

10. Computergestützte Rechtslinguistik (CAL²) – Das Gewirk von Sprache und Dogmatik des Rechts am Beispiel des JuReko-Referenzkorpus*

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

Mitarbeiter: Isabelle Gauer¹, Yinchun Bai¹, Julia Kanthak¹

¹ Institut für Medienkulturwissenschaft, Universität Freiburg

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

Hintergrund

Das Kollegprojekt „Juristisches Referenzkorpus (JuReko)“ widmet sich interdisziplinär den Möglichkeiten und Grenzen einer „Ausmessung“ juristischer Diskurse. Dabei werden neuere Erkenntnisinteressen und Methoden zweier Disziplinen – der Rechts- und Sprachwissenschaft – vereint und Wege zur Entwicklung einer computer- und korpusgestützten Rechtslinguistik besprochen. Die Begegnung einer evidenzbasierten Jurisprudenz mit der empirischen Computerlinguistik ermöglicht es, in großen Textmengen Diskurssedimente aufzuspüren und dadurch die Entstehung und Entwicklung dogmatischer Lehren, wissenschaftlicher Schulen und sozialer Netzwerke zu untersuchen. 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 und am Thema „Sprache und Recht“ interessierten Fachkollegen gelegt wird.

Projektstand

Die ersten beiden Phasen des Projekts sind abgeschlossen: Die im Rahmen der Korpuskonzeption ausgewählten Texte (6.300 Gesetzestexte, 370.000 Gerichtsentscheidungen, 43.000 wissenschaftliche Aufsätze) wurden in der anfangs entwickelten Datenbankstruktur erfasst und ihre Metadaten unter großem Arbeitsaufwand manuell nachkorrigiert, um eine hochvalide Datengrundlage zu erhalten. Die Volltexte der erfassten Dokumente wurden mittels der Software jTidy in ein TEI-P5-konformes xml-Format konvertiert, xsl-transformiert und mit Hilfe des TreeTaggers (Schmid 1994) annotiert, also computergestützt mit Zusatzinformationen zu Wortarten, Lemmata und Textstruktur angereichert.

* Früherer Titel des Projekts: „Vom corpus iuris zu den corpora iurum. Konzeption und Erschließung eines juristischen Referenzkorpus (JuReko)“

C. Förderung des wissenschaftlichen Nachwuchses

Diese computerlinguistischen Arbeiten wurden Ende 2015 abgeschlossen und im März 2016 auf der dritten Jahrestagung des Verbands „Digital Humanities im deutschsprachigen Raum e. V.“ in Leipzig der Forschungsöffentlichkeit vorgestellt. Das Kernkorpus zur empirischen Erforschung zeitgenössischer Rechtssprache, -kommunikation und -methodik steht damit in folgendem Umfang bereit:

Textsorte	Quelle	Zeitraum	Zahl Texte	Zahl Tokens	Datenmenge
Gesetze	Bund, BMJV	2015	6.366	19.460.327	325 MB
Rechtssprechung	BGH	1951–2015	40.176	58.181.286	940 MB
	BVerfG	1951–2014	4.474	10.766.647	179 MB
	AG/LG/OLG	1968–2015	85.229	225.340.131	3,883 GB
	andere	1882–2015	238.737	728.742.573	12,612 GB
Fachaufsätze	NJW/ZRP	1981–2012	9.334	50.045.433	808 MB
	MDR	1981–2015	2.688	11.949.257	189 MB
	JA/JuS	2000–2012	3.261	17.985.884	284 MB
	DÖV/NVwZ	1982–2012	4.585	28.632.615	467 MB
	ZStW/NStZ	1981–2013	2.371	15.580.589	249 MB
	ZGR/ZIP	2000–2014	729	7.716.853	124 MB
	AG/BB	1991–2014	10.126	58.734.088	961 MB
	andere	1984–2015	9.452	59.986.403	970 MB
Gesamt		1951–2015	375.381	1.293.122.086	21,991 GB

Die daran anschließende dritte Phase des Projekts (analytische Erschließung) läuft noch: Auf Basis semiautomatischer Verfahren wurden die erhobenen Texte analytisch erschlossen und dabei die Möglichkeiten und Grenzen einer rechtslinguistisch fundierten, computergestützten Korpusempirie für Sprach- und Rechtswissenschaft ausgelotet. Zur mitlaufenden epistemologischen Reflexion dieses Prozesses veranstaltete die Projektgruppe im März 2016 in Heidelberg eine internationale Tagung (dazu näher unten) und verfasste auf dieser Grundlage bis August 2016 ein gemeinsames Arbeitspapier unter dem Arbeitstitel „Computer Assisted Legal Linguistics: Corpus Analysis as a New Tool for Legal Studies“, das als Programmschrift für die weitere analytische Erschließung dienen soll. Es wurde nach positiver Peer-Review im Januar 2017 von der interdisziplinären Fachzeitschrift *Law & Social Inquiry (LSI)* zur Veröffentlichung angenommen.

Allmählich tritt das Projekt damit in seine vorerst letzte Phase (Evaluation und Dokumentation) ein: Die auf der Konferenz im März 2016 präsentierten Studien wurden inzwischen verschriftlicht und werden derzeit zur externen Begutachtung und Veröffentlichung im *International Journal of Language & Law (JLL)* vorbereitet, nachdem ein erstes Tagungsfazit bereits im Juli 2016 in der Open-Access-Zeitschrift *The Winnower* publiziert worden war. Desweiteren wurden die Projektergebnisse im Dezember 2016 auf der internationalen Rechtsinformatik-Konferenz

10. Computergestützte Rechtslinguistik (WIN-Programm)

(JURIX) in Nizza und im Februar 2017 auf der „Law & Corpus Linguistics“-Konferenz der Brigham Young University (BYU) in Utah vorgestellt. Das dafür verfasste Manuskript unter dem Arbeitstitel „Evidence-Based Jurisprudence and Legal Linguistics – Unlikely Blends Made in Germany“ wird im Laufe des Jahres 2017 in der US-amerikanischen Rechtszeitschrift *BYU Law Review* veröffentlicht. Weitere Möglichkeiten der Evaluation und Dokumentation werden derzeit erörtert, auch mit Blick auf eine etwaige Anschlussfinanzierung für die öffentlichkeitswirksame Weiterentwicklung des JuReko-Projekts.

Konferenzbericht: Tagung am 18./19. März 2016 in Heidelberg

Die Projektgruppe lud im März 2016 renommierte Fachleute der internationalen Rechtslinguistik zu einer Tagung über das „Gewirk von Sprache und Recht“ (*The Fabric of Language and Law*) ein. Dieser Einladung folgten etwa vierzig Wissenschaftler aus Deutschland, der Schweiz, Italien, Polen, Spanien und den USA – die meisten aus den Sprach- und Rechtswissenschaften sowie der Philosophie und der Informatik. Unter dem Motto der Tagung – „Muster entdecken durch Juristische Korpuslinguistik“ (Discovering Patterns through Legal Corpus Linguistics) – erörterten die Vortragenden miteinander und mit anderen Teilnehmern, wie Computerlinguistik systematische Muster im juristischen Diskurs aufdecken kann, die unser Verständnis von Recht, Gerechtigkeit und Rechtspolitik erweitern.

Das Programm der Tagung (online unter www.cal2.eu) bestand aus fünf zweistündigen Sitzungen mit jeweils zwei Vorträgen von je 30 bis 40 Minuten, gefolgt von 20- bis 30-minütiger Aussprache. Die Vorträge erstreckten sich über zwei Tage: In der ersten Sitzung vom 18. März 2016 (Patterns in Language and Law) führten der Gastredner Professor Dr. Lawrence M. Solan (Brooklyn) und Rechtsanwalt Dr. Stephen Mouritsen (New York) in das Potential und die Probleme der Analyse von Mustern in der Rechtssprache ein. Dem folgte eine zweite Sitzung, in der Dr. habil. Łucja Biel (Warschau) und Professor Dr. Stanisław Goźdz-Roszkowski (Łódź) über Muster in Diskursen und Übersetzungen referierten (Patterns in Discourses and Translations). Die dritte Sitzung beendete den ersten Konferenztag mit Reflexionen über computergestützte Methoden und Korpora (Computational Assistance and Corpora), vorgetragen von Dr. Stefan Höfler (Zürich) und Dr. Ruth Breeze (Navarra). Den zweiten Konferenztag, 19. März 2016, eröffneten Dr. María José Marín Pérez (Murcia) und Dr. Giulia Venturi (Pisa) mit einer Vorstellung ihrer jeweiligen korpuslinguistischen Projekte (Corpus Projects I), gefolgt von einer weiteren Sitzung (Corpus Projects II) mit beispielhaften Referenzprojekten von Professorin Dr. Rema Rossini Favretti (Bologna) und den Tagungsveranstaltern, die die Vortragsreihe mit ihrem Blick auf das deutsche juristische Referenzkorpus JuReko und einer Perspektive zur internationalen Vernetzung beendeten (The German Reference Corpus (JuReko) – Towards an International Research Network).

C. Förderung des wissenschaftlichen Nachwuchses

Abgerundet wurde die Tagung schließlich durch eine Podiumsdiskussion, moderiert von Professor Dr. Dieter Stein (Düsseldorf) als einem Gründungsmitglied der internationalen rechtslinguistischen Vereinigung, ILLA (International Language and Law Association). Neben ihm auf dem Podium saßen die Vortragsredner Larry Solan (ebenfalls ILLA-Mitgründer) und Łucja Biel, gemeinsam mit dem Rechtstheoretiker Professor Dr. Andreas Abegg (Winterthur/Zürich). Die Diskutanten fassten ihre jeweiligen Lehren aus der Tagung zusammen und erörterten Möglichkeiten, Recht durch Sprachforschung transparenter zu machen. Ein Protokoll der Diskussion, die wertvolle Impulse für die Zukunft der Rechts(korpus)-linguistik liefern könnte, wurde in der Open-Access-Zeitschrift *The Winnower* veröffentlicht, ein Tagungsbericht folgt Anfang 2017 in der Zeitschrift *Archiv für Rechts- und Sozialphilosophie (ARSP)*.

Perspektive: Vom Akademieprojekt zum internationalen Forschungsnetzwerk

Das um das JuReko herum aufgebaute internationale Forschungsnetzwerk (*International Research Group Computer Assisted Legal Linguistics, CAL²-Group, www.cal2.eu*) hat verschiedene rechtslinguistische Korpusprojekte zusammengeführt und Wissenschaftlerinnen und Wissenschaftler aus verschiedenen Teilen der Welt in Kontakt gebracht. Zudem ist das CAL²-Netzwerk mittlerweile maßgeblich beteiligt an der Neubelebung der internationalen rechtslinguistischen Vereinigung (ILLA) und der Ausrichtung ihres nächsten Fachkongresses im September 2017 in Freiburg, sowie am Fellowprojekt „Die Namen der Justiz“ der Wikimedia-Stiftung und des Stifterverbands für die Deutsche Wissenschaft. Diese weiterführenden Projekte wurden jeweils durch gesondert eingeworbene Drittmittel finanziert, wobei das JuReko-Projekt das verbindende Herzstück der wissenschaftlichen Projektarbeit darstellt.

Publikationen

Gauer, Isabelle, Hanjo Hamann und Friedemann Vögel: Das juristische Referenzkorpus (JuReko) – Computergestützte Rechtslinguistik als empirischer Beitrag zu Gesetzgebung und Justiz, in: Burr (Hrsg.), DHd 2016: Modellierung – Vernetzung – Visualisierung. Die Digital Humanities als fächerübergreifendes Forschungsparadigma, Leipzig 2016, S. 129–131 (online unter www.t1p.de/01aq).

Hamann, Hanjo, Friedemann Vögel und Isabelle Gauer: Computer Assisted Legal Linguistics (CAL²), in: Bex/Villata (Hrsg.), Legal Knowledge and Information Systems. JURIX 2016: The Twenty-Ninth Annual Conference. Amsterdam 2016, S. 195–198.

Hamann, Hanjo, Friedemann Vögel und Isabelle Gauer: The Fabric of Language and Law. Discovering Patterns Through Legal Corpus Linguistics. Special Issue of the International Journal of Language & Law (JLL), 2017 (im Erscheinen).

Vögel, Friedemann, Calculating legal meanings? Drawbacks and opportunities of corpus assisted legal linguistics to make the law (more) explicit. In: Dieter Stein und Janet Giltrow

11. Die Vermessung der Welt (WIN-Programm)

- (Hg.): *The pragmatic turn in law. Inference and Interpretation* (Mouton Series in Pragmatics), New York/Boston im Erscheinen 2017 (online unter www.tlp.de/3f4u)
- Vögel, Friedemann, Hanjo Hamann und Isabelle Gauer: *Computer Assisted Legal Linguistics: Corpus Analysis as a New Tool for Legal Studies*, in: *Law & Social Inquiry* (im Erscheinen).
- Vögel, Friedemann, Hanjo Hamann, Dieter Stein, Andreas Abegg, Łucja Biel und Lawrence M. Solan, „Begin at the beginning“. *Lawyers and Linguists Together in Wonderland*, in: *The Winner* Bd. 3 (2016), Beitrag 4.919, S. 1–9 (online unter www.tlp.de/afdy).
- Vögel, Friedemann, Isabelle Gauer, Hanjo Hamann: *Juristische Semantik messend verstehen. JUSLAB – Eine computergestützte Forschungs- und Experimentierplattform als Beitrag zu einer datengestützten Rechtslinguistik*. In: Friedemann Vogel (Hg.): *Recht ist kein Text: Studien zur Sprachlosigkeit im verfassten Rechtsstaat*. Berlin 2017 (im Erscheinen)

11. Die Vermessung der Welt: Religiöse Deutung und empirische Quantifizierung im mittelalterlichen Europa

Kollegiat: Dr. Christoph Mauntel¹

Mitarbeiterin: Elena Ziegler

¹ Graduiertenkolleg „Religiöses Wissen im vormodernen Europa (800–1800)“, Universität Tübingen

Das Projekt „Die Vermessung der Welt“ untersucht, welche Rolle dem Messen und Zählen als Beschreibungsmethode und Erklärungsmodell bei der Erfassung der Welt im mittelalterlichen Europa zukam. Der letzte Projektbericht endete an dieser Stelle im Jahrbuch 2015 mit dem Hinweis, dass erste Erklärungsansätze für die Präsenz quantifizierender Angaben zur Beschreibung der Welt an einem größeren Korpus hinterfragt werden sollten, „ohne dabei das Spannungsverhältnis zwischen vermeintlicher Präzision und übergeordneter Symbolik aus dem Blick zu verlieren, in dem derartige Zahlenangaben stehen können.“ Der methodische Ansatz scheint mir nach wie vor richtig, es stellt sich jedoch die Frage, ob das angesprochene Spannungsverhältnis nicht vielleicht nur ein scheinbares, oder zumindest situatives ist. Zwar lassen sich in den bisher analysierten Quellen (in diesem Fall: spätmittelalterliche Reiseberichte) ohne Frage sowohl ein auf die Abbildung der Wirklichkeit zielender Gebrauch von numerischen Angaben nachweisen als auch ein symbolischer. Die Frage jedoch ist, inwiefern diese beiden Varianten im Gegensatz zueinander stehen oder nicht schlicht zwei verwandte Möglichkeiten der Welterfassung sind.

Die Meistererzählung der Moderne, nach der davon ausgegangen wird, dass die Welt nur mithilfe rationaler Beobachtung erfasst werden kann, und die diesen Erkenntnisweg strikt von spirituellen Zugängen trennt, hat hier bis heute ganze Arbeit geleistet: Religion und Rationalität, Glaube und Wissen erscheinen so als