

Kommentiertes Digitales Korpus Deutscher Rechts- und Gesetzestexte

- D-LEX -

Florian Kuhn

Universität Potsdam

Institut für Linguistik

AB Angewandte Computerlinguistik

fkuhn@uni-potsdam.de

Dankmar W. Enke

Universität Tübingen

Seminar für Sprachwissenschaft

AB Quantitative Linguistik

dankmar.enke@uni-tuebingen.de

Antje Baumann

Bundesministerium der Justiz

Referat Rechtsprüfung, Sprachberatung,

Allgemeines Verwaltungsrecht

baumann-an@bmj.bund.de

&

Humboldt-Universität zu Berlin

Institut für deutsche Sprache und Linguistik

Universität Passau, 1. Jahrestagung der
Digital Humanities im deutschsprachigen Raum (DHd), 25. - 28.03.2014

1 Einführung

Die Arbeit am und mit dem Recht hat seit jeher nicht nur Juristen, sondern auch Akteure anderer Fachbereiche beschäftigt. Allerdings zeigt sich mit Blick auf die vergangenen Jahrzehnte, dass ein Austausch zwischen diesen Disziplinen kaum stattgefunden hat. Dabei nimmt die Rechtstheorie eine Sonderstellung ein; sie vermochte es Ende der sechziger und Anfang der siebziger Jahre des letzten Jahrhunderts, Impulse aus Sprachphilosophie und Soziologie in juristische Fragestellungen einzubeziehen.

Besonders Juristen und Sprachwissenschaftler pflegen seit Beginn der achtziger Jahre interdisziplinären Austausch im Rahmen von Arbeitsgruppen zur Juristischen Linguistik, gemeinsamen Veröffentlichungen, bei der Einbindung von Linguisten in den Gesetzgebungsprozess sowie in gemeinsamen Studiengängen.

Vor dem Hintergrund unseres digitalen Zeitalters zeichnet sich auch in der Rechtstheorie eine neue Rezeption ab, denn das Interesse an medientheoretischen Konzepten wächst. Rechtslinguisten können und sollen dieses Interesse ihrer Juristischen Kollegen aufgreifen. Auf Grundlage computerlinguistischer und informatischer Anwendungen können sie gemeinsam zu einer empirischen und methodologischen Fundierung beitragen. Diese soll in ein kommentiertes digitales Korpus deutscher Rechts- und Gesetzestexte münden, das wir in unserem Vortrag präsentieren.

2 Sprachprüfung im Deutschen Bundestag

Da die Regelungsinhalte deutscher Gesetze und deren Sprache oft unverständlich sind, führte die Kritik daran zu einer Neuorganisation der Sprachprüfung als Teil der Rechtsprüfung. Seit 2009 prüft der Redaktionsstab Rechtssprache beim Bundesministerium der Justiz alle Gesetz- und Verordnungsentwürfe der Bundesministerien auf sprachliche Richtigkeit und Verständlichkeit. Dass Linguisten einen fremden Blick auf entstehendes Recht haben und andere Fragen stellen, ist gewollt – und verlangt von juristischen und anderen Fachleuten eine neue Art des Sprechens über Texte. Die Sprachprüfung von Gesetzentwürfen verstärkt die (zum Schaffen von neuem Recht prinzipiell nötige) Rechtsvergleichung – wozu die Sprachberater aber auch neuartige Arbeitsmittel benötigen: Juristische Datenbanken sind jedoch bisher weder frei zugänglich und nicht für sprachliche Fragen ausgelegt, die Ergebnisse derartiger Anfragen fallen also zurzeit noch unbefriedigend aus. Gebraucht werden Aussagen über den Gebrauch von Wörtern oder Wortverbindungen (Häufigkeit, Kollokationen, Kontext, Textsorte etc.); dieser pragmatische Aspekt der Sprachbetrachtung fehlt weitgehend in Wörterbüchern und derzeitigen Terminologie-Datenbanken. Aus diesem Zusammenhang erwächst das Bedürfnis nach korpuspragmatischen Methoden und deren Anwendung in digitalisierter Form, um in der Zusammenarbeit mit der juristischen Fachebene möglichst gute Argumente einzusetzen. Ziel bleibt hier somit ein Zugänglichmachen der Texte für Nichtjuristen unter Wahrung fachlicher Präzision sowie die allgemeine Wahrung der sprachlichen Wohlgeformtheit unabhängig jedweder Fachtextsorte. Anzumerken ist, dass die Zugänglichkeit immer mit Bezug auf die tatsächliche Adressatengruppe geschieht. Nicht zuletzt soll somit auch der zeitliche Aufwand bei der Gesetzesformulierung optimiert werden.

3 Primärkorpus zu Rechts- und Gesetzestexten

Um das Korpus für eine zukünftige Verwendung verfügbar zu machen, ist die Repräsentation in einem offenen und vielseitigen Format unablässig. So ist etwa die Kompatibilität zu Architekturen wie ANNIS vorgesehen, da hierdurch auch ein möglicher Austausch mit Studien anderer Textsorten gewährleistet wird. Neben dieser Architektur, die vor allem für computerlinguistische Analysen relevant ist, wird auch ein möglicher Export nach offenen XML-Formaten zur Repräsentation juristischer Daten möglich sein. Ein solches Format stellt Metalex dar, das an mehreren europäischen Universitäten entwickelt und gepflegt wird.

Aufbauend auf bereits erfolgten Vorarbeiten (Kuhn 2009, 2010) werden die für eine semantische Kategorisierung von Rechtstermini textlinguistisch relevanten Informationen aus einem Sekundärkorpus mit Gerichtsentscheidungen durch eine automatenbasiertes mehrstufiges Parsingverfahren erschlossen. Termini, die hier in ihrem unmittelbaren Anwendungskontext auftreten, können im Primärkorpus durch Querverweise zu diesen Kontexten bereichert werden.

4 Digital Humanities und Digital Law?

Durch eine Architektur wie ANNIS kann ein Rechts- und Gesetzeskorpus in einem offenen Format digital zugänglich gemacht werden. Hierdurch wird es auch möglich, einen Austausch mit anderen geisteswissenschaftlichen Inhalten zu fördern, da dies durch eine textsortenunabhängige, linguistisch motivierte Repräsentation gegeben ist. Ebenfalls wird durch das offene Format eine von kommerziellen Interessen unabhängige Niederlegung der Daten möglich.

5 Zusammenfassung und Ausblick

Zur Entstehung des beschriebenen Forschungsprojektes bedarf es einerseits des Aufbaus und der Zugänglichmachung geeigneter Rechts- und Gesetzeskorpora, die das gesamte Spektrum dogmatischer Textsorten beinhalten (Entscheidungen der Oberinstanzgerichte sowie rechtswissenschaftliche Kommentarliteratur). Andererseits werden etablierte korpuslinguistische Verfahren, Werkzeuge und Algorithmen dahingehend angepasst und weiterentwickelt, dass

sie für die spezialisierte juristische Textpraxis unter verschiedenen institutionellen Rahmenbedingungen angewendet werden können .

Die bereits oben aufgeworfenen interdisziplinären Berührungspunkte lassen sich nur über einen fruchtbaren, offenen Dialog unterschiedlicher Disziplinen, insbesondere zwischen Rechtswissenschaft, Sprachwissenschaft, Computerlinguistik, Informatik und Medienwissenschaft sowie zwischen Wissenschaftlern und Praktikern der jeweiligen Bereiche bearbeiten. An einer solchen Zusammenarbeit besteht reges Interesse. Verbundprojekte, vor allem im angelsächsischen Raum, belegen dieses Bedürfnis und zeigen bereits erste Ergebnisse. Wir freuen uns, im Rahmen der ersten Jahrestagung der Digital Humanities im deutschsprachigen Raum unser Vorhaben vorzustellen und in Diskussionen fruchtbare Verbindungen zu den angrenzenden Fachbereichen herauszuarbeiten.

References

- [1] **Busse**, Dietrich 2000. *Textlinguistik und Rechtswissenschaft*. Berlin: de Gruyter.
- [2] **Carstensen**, Kai-Uwe, Christian Ebert, Cornelia Endriss, Susanne Jekat und Ralf Klambunde (Hrsg.) 2004. *Computerlinguistik und Sprachtechnologie*. Heidelberg: Springer.
- [3] **Engberg**, Jan 1992. Signalfunktion und Kodierungsgrad von sprachlichen Merkmalen in Gerichtsurteilen. *Hermes* 6, 65–82.
- [4] **Hachey**, Ben und Claire Grover 2006. Extractive summarisation of legal texts. *Artificial Intelligence and the Law* 14.
- [5] **Hopcroft**, John E. und Jeffrey Ullman 1979. *Introduction to Automata Theory*. Addison Wesley.
- [6] **Kuhn**, Florian 2010a. A description language for content zones of german court decisions. *Workshop Programme LREC 2010*.
- [7] **Kuhn**, Florian 2010b. A framework for graph-based parsing of German Private Law Decisions. *Business Information Systems Workshops*. Heidelberg: Springer, 292-297.
- [8] **Sipser**, Michael 2006. *Introduction to the Theory of Computation*. Boston: Wadsworth.