

Korpus der Entscheidungen des Bundesverfassungsgerichts

L.L.Con.-Korpus – Dokumentation

Luisa Wendel, Christoph Möllers

15. Dezember 2023

1 Einleitung

Dieses Korpus von Entscheidungen des Bundesverfassungsgerichts entstand am Lehrstuhl für Öffentliches Recht, insb. Verfassungsrecht, und Rechtsphilosophie (Prof. Dr. Christoph Möllers, LL.M. (Chicago)), Juristische Fakultät, Humboldt-Universität zu Berlin im Rahmen des aus Leibniz-Mitteln finanzierten Projekts „Leibniz Linguistic Research into Constitutional Law“¹ in den Jahren 2017-2023.

Es enthält 10.980 Entscheidungen aus den Jahren 1951 bis 2022 nebst Annotationen von einigen juristisch relevanten Teiltextrn (Ordner `xml`) und Metadaten. Die Metadaten wurden bereits separat veröffentlicht, aktuell in Version 2.7.1.²

Diese Dokumentation gibt einen Überblick über die Zusammensetzung des Korpus, die Herkunft und Eigenschaften der Primärtexte (2), die Entstehung der Annotationen (3), Lizenzfragen (4) und einige Möglichkeiten der Auswertung (5). Detaillierte Informationen zum Prozess der Erstellung des Korpus sind Inhalt einer (noch unveröffentlichten) Dissertation³ und weiterer Repositorien, die im Zusammenhang mit diesem Korpus und der Dissertation veröffentlicht werden.⁴

¹L.L.Con., <https://www.lehrstuhl-moellers.de/llcon>, letzter Abruf am 14. Dezember 2023.

²Wendel, Metadaten zu Entscheidungen des Bundesverfassungsgerichts. Siehe dort für nähere Erläuterungen zu den Metadaten.

³Wendel, Entscheidungen des Bundesverfassungsgerichts.

⁴Insbesondere: Wendel, Skripte zur Erstellung und Fortführung des Korpus der Entscheidungen des Bundesverfassungsgerichts.

2 Primärtexte

Die Primärtexte, die für das Korpus verarbeitet wurden, stammen aus zwei Quellen. Die Entscheidungen ab 1998 wurden von der Internetpräsenz des Bundesverfassungsgerichts⁵ heruntergeladen. Erst ab 1998 wurden die Entscheidungen systematisch auf der Internetpräsenz veröffentlicht. Das Korpus wurde über einen längeren Zeitraum aufgebaut und die Entscheidungen wurden jeweils zu bestimmten Zeitpunkten gespeichert. Es wurde später festgestellt, dass das BVerfG manche der Entscheidungen auf der Internetpräsenz nachträglich verändert hat, teilweise auch erst nach vielen Jahren. Da die Entscheidungen im Korpus zu diesem Zeitpunkt schon in verschiedener Weise bearbeitet worden waren (teilweise manuell), wurde darauf verzichtet, die geänderten Entscheidungen systematisch zu untersuchen und neu zu bearbeiten. In der Tabelle `info/Speicherdatum.csv` ist festgehalten, welche Entscheidung zu welchem Zeitpunkt gespeichert wurde, so dass Nutzer des Korpus bei Bedarf Änderungen an einzelnen Entscheidungen vornehmen können.

Bis zum Jahr 1997 sind nur solche Entscheidungen im Korpus enthalten, die in der amtlichen Sammlung veröffentlicht wurden. Die Ausgangsdateien wurden dem L.L.Con.-Projekt freundlicherweise vom Verlag Mohr Siebeck GmbH & Co. KG zur Verfügung gestellt. Gerichtsentscheidungen sind amtliche Werke im Sinne des § 5 Abs. 1 UrhG und damit urheberrechtsfrei.⁶ Diese älteren Ausgangsdateien enthalten diverse Fehler, die vermuten lassen, dass die gedruckte Sammlung eingescannt wurde zu einer Zeit, als die automatische Zeichenerkennung noch weniger ausgereift war (beispielsweise Zeichen, die in der deutschen Sprache ungewöhnlich sind).

Die ursprünglichen HTML-Dateien wurden vorverarbeitet, insbesondere wurde irrelevanter HTML-Markup entfernt. Dadurch wird das Format von Tabellen und auch Formatierungen wie Unterstreichungen oder hochgestellte Zeichen nicht erhalten. Bei den Entscheidungen ab 1998 wurden Maßnahmen ergriffen, um die Randnummern zu erhalten.⁷ Sofern in diesen Entscheidungen im Rubrum Angaben zu den Bevollmächtigten enthalten waren, wurden diese entfernt, dort steht stattdessen `[BEVOLLM. ENTFERNT]`. Dies reflektiert die jüngere Praxis des BVerfG, auf die Angabe der Bevollmächtigten zu verzichten.

Bei den Entscheidungen bis 1997 wurden insbesondere Seitenzahlen und Fußnoten entfernt, damit der Fließtext nicht unterbrochen wird.⁸ Auch wurde versucht, durch Seitenumbrüche getrennte Wörter zu rekonstruieren, indem der Bindestrich und der Zeilenumbruch gelöscht wurden. Dies führte an anderen Stellen dazu, dass Begriffe, die abgetrennte Wortteile enthalten, fehlerhaft zusammengesetzt wurden.⁹

⁵<https://bundesverfassungsgericht.de/e/> (letzter Abruf am 14. Dezember 2023)

⁶Für urheberrechtliche und datenbankrechtliche Details siehe *Rolfes/Wendel*, Die Veröffentlichung von Korpora amtlicher Werke zu Forschungszwecken aus urheber- und datenbankrechtlicher Sicht.

⁷Siehe dazu unten 3.5.

⁸Siehe zu Fußnoten unten 3.9.

⁹Ein fiktives Beispiel wäre „Senatsund Kammerentscheidungen“ statt „Senats- und Kammerentscheidungen“.

Obwohl möglichst wenig in den Primärtext eingegriffen wurde, ist also davon auszugehen, dass die Texte Fehler enthalten und eine Argumentation in juristischen Einzelfällen sollte nicht auf das Korpus gestützt werden.

3 Annotationen

Das Korpus enthält Annotationen für einige juristisch interessante Teiltex-te. Die entsprechenden Passagen in den Texten werden durch XML-Tags gekennzeichnet. Mithilfe der `style.css`-Datei können die Dateien im Browser mit farblicher Hervorhebung der Teiltex-te dargestellt werden.¹⁰

Abbildung 1 zeigt die Annotationsstruktur der Dokumente. Nicht alle Annotationen sind in jeder einzelnen Entscheidungen vertreten. Die Annotationen `<bverfgdokument>`, `<entscheidung>`, `<leitsaetze>`, `<rubrum>`, `<tenor>`, `<gruende>`, `<fnentscheidung>`, `<abwmeinungen>` und `<fnabwmeinungen>` kommen maximal einmal pro Entscheidung vor. Die Annotationen `<ebene1>`, `<absatz>` und `<fussnote>` können mehrfach vorkommen.

Die Annotationen wurden so weit wie möglich automatisch erstellt. Dabei wurden für fast alle Annotationen regelbasierte Verfahren (z.B. unter Verwendung von regulären Ausdrücken) verwendet. Bei der Annotation von Tatbestand und Entscheidungsgründen wurden auch Verfahren des maschinellen Lernens verwendet. Dort, wo die automatische Annotation nicht funktioniert hat und dies aufgefallen ist, wurde manuell annotiert oder korrigiert. Details zur Vorgehensweise lassen sich der erwähnten Dissertation und den weiteren Repositorien, die im Zusammenhang mit der Dissertation und dem Korpus veröffentlicht werden, entnehmen.

3.1 Gesamtdokument

Das Gesamtdokument wird mit einer `<bverfgdokument>`-Annotation umfasst, dies ist das Wurzelement.

3.2 Entscheidung

Die Annotation `<entscheidung>` kennzeichnet die eigentliche Entscheidung und dient damit zur Abgrenzung von der Annotation der abweichenden Meinung.

Der folgende Ausschnitt aus einer Entscheidung¹¹ zeigt den Anfang einer Entscheidung mit XML-Informationen, Stylesheet-Einbindung und die XML-Elemente für die Annotationen `<bverfgdokument>`, `<entscheidung>` und `<leitsaetze>`.

¹⁰Die `style.css`-Datei ist hier sehr einfach gehalten und kann bei Bedarf an die Darstellungsvorlieben der Korpusnutzerinnen angepasst werden.

¹¹Basierend auf: BVerfG, Beschluss des Zweiten Senats vom 13. Februar 2020 – 2 BvR 739/17 – https://www.bverfg.de/e/rs20200213_2bvr073917.html.

bverfgdokument

entscheidung

leitsaetze

rubrum

tenor

gruende

ebene1

absatz

absatz

ebene1

absatz

absatz

fnentscheidung

fussnote

abwmeinungen

absatz

absatz

fnabwmeinungen

fussnote

Abbildung 1: Annotationsstruktur der Dokumente

```

1 <?xml version="1.0" encoding="utf-8"?>
2 <?xml-stylesheet href="..\style.css"?>
3 <bverfgdokument>
4 <entscheidung>
5 <leitsaetze>
6 Leitsätze
7 zum Beschluss des Zweiten Senats vom 13. Februar 2020

```

3.3 Leitsätze, Rubrum und Tenor

Die Annotationen `<leitsaetze>`, `<rubrum>` und `<tenor>` kennzeichnen die Leitsätze, das Rubrum und den Tenor einer Entscheidung. Diese Textabschnitte selbst wurden bislang nicht detailliert aufbereitet. In den Ausgangsdateien werden Nummerierungen in diesen Abschnitten teilweise durch HTML-Markup dargestellt und sind daher im reinen Text nicht mehr erkennbar. Ein etwaiges Inhaltsverzeichnis ist noch Teil des Tenors. Das BVerfG verfasst zuweilen Tenorbegründungen, d.h. Begründungen, bei denen keine optische Trennung zwischen Tenor und Gründen stattfindet und bei denen manche Sätze Tenor und Gründe enthalten. Hierfür wurde keine eigene Annotationskategorie vorgesehen, sondern der Text wurde (manuell) in Tenor- und Gründe aufgeteilt, sofern diese Entscheidungen auffielen.

Der folgende Ausschnitt¹² zeigt die Annotation des Ende des Rubrums, des Tenors und des Anfangs der Gründe:

```

37 König,
38 Maidowski,
39 Langenfeld
40 am 24. Juli 2018 beschlossen:
41 </rubrum>
42 <tenor>
43 Die Verfassungsbeschwerde wird verworfen.
44 </tenor>
45 <gruende><ebene1 nr="0" zeichen="NA" tbeg="tb"><absatz ebene1nr="0"
    ebene1zeichen="NA" absatzID="0" rn="NA" tbeg="tb">
46 G r ü n d e :

```

3.4 Gründe

Die Annotation `<gruende>` umfasst die Gründe einer Entscheidung. Dieser Abschnitt wurde am stärksten bearbeitet.

In der Regel enthält dieser Abschnitt Annotationen des Typs `<absatz>`. Es sollte sich kein Text außerhalb von Absatz-Annotationen befinden. Des weiteren wurde die formale

¹²Basierend auf: BVerfG, Beschluss des Zweiten Senats vom 24. Juli 2018 – 2 BvR 1961/09 – https://www.bverfg.de/e/rs20180724_2bvr196109.html.

Gliederungsstruktur möglichst automatisch annotiert, allerdings nur auf der obersten Ebene (oft als „A.“-Teil, „B.“-Teil, „C.“-Teil usw. bezeichnet). Diese ist jedoch nicht in allen Entscheidungen vorhanden bzw. wurde nicht immer korrekt erkannt (siehe dazu sogleich 3.6).

3.5 Absatz

Die Annotation `<absatz>` kennzeichnet einen Absatz innerhalb einer BVerfG-Entscheidung (bzw. innerhalb einer abweichenden Meinung, siehe dazu sogleich 3.8). Der Absatz kann leer sein oder lediglich ein Gliederungszeichen oder die Überschrift *Gründe:* enthalten. Am Anfang und am Ende kann der Absatz *whitespace* (z.B. Leerzeichen, Zeilenumbrüche) enthalten. Diese müssen für Auswertungen oder Weiterbearbeitungen ggf. entfernt werden.

Die `<absatz>`-Annotation kann folgende Attribute enthalten:

- **ebene1nr** – Die laufende Nummer des Gliederungsabschnitts, zu dem der Absatz gehört.
- **ebene1zeichen** – Das Zeichen, mit dem der Gliederungsabschnitt, zu dem der Absatz gehört, bezeichnet ist. Möglich sind Großbuchstaben (A, B, C ...), römische Zahlen (I, II, III ...), arabische Zahlen (1, 2, 3 ...) oder auch NA, d.h., es gibt kein Gliederungszeichen.
- **absatzID** – Eine fortlaufende Nummer des Absatzes innerhalb des Gliederungsabschnitts, zu dem der Absatz gehört. Bei Entscheidungen ohne Annotation der formalen Gliederung: fortlaufende Nummer innerhalb der Gründe.
- **rn** – Die Randnummer, die der Absatz vom BVerfG erhalten hat. Mögliche Ausprägungen sind einzelne Ziffern, mehrere durch Bindestrich getrennte Ziffern („264-265“), NA oder eine leere Zeichenkette („“). Diese wurde im Rahmen der Vorverarbeitung der Rohdaten extrahiert und möglichst dem richtigen Absatz zugeordnet. Hier sind Fehler nicht ausgeschlossen. Diese Information ist nur für jüngere Entscheidungen (ab 1998) verfügbar.
- **tbeg** – Die Information, ob ein Absatz zum Tatbestand oder zu den Entscheidungsgründen gehört (siehe dazu 3.7).

Jeder Absatz lässt sich also eindeutig identifizieren über die Kombination aus Dateiname, **ebene1nr** und **absatzID**, bzw. bei Entscheidungen ohne annotierte Gliederungsabschnitte über Dateiname und **absatzID**.

3.6 Formale Textgliederung (1. Ebene)

Im Allgemeinen wird angenommen, dass eine BVerfG-Entscheidung aus einem „A.-Teil“ (dem Tatbestand), einem „B.-Teil“ (der Zulässigkeit) und einem „C.-Teil“ (der Begründetheit) besteht. In der Praxis gibt es jedoch zahlreiche Abweichungen davon. Außerdem sind die Gliederungszeichen teilweise nicht oder falsch wiedergegeben. Im Rahmen der Korpuserstellung konnte also nicht davon ausgegangen werden, dass die Gliederungszeichen mit einfachen Regeln automatisiert erkannt werden können und von ihnen auf die Zugehörigkeit eines Absatzes zu einem bestimmten Teilttext geschlossen werden kann. Dennoch wurde versucht, die formale Gliederungsstruktur automatisch zu erkennen und zu annotieren, da diese jedenfalls als erster Anhaltspunkt für die Identifizierung der relevanten Teilttexte dienen kann. Allerdings wurde nur die erste/oberste Ebene der Gliederungsstruktur betrachtet, da die tieferen Ebenen noch fehleranfälliger sind. Die Gliederungsabschnitte auf der ersten Ebene wurden jeweils mit einer `<ebene1>`-Annotation versehen.

Diese kann folgende Attribute enthalten:

- `nr` – Die laufende Nummer des Gliederungsabschnitts.
- `zeichen` – Das Zeichen, mit dem der Gliederungsabschnitt bezeichnet ist. Möglich sind Großbuchstaben (A, B, C ...), römische Zahlen (I, II, III ...), arabische Zahlen (1, 2, 3 ...) oder auch NA, d.h., es gibt kein Gliederungszeichen.
- `tbeg` – Die Information, ob dieser Gliederungsabschnitt zum Tatbestand und/oder zu den Entscheidungsgründen gehört (siehe dazu 3.7). Mögliche Ausprägungen: `tb`, `eg` oder `tbeg`.

Die `<ebene1>`-Annotation sollte nur als Hilfsmittel verstanden werden. Auswertungen sollten besser anhand von Absätzen stattfinden.

3.7 Tatbestand und Entscheidungsgründe

Tatbestand und Entscheidungsgründe sind nicht durch eigenen XML-Markup gekennzeichnet, sondern durch Attribute innerhalb der `<absatz>`-Annotationen. Das Attribut `tbeg="tb"` bedeutet, dass der Absatz zum Tatbestand gehört, `tbeg="eg"` bedeutet, dass der Absatz zu den Entscheidungsgründen gehört. Innerhalb der `<ebene1>`-Annotation gibt es ebenfalls das Attribut `tbeg`, dieses kann zusätzlich die Ausprägung `tbeg="tbeg"` haben, für Gliederungsabschnitte, die sowohl Tatbestands-Absätze also auch Entscheidungsgründe-Absätze enthalten. Es ist möglich, dass in einer Entscheidung kein Tatbestand annotiert wurde, z.B. weil der Tatbestand nur in wenigen Sätzen beschrieben wurde und in einem Absatz, der den Entscheidungsgründen zugeordnet wurde, enthalten ist. Andere Entscheidungen enthalten einen sehr kurzen „Tatbestand“, der nur aus der Überschrift „Gründe:“ besteht. Bei der Auswertung von Tatbeständen muss daher entschieden werden, ob diese sehr kurzen Tatbestände ausgeschlossen werden.

Tabelle 1: Angewandte Klassifikationsverfahren

	AS	nichtAS
Entscheidungen insgesamt (mit Gründen)	3.460	7.082
Tatbestand und Entscheidungsgründe		
annotiert mit:		
regelbasierten Verfahren	2.421	5.967
Naive Bayes und tf-idf	548	0
SVM und word embeddings	0	827
Rückgriff auf den 2. Absatz (subsidiär)	182	0
manuell	309	293

Der folgende Ausschnitt¹³ zeigt den Beginn eines Gliederungsabschnitts und die ersten beiden Absätze darin:

```

907 [...]</ebene1><ebene1 nr="2" zeichen="B" tbeg="eg"><absatz ebene1nr="2"
    ebene1zeichen="B" absatzID="0" rn="NA" tbeg="eg"> B.
908 </absatz><absatz ebene1nr="2" ebene1zeichen="B" absatzID="1" rn="91" tbeg="eg">
909 Die Verfassungsbeschwerde ist zulässig, soweit sie eine Verletzung des
    Beschwerdeführers in seinem Recht aus Art. 38 Abs. 1 Satz 1 in Verbindung
    mit Art. 20 Abs. 1 und Abs. 2, Art. 79 Abs. 3 GG durch Verstoß gegen das
    Erfordernis einer qualifizierten Mehrheit für das EPGÜ-ZustG nach Art. 23
    Abs. 1 Satz 3 in Verbindung mit Art. 79 Abs. 2 GG rügt (I.). Im Übrigen ist
    sie unzulässig (II.).
910 </absatz><absatz ebene1nr="2" ebene1zeichen="B" absatzID="2" rn="NA" tbeg="eg">
911 I.
912 </absatz>[...]
```

Die Information, ob ein Absatz zum Tatbestand oder zu den Entscheidungsgründen gehört, wurde zunächst in 302 Entscheidungen aus der amtlichen Sammlung und in 150 Entscheidungen von außerhalb der amtlichen Sammlung manuell ermittelt.¹⁴ Die manuell annotierten Entscheidungen wurden einerseits verwendet, um Regeln für die Erkennung des Übergangs vom Tatbestand zu den Entscheidungsgründen zu ermitteln, und andererseits um verschiedene statistische Verfahren des maschinellen Lernens zu trainieren und zu testen. Darauf basierend konnte ein Großteil der Entscheidungen automatisch annotiert werden. Entscheidungen, bei denen keiner der automatischen Klassifikatoren den Übergang zu den Entscheidungsgründen feststellen konnte, wurden abschließend manuell annotiert.

Eine abschließende Evaluation der automatischen Annotation anhand von je 100 AS- und 100 nichtAS-Entscheidungen ergab, dass in 93% der AS-Entscheidungen und in 87%

¹³Basierend auf: BVerfG, Beschluss des Zweiten Senats vom 13. Februar 2020 – 2 BvR 739/17 – https://www.bverfg.de/e/rs20200213_2bvr073917.html.

¹⁴Genauer gesagt wurden statt der Entscheidungsgründe die Teiltexthe Zulässigkeit, Begründetheit und weitere Teiltexthe, die zu den Entscheidungsgründen gehören, annotiert. Die Identifizierung dieser Teiltexthe im Gesamtkorpus wurde bisher nicht automatisiert.

Tabelle 2: Evaluation der automatischen Annotation

		n	davon korrekt	annotiert
AS	Entscheidungen	100	93	0,9300
	Absätze	5.795	5.771	0,9959
nichtAS	Entscheidungen	100	87	0,8700
	Absätze	2.337	2.264	0,9688

der nichtAS-Entscheidungen der Übergang vom Tatbestand zu den Entscheidungsgründen korrekt ermittelt werden konnte. Sofern der Übergang in einer Entscheidung nicht korrekt ermittelt wurde, betraf dies nur einzelne Absätze, so dass insgesamt mehr als 99% der Absätze aus den AS-Entscheidungen und ca. 97% der Absätze der nichtAS-Entscheidungen korrekt zugeordnet wurden (siehe Tabelle 2).

Den Tabellen `info/TBEG_Annotationsart_AS.csv` und `info/TBEG_Annotationsart_nichtAS.csv` lässt sich entnehmen, auf welche Art in den einzelnen Entscheidungen die Feststellung des Übergangs von Tatbestand zu Entscheidungsgründen erfolgt ist. So können Untersuchungen bei Bedarf auf Entscheidungen beschränkt werden, die mit einem bestimmten Verfahren annotiert wurden.

3.8 Abweichende Meinung

Etwaige vorhandene abweichende Meinungen sind durch eine Annotation `<abwmeinungen>` gekennzeichnet. Zwischen mehreren abweichenden Meinungen zu einer Entscheidung wird dabei nicht unterschieden. Abweichende Meinungen können ebenfalls `<absatz>`-Annotationen enthalten. Diese wurden bisher nicht aufbereitet.

Der folgende Ausschnitt¹⁵ zeigt das Ende der `<entscheidung>`-Annotation und den Beginn der `<abwmeinungen>`-Annotation:

```

1080 Kessal-Wulf
1081 König
1082 Maidowski
1083 Langenfeld
1084 </absatz></ebene1></gruende>
1085 </entscheidung>
1086 <abwmeinungen><absatz rn="NA">
1087 Abweichende Meinung der Richterinnen König und Langenfeld
1088 sowie des Richters Maidowski
1089 </absatz><absatz rn="">

```

¹⁵Aus Entscheidung BVerfG, Beschluss des Zweiten Senats vom 13. Februar 2020 – 2 BvR 739/17 – https://www.bverfg.de/e/rs20200213_2bvr073917.html.

3.9 Fußnoten

In den älteren Entscheidungen (bis 1997) sind Fußnoten keine Seltenheit. Fußnoten und Fußnotenmarken mussten aus dem Fließtext entfernt werden. Die Information aus der Fußnote wurde in einem speziellen Abschnitt abgelegt. Die Annotationen `<fnentscheidung>` bzw. `<fnabwmeinungen>` enthalten alle Fußnoten zur Entscheidung bzw. zur abweichenden Meinung. Alle Fußnoten ab der Seite, auf der die abweichende Meinung beginnt, wurden der abweichenden Meinung zugeordnet, was im Einzelfall falsch sein kann. Jede Fußnote erhält eine eigene `<fussnote>`-Annotation, die auch ein Attribut `seite` enthält, in welchem die Seite angegeben ist, auf der die Fußnote steht. Die Information über den konkreten Satz oder das Wort, auf das sich die Fußnote bezieht, wurde allerdings nicht bewahrt, da dies nicht automatisch möglich war.

4 Lizenz

Die Sammlung der bis 1997 veröffentlichten Entscheidungen, d.h. der Dateien, deren Dateiname mit „BVerfGE“ beginnt, steht unter einer CC BY-NC-SA 4.0 (Attribution-NonCommercial-ShareAlike 4.0 International) Lizenz.

Alle sonstigen Dateien, d.h. die Dateien, deren Name nicht mit „BVerfGE“ beginnt, die Dateien im Ordner `info`, die Metadaten und diese Dokumentation stehen unter einer CC BY-SA 4.0 (Attribution-ShareAlike 4.0 International) Lizenz.

5 Möglichkeiten für Auswertung und Weiterentwicklung

Da das Korpus sich aus zwei verschiedenen Quellen speist und die Datenabdeckung vor und nach 1998 unterschiedlich ist, dürfte es für viele Anwendungen nicht sinnvoll sein, alle Entscheidungen zusammen zu betrachten. Stattdessen liegt es nahe, zwischen Entscheidungen aus der amtlichen Sammlung und Entscheidungen von außerhalb der amtlichen Sammlung zu unterscheiden. Letztere liegen dann erst ab 1998 vor. In anderen Fällen kann es sinnvoll sein, nur alle Entscheidungen ab 1998 zu betrachten.

Mithilfe der Metadaten lassen sich verschiedene Subkorpora zusammenstellen bzw. bestimmte Entscheidungen ausschließen. So zeigt die Spalte `gruende` ob eine Entscheidung über Gründe verfügt (1) oder nicht (0). Die Spalte `inAS` zeigt, ob eine Entscheidung in der amtlichen Sammlung veröffentlicht wurde (1) oder nicht (0).¹⁶ Weitere Subkorpora könnten etwa nach Verfahrensart oder Spruchkörper gebildet werden.

¹⁶Für Details wird auf die Dokumentation der separat veröffentlichten Metadaten verwiesen: *Wendel*, Metadaten zu Entscheidungen des Bundesverfassungsgerichts.

Tabelle 3: Vorkommen der verschiedenen Teiltexthe bzw. Annotationen in den beiden Teilkorpora

	AS	Anteil	nichtAS	Anteil
Entscheidungen insgesamt	3.484		7.496	
Leitsätze	2.521	0,7236	5	0,0007
Rubrum	3.484	1,0000	7.496	1,0000
Tenor	3.482	0,9994	7.488	0,9989
Gründe	3.460	0,9931	7.087	0,9454
abweichende Meinungen	178	0,0511	1	0,0001
Gliederungsabschnitt (ebene1)	3.301	0,9475	5.715	0,7624
Tatbestand	3.452	0,9908	6.478	0,8642
Entscheidungsgründe	3.460	0,9931	7.086	0,9453

Die Annotationen bzw. die XML-Elemente können genutzt werden, um bestimmte Textstellen anzusteuern, die untersucht werden sollen. Dazu können Werkzeuge wie die Python-Bibliothek *Beautiful Soup*¹⁷ verwendet werden. So können beispielsweise abweichende Meinungen von der Untersuchung der Entscheidungen ausgeschlossen werden, da sie nicht Teil der Entscheidung sind. Sprachliche Eigenschaften der Tatbestände können mit denen der Entscheidungsgründe verglichen werden oder gezielt Stichproben aus dem einen oder anderen Teiltexthe gezogen werden. Korpuslinguistische und NLP-Tools können genutzt werden, um die Teiltexthe mit linguistischen Annotationen anzureichern.

Tabelle 3 fasst zusammen, in wie vielen Entscheidungen der beiden Teilkorpora die genannten Teiltexthe bzw. Annotationsarten mindestens einmal zu finden sind. Hier ist schon ersichtlich, dass in Entscheidungen, die nicht in der amtlichen Sammlung enthalten sind, häufiger ein eigenständiger, abgegrenzter Tatbestand fehlt als in Entscheidungen aus der amtlichen Sammlung.

Es ist wünschenswert, weitere Teiltexthe in den Entscheidungsgründen zu annotieren, etwa Zulässigkeit, Begründetheit, Maßstabsteil o.Ä. Dies erfordert neben einer theoretisch fundierten Definition dieser Teiltexthe die manuelle Annotation einer Stichprobe um Trainingsdaten zu generieren. Die XML-Tags und Attribute können dann genutzt werden, um die Entscheidungen in ein Format zu überführen, das sich gut annotieren lässt. Im Rahmen des L.L.Con.-Projekts wurde gerne auf ein Tabellenkalkulationsformat zurückgegriffen, bei dem jeder Zeile ein Satz oder ein Absatz zugeordnet ist, der durch die Eigenschaften `dateiname`, `ebene1nr` und `absatzID` (sowie ggf. einer `satzID`) identifiziert wird. Anhand dieser Attribute lassen sich die manuellen Annotationen dann auch in die XML-Dateien zurückschreiben, beispielsweise als zusätzliche Attribute in den `<absatz>`-Annotationen. Die XML-Tags und Attribute können weiterhin dafür genutzt werden, den menschlichen Annotator:innen nur die Teiltexthe vorzulegen, in denen sie tatsächlich annotieren sollen. Dies kann die Arbeit erleichtern und helfen, Annotationsfehler zu minimieren.

¹⁷ *Richardson*, *Beautiful Soup*.

Für die aktuelle Version des Korpus wurden im Wesentlichen nur die Gründe der Entscheidung aufbereitet. Ein zukünftiges Forschungsprojekt könnte sich beispielsweise der Aufbereitung des Rubrums oder des Tenors widmen. Aufgrund der vorhandenen Annotationen können diese Teiltex te dann leicht durch überarbeitete Teiltex te ausgetauscht werden.

Literatur

Richardson, Leonard: Beautiful Soup, Version 4.9.0, 2004–2020, URL: <https://www.crummy.com/software/BeautifulSoup/> (besucht am 17.08.2023).

Rolfes, Louis/Luisa Wendel: Die Veröffentlichung von Korpora amtlicher Werke zu Forschungszwecken aus urheber- und datenbankrechtlicher Sicht, in: *Zeitschrift für Urheber- und Medienrecht* 62.8/9 2018, S. 590–600.

Wendel, Luisa: Entscheidungen des Bundesverfassungsgerichts, Eine korpusgestützte Untersuchung von Teiltex tsorten und Registereigenschaften, Diss., Berlin: Humboldt-Universität zu Berlin, 6. Nov. 2023, eingereicht.

Wendel, Luisa: Metadaten zu Entscheidungen des Bundesverfassungsgerichts, Version 2.7.1, 14. Dez. 2023, DOI: 10.5281/zenodo.10378324.

Wendel, Luisa: Skripte zur Erstellung und Fortführung des Korpus der Entscheidungen des Bundesverfassungsgerichts, Version 1.0, 15. Dez. 2023, DOI: 10.5281/zenodo.10369226.