

BVerfG-Korpus v.1.0

Dokumentation

Christoph Möllers, Anna Shadrova, Luisa Wendel
Humboldt-Universität zu Berlin

4. März 2021

1 Allgemeine Informationen und Aufbau

Das Korpus besteht aus Entscheidungen des Bundesverfassungsgerichts, die in den Jahren 1951–1997 in der Amtlichen Sammlung veröffentlicht wurden. Zusätzlich enthält es Entscheidungen aus den Jahren 1998–2019, die auf der Webseite des Gerichts veröffentlicht wurden.

Die Korpusdokumente liegen als CoNLL-Dateien mit Wortarten- und syntaktischen Abhängigkeitsannotationen vor (s.u.) und als xls-Tabellen mit zusätzlichen Textabschnittsannotationen. Die Tokenisierung entspricht der Treetagger-Tokenisierung. In Version 1.0 sind ca. 90% der CoNLL-Dateien aus der Amtlichen Sammlung und ca 95% der CoNLL-Dateien der Entscheidungen von der Webseite des Gerichts in xls übertragen und mit Textabschnittsnotationen versehen. Zusätzlich liegt eine Metadatentabelle vor.

2 Linguistische Annotationen

Es wurden automatische Wortartentags (STTS: Stuttgart-Tübingen Tagset, Schiller et al. (1995)) mit dem Treetagger erstellt (Schmid, 1994). Die Dateien wurden mit dem Maltparser (Nivre et al., 2006) und der Abhängigkeitsgrammatik von (Foth, 2006) automatisch geparst. Es gab keine manuellen Korrekturen.

3 Annotationen für Gliederungsstrukturen

Folgende Annotationen wurden mit Python erstellt, dabei wurde insbesondere auf die Beautiful Soup-Bibliothek Richardson (2007) zurückgegriffen.

entscheidung: Umfasst Leitsätze, Rubrum, Tenor und Gründe, aber nicht die abweichende Meinung.

leitsaetze: Ab 1998 wurde der Beginn mittels regulärem Ausdruck identifiziert. Entscheidungen mit Leitsätzen sind eine Minderheit, so dass auf eine Fehlerabschätzung

zur Ermittlung falsch-negativer Ergebnisse bisher verzichtet wurde. Bis 1997 sind die Leitsätze der verbliebene Text vor dem Rubrum nach Entfernen der Metadaten.

rubrum: Der Beginn des Rubrums wurde mittels regulärem Ausdruck identifiziert. Die Annotation endet vor Beginn der *tenor*-Annotation.

tenor: Der Beginn des Tenors wurde mittels regulärem Ausdruck identifiziert. Die Annotation endet vor Beginn der *gruende*-Annotation oder, wenn eine solche nicht vorhanden ist, mit dem Ende des Dokuments. Mittels einer Fehlerabschätzung wurden Dokumente identifiziert, in denen der Tenor manuell annotiert wurde. Die beiden Gutachtenverfahren haben keinen Tenor. Im Falle von sog. Tenorbegründungen enthält die *tenor*-Annotation auch die Begründung.

gruende: Der Beginn der Gründe wurde mittels regulärem Ausdruck identifiziert. Die Annotation endet vor Beginn der *abwmeinungen*-Annotation oder, wenn eine solche nicht vorhanden ist, mit dem Ende des Dokuments. Einige Entscheidungen haben tatsächlich keine Gründe, in anderen Fällen wurden die Gründe nicht gefunden. Mittels einer Fehlerabschätzung wurden Dokumente identifiziert, in denen die Gründe manuell annotiert wurden. Bei sog. Tenorbegründungen findet sich die Begründung in der *tenor*-Annotation.

abwmeinungen: Der Beginn der abweichenden Meinungen wurde mittels regulärem Ausdruck identifiziert. Ende der abweichenden Meinungen ist das Ende des Dokuments.

absatz (ab 1998): In den HTML-Dateien von der BVerfG-Webseite stehen die Randnummern der einzelnen Absätze im Text. Die zugehörigen Absätze erhielten jeweils eine *absatz*-Annotation mit der zugehörigen Randnummer als Attribut. Die Randnummern im Text wurden entfernt.

3.1 Annotationen von Fußnoten

fussnote (bis 1997): Eine Fußnote. Diese wurden mittels regulärer Ausdrücke ermittelt und der Entscheidung oder abweichenden Meinung zugeordnet. Attribut *seite* enthält die Seite, auf der die Fußnote erscheint.

fnentscheidung (bis 1997): Abschnitt mit etwaigen Fußnoten aus der Entscheidung.

fnabwmeinungen (bis 1997): Abschnitt mit etwaigen Fußnoten aus den Abweichenden Meinungen.

3.2 Annotationen von Grundgesetznormen, Grundrechtsverletzungen und -rügen

Die folgenden Annotationen wurden mit GATE Cunningham et al. (2013) erstellt, dabei kam auch der Treetagger Schmid (1994) zum Einsatz. Wir verwendeten Teile der GATE-Komponenten ANNIE Cunningham et al. (2002), TaggerFramework und Tools und passten sie für unsere Bedürfnisse an.¹ Für die Erstellung der Annotationsroutinen wurde der GATE Developer verwendet, für die Anwendung auf das gesamte Kor-

¹Dabei kommen sogenannte JAPE-Regeln zum Einsatz: <https://gate.ac.uk/sale/tao/splitch8.html> (letzter Abruf am 27.02.2021). Für Details siehe Wendel (2020a) bzw. Wendel (2020b).

pus der GATE Cloud Paralleliser (GCP) (<https://gate.ac.uk/gcp/>, letzter Abruf am 27.02.2021).

gegenstand: Existiert in Verfahren, in denen nur über eine einstweilige Anordnung, Prozesskostenhilfe, ein Ablehnungsgesuch oder den Gegenstandswert entschieden wird. Wird anhand eines Schlagwortverzeichnis identifiziert.

ggnorm: Zitat einer Grundgesetznorm. Wird mit regulären Ausdrücken (in sog. JAPE-Regeln) identifiziert.

grundrechtWort: Bezeichnung eines Grundrechts mit einer wörtlichen Bezeichnung anstelle eines konkreten Zitats. Wird anhand eines Schlagwortverzeichnis identifiziert.

fstverletzung: Tenorsatz, der die Verletzung eines Grundrechts / einer Grundgesetznorm kennzeichnet. Wird mit regulären Ausdrücken (in sog. JAPE-Regeln) identifiziert.

bfruegt: Rüge einer Grundrechtsverletzung in einer Verfassungsbeschwerde. Wird anhand von Schlagwortlisten (z.B. Konjunktiv, bestimmte Formulierungen der indirekten Rede) und mit regulären Ausdrücken (in sog. JAPE-Regeln) identifiziert.

4 Metadaten

dateiname: Eine eindeutige Bezeichnung der Entscheidung.

aktenzeichen: Die Aktenzeichen der Verfahren, zu denen die Entscheidung ergangen ist und die im Rubrum genannt sind. Mehrere Aktenzeichen sind meistens einzeln aufgelistet und durch Semikola getrennt. In ca. 100 Entscheidungen von vor 1997 sind allerdings mehrere Aktenzeichen als Folgen von Eingangsnummern bezogen auf eine oder wenige Verfahrensarten und ein oder wenige Eingangsjahre wiedergegeben, beispielsweise als ”2 BvR 909, 934, 935, 936, 938, 941, 942, 947/82, 64/83 und 142/84”.² Diese wurden so in die Tabelle übernommen und bislang nicht umgeformt.

inAS: Diese Variable nimmt den Wert 1 an, wenn die Entscheidung in der amtlichen Sammlung vorhanden ist und den Wert 0, wenn dies nicht der Fall ist.

fundstelle: Die Fundstelle in der amtlichen Sammlung, sofern die Entscheidung in der amtlichen Sammlung enthalten ist, ansonsten ”NA”. Bis 1997 wurde teilweise auf Tabelle auf der Internetseite des Bundesverfassungsgerichts³ zurückgegriffen, ansonsten wurde sie aus den Angaben im htm(l)-Dokument der Entscheidung ermittelt.

band, ersteSeite, letzteSeite: Diese Variablen enthalten die einzelnen Bestandteile der Fundstelle.

jahr, monat, tag: Das Datum der Entscheidung.

²BVerfGE 75, 108.

³<https://www.bundesverfassungsgericht.de/DE/Entscheidungen/Entscheidungen/Amtliche%20Sammlung%20BVerfGE.html> (letzter Abruf 28.02.2021).

entscheidungsart: Diese Variable kann die Ausprägungen "Beschluss", "Urteil" oder "Gutachten" annehmen. Die Ausprägung "Beschluss" wird auch für solche Dokumente verwendet, in denen lediglich die Gründe zu einem früher gefassten Beschluss beigefügt werden. Die Ausprägung "Urteil" wird auch für Teilurteile verwendet.

spruchkoerper: Der Spruchkörper, der die Entscheidung getroffen hat. Mögliche Werte sind "I" (Erster Senat), "II" (Zweiter Senat), die Kammer eines Senats nach dem Schema "I-2" für die 2. Kammer des Ersten Senats,⁴ "Beschwerdekammer" und "Plenum". Für den Zeitraum bis 1997 sind teilweise Entscheidungen den Senaten zugeordnet, die von einer Kammer bzw. einem Vorprüfungsausschuss entschieden wurden. Dies wurde noch nicht korrigiert.

gruende: Diese Variable nimmt den Wert 1 an, wenn die Entscheidung Gründe enthält und den Wert 0, wenn dies nicht der Fall ist. Erfolgt die Begründung nur in einem Tenorsatz, werden keine Gründe annotiert, also enthält die Entscheidung hier eine 0.

abwmeinungen: Diese Variable nimmt den Wert 1 an, wenn die Entscheidung mindestens eine abweichende Meinung enthält und den Wert 0, wenn dies nicht der Fall ist.

Registerzeichen (bvr, bvl, bvq, bvc, bve, bvf, bvo, bva, bvb, bvd, bvg, bvh, bvj, bvk, bvm, bvn, bvp, bvt, pbvs, pbvu, pbvv, vz): Diese Variablen stehen für die Registerzeichen der verschiedenen Verfahrensarten. Sie nehmen jeweils den Wert "TRUE" an, wenn die Entscheidung mindestens ein Verfahren der entsprechenden Verfahrensart betrifft und "FALSE", wenn dies nicht der Fall ist. Die Information beruht auf der Variable *aktenzeichen* und dient dem erleichterten Zugriff auf Entscheidungen zu bestimmten Verfahrensarten.

manuellErgaenzt: Diese Variable kennzeichnet, ob manuelle Korrekturen oder Ergänzungen an dem betreffenden Datensatz vorgenommen wurden. Die Ausprägung "0" bedeutet, dass keine Änderungen vorgenommen wurden, ansonsten wird angegeben, welche Variable korrigiert wurde. Die Ausprägung "1" bedeutet, dass möglicherweise Korrekturen aus einer früheren, unveröffentlichten, Version der Metadaten übernommen wurden, für die nicht vermerkt wurde, welche Variable betroffen war.

Referenzen

Cunningham, H., Maynard, D., Bontcheva, K., and Tablan, V. (2002). GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. In *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02)*.

⁴Im Einzelnen: I-1, I-2, I-3, I-4, II-1, II-2, II-3, II-4.

- Cunningham, H., Tablan, V., Roberts, A., and Bontcheva, K. (2013). Getting More Out of Biomedical Documents with GATE’s Full Lifecycle Open Source Text Analytics. 9:1–16.
- Foth, K. A. (2006). Eine umfassende Constraint-Dependenz-Grammatik des Deutschen.
- Nivre, J., Hall, J., and Nilsson, J. (2006). Maltparser: A data-driven parser-generator for dependency parsing. In *LREC*, volume 6, pages 2216–2219.
- Richardson, L. (2007). Beautiful soup documentation. *April*.
- Schiller, A., Thielen, C., Teufel, S., and Stöckert, C. (1995). STTS (Stuttgart-Tübingen Tagset).
- Schmid, H. (1994). TreeTagger-a language independent part-of-speech tagger. <http://www.ims.uni-stuttgart.de/projekte/complex/TreeTagger/>.
- Wendel, L. (2020a). Welche Grundrechte führen zum Erfolg?
- Wendel, L. (2020b). Welche Grundrechte führen zum Erfolg? pages 668–679.