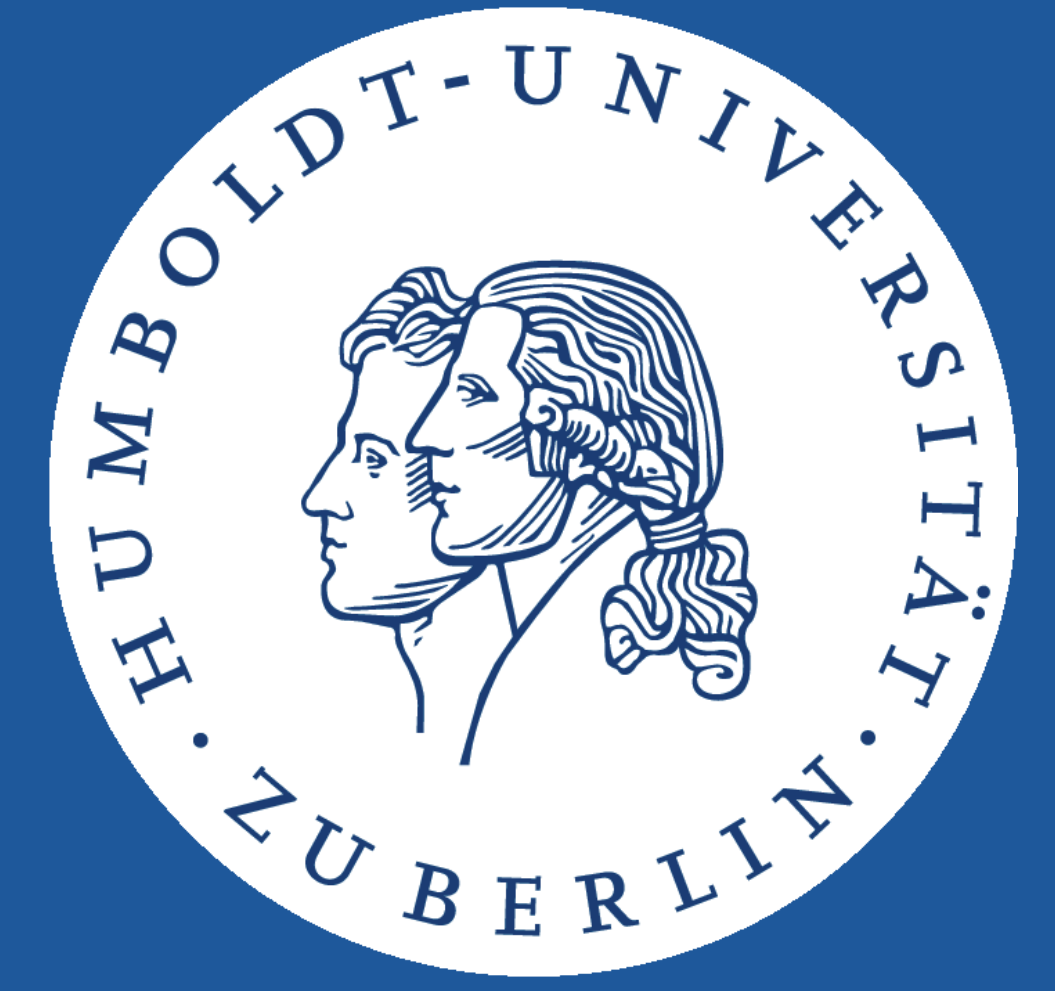# A Corpus of Federal Constitutional Court Decisions: BVerfGE-Korpus 1.0

Anna Shadrova & Alexander Tischbirek
Project Leibniz Linguistic Research into Constitutional Law (L.L.Con), chaired by Christoph Möllers
Humboldt-Universität zu Berlin
Digital Methods and Resources in Legal History 2021

The Corpus

- Contains 3354 decisions published in the Official Series (Amtliche Sammlung) in the years 1951-2019 and an additional 6462 decisions from the years 1997-2019 published on the GFCC website (total 9816 decisions, ~50 million tokens + metadata)
- Multi-layer annotations include
  - content sections (Leitsätze, Rubrum, Tenor, Gründe, abweichende Meinung)
  - identification of basic rights
  - linguistic layers: part-of-speech tags with Treetagger (Schmid 1995 with the Stuttgart-Tübingen Tagset, Schiller et al. 1996), MaltParser dependency parses (Nivre et al. 2006) with Foth's (2006) dependency grammar tagset
- Available in various formats under CC-Attribution-Non-Commercial 4.0 via Zenodo (doi: 10.5281/zenodo.4551408)

Ongoing Developments & Future Research

- Extension of network analysis to European contexts (Kirchmair et al. in prep)
- Text-linguistic analysis of GFCC developments (Wendel in prep.)
- Historical comparison of GFCC jurisdiction rooted in freedom rights and equality rights (Tischbirek in prep.)
- Analysis of word bundles and collocatios (Reule in prep.)
- Historical development of legal concepts and argumentative structure in GFCC jurisdiction
- Implementation as a graph database for integrated search functionality with knowledge graphs of legal entities

References

Foth, K. A. (2006). Eine umfassende Constraint-Dependenz-Grammatik des Deutschen.
Ighreiz, A., C. Möllers, L. Rolfes, A. Shadrova, and A. Tischbirek (in press): Selbst- und Fremdkanonisierung der Rechtsprechung des Bundesverfassungsgerichts. Archiv des öffentlichen Rechts.
Nivre, J., Hall, J., & Nilsson, J. (2006). Maltparser: A data-driven parser-generator for dependency parsing. In LREC (Vol. 6, pp. 2216-2219).
Schiller, A., Thielen, C., Teufel, S., & Stöckert, C. (1995). STTS (Stuttgart-Tübingen Tagset).
Schmid, Helmut (1995): Improvements in Part-of-Speech Tagging with an Application to German. Proceedings of the ACL SIG-DAT-Workshop. Dublin, Ireland.C. (1995).
Wendel, L., A. Shadrova, and A. Tischbirek (submitted): From Modeled Topics to Areas of Law: A Comparative Analysis of Types of Proceedings in the German Federal Constitutional Court.

Example Studies

Investigating Canonization through Self-Citation Networks

In Ighreiz et al. (in press), we analyzed the GFCC's self-citation network of the years 1951-2017 and compared it to decisions frequently mentioned in casebooks and other curated collections of GFCC decisions as they are used in academia and teaching.

We show that - unlike in the American legal tradition - there is a major overlap between the academic and the practical canon. This is somewhat surprising given that we also show that there can be rather divergent ways and reasons for a decision to become canonical.
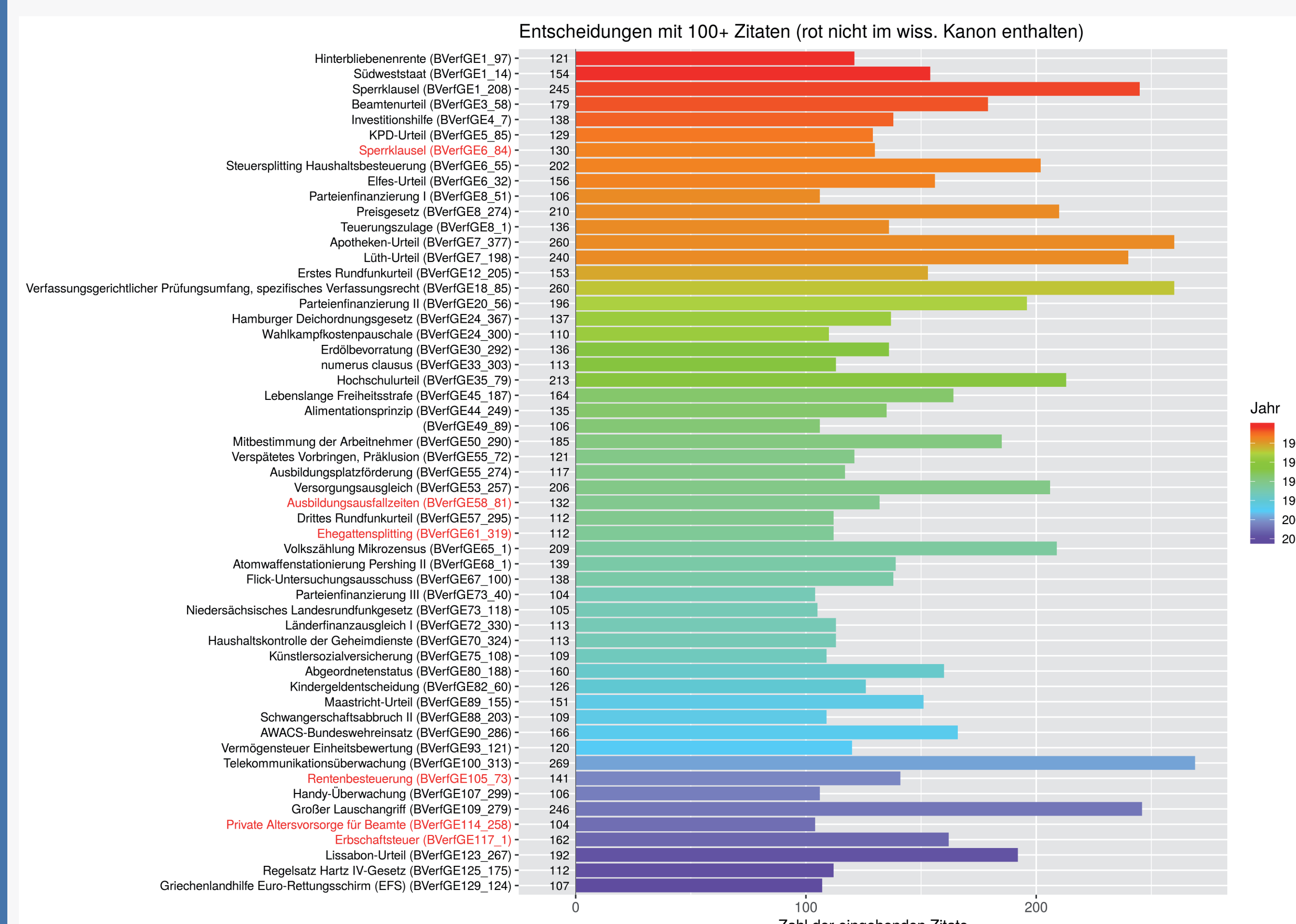

Fig. 1: Decisions with 100 or more incoming citations, those marked red are not included in the curated canon

Areas of Law and Types of Proceeding: A Topic-Modeling-based Approach

In Wendel et al. (submitted), we show that areas of law such as social, tax, or criminal law, are unequally represented in the GFCC's caseload by type of proceeding: The more technical and somewhat instable areas of social and tax law are significantly overrepresented in referrals for judicial review, i.e. taken to the GFCC by lower courts, whereas areas of law such as criminal law, which are characterized by well-developed case law and judicial doctrine, appear significantly more in constitutional complaints.

We show this by way of a topic-modeling-inspired approach: We first run topic models to generate sets of keywords belonging to each area of law and then perform an exact keyword search in a sampling approach.

With this, we are able to avoid several of the methodological problems inherent to the application of topic modeling in research contexts while still maintaining analytical access to large amounts of data.

Fig. 2: Proportion of decisions that contain at least 10 topic-related terms by procedure and year