Corpus of the decisions of the German Federal Constitutional Court

L.L.Con.-Corpus – Documentation (Overview)

Luisa Wendel

15 December 2023

1 Einleitung

This documentation accompanies a corpus of decisions of the German Federal Constitutional Court (GFCC) which was compiled at the Chair of Public Law and Legal Philosophy (Prof. Dr. Christoph Möllers, LL.M. (Chicago)), Faculty of Law, Humboldt-Universität zu Berlin as part of the project "Leibniz Linguistic Research into Constitutional Law"[1] between 2017 and 2023.

It contains 10.980 decisions of the years 1951 to 2022 with sectioning annotations that might be interesting for legal research (directory `xml`) and metadata.[2]

The documentation gives an overview of the composition of the corpus, the origin and characteristics of the texts and the annotation process. The german version of the documentation contains more detailed information. A more detailed description of the creation of the corpus is the subject of an unpublished Ph.D. thesis.[3] Code used in the preparation of the corpus is published separately.[4]

2 Primary texts

Primary texts of decisions from 1998 on have been downloaded from the GFCCs website.[5] The table `info/Speicherdatum.csv` contains the date the decision was saved for each

---

[1] L.L.Con., `https://www.lehrstuhl-moellers.de/llcon`, (last visited 14th December 2023).

[2] See *Wendel*, Metadaten zu Entscheidungen des Bundesverfassungsgerichts for more information on the metadata.

[3] *Wendel*, Entscheidungen des Bundesverfassungsgerichts.

[4] Notably here: *Wendel*, Skripte zur Erstellung und Fortführung des Korpus der Entscheidungen des Bundesverfassungsgerichts.

[5] `https://bundesverfassungsgericht.de/e/` (last visited 14th December 2023).

decision. This might be relevant, as the GFCC sometimes retroactively changes decisions published on the website.

For the period of 1951-1997 only decisions which have been published in the official report series (ORS) are included in the corpus. These have been kindly provided to the L.L.Con.-Project by the publisher Mohr Siebeck GmbH & Co. KG. German court decisions are in the public domain (§ 5 Abs. 1 UrhG).[6]

The decisions have been preprocessed. In particular, irrelevant HTML-markup has been removed. Decisions of 1998 and later preserved the paragraph numbers. In these decisions the information about the representatives has been removed from the decision caption (`[BEVOLLM. ENTFERNT]`).

The primary text was modified as little as possbile. Nevertheless, a legal argumentation in individual legal disputes should not be based on this corpus.

3 Annotations

The decisions contain annotations of text sections that might be of interest for legal scholars. These can be highlighted with the help of the `style.css`-file when opening the .xml-files in a browser.

Figure 1 shows the annotation structure of the documents. Not all annotations are present in every single decision. The annotations `<bverfgdokument>`, `<entscheidung>`, `<leitsaetze>`, `<rubrum>`, `<tenor>`, `<gruende>`, `<fnentscheidung>`, `<abwmeinungen>` and `<fnabwmeinungen>` appear at most once per decision. The annotations `<ebene1>`, `<absatz>` and `<fussnote>` may appear more than one time.

The annotations have been created automatically where possible. For the most part, rule-based approaches have been used. Where automatic annotation was not successful, the annotations were done manually.

The following annotation categories exist:

- `<bverfgdokument>` – root element
- `<entscheidung>` – the decision itself
- `<leitsaetze>` – headnotes
- `<rubrum>` – caption of the decision (includes names of parties, judges, date of decision ...)
- `<tenor>` – judgment, may include a table of contents
- `<gruende>` – reasons (facts and reasoning)

---

[6]For details on relevant intellectual property rights see *Rolfes/Wendel*, Die Veröffentlichung von Korpora amtlicher Werke zu Forschungszwecken aus urheber- und datenbankrechtlicher Sicht.
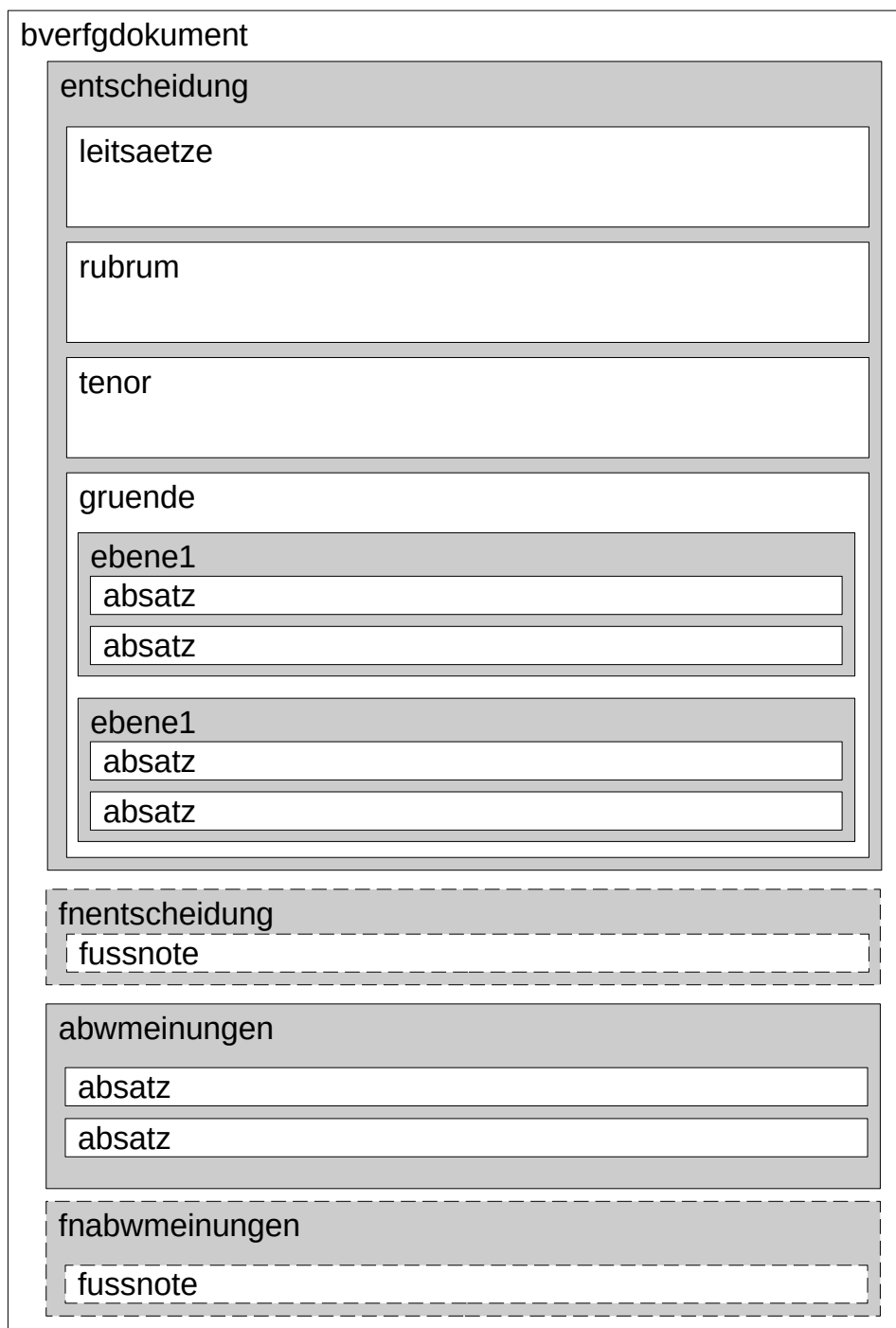
Figure 1: Structure of annotations in documents

- `<absatz>` – paragraph within a decision or separate opinion. May be empty. May contain additional whitespace at beginning or end. Attributes:

  - `ebene1nr` – number of typographical section the paragraph belongs to

  - `ebene1zeichen` – letter or number denoting the typographical section, may be a capital letter (A, B, C ...), Roman numeral (I, II, III ...), Arabic numeral (1, 2, 3 ...) or NA.

  - `absatzID` – number of the paragraph within the typographical section (or within the reasons, if no typographical sectioning exists)

  - `rn` – paragraph number as counted by GFCC, may be a number, a range of numbers ("264-265"), NA or an empty string.

  - `tbeg` – information on whether the paragraph belongs to fact section (`tb`) or reasoning section (`eg`).

- `<ebene1>` – typographical section ("A-Part", "B-Part" ...) on the first level of the sectioning. Should be regarded as an aid for annotation processes only, analysis should be based on paragraphs. Attributes:

  - `nr` – number of typographical section

  - `zeichen` – letter or number denoting the typographical section

  - `tbeg` – information on whether section belongs to fact section or reasoning section. Possible values: `tb`, `eg` or `tbeg` (contains both types of paragraphs).

- `<abwmeinungen>` – separate opinions for this decision

- `<fnentscheidung>` – footnotes for decision (only available for decisions from 1951-1997)

- `<fnabwmeinungen>` – footnotes for separate opinion (only available for decisions from 1951-1997)

- `<fussnote>` – individual footnote (only available for decisions from 1951-1997)

3.1 Fact section and reasoning section

This information has been manually annotated in a sample of decisions. These were then used to train rule-based and machine learning classifiers which were used to annotate the majority of decisions. A final evaluation of 100 decisions from the official report series (ORS-decisions) and 100 decisions from outside of the ORS (nonORS-decisions) showed that in 93% of ORS-decisions and 87% of nonORS-decisions the beginning of the reasoning section has been identified correctly. This leads to 99% correctly classified paragraphs in ORS-decisions and 97% in nonORS-decisions.

Tables `info/TBEG_Annotationsart_AS.csv` and `info/TBEG_Annotationsart_nichtAS.csv` show in which way each of the decision has been annotated.

4 License

Collection of decisions until 1997 (XML-files with file name beginning with "BVerfGE"): CC BY-NC-SA 4.0 (Attribution-NonCommercial-ShareAlike 4.0 International).

All other decisions (XML-files with file name not beginning with "BVerfGE"), files in `info`-directory, metadata and documentation: CC BY-SA 4.0 (Attribution-ShareAlike 4.0 International).

References

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, pp. 590–600.

Wendel, Luisa: Entscheidungen des Bundesverfassungsgerichts, Eine korpusgestützte Untersuchung von Teiltextsorten und Registereigenschaften, PhD thesis, Berlin: Humboldt-Universität zu Berlin, Nov. 6, 2023, submitted.

Wendel, Luisa: Metadaten zu Entscheidungen des Bundesverfassungsgerichts, version 2.7.1, Dec. 14, 2023, DOI: `10.5281/zenodo.10378324`.

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