

Was heißt eigentlich ‚offen‘? Eine korpuslinguistische Untersuchung am Beispiel des bibliothekarischen Diskurses der SLUB Dresden

Meier-Vieracker, Simon

simon.meier-vieracker@tu-dresden.de
TU Dresden, Deutschland

Weigelt, Lucie

lucie.weigelt@tu-dresden.de
TU Dresden, Deutschland

Dutschke, René

rene.dutschke@tu-dresden.de
TU Dresden, Deutschland

Lasch, Alexander

alexander.lasch@tu-dresden.de
TU Dresden, Deutschland

Scherbaum, Stefan

stefan.scherbaum@tu-dresden.de
TU Dresden, Deutschland

Seemann, Sophia

sophia_marie.seemann@tu-dresden.de
TU Dresden, Deutschland

Pfeifer, Ulrike

ulrike_marie.pfeifer1@tu-dresden.de
TU Dresden, Deutschland

Hintergrund

Forschung zum Thema Open Science fokussiert typischerweise technische oder auch regulatorische und wissenschaftspolitische Aspekte (Vicente-Saez und Martinez-Fuentes 2018). Der Begriff der Openness bzw. der Offenheit selbst, der im Diskurs um digitale Forschung als zentrales Schlagwort fungiert, wird dabei eher

selten in seiner Semantik reflektiert. Dabei führt der Begriff der Offenheit reichhaltige, auch alltagssprachlich verankerte Assoziationspotenziale mit sich, aus denen sich die ausgesprochen optimistische, vielleicht sogar utopisch-überhöhende Rahmung von Open Science (Dinkel und Franzen 2015; Tkacz 2012) und Open Humanities maßgeblich speist.

Im interdisziplinären Projekt „Digitalisierung als Disruption von Wissenssystemen – Opening Knowledge“ (DiaDisk; Laufzeit 10/2021 – 03/2025) im Rahmen des EXU-Verbundes „Disruption and Societal Change“ an der TU Dresden fragen wir nach den disruptiven Auswirkungen der Digitalisierung in den für die Wissensgesellschaft zentralen Institutionen Universität, Bibliothek und Schule. Wir gehen davon aus, dass sich gerade im Prinzip der Offenheit und seiner diskursiven Verhandlungen Deutungsmuster manifestieren, welche die Disruptionen, die digitale Technologien für die traditionellen Routinen in Bildung und Wissenschaft mit sich bringen, positiv als Chancen, Transformationen usw. rahmen (Koch, Nanz, und Pause 2016, 19). Diese Deutungsmuster nehmen wir in einem der linguistischen Arbeitspakete des Projektes, das sich den Open-Science-Aktivitäten der Sächsischen Landesbibliothek – Staats- und Universitätsbibliothek Dresden (SLUB) widmet, aus einer korpuslinguistischen Perspektive in den Blick. Wir untersuchen, welche komplexen diskurssemantischen Profilierungen des Begriffs der Offenheit sich in diesem konkreten Diskurs nachweisen lassen, die seinen Schlagwortcharakter (Schröter 2011) grundieren. Damit soll eine Schärfung des Begriffs ermöglicht werden, die es erlaubt, die oft strikt nach Pro und Kontra geführten Debatten über Openness differenzierter zu führen, nicht zuletzt in wissenschaftspolitischen Kontexten.

Daten und Methoden

Grundlage unserer Analysen bilden verschiedene Publikationen (Forschungsbeiträge, Geschäftsberichte, aber auch Blogbeiträge und Tweets) sowie Dokumente (Drittmittelanträge und Strategiepapiere) aus dem Umfeld der SLUB, die dezidiert aus der Perspektive der Institution formuliert sind. Wie prominent im Strategiepapier SLUB 2025 festgehalten ist, positioniert sich die SLUB proaktiv in der deutschsprachigen Bibliotheks- und Wissenschaftslandschaft als „Motor für offene Wissenschaft und Gesellschaft“ (Bonte und Muschalek 2019). Sie nutzt in ihren Selbstdarstellungen strategisch die vielfältigen, zumeist positiven Konnotationen des Begriffs(feldes) der Offenheit, weshalb sich diese für unsere Fragestellung besonders gut eignen.

Das laufend zu erweiternde Korpus (1,59 Millionen Tokens, Stand Juli 2022) stellen wir im Projekt in morphosyntaktisch annotierter Form über verschiedene digitale Analyseumgebungen wie die Korpusanalyseplattform CQPweb (Hardie 2012) und die kollaborative Annotationsumgebung INCEpTION (Castilho u. a. 2018) zur Verfügung. Für die Analyse nutzen wir zum einen korpuslinguistische Verfahren der datengeleiteten diskurssemantischen Analyse wie Kollokations- und Ngram-Analysen (Bubenhof 2017), aber auch Word Embeddings als Verfahren der distributionellen Semantik mit der Software word2vec (Mikolov u. a. 2013; Kozłowski,

Taddy, und Evans 2019), um die Gebrauchsprofile einschlägiger Lexeme zu erschließen. Zum anderen wählen wir mit der Frame-Semantik (Ziem 2020) einen stärker theoriegeleiteten Ansatz. Dabei handelt es sich um eine semantische Theorie, die die Bedeutung von sprachlichen Ausdrücken in Bezug auf das in untereinander vernetzten Frames organisierte Weltwissen der Sprechenden beschreibt. Frames sind dabei schematische Repräsentationen von Situationen oder Konstellationen, vor deren Hintergrund dann sprachliche Ausdrücke verstanden werden (Busse 2012). Ausgehend von der lexikographischen Ressource FrameNet (<https://framenet.icsi.berkeley.edu/>) annotieren wir im Textmaterial die semantischen Valenzen einschlägiger Ausdrücke und der durch sie evozierten Frames, so dass die unterschiedlichen Lesarten etwa von *offen* und die assoziierten Formulierungsmuster präzise erfasst werden können.

Erste Ergebnisse

Erste Ergebnisse zeigen, dass im untersuchten Diskurs das Begriffsfeld der Offenheit systematisch zwischen einer eher technisch auf digitale Daten und ihre Zugänglichkeit bezogenen Lesart, einer auch politisch aufgeladenen, auf Partizipation und Inklusion abzielenden Lesart sowie einer auf Offenheit als epistemische Tugend abzielenden Lesart changiert. Typische Kollokate (Assoziationsmaß Log Dice) von *offen* sind etwa *Daten*, *Schnittstelle* und *Standards* im Sinne der technischen Lesart, *Kreativraum* und *Austausch* im Sinne der partizipativen Lesart, aber auch Abstrakta wie *Wissen* und *Neugier*. Paarformeln und Aufzählungen wie *offene und freie Wissensgesellschaft* oder *Offenheit, intellektuelle Freiheit und Redlichkeit* zeigen, dass das Begriffsfeld der Offenheit mit anderen Schlagwörtern in Interaktion gebracht und so zusätzlich semantisch als ethisch gehaltvolle Zielnorm aufgeladen wird. Auch die Berechnung von Word Embeddings weist in diese Richtung. Als sog. Nearest Neighbors von *offen*, die im untersuchten Diskurs also semantische Ähnlichkeit aufweisen, werden *nachhaltig*, *transparent*, *interdisziplinär*, *nachnutzbar*, aber auch *kreativ*, *vernetzt*, *modern* und *innovativ* ausgegeben (Abb. 1); zu *Offenheit* werden *Transparenz*, *Verbreitung*, *barrierefrei*, aber auch *demokratisch* ausgegeben (die Visualisierung des Modells ist unter <https://kurzelinks.de/dia-disk> öffentlich zugänglich; auch auf Teile der Korpora kann auf Anfrage Zugriff gewährt werden).

②

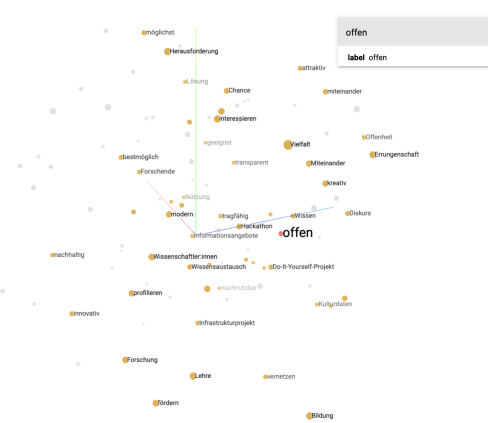


Abb. 1

Für die framesemantische Annotation bietet sich u.a. der Frame Openness an, der auf das Moment der Zugänglichkeit abzielt und sowohl wörtliche als auch metaphorische Verwendungen zulässt. Dabei zeigt sich in einer Pilotierung am Beispiel der Forschungsbeiträge im Korpus (51 Texte), dass die Frame-Elemente THEME (für wen ist etwas offen?) und BARRIER (was verhindert potentiell den Zugang?) nur in 11% bzw. 3% der insgesamt 72 Fälle explizit besetzt werden, so dass der Begriff der Offenheit vage gehalten und so seine freien Assoziationspotenziale besonders gut entfalten kann (Abb. 2).

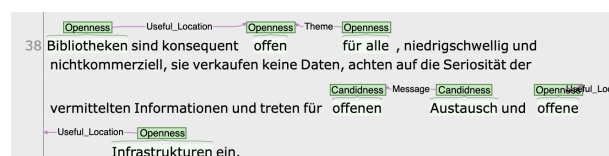


Abb. 2

Ausblick

Unsere Ergebnisse versprechen einen detaillierten und empirisch gestützten Blick auf das für die Digital Humanities so bedeutsame Prinzip der Offenheit weniger in seinen technischen als in seinen diskursiven Aspekten. Im Projekt werden unsere linguistischen Analysen ergänzt werden durch psychologische, experimentelle Erhebungen individueller Konstruktsysteme (Kelly 2005) von Akteur:innen im Feld der Open Science, die wir mit unseren Befunden zu diskursiven Deutungsmustern triangulieren werden. Zudem werden wir die Korpusgrundlage auf Publikationen anderer Akteur:innen im bibliothekarischen Diskurs ausweiten.

Bibliographie

Bonte, Achim und Antonie Muschalek, Hrsg. 2019. *SLUB 2025. Wissen teilen - Menschen verbinden. Strategie der Sächsischen Landesbibliothek - Staats- und Uni-*

versitätsbibliothek Dresden . Dresden: SLUB. <https://nbn-resolving.org/urn:nbn:de:bsz:14-qucosa2-357501>.

Bubenhof, 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 Boullosa 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/>.

Dickel, Sascha und Martina Franzen . 2015. „Digitale Inklusion: Zur Sozialen Öffnung Des Wissenschaftssystems“. *Zeitschrift für Soziologie* 44 (5): 330–47. <https://doi.org/10.1515/zfsoz-2015-0503>.

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>.

Koch, Lars, Tobias Nanz und Johannes Pause . 2016. „Imaginationen der Störung: ein Konzept“. *Behemot* 9 (1): 6–23. <https://doi.org/10.6094/BEHEMOTH.2016.9.1.885>.

Kozlowski, 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>.

Schröter, Melani . 2011. „Schlagwörter im politischen Diskurs“. *Mitteilungen des Deutschen Germanistenverbandes* 58 (3): 249–57. <https://doi.org/10.14220/mdge.2011.58.3.249>.

Tkacz, Nathaniel. 2012. „From Open Source to Open Government: A Critique of Open Politics“. In *ephemera* 12(4): 386–405.

Vicente-Saez, Ruben und Clara Martinez-Fuentes . 2018. „Open Science Now: A Systematic Literature Review for an Integrated Definition“. *Journal of Business Research* 88 (Juli): 428–36. <https://doi.org/10.1016/j.jbusres.2017.12.043>.

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.