

## **Dr. Andreas Nolda**

Berlin-Brandenburgische Akademie der Wissenschaften  
Zentrum für digitale Lexikographie der deutschen Sprache  
<https://andreas.nolda.org>

## **DWDSmor**

A toolbox for morphological analysis and generation in German,  
based on the DWDS lexicon and an SMOR-style grammar

## **eLex 2023**

29 June

## DWDSmor in a nutshell

DWDSmor is a toolbox for creating and applying a set of *finite-state automata* for morphological analysis and generation in written German.

## DWDSmor in a nutshell

DWDSmor is a toolbox for creating and applying a set of *finite-state automata* for morphological analysis and generation in written German.

The automata are compiled from an *SMOR-style grammar* in *SFST format* and a corresponding *lexicon*.

- Schmid, Helmut, Arne Fitschen, and Ulrich Heid (2004). SMOR: A German computational morphology covering derivation, composition, and inflection. In LREC 2004: Fourth International Conference on Language Resources and Evaluation, ed. by Maria T. Lino *et al.*, European Language Resources Association, 1263–1266. <http://www.lrec-conf.org/proceedings/lrec2004/pdf/468.pdf> [18 July 2019].
- Schmid, Helmut (2006). A programming language for finite state transducers. In *Finite-State Methods and Natural Language Processing: 5th International Workshop, FSMNLP 2005, Helsinki, Finland, September 1–2, 2005*, ed. by Anssi Yli-Jyrä, Lauri Karttunen, and Juhani Karhumäki, Lecture Notes in Artificial Intelligence 4002, Berlin: Springer, 1263–1266.

## DWDSmor in a nutshell

DWDSmor is a toolbox for creating and applying a set of [finite-state automata](#) for morphological analysis and generation in written German.

The automata are compiled from an SMOR-style grammar in SFST format and a corresponding [lexicon](#).

The lexicon is derived at build time from [XML sources](#) of the online dictionary “[Digitales Wörterbuch der deutschen Sprache](#)” (DWDS).

DWDS articles sources contain, in principle, all of the required information: lemma spellings, part-of-speech classes, inflected *Eckformen*, links to word-formation bases, etc.

In this way, DWDSmor also benefits from weekly updates of the DWDS reflecting the latest lexical trends in the German language.

<https://www.dwds.de>

## Components

DWDSmor has the following components:

- a [morphological grammar](#) of written German in [SFST format](#)

## Components

DWDSmor has the following components:

- a morphological grammar of written German in SFST format
- XSLT stylesheets and Makefiles for generating a corresponding [morphological lexicon](#) in [SFST format](#) from DWDS article sources in XML format

## Components

DWDSmor has the following components:

- a morphological grammar of written German in SFST format
- XSLT stylesheets and Makefiles for generating a corresponding morphological lexicon in SFST format from DWDS article sources in XML format
- Makefiles for compiling `automata` from the grammar and the lexicon for morphological analysis and generation in written German (`dwdsmor.a`, `dwdsmor-index.a`, `dwdsmor-root.a`)

## Components

DWDSmor has the following components:

- a morphological grammar of written German in SFST format
- XSLT stylesheets and Makefiles for generating a corresponding morphological lexicon in SFST format from DWDS article sources in XML format
- Makefiles for compiling automata from the grammar and the lexicon for morphological analysis and generation in written German (`dwdsmor.a`, `dwdsmor-index.a`, `dwdsmor-root.a`)
- a Python script for the [morphological analysis](#) of German words (`dwdsmor.py`)
- a Python script for the [generation of inflectional paradigms](#) of German words (`paradigm.py`)



## Components

DWDSmor has the following components:

- a morphological grammar of written German in SFST format
- XSLT stylesheets and Makefiles for generating a corresponding morphological lexicon in SFST format from DWDS article sources in XML format
- Makefiles for compiling automata from the grammar and the lexicon for morphological analysis and generation in written German (`dwdsмор . a`, `dwdsмор-іndex . a`, `dwdsмор-root . a`)
- a Python script for the morphological analysis of German words (`dwdsмор . py`)
- a Python script for the generation of inflectional paradigms of German words (`paradigm . py`)
- several unit tests, comparing, *inter alia*, the [coverage](#) of `dwdsмор . a` against a gold standard ([TüBa-D/Z](#))

## Development

DWDSmor is in [active development](#) and available from [GitHub](#).

<https://github.com/zentrum-lexikographie/dwdsmor>

## Development

DWDSmor is in [active development](#) and available from [GitHub](#).

Currently, the DWDSmor grammar supports [all major German inflection classes](#) as well as [some productive word-formation patterns](#).

<https://github.com/zentrum-lexikographie/dwdsmor>

## Development

DWDSmor is in active development and available from [GitHub](#).

Currently, the DWDSmor grammar supports all major German inflection classes as well as some productive word-formation patterns.

The GitHub repo provides a [sample lexicon](#) in DWDS XML format, from which DWDSmor automata with limited coverage can be compiled for testing purposes.

DWDSmor automata compiled from full DWDS sources are available on request.

<https://github.com/zentrum-lexikographie/dwdsmor>

## Institutions

DWDSmor is developed at the [Zentrum für digitale Lexikographie der deutschen Sprache \(ZDL\)](#) by Andreas Nolda and Gregor Middell.



## Institutions

DWDSmor is developed at the [Zentrum für digitale Lexikographie der deutschen Sprache \(ZDL\)](#) by Andreas Nolda and Gregor Middell.

The ZDL is a joint institution of the [Berlin-Brandenburg Academy of Sciences and Humanities \(BBAW\)](#) and the academies in Göttingen, Leipzig, and Mainz under the umbrella of the Union der deutschen Akademien der Wissenschaften.

It cooperates with the Institut für deutsche Sprache (IDS) in Mannheim and is funded by the Bundesministerium für Bildung und Forschung (BMBF).



Akademie der Wissenschaften  
zu Göttingen



Akademie der Wissenschaften  
und der Literatur | Mainz

IDS

LEIBNIZ-INSTITUT FÜR  
DEUTSCHE SPRACHE



## Institutions

DWDSmor is developed at the [Zentrum für digitale Lexikographie der deutschen Sprache \(ZDL\)](#) by Andreas Nolda and Gregor Middell.



## **Outline**

- 1 Inflection
- 2 Word formation
- 3 Applications



## Outline

- 1 Inflection
- 2 Word formation
- 3 Applications

## DWDS dictionary

🔍
?

Bank<sup>1</sup>

Bank<sup>2</sup>

### Bank, die

*Grammatik* Substantiv (Femininum) · Genitiv Singular: **Bank** · Nominativ Plural: **Bänke**

*Aussprache* 🗣️ [bʌŋk]

*Wortbildung* mit ›Bank‹ als Erstglied: ↗ [Bankdrücken](#) ... [6 weitere](#) · mit ›Bank‹ als Letztglied: ↗ [Abgeordnetenbank](#) ... [120 weitere](#)

*Mehrwortausdrücke* ↗ durch die Bank · ↗ etw. auf die lange Bank schieben

### Bedeutungsübersicht



1. Sitz für mehrere Personen nebeneinander, meist aus Holz  
[umgangssprachlich, übertragen] ...  
ohne Ausnahme
2. Handwerkstisch
3. Zusammenballung, Anhäufung
  - a) von Sand, Fels, Schlamm, Tieren in Gewässern
  - b) von Gestein, Kohle in der Erde
  - c) von Wolken, Dunst am Himmel

<https://www.dwds.de>

## DWDS dictionary

🔍
?

Bank<sup>1</sup>

Bank<sup>2</sup>

### Bank, die

*Grammatik* Substantiv (Femininum) · Genitiv Singular: **Bank** · Nominativ Plural: **Banken**

*Aussprache* 🔊 [bank]

*formal verwandt mit* ↗ [Bancomat](#), ↗ [Bankomat](#)

*Wortbildung* mit ›Bank‹ als Erstglied: ↗ [Bankabschluss](#) ... [145 weitere](#) · mit ›Bank‹ als Letztglied: ↗ [Aktienbank](#) ... [70 weitere](#)

*Mehrwortausdrücke* ↗ [eine Bank sein](#) · ↗ [sichere Bank](#) · ↗ [todsichere Bank](#)

### Bedeutungsübersicht



1. [spezieller] Unternehmen, das gewerbsmäßig Geldgeschäfte und Börsengeschäfte betreibt
  - [metonymisch] einzelne Filiale einer Bank
2. Einrichtung, in der etwas in großer Menge gesammelt, aufbewahrt, verwaltet wird und entnommen werden kann
3. [spezieller, nur Glücksspiel] ...
  - a) Geldeinsatz desjenigen Spielers, der bei Bankhalter-Spielen allein gegen alle übrigen spielt
  - b) [metonymisch] Synonym zu Bankhalter
  - c) [umgangssprachlich] (sichere Bank)
  - d) [umgangssprachlich] (eine Bank sein)

<https://www.dwds.de>

## DWDS article sources and DWDSmor lexical entries

DWDS article sources provide all the information required for DWDSmor lexical entries: lemma spellings, homograph indices, part-of-speech classes, nominal gender categories, and inflected *Eckformen*.

DWDS article source in XML format:

```
<DWDS xmlns="http://www.dwds.de/ns/1.0">
  <Artikel ...>
    <Formangabe ...>
      <Schreibung hidx="1">Bank</Schreibung>
      <Grammatik>
        <Wortklasse>Substantiv</Wortklasse>
        <Genus>fem.</Genus>
        <Genitiv>-</Genitiv>
        <Plural>Bänke</Plural>
      </Grammatik>
      ...
    </Formangabe>
    ...
  </Artikel>
</DWDS>
```

DWDSmor lexical entry in SFST format:

```
<Stem>Bank<IDX1><NN><base><native><NFem_0_<em>
```

DWDS article source in XML format:

```
<DWDS xmlns="http://www.dwds.de/ns/1.0">
  <Artikel ...>
    <Formangabe ...>
      <Schreibung hidx="2">Bank</Schreibung>
      <Grammatik>
        <Wortklasse>Substantiv</Wortklasse>
        <Genus>fem.</Genus>
        <Genitiv>-</Genitiv>
        <Plural>-en</Plural>
      </Grammatik>
      ...
    </Formangabe>
    ...
  </Artikel>
</DWDS>
```

DWDSmor lexical entry in SFST format:

```
<Stem>Bank<IDX2><NN><base><native><NFem_0_en>
```

## DWDS article sources and DWDSmor lexical entries

DWDS article sources provide all the information required for DWDSmor lexical entries: lemma spellings, [homograph indices](#), part-of-speech classes, nominal gender categories, and inflected *Eckformen*.

DWDS article source in XML format:

```
<DWDS xmlns="http://www.dwds.de/ns/1.0">
  <Artikel ...>
    <Formangabe ...>
      <Schreibung hidx="1">Bank</Schreibung>
      <Grammatik>
        <Wortklasse>Substantiv</Wortklasse>
        <Genus>fem.</Genus>
        <Genitiv>-</Genitiv>
        <Plural>Bänke</Plural>
      </Grammatik>
      ...
    </Formangabe>
    ...
  </Artikel>
</DWDS>
```

DWDSmor lexical entry in SFST format:

```
<Stem>Bank<IDX1><NN><base><native><NFem_0_<em>
```

DWDS article source in XML format:

```
<DWDS xmlns="http://www.dwds.de/ns/1.0">
  <Artikel ...>
    <Formangabe ...>
      <Schreibung hidx="2">Bank</Schreibung>
      <Grammatik>
        <Wortklasse>Substantiv</Wortklasse>
        <Genus>fem.</Genus>
        <Genitiv>-</Genitiv>
        <Plural>-en</Plural>
      </Grammatik>
      ...
    </Formangabe>
    ...
  </Artikel>
</DWDS>
```

DWDSmor lexical entry in SFST format:

```
<Stem>Bank<IDX2><NN><base><native><NFem_0_en>
```

## DWDS article sources and DWDSmor lexical entries

DWDS article sources provide all the information required for DWDSmor lexical entries: lemma spellings, homograph indices, **part-of-speech classes**, nominal gender categories, and inflected *Eckformen*.

DWDS article source in XML format:

```
<DWDS xmlns="http://www.dwds.de/ns/1.0">
  <Artikel ...>
    <Formangabe ...>
      <Schreibung hidx="1">Bank</Schreibung>
      <Grammatik>
        <Wortklasse>Substantiv</Wortklasse>
        <Genus>fem.</Genus>
        <Genitiv>-</Genitiv>
        <Plural>Bänke</Plural>
      </Grammatik>
      ...
    </Formangabe>
    ...
  </Artikel>
</DWDS>
```

DWDSmor lexical entry in SFST format:

```
<Stem>Bank<IDX1><NN><base><native><NFem_0_>e
```

DWDS article source in XML format:

```
<DWDS xmlns="http://www.dwds.de/ns/1.0">
  <Artikel ...>
    <Formangabe ...>
      <Schreibung hidx="2">Bank</Schreibung>
      <Grammatik>
        <Wortklasse>Substantiv</Wortklasse>
        <Genus>fem.</Genus>
        <Genitiv>-</Genitiv>
        <Plural>-en</Plural>
      </Grammatik>
      ...
    </Formangabe>
    ...
  </Artikel>
</DWDS>
```

DWDSmor lexical entry in SFST format:

```
<Stem>Bank<IDX2><NN><base><native><NFem_0_en>
```

## DWDS article sources and DWDSmor lexical entries

DWDS article sources provide all the information required for DWDSmor lexical entries: lemma spellings, homograph indices, part-of-speech classes, [nominal gender categories](#), and [inflected \*Eckformen\*](#).

DWDS article source in XML format:

```
<DWDS xmlns="http://www.dwds.de/ns/1.0">
  <Artikel ...>
    <Formangabe ...>
      <Schreibung hidx="1">Bank</Schreibung>
      <Grammatik>
        <Wortklasse>Substantiv</Wortklasse>
        <Genus>fem.</Genus>
        <Genitiv>-</Genitiv>
        <Plural>Bänke</Plural>
      </Grammatik>
      ...
    </Formangabe>
    ...
  </Artikel>
</DWDS>
```

DWDSmor lexical entry in SFST format:

```
<Stem>Bank<IDX1><NN><base><native><NFem_0_<em>
```

DWDS article source in XML format:

```
<DWDS xmlns="http://www.dwds.de/ns/1.0">
  <Artikel ...>
    <Formangabe ...>
      <Schreibung hidx="2">Bank</Schreibung>
      <Grammatik>
        <Wortklasse>Substantiv</Wortklasse>
        <Genus>fem.</Genus>
        <Genitiv>-</Genitiv>
        <Plural>-en</Plural>
      </Grammatik>
      ...
    </Formangabe>
    ...
  </Artikel>
</DWDS>
```

DWDSmor lexical entry in SFST format:

```
<Stem>Bank<IDX2><NN><base><native><NFem_0_<en>
```

### Inflection classes

DWDSmor lexical entries are derived from DWDS article sources by XSLT stylesheets, which map DWDS lexical information to SMOR inflection classes:

```
<xsl:variable name="noun-class-mapping">
  ...
  <class gender="fem."
    genitive-singular="-"
    nominative-plural="-e">NFem_0_§e</class>
  <class gender="fem."
    genitive-singular="-"
    nominative-plural="-en">NFem_0_en</class>
  ...
</xsl:variable>
```



### Inflection classes

DWDSmor lexical entries are derived from DWDS article sources by XSLT stylesheets, which map DWDS lexical information to SMOR inflection classes:

```
<xsl:variable name="noun-class-mapping">
  ...
  <class gender="fem."
        genitive-singular="-"
        nominative-plural="-e">NFem_0_§e</class>
  <class gender="fem."
        genitive-singular="-"
        nominative-plural="-en">NFem_0_en</class>
  ...
</xsl:variable>
```

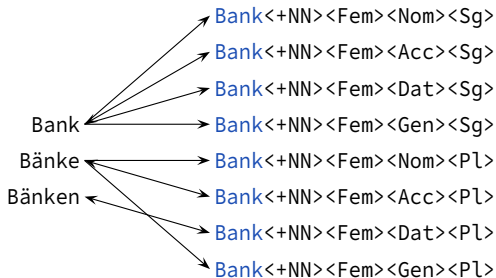
### Inflection classes

DWDSmor lexical entries are derived from DWDS article sources by XSLT stylesheets, which map DWDS lexical information to SMOR inflection classes:

```
<xsl:variable name="noun-class-mapping">
  ...
  <class gender="fem."
    genitive-singular="-"
    nominative-plural="-e">NFem_0_§e</class>
  <class gender="fem."
    genitive-singular="-"
    nominative-plural="-en">NFem_0_en</class>
  ...
</xsl:variable>
```

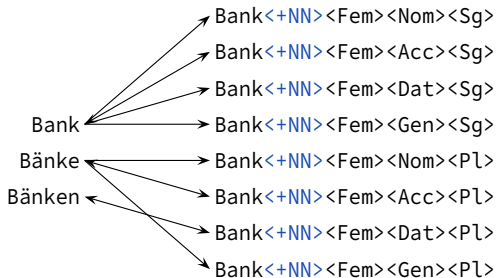
## Inflection in DWDSmor

The DWDSmor automata relate surface forms to analysis strings, which provide a [lemma](#), a part-of-speech class, and a set of morphosyntactic categories:



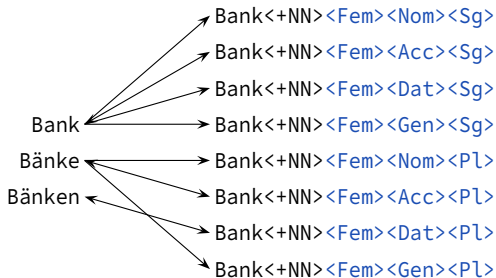
## Inflection in DWDSmor

The DWDSmor automata relate surface forms to analysis strings, which provide a lemma, a *part-of-speech class*, and a set of morphosyntactic categories:



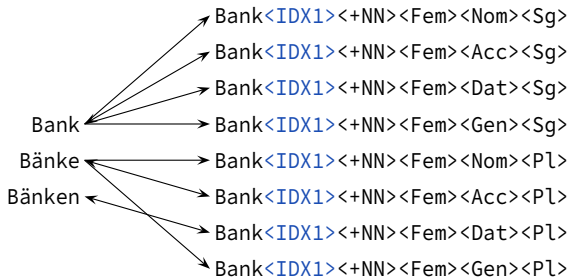
## Inflection in DWDSmor

The DWDSmor automata relate surface forms to analysis strings, which provide a lemma, a part-of-speech class, and a set of **morphosyntactic categories**:



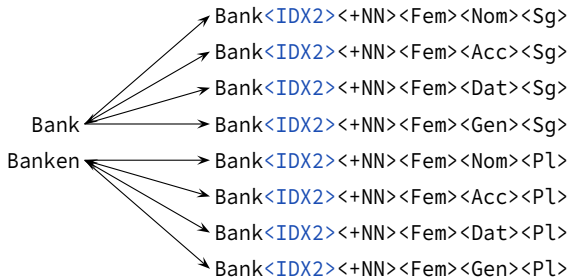
## Inflection in DWDSmor

The analysis strings of the automaton `dwdsmor-index.a` also provide [homograph indices](#) (if any):



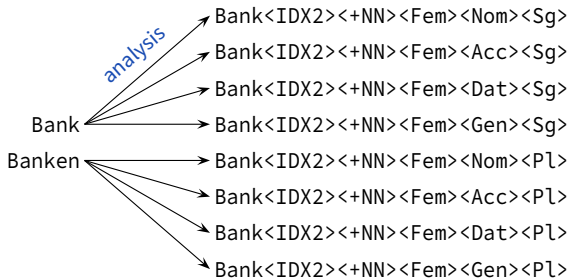
## Inflection in DWDSmor

The analysis strings of the automaton `dwdsmor-index.a` also provide [homograph indices](#) (if any):



## Analysis and generation

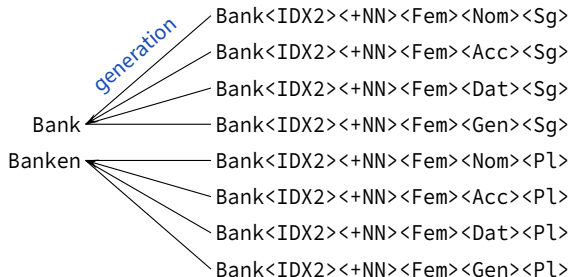
In *analysis mode*, the relation is read from left to right. In *generation mode*, it is read from right to left.





## Analysis and generation

In analysis mode, the relation is read from left to right. In **generation mode**, it is read from right to left.



## Outline

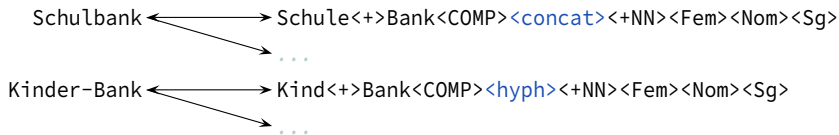
- 1 Inflection
- 2 Word formation
- 3 Applications





## Word formation in DWDSmor

The DWDSmor automaton `dwdsmor-root.a` analyses word-formation products in terms of base lemmas, formation processes, and **formation means**, in the sense of the Pattern-and-Restriction Theory (Nolda 2022):



Nolda, Andreas (2022). Headedness as an epiphenomenon: Case studies on compounding and blending in German. In *Headedness and/or Grammatical Anarchy?*, ed. by Ulrike Freywald, Horst Simon, and Stefan Müller, Empirically Oriented Theoretical Morphology and Syntax 11, Berlin: Language Science Press, 343–376. <https://zenodo.org/record/7142720/files/336-FreywaldSimonMüller-2022-11.pdf> [23 Oct. 2022].



## Word-formation stems

Compounding stems with or without linking elements can be inferred from links to word-formation bases in DWDS article sources.

DWDS article source in XML format:

```
<DWDS xmlns="http://www.dwds.de/ns/1.0">
  <Artikel ...>
    <Formangabe ...>
      <Schreibung>Schulbuch</Schreibung>
      ...
    </Formangabe>
    <Verweise Typ="Komposition" ...>
      <Verweis Typ="Erstglied">
        <Ziellemma>Schule</Ziellemma>
        ...
      </Verweis>
      <Verweis Typ="Letztglied">
        <Ziellemma>Buch</Ziellemma>
        ...
      </Verweis>
    </Verweise>
    ...
  </Artikel>
</DWDS>
```

DWDSmor lexical entry in SFST format:

```
<Stem>Schule:<><NN><comp><native>
```

DWDS article source in XML format:

```
<DWDS xmlns="http://www.dwds.de/ns/1.0">
  <Artikel ...>
    <Formangabe ...>
      <Schreibung>Kinderbuch</Schreibung>
      ...
    </Formangabe>
    <Verweise Typ="Komposition" ...>
      <Verweis Typ="Erstglied">
        <Ziellemma>Kind</Ziellemma>
        ...
      </Verweis>
      <Verweis Typ="Letztglied">
        <Ziellemma>Buch</Ziellemma>
        ...
      </Verweis>
    </Verweise>
    ...
  </Artikel>
</DWDS>
```

DWDSmor lexical entry in SFST format:

```
<Stem>Kind:<:<FB><:<e>:<r><NN><comp><native>
```

## Outline

- 1 Inflection
- 2 Word formation
- 3 Applications



## DWDS corpora

"auf die Bank" 🔍 ? 🔄

**Korpus:** Referenz- und Zeitungskorpus

**Start:** 1994 **Ende:** 2005

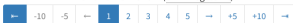
**Textklassen:**  Belletristik  Wissenschaft  Gebrauchsliteratur  Zeitung

**Anzeige:**  KWIC  voll  maximal

**Sortierung:** Datum absteigend

**Treffer pro Seite:** 10

1-10 von 3728 Treffern (3741 insgesamt)



- 1: Berliner Zeitung, 19.12.2005  
In der Schlussphase sei **auf der Bank** die "Panik auf der Titanic" ausgebrochen, berichtete Hoeneß.
- 2: Berliner Zeitung, 19.12.2005  
Vier Wochen später hieß es schon: "Köln im Keller", und Rapolder, der es gewagt hatte, den Stadtheiligen Lukas Podolski gegen Hannover **auf die Bank** zu setzen, war spätestens von da an nur noch ein Trainer auf Bewährung.
- 3: Die Zeit, 15.12.2005, Nr. 51  
Die Zuhörer im großen Saal der Svenska Akademien, der sich über die Breite eines schönen Barockbaus in Stockholms Altstadt Gamla Stan dahinzieht, applaudierten mit kontrolliertem Nachdruck, vielleicht, weil sie die Gefühle bändigen wollten, die diese Rede in ihnen ausgelöst hatte, die sie **auf den Bänken** mit den Plüschpolstern unter den großen Lüstern sitzen konnten und auf die Leinwände schauten, die zwischen den goldgeschmückten Säulen des Saales hingen und auf denen der alte Mann zu sehen war.
- 4: Die Zeit, 08.12.2005, Nr. 50  
Ex-Arbeitsminister Walter Riester setzte sich im Reichstag **auf die Bänke** der Linken, um mit alten Bekannten von der Gewerkschaft zu plaudern.
- 5: Berliner Zeitung, 06.12.2005  
Goran Ivanisevic, der sich als vierter Mann des Teams **auf der Bank** heiser brüllte bei der Unterstützung von Ljubovic und Ancic, sagt, er sei bereit, die Rolle des Teamchefs bald zu übernehmen.

<https://www.dwds.de/r>

## Corpus annotation

Output of `dwdsmor.py` for *auf den Bänken*:

<i>Wordform</i>	<i>Lemma</i>	<i>POS</i>	<i>Gender</i>	<i>Case</i>	<i>Number</i>	<i>Inflection</i>	<i>Function</i>	<i>...</i>
auf	auf	ADV						
auf	auf	PREP						
den	die	REL	Masc	Acc	Sg	St	Subst	
den	die	DEM	Masc	Acc	Sg	St	Subst	
den	die	DEM	NoGend	Dat	Pl	St	Attr	
den	die	DEM	Masc	Acc	Sg	St	Attr	
den	die	ART	Masc	Acc	Sg	St	Subst	
den	die	ART	NoGend	Dat	Pl	St	Attr	
den	die	ART	Masc	Acc	Sg	St	Attr	
Bänken	Bank	NN	Fem	Dat	Pl			

## Corpus annotation

Output of `dwdsмор.py` for *auf den Bänken*:

Wordform	Lemma	POS	Gender	Case	Number	Inflection	Function	...
auf	auf	ADV						
auf	auf	PREP						
den	die	REL	Masc	Acc	Sg	St	Subst	
den	die	DEM	Masc	Acc	Sg	St	Subst	
den	die	DEM	NoGend	Dat	Pl	St	Attr	
den	die	DEM	Masc	Acc	Sg	St	Attr	
den	die	ART	Masc	Acc	Sg	St	Subst	
den	die	ART	NoGend	Dat	Pl	St	Attr	
den	die	ART	Masc	Acc	Sg	St	Attr	
Bänken	Bank	NN	Fem	Dat	Pl			

supplementing tagger output with morphosyntactic information:

```
<w lemma="auf" pos="APPR">auf</w>
<w lemma="die" pos="ART" msd="NoGend.Dat.Pl.St.Attr">den</w>
<w lemma="Bank" pos="NN" msd="Fem.Dat.Pl">Bänken</w>
```

## Inflectional paradigms

### Deklinationstabelle für »Bank<sup>1</sup>« (Substantiv)

Genus: Femininum

	Singular	Plural
<b>Nominativ</b>	Bank	Bänke
<b>Akkusativ</b>	Bank	Bänke
<b>Dativ</b>	Bank	Bänken
<b>Genitiv</b>	Bank	Bänke

Output of `paradigm.py --lemma-index 1` for *Bank*:

```
Case Number Paradigm Forms ...
Nom Sg Bank
Acc Sg Bank
Dat Sg Bank
Gen Sg Bank
Nom Pl Bänke
Acc Pl Bänke
Dat Pl Bänken
Gen Pl Bänke
```

## Inflectional paradigms

### Deklinationstabelle für »Bank<sup>2</sup>« (Substantiv)

Genus: Femininum

	Singular	Plural
<b>Nominativ</b>	Bank	Banken
<b>Akkusativ</b>	Bank	Banken
<b>Dativ</b>	Bank	Banken
<b>Genitiv</b>	Bank	Banken

Output of `paradigm.py --lemma-index 2` for *Bank*:

```
Case Number Paradigm Forms ...
Nom Sg Bank
Acc Sg Bank
Dat Sg Bank
Gen Sg Bank
Nom Pl Banken
Acc Pl Banken
Dat Pl Banken
Gen Pl Banken
```

## Outline

- 1 Inflection
- 2 Word formation
- 3 Applications

Coverage test

## Nominal part-of-speech classes in TüBa-D/Z

POS tag	interpretation	coverage
NN	ordinary nouns	90.31 %
NE	proper names	7.74 %
ADJA	prenominal adjectives	89.21 %
ADJD	other adjectives	92.78 %
CARD	cardinals	90.84 %
ART	articles	99.86 %
PDAT	prenominal demonstrative pronouns	96.39 %
PDS	other demonstrative pronouns	97.54 %
PPOSAT	prenominal possessive pronouns	99.11 %
PPOSS	other possessive pronouns	100.00 %
PPER	personal pronouns	98.87 %
PRF	reflexive pronouns	99.50 %
PRELAT	prenominal relative pronouns	99.84 %
PRELS	other relative pronouns	95.08 %
PIAT	prenominal indefinite pronouns	95.65 %
PIDAT	prenominal indefinite prenominal with determiner	99.35 %
PIS	other indefinite pronouns	96.76 %
PWAT	prenominal interrogative pronouns	89.78 %
PWAV	adverbial interrogative or relative pronouns	99.60 %
PWS	other interrogative pronouns	98.99 %

as of 26 June 2023

## Verbal part-of-speech classes in TüBa-D/Z

POS tag	interpretation	coverage
VVFIN	non-auxiliary finite verbs	93.89 %
VVIMP	non-auxiliary imperative verbs	90.65 %
VVINFINF	non-auxiliary infinitives	92.39 %
VVIZU	non-auxiliary infinitives with <i>zu</i>	80.74 %
VVPP	non-auxiliary past participles	90.16 %
VMFIN	finite modal verbs	99.85 %
VMINF	modal infinitives	100.00 %
VMPP	modal past participles	100.00 %
VAFIN	finite auxiliary verbs	99.78 %
VAIMP	imperative auxiliary verbs	100.00 %
VAINFINF	auxiliary infinitives	99.89 %
VAPP	auxiliary past participles	99.92 %

as of 26 June 2023



## Other part-of-speech classes in TüBa-D/Z

POS tag	interpretation	coverage
ADV	adverbs	97.09 %
PROP	pronominal adverbs	98.50 %
APPR	prepositions	99.63 %
APPRART	contracted prepositions and articles	99.73 %
APPO	postpositions	91.96 %
APZR	right part of circumpositions	94.63 %
KOUI	subordinating conjunctions for infinitives with <i>zu</i>	99.92 %
KOUS	subordinating conjunctions for clauses	99.44 %
KON	coordinating conjunctions	97.93 %
KOKOM	comparative particles	99.78 %
PTKNEG	negation particles	99.85 %
PTKVZ	preverbal particles	96.04 %
PTKA	preadjectival particles <i>am</i> and <i>zu</i>	100.00 %
PTKZU	pre-infinitival particle <i>zu</i>	99.99 %
PTKANT	response particles	96.50 %
ITJ	interjections	65.21 %
FM	foreign material	0.00 %
TRUNC	elliptical word parts	0.00 %
XY	non-words	0.00 %
\$.	sentence-final punctuation characters	100.00 %
\$,	commas	100.00 %
\$(	other punctuation characters	99.06 %

as of 26 June 2023

## Summary

domain	coverage
all tokens	90.84 %
tokens without POS tags NE, FM, XY, and TRUNC	95.94 %

as of 26 June 2023