

Heidelberger Akademie der Wissenschaften

Jahrbuch 2018



HEIDELBERG 2019

8. CAL²Lab – Erkundung der Rechtssprache in einer computergestützten Forschungsumgebung

Kollegiaten: Prof. Dr. Friedemann Vogel¹, Dr. Dr. Hanjo Hamann²

Mitarbeiterinnen: Isabelle Gauer³, Yinchun Bai³

¹ Lehrstuhl für Germanistik: Sozio- und Diskurslinguistik, Universität Siegen

² Max-Planck-Institut zur Erforschung von Gemeinschaftsgütern, Bonn, und Fellow des Jungen ZiF (Zentrum für interdisziplinäre Forschung), Bielefeld

³ Institut für Medienkulturwissenschaft, Albert-Ludwigs-Universität Freiburg

Hintergrund

Das Kollegprojekt „Juristisches Referenzkorpus (JuReko)“ vereint neuere Erkenntnisinteressen und Methoden zweier Disziplinen – der Rechts- und Sprachwissenschaft –, um Wege zur Entwicklung einer computer- und korpusgestützten Rechtslinguistik zu finden und die interdisziplinären Möglichkeiten und Grenzen einer „Ausmessung“ juristischer Diskurse auszuloten. Die Begegnung einer evidenzbasierten Jurisprudenz mit der empirischen Computerlinguistik spürt „Diskurs sedimenten“ in großen Textmengen nach und soll dadurch die Entstehung und Entwicklung dogmatischer Lehren, wissenschaftlicher Schulen und sozialer Netzwerke untersuchen helfen. Damit schlägt das Projekt Brücken zwischen neuerer empirischer Rechtsforschung auf der einen und computergestützter und rechtsmethodisch geschulter Korpuslinguistik auf der anderen Seite. Dieser Brückenschlag erfolgt zugleich zwischen den Kulturen, indem besonderer Wert auf die Vernetzung mit international führenden Forschern zum Thema „Sprache und Recht“ gelegt wird.

Projektstand

Im vierten Jahr des Kollegprojekts wurde das in den Vorjahresberichten erläuterte Referenzkorpus um eine Benutzeroberfläche erweitert, die mittelfristig als öffentliche Forschungsinfrastruktur für interessierte Rechtslinguist(inn)en zur Verfügung stehen soll. Dazu wird derzeit eine Online-Plattform entwickelt, die einen selektiven und geschützten Zugriff auf das Referenzkorpus verknüpft mit dem Angebot umfassender Metriken zur juristischen Sprache und Begriffsbildung. Im Fokus steht die kontextuelle Bestimmtheit bzw. Unbestimmtheit juristischer Ausdrücke sowohl diachron (Begriffswandel über die Zeit) als auch synchron (Querschnitt durch juristische Schulen, Medien, Textsorten, Rechtsbereiche u. a.).

Konkret wurden die im Korpus nach Deduplikation enthaltenen knapp 1 Mrd. Wortformen computerlinguistisch so aufbereitet, dass 200.000 bedeutungstragende Hauptwörter (Substantiv-, Verb- und Adjektivlemmata, ohne Namen und Abkürzungen) ausgewählt wurden. Diese wurden in Verbindung mit ihrer

D. Förderung des wissenschaftlichen Nachwuchses

Wortartenzugehörigkeit gespeichert, wodurch zumindest eine grobe Disambiguierung nach Substantiv/Verb/Adjektiv vorgenommen wurde. Da die Zuordnung der Wortarten nicht rein computertechnisch lösbar ist, wurden alle Zweifelsfälle manuell geprüft. Für die gewonnenen Kombinationen aus Lemmata und Wortarten (= Listenlemmata) wurden sogenannte Kontextprofile errechnet, d. h. es wurden statistische Daten in Relation zu Vorkommen pro Jahr, Textsorte, Quartil, Gericht, Zeitschrift und Autor erhoben. Zusätzlich wurden Mehrworteinheiten mit bis zu 5 Wortformen und diejenigen Begleitwörter berechnet (Nachbarn mit einer Umgebung von ± 8), mit denen sich die Zielwörter am signifikant häufigsten umgeben, berechnet – auch wieder in Relation zu den vorher genannten Metadaten. Die Berechnung der Daten erfolgte durch Java-Programme, die unter anderem die Rechenleistung eines Computerclusters (High Performance Computing) nutzten. Die Ergebnisse wurden zur schnelleren Abfrage in einer relationalen Datenbank mit 104 Tabellen gespeichert.

Die daraus gewonnenen Daten sind über eine Onlineplattform (frei verfügbar ab vssl. Mitte 2019) durchsuchbar, wobei nicht nur mit dem Lemma, sondern auch mit einer konkreten Wortform gesucht werden kann. Zudem kann die Suche über die Metadaten eingeschränkt werden (Abb. 1). Damit ist beispielsweise eine Suche nach dem Vorkommen eines Wortes während einer bestimmten Zeitperiode in einer bestimmten Textsorte möglich.

Die statistischen Daten (Kontextprofile) werden dargestellt als ausklappbare Listen (Abb. 2), sortiert nach der relativen Häufigkeit im (Sub-) Korpus. Daneben werden bis zu 1.000 Textauszüge pro Suchwort in Form einer Key-Word-In-Context-Darstellung angezeigt, um dem Benutzer konkrete Verwendungsbeispiele der

Listenlemmata an die Hand zu geben (Abb. 3).

Während diese Arbeiten liefen, wurde Kollegiat Friedemann Vogel auf einen Lehrstuhl (W3) an der Universität Siegen berufen, wo er seit dem Frühjahr 2018 die Sozio- und Diskurslinguistik in Lehre und Forschung vertritt, mit besonderem Schwerpunkt auf Rechtslinguistik

Abb. 1: Suchmaske CAL²Lab

D. Förderung des wissenschaftlichen Nachwuchses

der WIN-Forschungsgruppe: So war sie am 7./8. September 2018 beim zweiten Jahrestreffen der International Language and Law Association (ILLA) in Kopenhagen – dem *Focus Workshop „Computers, Language, and Law: Spotlight on Blind Spots“* – gleich doppelt vertreten: Kollegiat Friedemann Vogel, seit 2018 auch Präsident der ILLA, diskutierte auf dem Eröffnungspodium, Kollegiat Hanjo Hamann als eingeladener Diskutant auf dem Abschlusspodium mit Fachkolleg(inn)en aus Dänemark und den USA über den Stand und die Perspektiven der Rechtslinguistik in Europa und der Welt.

Ausblick

Im angelaufenen fünften Projektjahr wird die Benutzeroberfläche weiterentwickelt und bis zur Freischaltung überarbeitet werden. Zudem wird das Kernkorpus, dessen Aufsatzsammlung bisher unterschiedlich aktuell und zum Teil noch auf 2012 datiert ist, umfassend aktualisiert. Die Anzahl enthaltener Aufsatztexte wird hierdurch von zuletzt 42.546 um knapp 16.000 Texte (37,6 %) anwachsen und dann einheitlich den Stichtag 31. Juli 2018 aufweisen.

Schließlich wird im fünften Projektjahr der bei Antragstellung vorgesehene Abschlussworkshop stattfinden, der unter dem Titel „Empirische Sprachgebrauchsermittlung im Recht: Methoden & Perspektiven“ wissenschaftlich tätige Korpuslinguisten mit Rechtspraktikern zusammenbringen soll, um den konkreten Erkenntnismehrwert der im Kollegprojekt entwickelten Methoden auszuloten. Als Teilnehmer konnten bisher neben diversen Professoren beider beteiligten Disziplinen auch ein Vorsitzender Richter am Bundesverwaltungsgericht, ein pensionierter Vorsitzender Richter am Bayerischen Obersten Landesgericht und mehrere Fachanwälte gewonnen werden. Diese Experten sollen im Rahmen des Abschlussworkshops die Analysen von Korpuslinguisten zu konkreten interdisziplinären Fragestellungen kommentieren und hierdurch ihre Praxisexpertise in die wissenschaftliche Erkenntnisgewinnung einfließen lassen. Die daraus zu erhoffende Abschlusspublikation wird dem Kollegprojekt als Ausweis des Erreichten und Grundlage für weitere Forschung dienen.

Projektrelevante Publikationen im Berichtszeitraum

- Hamann, Hanjo: Normativität und Gerechtigkeit als emergente Sprachphänomene. [Rezension zu] Claudia Woldt, Normativität als linguistische Schnittstelle, ARSP, Jg. 104 (2018), S. 136–141.
- Hamann, Hanjo und Daniel Hürlimann: Aktivisten setzen Datenbankschutz für BGBI auf den Prüfstand. Erstmals offene Gesetze für Deutschland?, LTO, Jg. 9 (2018), Nr. 32689, S. 1–4.
- Hamann, Hanjo und Friedemann, Vogel: Evidence-Based Jurisprudence meets Legal Linguistics – Unlikely Blends Made in Germany, *BYU L. Rev.*, Jg. 43 (2018), S. 1473–1501.

9. „Working Numbers“ (WIN-Programm)

Vogel, Friedemann, Hanjo Hamann und Isabelle Gauer: Computer-Assisted Legal Linguistics: Corpus Analysis as a New Tool for Legal Studies, *Law & Social Inquiry*, Jg. 43 (2018), S. 1340–1363.

9. „Working Numbers“: *Science and Contemporary Politics*

Kollegiat: PD Dr. Markus J. Prutsch¹

Mitarbeiter: Lars Lehmann

In Zusammenarbeit mit Dr. Georg von Graevenitz, Dr. Kathrine von Graevenitz, Dr. Kelly L. Grotke, Dr. Stephen W. Hastings-King

¹ Europäisches Parlament/Universität Heidelberg

1. Die Aktivitäten des Forschungsprojektes im Überblick

Das WIN-Projekt „Working Numbers“: *Science and Contemporary Politics* ging 2018 in sein zweites Jahr. Es schließt nahtlos an die Forschungstätigkeiten des im Jahr 2017 ausgelaufenen Projektes „Wissen(schaft), Zahl und Macht“ an und beleuchtet das spannungsreiche Verhältnis von Wissenschaft und Politik, wobei ein besonderer Schwerpunkt auf die Rolle von Zahlen und Quantifizierungen im politischen Betrieb gelegt wird.

Im Jahr 2018 konzentrierte sich die Arbeit im Projekt zum einen auf die Weiterverarbeitung und Druckvorbereitung der bislang gewonnenen Forschungsergebnisse. In diesem Zusammenhang wurde im Frühjahr 2018 ein finales Begutachtungsverfahren für die Resultate der insgesamt zwölf Teilprojekte durchgeführt. Im Zuge dessen erhielten alle beteiligten Wissenschaftler¹ abschließende Empfehlungen für die schriftliche Ausarbeitung ihrer jeweiligen Teilprojektergebnisse. Auf der Basis der finalen Beiträge begannen ab Jahresmitte 2018 sodann die editorischen Arbeiten an der Publikation „Science, Numbers and Politics“, die die Projektergebnisse in Buchform einem breiten Publikum zugänglich machen soll, und für die mit Palgrave Macmillan einer der renommiertesten internationalen Wissenschaftsverlage gewonnen werden konnte. Im intensiven Austausch mit dem Verlag konnten diese editorischen Arbeiten im November 2018 zum Abschluss gebracht werden, sodass die Voraussetzungen für die geplante Veröffentlichung des Bandes in der ersten Jahreshälfte 2019 erfüllt sind.

Zum anderen lag das Hauptaugenmerk im Jahr 2018 darauf, konkrete Schlussfolgerungen für einen angemessenen Gebrauch von Zahlen und Quan-

¹ Aus Gründen der leichteren Lesbarkeit wird im vorliegenden Text ausschließlich die männliche Sprachform verwendet. Diese sprachliche Vereinfachung ist als geschlechtsneutral zu verstehen und impliziert keine Benachteiligung des weiblichen Geschlechts.