science oft optimistisch, vielleicht sogar utopisch-überhöhend gerahmt
- Linguistisches Arbeitspaket nimmt Open-Science-Aktivitäten der SLUB in den Blick und untersucht die diskursive Verhandlung des Begriffs der Offenheit

Ziele

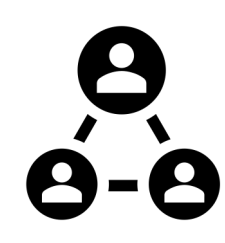
- Schärfung des semantisch oft vagen Begriffs der Offenheit bzw. Openness, um differenzierte Debatten über Open Science jenseits strikter Pro-Kontra-Positionierungen zu fördern
- Herausarbeitung der komplexen diskurssemantischen Profilierungen des oft schlagwortartig verwendeten Begriffs

Erste Ergebnisse

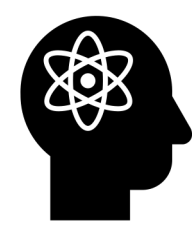
Das Begriffsfeld der Offenheit changiert systematisch zwischen drei Lesarten:



Technisch: digitale Daten und ihre Zugänglichkeit (Kollokate von *offen*: Daten, Schnittstelle, Standards, Technologien)



Politisch: Partizipation und Inklusion (Kollokate von *offen*: Zugang, Kreativraum, Austausch, Gesellschaft, kooperativ)



Ethisch: Offenheit als epistemische Tugend (Kollokate von *offen*: Wissen, Neugier, Wert)

Daten

- Untersuchungskorpus aus Publikationen aus dem Umfeld der Sächsischen Landesbibliothek – Staats- und Universitätsbibliothek Dresden (SLUB)
- 1,59 Mio. Token, lemmatisiert, annotiert
- SLUB positioniert sich proaktiv als „Motor für offene Wissenschaft und Gesellschaft“ Bonte und Muschalek 2019)



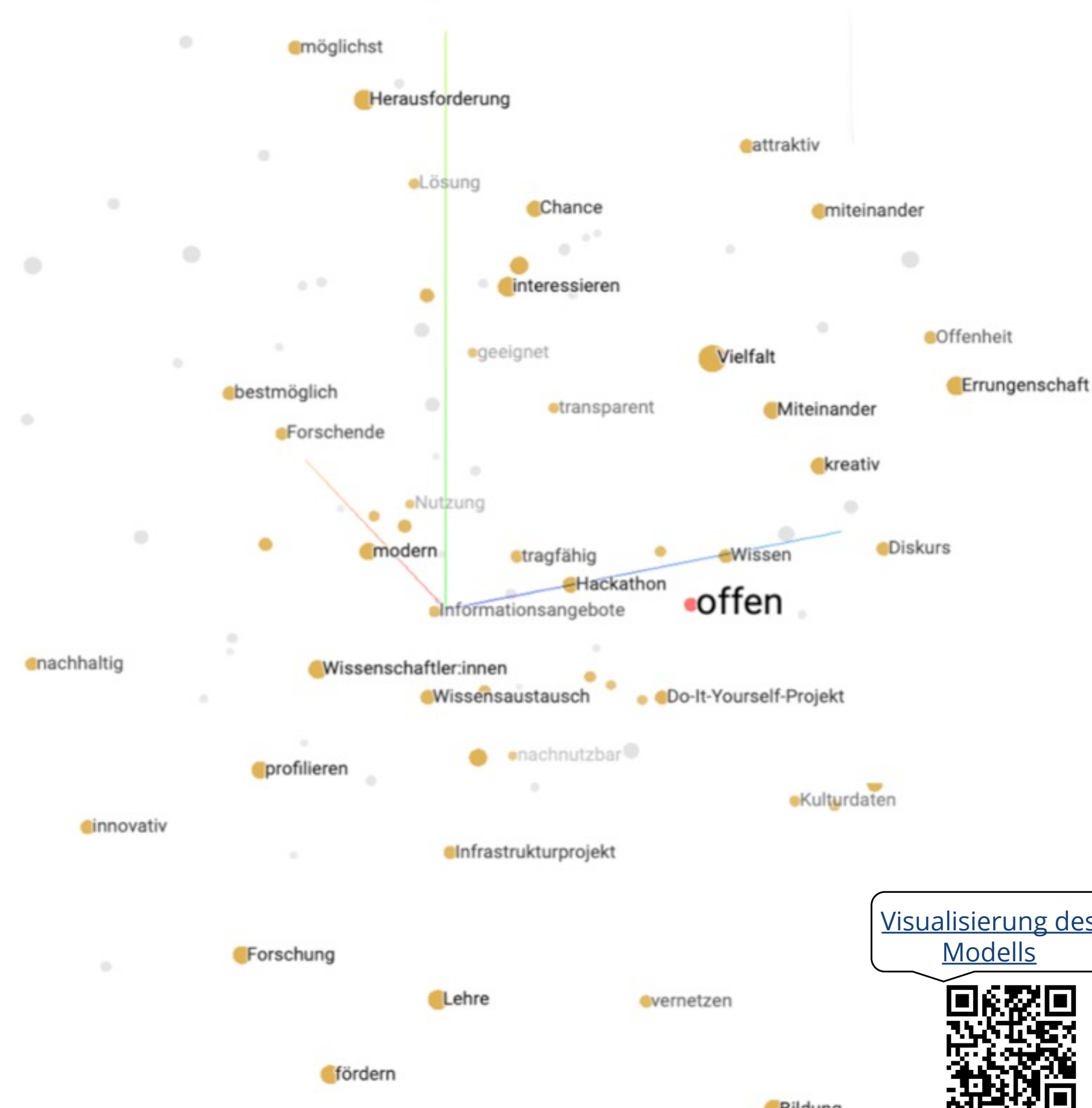
Methode

- Datengeleitete diskurssemantische Analysen wie die Berechnung von **Kollokationen** und **Ngrammen** (Bubenhöfer 2017) mit CQPweb (Hardie 2012)
- **Word Embeddings** als Verfahren der distributionellen Semantik, um Gebrauchsprofile einschlägiger Lexeme zu berechnen mit Software word2vec (Mikolov u. a. 2013; Kozłowski, Taddy, und Evans 2019)
- qualitative **framesemantische Annotationen** in der kollaborativen Annotationsumgebung INCEPTION (Castilho u. a. 2018) und unter Rückgriff auf die lexikographische Ressource FrameNet (<https://framenet.icsi.berkeley.edu/>)
- Annotation der semantischen Valenzen spezifischer Ausdrücke und der durch sie evozierten Frames machen die unterschiedlichen Lesarten etwa von *offen* präzise erfassbar

38 Bibliotheken sind konsequent offen für alle, niedrigschwellig und nichtkommerziell, sie verkaufen keine Daten, achten auf die Seriosität der vermittelten Informationen und treten für offenen Austausch und offene Infrastrukturen ein.

Word Embeddings

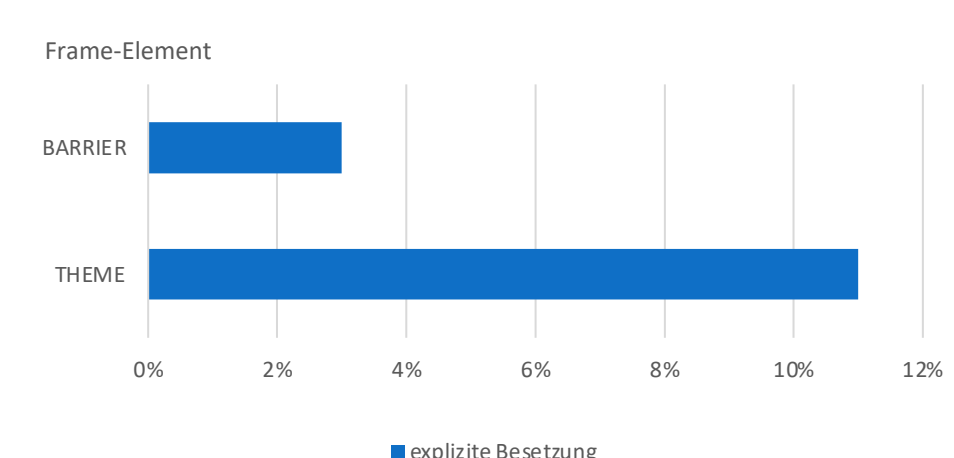
- Wörter mit semantischer Ähnlichkeit zu *offen* (sog. Nearest Neighbors) sind positiv konnotierte Lexeme wie *nachhaltig*, *transparent*, *interdisziplinär*, *nachnutzbar*, *kreativ*, *vernetzt*, *modern* und *innovativ* (Abb. rechts)



Typische Ngramme

- *offene und freie Wissensgesellschaft* oder *Offenheit, intellektuelle Freiheit und Redlichkeit*

- Prinzip Offenheit wird so als normative Zielvorgabe gerahmt



Framesemantischen Annotation

- Frame ‚Openness‘: Frame-Elemente **THEME** (für wen ist etwas offen?) und **BARRIER** (was verhindert potenziell den Zugang?) nur in 11% bzw. 3% der insgesamt 72 Fälle explizit besetzt (Abb. links).
- Vagheit des Begriffs der Offenheit bedingt seine vielfältigen Assoziationspotenziale

Ausblick

- Das für die Digital Humanities äußerst bedeutsame Prinzip der Offenheit kann in Hinblick auf seine diskursive Dimension empirisch gestützt präzisiert werden.
- Die linguistischen Analysen werden im Projekt mit der psychologischen, experimentellen Erhebung individueller Konstruktsysteme (Kelly 2005) von Akteur:innen im Feld der Open Science trianguliert.
- Die Datengrundlage wird momentan um Publikationen weiterer Akteur:innen im Bibliotheksdiskurs erweitert.

Bibliographie

- Bonte, Achim und Antonie Muschalek, Hrsg. 2019. *SLUB 2025. Wissen teilen - Menschen verbinden. Strategie der Sächsischen Landesbibliothek – Staats- und Universitätsbibliothek Dresden*. Dresden: SLUB. <https://nbn-resolving.org/urn:nbn:de:bsz:14-qucosa2-357501>.
- Bubenhöfer, Noah. 2017. „Kollokationen, n-Gramme, Mehrworteinheiten“. In *Handbuch Sprache in Politik und Gesellschaft*, hg. von Kersten Sven Roth, Martin Wengeler und Alexander Ziem, 69–93. Berlin, Boston: De Gruyter. <https://doi.org/10.1515/9783110296310-004>.
- Busse, Dietrich. 2012. „Frame-Semantik. Ein Kompendium“. Berlin u.a.: De Gruyter.
- Castilho, Richard Eckart de, Jan-Christoph Klie, Naveen Kumar, Beto Boulosa und Iryna Gurevych. 2018. „INCEPTION - Corpus-based Data Science from Scratch“. In *Digital Infrastructures for Research (DI4R)* 2018. <http://tubiblio.ulb.tu-darmstadt.de/106982/>.
- Hardie, Andrew. 2012. „CQPweb — Combining Power, Flexibility and Usability in a Corpus Analysis Tool“. *International Journal of Corpus Linguistics* 17 (3): 380–409. <https://doi.org/10.1075/ijcl.17.3.04har>.
- Kelly, George A. 2005. „A Brief Introduction to Personal Construct Theory“. In *International Handbook of Personal Construct Psychology*, herausgegeben von Fay Fransella, 3–20. Chichester, UK: Wiley. <https://doi.org/10.1002/0470013370.ch1>.
- Kozłowski, Austin C., Matt Taddy und James A. Evans. 2019. „The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings“. *American Sociological Review* 84 (5): 905–49. <https://doi.org/10.1177/0003122419877135>.
- Mikolov, Tomas, Kai Chen, Greg Corrado und Jeffrey Dean. 2013. „Efficient Estimation of Word Representations in Vector Space“. arXiv:1301.3781 [cs], Januar. <http://arxiv.org/abs/1301.3781>.
- Tkacz, Nathaniel. 2012. „From Open Source to Open Government: A Critique of Open Politics“. In *ephemera* 12(4): 386–405.
- Ziem, Alexander. 2020. „Wortbedeutungen als Frames: ein Rahmenmodell zur Analyse lexikalischer Bedeutungen“. In *Semantiktheorien II: Analysen von Wort- und Satzbedeutungen im Vergleich*, hg. von Jörg Hagemann und Sven Staffeldt, 27–56. Tübingen: Stauffenburg.