

TEI Format zur Annotation der sorbischen Korpora

Dieses Dokument beschreibt die Anwendung des TEI-P5-Formats als Modell und zur Kodierung der Inhalte der sorbischen Textkorpora.

Inhaltsverzeichnis

Ziel.....	2	div/@type.....	14
Stufen der strukturellen Kodierung.....	2	Grundsätzliche Blockelemente des Textes.....	14
Stufe 1: Unstrukturierter Text in TEI-XML-Container.....	2	head.....	14
Stufe 2: Text mit minimaler TEI-Struktur.....	2	head/@type.....	14
Stufe 3: Nach TEI strukturierter Text mit definierten Ausnahmen.....	3	byline.....	15
Stufe 4: Vollständig nach TEI strukturierter Text mit offenen Problemen.....	3	dateline.....	15
Stufe 5: Vollständig nach TEI strukturierter Text.....	3	epigraph.....	15
Übersicht.....	3	p.....	15
Nicht kodierte Angaben.....	4	Sonder(block)elemente des Textes.....	15
Elemente der Korpusebene.....	4	Bild.....	15
teiCorpus.....	4	figure.....	15
Elemente der Dokumentenebene.....	5	Gedicht.....	16
TEI.....	5	lg.....	16
TEI/@xml:id.....	5	lg/@type.....	16
teiHeader.....	5	l.....	16
Elemente des TEI-Headers.....	5	Liste.....	16
fileDesc.....	5	list.....	16
titleStmt.....	6	label.....	17
title.....	6	item.....	17
author.....	6	Tabelle.....	17
funder.....	6	table.....	17
publicationStmt.....	6	row.....	17
distributor.....	6	row/@role.....	17
sourceDesc.....	6	cell.....	18
encodingDesc.....	7	cell/@cols.....	18
projectDesc.....	7	cell/@rows.....	18
editorialDecl.....	7	Zitat.....	18
hyphenation.....	7	cit.....	18
hyphenation/@eol.....	7	quote.....	18
profileDesc.....	7	bibl.....	18
langUsage.....	7	Unstrukturierter Text.....	18
language.....	8	Elemente der Textformatierung.....	19
language/@ident.....	8	hi.....	19
textClass.....	8	hi/@rend.....	19
classCode.....	8	lb.....	19
classCode/@scheme.....	8	pb.....	19
revisionDesc.....	8	pb/@n.....	20
revisionDesc/@status.....	9	Vertikaler Abstand oder (Trenn-)Ornament.....	20
change.....	9	Sonstige besondere Elemente.....	20
change/@when.....	9	note.....	20
change/@who.....	10	note/@type.....	21
change/@status.....	10	note/@subtype.....	21
Struktur des Textes.....	10	note/@rend.....	21
text.....	10	note/@resp.....	21
text/@xml:lang.....	10	stage.....	22
front.....	10	ref.....	22
body.....	11	ref/@type.....	22
back.....	11	foreign.....	22
Elemente des Titelblatts.....	11	foreign/@xml:lang.....	23
titlePage.....	11	space.....	23
docTitle.....	11	Annotation der Sprache.....	23
titlePart.....	12	Textanalyse.....	24
titlePart/@type.....	12	w.....	24
docAuthor.....	12	w/@type.....	24
docDate.....	12	w/@subtype.....	24
docEdition.....	12	w/@orig.....	25
docImprint.....	12	w/@norm.....	25
byline.....	12	w/@lemma.....	25
Textaufteilung.....	13	pc.....	26
div.....	13	Änderungsprotokoll.....	27

Ziel

Das Ziel besteht im Allgemein darin, eine physische Textausgabe in ein digitales Textdokument (als reiner Unicode-Text ohne Formatierung) zu transformieren. Die semantisch bedeutungstragenden (typo)graphischen Phänomene werden mithilfe von entsprechenden XML-Strukturen kodiert.

Die XML-Kodierung wird nach den TEI Richtlinien¹ durchgeführt. Die TEI-Infrastruktur wurde nicht deshalb ausgewählt, weil sie optimal den Anforderungen entspräche (das tut sie leider nicht, manchmal reicht sie sogar nicht aus), sondern weil sie ein empfohlener Standard² ist. Aus dieser Sachlage ergibt sich eine Reihe von Schwierigkeiten, denen nur durch Kompromisslösungen zu beugen ist.

Stufen der strukturellen Kodierung

Bei der strukturellen Kodierung werden fünf Zustände (Stufen) definiert. Jeder der aufeinander folgenden Zustände repräsentiert einen standardmäßig kodierten Text, aber mit unterschiedlichem Detaillierungsgrad. Abhängig von Bedarf und möglichem Arbeitsaufwand wird ein erwünschter Zustand erreicht und dokumentiert.

Stufe 1: Unstrukturierter Text in TEI-XML-Container

Der Text genügt dem TEI-Schema, aber nur formell: die TEI-Kopfzeile (`/TEI/teiHeader`) ist vollständig, die Annotationen sind maximal vereinfacht und umfassen nur die undifferenzierten `p`-Tags unter dem `/TEI/text/body`-Tag. Die `p`-Tags (die nach dem richtigen TEI Absätze markieren) dienen als provisorische XML-Klammern für einen eigentlich unstrukturierten Text (für jeweils günstige Textmengen, z.B. einen Absatz, eine Zeile, einen Titel usw.).

Dieses provisorische Format ist nützlich für eine erste, teilweise automatisierte XML-Herstellung am Anfang der Bearbeitung.

Stufe 2: Text mit minimaler TEI-Struktur

Der Text hat die dem TEI-Format entsprechende XML-Struktur, aber berücksichtigt – neben der obligatorischen TEI-Kopfzeile – nur die Elemente, die geläufige Textbestandteile abgrenzen und benennen. Das bedeutet insbesondere, dass:

- Die Hauptstruktur des Textes (`text/front`, `text/body`, `text/back`) wird strukturiert.
- Die Artikel in einer Zeitung resp. die Texte in einer Textsammlung werden getrennt (`div[@type='article']` resp. `div[@type='work']`).
- Jeder Text resp. Artikel hat seinen Titel (`head`) und ggf. seinen Autor (`byline`).

1 TEI: P5 Guidelines, <http://www.tei-c.org/Guidelines/P5>

2 DFG-Praxisregeln „Digitalisierung“ [02/13], S. 28 (2.3.2 Strukturelle Metadaten für digitale Faksimiles), http://www.dfg.de/formulare/12_151/12_151_de.pdf

- Die Bilder und Bildunterschriften werden annotiert ([figure](#)).
- Die Seitennummerierung und Textflussbearbeitung werden angewandt.
- Umschlagtexte und -bilder können (explizit) ausgelassen werden.

Mit diesem Format ist ein Text fast vollständig nutzbar (die einzelnen Bestandteile lassen sich identifizieren und indexieren und der Textinhalt lässt sich durchsuchen), obwohl eine einwandfreie Präsentation (Textausgabe) unmöglich ist.

Stufe 3: Nach TEI strukturierter Text mit definierten Ausnahmen

Der Text verwendet alle für das Korpus vorgesehenen TEI-Elemente. Nur die in diesen Richtlinien explizit aufgezählten Ausnahmen sind möglich (z.B. eine komplexe Tabelle, ein Inserat oder mit dem Text auf ungünstige Weise verschränkte Zeichnungen) – diese Elemente bleiben unstrukturiert und werden als solche klar markiert (siehe Kapitel »Unstrukturierter Text«).

Eine besondere Ausnahme gilt für die nicht-strukturellen Elemente [hi](#), [foreign](#) und [ref](#) – auf den Stufen 3-5 sind sie nicht obligatorisch. Wo es technisch bzw. zeitökonomisch begründet ist, sollen sie jedoch annotiert werden, da das in Zukunft Arbeit spart.

Diese Bearbeitungsstufe ist arbeitsökonomisch begründet und spart Zeit z.B. bei großen numerischen Tabellen. Die Auslassungen entstehen aber nicht immer aus Sparsamkeit; man kann auch ganz unklaren oder anderen äußerst komplexen Strukturen begegnen, die sich nicht sinnvoll kodieren lassen und/oder eine ansonsten unerwünschte Modellerweiterung erfordern würden.

Stufe 4: Vollständig nach TEI strukturierter Text mit offenen Problemen

Der kodierte Text entspricht vollständig dem TEI-Modell und den TEI-Guidelines (im Umfang der in diesem Dokument gegebenen Spezifikation). Alle strukturierbaren Elemente wurden strukturiert. Dies schließt aber nicht aus, dass es im Text noch Anmerkungen ([note\[@typ='encoders-comment'\]](#)) gibt, die auf die Notwendigkeit weiterer Bearbeitung oder kleiner Textergänzungen verweisen. Diese Stufe erlaubt jedoch keine wesentlichen Strukturprobleme oder Textmängel mehr.

Stufe 5: Vollständig nach TEI strukturierter Text

Der Text ist in jeglicher Hinsicht fertig und benötigt keine weitere Bearbeitung, solange die Richtlinien unverändert bleiben.

Übersicht

In diesem Dokument werden die für Benutzung im Korpus ausgewählten TEI-Elemente

aufgeführt und unsere eigenen Sonderlösungen erläutert. Ausführlichere Vorschriften, eingehendere Gebrauchsanweisungen und Beispiele sind in den online veröffentlichten TEI-Richtlinien selbst aufzusuchen.

Bei jedem Element werden zusammengefasst: das entsprechende TEI-Modul, die bisher benutzten Unter- und Überelemente und der Verweis auf die „TEI: P5 Guidelines“ oder andere relevante Dokumentationen.

Das Schema ist (mit Absicht) sehr vereinfacht und benutzt für die Modellierung des eigentlichen Textes nur die folgenden Elemente: [bibl](#), [byline](#), [cell](#), [cit](#), [dateline](#), [div](#), [epigraph](#), [figure](#), [foreign](#), [head](#), [hi](#), [item](#), [l](#), [label](#), [lb](#), [lg](#), [list](#), [note](#), [p](#), [pb](#), [quote](#), [ref](#), [row](#), [space](#), [table](#) sowie die Attribute: [n](#), [rend](#), [resp](#), [role](#), [type](#), [subtype](#), [xml:lang](#). Um die Bearbeitung zu erleichtern, wird hier eine modifizierte Version des TEI-Schemas benutzt, die nur relevante Strukturen umfasst.

Sowohl die Auswahl der Elemente als auch die Dokumentation kann noch erweitert und vervollständigt werden.

Nicht kodierte Angaben

Die folgenden Angaben werden nicht kodiert (und sind daher in der digitalen Version nicht enthalten):

- Die Worttrennung (die Silbentrennung) wird nicht übernommen – die Wortteile werden vereinigt und das Trennzeichen wird gelöscht.
- Die Zeilentrennung (z.B. in einem Absatz oder Titel) wird nicht beachtet. Zu seltenen Ausnahmen siehe unter dem Element [lb](#). Zu Gedichtversen siehe unter dem Element [l](#). In XML wird eine Zeilentrennung (Zeilenumbruchzeichen) so wie ein Leerzeichen behandelt.
- Die Spaltentrennung (oder z.B. ein besonderer Spaltenumriss) wird nicht beachtet.
- Die Seitentrennung wird kodiert, aber nur als eine Positionsangabe (siehe unter dem Element [pb](#)).
- Sich wiederholende Textelemente am Seitenrand, -kopf oder -fuß werden nicht übernommen. Das betrifft auch Seitenzahlen (siehe oben).
- Außerordentliche Trennelemente zwischen den Textblöcken (z.B. ein großer Abstand zwischen den Absätzen, eine horizontale Linie oder andere Ornamente) werden ignoriert.
- Farben (z.B. des Textes) werden nicht kodiert, sofern sie keine Funktion haben. Über die Hervorhebung siehe unter dem Element [hi](#).
- Der Inhalt von Bildern und anderen graphischen Elementen wird nicht kodiert.

Elemente der Korpusebene

teiCorpus

TEI-Modul	core – Elements Available in All TEI Documents
Unterelement von	—
Überelement für	teiHeader , TEI
Beschreibung	4 Default Text Structure http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html

Ein technisches Mittel, um mehrere Korpusdokumente zusammenzufügen (wenn in einer XML-Datei mehrere Dokumente enthalten sein sollen).

Hat keine semantische Funktion (wie z.B. ein Umschlag für eine Serie oder einen Jahrgang).

Elemente der Dokumentenebene

(Die genaue Definition eines Korpusdokuments wird noch benötigt.)

TEI

TEI-Modul	textstructure – Default Text Structure
Unterelement von	teiCorpus , —
Überelement für	teiHeader , text
Beschreibung	4 Default Text Structure • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html

Umfasst ein einzelnes und vollständiges Korpusdokument, z.B. eine einzelne Buch- oder Zeitungsausgabe, die einem physischen Objekt entspricht.

Auch Beilagen (wie z.B. ein zyklischer, betitelter und nummerierter Zeitungsanhang) können als selbstständige Dokumente angesehen werden.

TEI/@xml:id

TEI-Modul	tei – The TEI Infrastructure
Beschreibung	1.3.1.1 Global Attributes • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ST.html#STGA

Ein obligatorischer, eindeutiger und einzigartiger Identifikator für das betreffende Korpusdokument.

Wird nur vom Korpuskoordinator vergeben.

teiHeader

TEI-Modul	header – The TEI Header
Unterelement von	teiCorpus , TEI
Überelement für	encodingDesc , fileDesc , profileDesc , revisionDesc
Beschreibung	2 The TEI Header • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html

Umfasst verschiedene Metaangaben.

Normalerweise sollten die Metaangaben an anderer Stelle (nicht im Dokument selbst) gespeichert werden, aber das [teiHeader](#)-Element ist obligatorisch und darf nicht leer bleiben. Die hier vorgeschlagene Struktur stellt den minimalen informativen Inhalt dieses Elementes dar.

Elemente des TEI-Headers

fileDesc

TEI-Modul	header – The TEI Header
Unterelement von	teiHeader
Überelement für	titleStmt , publicationStmt , sourceDesc
Beschreibung	2.2 The File Description • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html#HD2

Angaben zur digitalen Version des Korpusdokuments.

titleStmt

TEI-Modul	header – The TEI Header
Unterelement von	fileDesc
Überelement für	title , author , funder
Beschreibung	2.2.1 The Title Statement ◀ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html#HD21

Eine einsichtige Kennzeichnung des Dokuments.

title

TEI-Modul	core – Elements Available in All TEI Documents
Unterelement von	titleStmt
Überelement für	#PCDATA
Beschreibung	2.2.1 The Title Statement ◀ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html#HD21

Hier wird der Titel des Werks aufgenommen.

author

TEI-Modul	core – Elements Available in All TEI Documents
Unterelement von	titleStmt
Überelement für	#PCDATA
Beschreibung	2.2.1 The Title Statement ◀ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html#HD21

Hier wird der Autor des Werks eingetragen.

funder

TEI-Modul	header – The TEI Header
Unterelement von	titleStmt
Überelement für	#PCDATA
Beschreibung	2.2.1 The Title Statement ◀ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html#HD21

Nur unter [/teiCorpus/teiHeader/fileDesc/titleStmt](#) verwendet.

publicationStmt

TEI-Modul	header – The TEI Header
Unterelement von	fileDesc
Überelement für	distributor
Beschreibung	2.2.4 Publication, Distribution, Licensing, etc. ◀ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html#HD24

distributor

TEI-Modul	header – The TEI Header
Unterelement von	publicationStmt
Überelement für	#PCDATA
Beschreibung	2.2.4 Publication, Distribution, Licensing, etc. ◀ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html#HD24

Der Name des Sorbischen Instituts.

sourceDesc

TEI-Modul	header – The TEI Header
Unterelement von	fileDesc
Überelement für	bibl
Beschreibung	2.2.7 The Source Description ◀ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html#HD3

Ein Verweis auf die (analoge) Quelle, z.B. eine bibliographische Angabe, möglicherweise

aus der »Serbska bibliografija = Sorbische Bibliographie« übernommen.

encodingDesc

TEI-Modul	header – The TEI Header
Unterelement von	teiHeader
Überelement für	projectDesc , editorialDecl
Beschreibung	2.3 The Encoding Description • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html#HD5

Technische Beschreibung des Projekts und der Textbearbeitung.

projectDesc

TEI-Modul	header – The TEI Header
Unterelement von	encodingDesc
Überelement für	p
Beschreibung	2.3.1 The Project Description • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html#HD51

Eine allgemeine Notiz über das Korpusprojekt.

editorialDecl

TEI-Modul	header – The TEI Header
Unterelement von	encodingDesc
Überelement für	hyphenation , p
Beschreibung	2.3.3 The Editorial Practices Declaration • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html#HD53

Umfasst technische Bearbeitungsinformationen und -bemerkungen, die sich auf das ganze Dokument beziehen.

Es scheint, dass die nicht normierten Bemerkungen nur als ein [p](#)-Unterelement kodiert werden dürfen.

hyphenation

TEI-Modul	header – The TEI Header
Unterelement von	editorialDecl
Überelement für	p
Beschreibung	2.3.3 The Editorial Practices Declaration • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html#HD53

Eine normierte Angabe über die Behandlung der Silbentrennung.

hyphenation/@eol

TEI-Modul	header – The TEI Header
Beschreibung	hyphenation • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-hyphenation.html

Genutzter Wert: *none*.

profileDesc

TEI-Modul	header – The TEI Header
Unterelement von	teiHeader
Überelement für	langUsage , textClass
Beschreibung	2.4 The Profile Description • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html#HD4

Klassifikation des Dokuments.

langUsage

TEI-Modul	header – The TEI Header
Unterelement von	profileDesc

Überelement für	language
Beschreibung	2.4.2 Language Usage • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html#HD41

Zählt die Hauptsprachen des Dokuments auf (normalerweise Nieder- bzw. Obersorbisch und häufig auch Deutsch).

Die Sprachen, die in einzelnen kurzen Einschüben, Zitaten u.ä. verwendet werden, werden hier nicht angeführt.

Siehe auch Element [foreign](#) und Kapitel »Annotation der Sprache«.

language

TEI-Modul	header – The TEI Header
Unterelement von	langUsage
Überelement für	#PCDATA
Beschreibung	2.4.2 Language Usage • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html#HD41

Der Name der Sprache (auf Englisch).

language/@ident

TEI-Modul	header – The TEI Header
Beschreibung	data.language • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-data.language.html

Der Sprachcode nach ISO-639-3 (siehe [foreign/@xml:lang](#)).

textClass

TEI-Modul	header – The TEI Header
Unterelement von	profileDesc
Überelement für	classCode
Beschreibung	2.4.3 The Text Classification • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html#HD43

Das Genre des Dokuments.

classCode

TEI-Modul	header – The TEI Header
Unterelement von	textClass
Überelement für	#PCDATA
Beschreibung	2.4.3 The Text Classification • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html#HD43

Derzeit wird provisorisch die Genre-Klassifikation »DWDS-Kategorie 1«³ genutzt.

classCode/@scheme

TEI-Modul	header – The TEI Header
Beschreibung	data.pointer • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-data.pointer.html

Der Wert für »DWDS-Kategorie 1« ist:

<http://www.deutschestextarchiv.de/doku/klassifikation#dwds1main>.

revisionDesc

TEI-Modul	header – The TEI Header
Unterelement von	teiHeader
Überelement für	change
Beschreibung	2.6 The Revision Description • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html#HD6

Das Bearbeitungstagebuch.

³ <http://www.deutschestextarchiv.de/doku/klassifikation>

Die erste Eingabe in der **change**-Liste soll die Quelle des digitalen Textes beschreiben, z.B.:

```
<revisionDesc status="structure-4 lemmatized">
  <change when="2015-04-23" who="Grepect">scanns-based transcription</change>
  <change when="2016-08-26" who="MSI" status="structure-1">
    initial TEI-P5 encoding
  </change>
  <change when="2016-09-08" who="JSA">minor text corrections</change>
  <change when="2022-01-05" who="JSA" status="structure-4">TEI encoding</change>
  <change when="2022-01-12" who="MSA" status="structure-4 lemmatized">
    automated Corproc processing
  </change>
</revisionDesc>
```

revisionDesc/@status

TEI-Modul	tei – The TEI Infrastructure
Beschreibung	att.dataable.w3c • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.docStatus.html#tei_att.status

Das **@status**-Attribut stellt die Stufe des Bearbeitungsfortschritts fest.

Der besondere Wert – *disqualified* – bezeichnet ein fehlerhaftes Textdokument, das nicht weiter bearbeitet wird (praktisch bedeutet es, dass der Text zu entsorgen ist).

Die folgenden Werte beziehen sich auf die strukturelle Bearbeitung (nur eine wird gewählt):

- *structure-1* = Die Stufe 1 (Unstrukturierter Text in TEI-XML-Container) der strukturellen Textbearbeitung wurde erreicht.
- *structure-2* = Die Stufe 2 (Text mit minimaler TEI-Struktur) der strukturellen Textbearbeitung wurde erreicht.
- *structure-3* = Die Stufe 3 (Nach TEI strukturierter Text mit definierten Ausnahmen) der strukturellen Textbearbeitung wurde erreicht.
- *structure-4* = Die Stufe 4 (Vollständig nach TEI strukturierter Text mit offenen Fragen) der strukturellen Textbearbeitung wurde erreicht.
- *structure-5* = Die Stufe 5 (Vollständig nach TEI strukturierter Text) der strukturellen Textbearbeitung wurde erreicht.

Zusätzlich wird die linguistische Bearbeitung mit den folgenden Werten bezeichnet:

- *lemmatized* = Dokument wurde (tokenisiert und) lemmatisiert.
- *modernized* = Dokument wurde für die modernisierte Ausgabe vorbereitet.

change

TEI-Modul	header – The TEI Header
Unterelement von	revisionDesc
Überelement für	#PCDATA
Beschreibung	2.6 The Revision Description • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html#HD6

Eine kurze aber sinnvolle Beschreibung, was geändert wurde.

change/@when

TEI-Modul	tei – The TEI Infrastructure
Beschreibung	att.dataable.w3c • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.dataable.w3c.html#tei_att.when

Das Bearbeitungsdatum nach ISO-8601.
Das bevorzugte Format ist [2023-12-31T23:59:59](#).

change/@who

TEI-Modul	tei – The TEI Infrastructure
Beschreibung	att.ascribed ➤ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.ascribed.html#tei_att_who

Der Name der Bearbeiterin oder des Bearbeiters im Drei-Buchstaben-Format: der erste Buchstabe des Vornamens, der erste und letzte des Nachnamens (z.B. *MSI* für Marcin Szczepański).

change/@status

TEI-Modul	tei – The TEI Infrastructure
Beschreibung	att.dataable.w3c ➤ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.docStatus.html#tei_att_status

Das [@status](#)-Attribut zeigt, dass mit dieser Änderung das XML-Dokument die entsprechende Stufe des Bearbeitungsfortschritts erreicht hat. Die Werte sind gleich wie unter dem [revisionDesc/@status](#).

Wenn ein Arbeitsschritt den Status des Dokuments verändert, wird der [change](#)-Eintrag mit der neuen [@status](#)-Angabe ergänzt. Der erreichte Status des Dokuments wird außerdem im übergeordneten [revisionDesc/@status](#) aktualisiert.

Struktur des Textes

text

TEI-Modul	textstructure – Default Text Structure
Unterelement von	TEI
Überelement für	front , body , back
Beschreibung	4 Default Text Structure ➤ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html

Der Inhalt des Korpusdokuments.

text/@xml:lang

TEI-Modul	tei – The TEI Infrastructure
Beschreibung	att.global ➤ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.global.html#tei_att_xml-lang

Das Attribut bestimmt hier die Hauptsprache des Dokuments und ist für dieses Element obligatorisch.

Siehe auch Kapitel »Annotation der Sprache«.

front

TEI-Modul	textstructure – Default Text Structure
Unterelement von	text
Überelement für	div , titlePage , p , pb
Beschreibung	4.5 Front Matter ➤ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html#DSFRONT

Alles, was vor dem Haupttext steht: Texte und Bilder aus dem Umschlag, Titelblatt (in einem Buch), Titelpfopf (in einer Zeitung), Inhaltsverzeichnis, Vorwort des Herausgebers u.ä.

Sonderfall: ein Zeitungstitelpfopf wird hier umgesetzt.

Die Angaben (grundsätzlich auf dem Titelblatt), die weder zu `titlePage` noch zum `div`-Element passen (z.B. eine Umschlagbildbeschreibung oder der Zeitungspreis), werden als `p`-Elemente modelliert.

Das erste Unterelement soll immer eine Seitenzahl (`pb`-Element) sein.

body

TEI-Modul	textstructure – Default Text Structure
Unterelement von	<code>text</code>
Überelement für	<code>div</code> , <code>pb</code>
Beschreibung	4 Default Text Structure • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html

Haupttext des Korpusdokuments.

Das erste Unterelement soll immer eine Seitenzahl (`pb`-Element) sein.

back

TEI-Modul	textstructure – Default Text Structure
Unterelement von	<code>text</code>
Überelement für	<code>div</code> , <code>pb</code>
Beschreibung	4.7 Back Matter • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html#DSBACK

Alles, was hinter dem Haupttext steht: Nachwort, Inhaltsverzeichnis, Index, Anhänge und Appendix, Impressum, Texte und Bilder aus dem Umschlag u.ä.

Sonderfall: eine Bildeinlage (Bildblätter, die zwischen die Textseiten eingesetzt sind), die nicht zum Textfluss gehört, wird hierher umgesetzt.

Sonderfall: ein Zeitungsimpressum wird hierher umgesetzt.

Bei der Umsetzung müssen die Seiten richtig nummeriert (und evtl. Seitenzahlen wiederholt) werden.

Das erste Unterelement soll immer eine Seitenzahl (`pb`-Element) sein.

Elemente des Titelblatts

titlePage

TEI-Modul	textstructure – Default Text Structure
Unterelement von	<code>front</code>
Überelement für	<code>docTitle</code> , <code>docAuthor</code> , <code>docDate</code> , <code>docEdition</code> , <code>docImprint</code> , <code>byline</code> , <code>pb</code>
Beschreibung	4.6 Title Pages • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html#DSTITL

Eine semantische Gruppierung der Titelangaben aus dem Titelblatt (in einem Buch) oder dem Titelkopf (in einer Zeitung).

Hinweis: Zum `titlePage` (in einem Buch) gehört der Inhalt der Seite (einer oder mehrerer), die den Titel und/oder den Autorennamen umfasst. Der sonstige Vorderinhalt wird im `front`-Element, aber außer des `titlePage`-Elements modelliert.

docTitle

TEI-Modul	textstructure – Default Text Structure
Unterelement von	<code>titlePage</code>
Überelement für	<code>titlePart</code>
Beschreibung	4.6 Title Pages • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html#DSTITL

Titelsektion.

titlePart

TEI-Modul	textstructure – Default Text Structure
Unterelement von	docTitle
Überelement für	#PCDATA , pc , w
Beschreibung	4.6 Title Pages • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html#DSTITL

Ein Titel oder Titelteil.

titlePart/@type

TEI-Modul	textstructure – Default Text Structure
Beschreibung	4.6 Title Pages • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html#DSTITL

Die genutzte Werte: *main* für Haupttitel, *sub* für Untertitel, *desc* für einen erklärenden Teil des Titels (z.B. »Wuběrk proze«, »ORGAN DOMOWINY«).

docAuthor

TEI-Modul	textstructure – Default Text Structure
Unterelement von	titlePage
Überelement für	#PCDATA , pc , w
Beschreibung	4.6 Title Pages • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html#DSTITL

Name des Verfassers oder der Verfasserin.

docDate

TEI-Modul	textstructure – Default Text Structure
Unterelement von	titlePage , docImprint
Überelement für	#PCDATA , pc , w
Beschreibung	4.6 Title Pages • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html#DSTITL

Ausgabedatum.

docEdition

TEI-Modul	textstructure – Default Text Structure
Unterelement von	titlePage
Überelement für	#PCDATA , pc , w
Beschreibung	4.6 Title Pages • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html#DSTITL

Nummer der Ausgabe; in einer Zeitung: Nummer, Jahrgang u.ä. Kann auch ein Teil von [docImprint](#) sein.

docImprint

TEI-Modul	textstructure – Default Text Structure
Unterelement von	titlePage
Überelement für	#PCDATA , docDate , lb , pc , w
Beschreibung	4.6 Title Pages • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html#DSTITL

Verlags- und Ortsangaben.

byline

TEI-Modul	textstructure – Default Text Structure
Unterelement von	titlePage
Überelement für	#PCDATA , lb , pc , w
Beschreibung	4.6 Title Pages • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html#DSTITL

Andere Angaben zu den beteiligten Personen, z.B. Redakteur, Illustrator u.ä.

Textaufteilung

Die Bestandteile auf jeder Textebene (die Artikel, Kapitel, Unterkapitel, Absätze, Wörter usw.) folgen nach Möglichkeit der originalen Anordnung. In Sonderfällen wird die Reihenfolge nicht gewahrt (diese Operationen werden nicht dokumentiert):

- Wenn ein einzelner Bestandteil (ein Wort, Bild, Bild mit einer Unterschrift) zwischen den Seiten geteilt ist, wird er an seinem Anfang vollständig modelliert. Beispiel: nicht »pśe- `<pb/>` njasone słowo«, sondern »pśenjasone `<pb/>` słowo«.
- Wenn der Textfluss durch einen Einschub (ein Bild, Text im Rahmen) unterbrochen ist, muss die Bearbeiterin oder der Bearbeiter entscheiden, wo der Einschub modelliert wird. Die einzelnen Bilder (`figure`-Elemente) können im Text bleiben, sie können aber auch verschoben werden: in die Position nach dem Absatz oder an den Anfang oder das Ende der Sektion. Der eingeschobene Text, wenn er kein Zitat oder keine Notiz ist, kann nicht im anderen Text bleiben und muss immer sinnvoll versetzt werden.
- Wenn ein Text (vor allem ein Zeitungsartikel) unterbrochen ist und auf einer anderen Seite fortgesetzt wird (häufig mit einem expliziten Fortsetzungshinweis), dann wird er als Ganzes genommen und vollständig an seinem Anfang modelliert. Die Seitenbezeichnungen müssen dann achtsam eingetragen und, wenn nötig, auch wiederholt werden. Ein sich wiederholender Titel ist auszulassen (das wird mit `note[@type='encoders-comment']` dokumentiert). Beispiel:

```
<pb n="1"/>
<div type="article">
  <head>Ein Titel</head>
  <p>Anfang auf 1. Seite...</p>
  <pb n="2"/>
  <note type="encoders-comment" resp="MSI">
    Duplicated article title »Ein Titel« has been removed.
  </note>
  <p>... Ende auf 2. Seite</p>
</div>
<pb n="1"/>
<div type="article">
  <p>Ein nächster Artikel von der 1. Seite</p>
</div>
<pb n="2"/>
<div type="article">
  <p>Der erste volle Artikel von der 2. Seite</p>
</div>
```

div

TEI-Modul	textstructure – Default Text Structure
Unterelement von	<code>front</code> , <code>body</code> , <code>back</code> , <code>div</code>
Überelement für	<code>div</code> , <code>head</code> , <code>byline</code> , <code>epigraph</code> , <code>p</code> , <code>figure</code> , <code>list</code> , <code>lg</code> , <code>table</code> , <code>note</code> , <code>pb</code>
Beschreibung	4.1 Divisions of the Body • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html#DSDIV

Das Grundelement für die Inhaltsstrukturierung.

div/@type

TEI-Modul	tei – The TEI Infrastructure
Beschreibung	4.5 Front Matter • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html#DSFRONT 4.7 Back Matter • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html#DSBACK 4.1.1 Un-numbered Divisions • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html#DSDIV1 Recommended Practices: div • http://wiki.tei-c.org/index.php/Recommended_Practices#.3Cdiv.3E

Die genutzten Werte:

- *part* für einen Teil eines Buchs, der Kapitel oder weitere Teile umfasst
- *work* für ein einzelnes Werk in einer Anthologie (vorläufig definiert als eine Sammlung von Texten, die einen explizit genannten Herausgeber hat)
- *chapter* für ein Kapitel, Unterkapitel oder z.B. einen Teil des Artikels
- *section* für eine Rubrik oder eine Serie von Artikeln in einer Zeitung
- *article* für einen Artikel in einer Zeitung (in einem Periodikum)
- *text-frame* für einen eingeschobenen oder hinzugefügten Text in einem Rahmen
- *advertisement* für eine Anzeige oder Werbung (in einer Zeitung), anders gesagt: ein bezahlter Text nicht vom Verfasser
- *slogan* für einen selbstständigen Slogan, einen Spruch oder ein Motto, der/das anderenfalls als ein Artikel oder Kapitel modelliert werden müsste (vgl. auch *epigraph*)
- *notes* für ein Kapitel mit Bemerkungen/Erläuterungen
- *contents* für das Inhaltsverzeichnis
- *figures* für eine Sammlung von Illustrationen
- *colophon* für ein Impressum oder ähnliche technische Angaben.
- *dummy* – Manchmal lässt TEI (fehlerhaft?) keinen Textinhalt zu. Z.B. darf ein Kapitel, das mit *p* anfängt und dann Unterkapitel (`div[@type='chapter']`) umfasst, nicht mit *p*-Elementen enden; diese Schlussabsätze werden als Attrappe mit `div[@type='dummy']` modelliert.

Grundsätzliche Blockelemente des Textes

head

TEI-Modul	core – Elements Available in All TEI Documents
Unterelement von	<code>div</code> , <code>figure</code> , <code>lg</code> , <code>list</code> , <code>table</code>
Überelement für	<code>#PCDATA</code> , <code>lb</code> , <code>pc</code> , <code>w</code>
Beschreibung	4.2.1 Headings and Trailers • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html#DSHD

Umfasst einen Titel oder Untertitel (einer Textsektion oder eines Gedichtes), oder auch eine Beschriftung (eines Bildes, einer Tabelle) oder eine Überschrift (einer Liste).

head/@type

TEI-Modul	tei – The TEI Infrastructure
Beschreibung	4.2.1 Headings and Trailers • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html#DSHD

Die genutzten Werte: *main* für Haupttitel, *sub* für Untertitel, *desc* für einen erklärenden Teil des Titels.

byline

TEI-Modul	textstructure – Default Text Structure
Unterelement von	div , figure , lg
Überelement für	#PCDATA , lb , pc , w
Beschreibung	4.2.2 Openers and Closers ➤ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html#DSOC

Umfasst die Angabe zur Autorschaft.

date line

TEI-Modul	textstructure – Default Text Structure
Unterelement von	div , figure , lg
Überelement für	#PCDATA , lb , pc , w
Beschreibung	4.2.2 Openers and Closers ➤ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html#DSOC

Umfasst die explizit gegebene Angabe zu Ort und Datum: wo und wann ist der Text entstanden.

epigraph

TEI-Modul	textstructure – Default Text Structure
Unterelement von	div , lg
Überelement für	cit , p
Beschreibung	4.2.3 Arguments, Epigraphs, and Postscripts ➤ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html#DSAE

Umfasst ein Zitat oder eine kurze Widmung am Anfang des Textes oder Kapitels.

p

TEI-Modul	core – Elements Available in All TEI Documents
Unterelement von	div , front , item , note , quote
Überelement für	#PCDATA , cit , figure , foreign , hi , lb , lg , list , note , pb , pc , w
Beschreibung	3.1 Paragraphs ➤ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html#COPA

Umfasst einen Absatz.

Konvention: Ein Absatz ([p](#)) vor dem Titel ([head](#)) ist mit TEI leider nicht modellierbar. Wenn es die strukturelle Semantik des Textes nicht stört, müssen in solchen Fällen die [p](#)- und [head](#)-Elemente ausgetauscht werden (anders gesagt: [head](#) wird an den Anfang vorgezogen).

Konvention: Ist der erste Absatz mit einer anderen Schrift / in einem anderen Format gedruckt, muss das nicht unbedingt mit einem [hi](#)-Element kodiert werden.

Sonder(block)elemente des Textes

Die hier verzeichneten Elemente können sowohl im fortlaufenden Text (z.B. in einem [p](#)-Element) stehen, als auch einen gesonderten Textblock ausmachen (z.B. zwischen zwei [p](#)-Elementen).

Bild

figure

TEI-Modul	figures – Tables, Formulæ, Graphics and Notated Music
Unterelement von	div , p

Überelement für	byLine , head , p , (pb , note)
Beschreibung	14.4 Specific Elements for Graphic Images ➤ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/FT.html#FTGRA

Umfasst ein Bild oder eine andere nicht-textuelle Illustration.

Konvention: Bildunterschriften stehen im Druck bisweilen auf einer anderen Seite als das Bild selbst. In der XML-Kodierung treten sie allerdings immer an der Position des (Anfangs des) Bildes auf. Anders gesagt, die Bildunterschrift wird von ihrer eigentlichen Position umgesetzt an die Position des Bildes.

Konvention: Die Bildunterschrift wird mit einem [head](#)- (wenn es ein Titel und kein Satz ist) oder einem [p](#)-Element (wenn es ein Satz oder ein längerer Text ist) modelliert.

Gedicht

Konvention: In Zeitungsartikeln ([div\[@type='article'\]](#)), die nur ein (oder mehrere) Gedicht(e) umfassen, gehört der Titel ([head](#)) zum Gedicht ([lg\[@type='poem'\]](#)) aber der Autor ([byline](#)) zum Artikel ([div\[@type='article'\]](#)).

lg

TEI-Modul	core – Elements Available in All TEI Documents
Unterelement von	div , lg , p
Überelement für	byLine , head , lg , l , (pb note)
Beschreibung	3.12 Passages of Verse or Drama ➤ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html#CODV

Umfasst ein Gedicht oder ähnlichen Zeileninhalt.

lg/@type

TEI-Modul	tei – The TEI Infrastructure
Beschreibung	att.typed ➤ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.typed.html#tei_att.type

Die genutzten Werte:

- [poem](#) für ein Gedicht; auch für einen selbständigen Abschnitt
- [part](#) für einen Teil eines längeren Gedichts (ähnlich wie Kapitel in einem Roman)
- [stanza](#) für eine Strophe
- [fragment](#) für einen fragmentarischen, nicht selbständigen Auszug; auch für unklare Fälle

l

TEI-Modul	core – Elements Available in All TEI Documents
Unterelement von	lg
Überelement für	#PCDATA , foreign , pc , w
Beschreibung	3.12 Passages of Verse or Drama ➤ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html#CODV

Umfasst eine einzelne Zeile eines Gedichts.

Liste

list

TEI-Modul	core – Elements Available in All TEI Documents
Unterelement von	div , item , p

Überelement für	head , label , item (pb , note)
Beschreibung	3.7 Lists ➤ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html#COLI

Umfasst eine Liste (nur wenn sie im Original deutlich als Liste formatiert ist).

label

TEI-Modul	core – Elements Available in All TEI Documents
Unterelement von	list (mit space zusammen auch von item , p)
Überelement für	#PCDATA , pc , w
Beschreibung	3.7 Lists ➤ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html#COLI

Umfasst das Aufzählungszeichen oder das, was als Kopf für den Eintrag der Liste interpretiert wird.

Das Element [label](#) kann auch als eine Beschriftung (Etikett) für das Element [space](#) dienen.

item

TEI-Modul	core – Elements Available in All TEI Documents
Unterelement von	list
Überelement für	#PCDATA , foreign , hi , lb , list , note , p , pb , pc , ref , w
Beschreibung	3.7 Lists ➤ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html#COLI

Umfasst einen einfachen Eintrag der Liste, oder mehrere Text-Absätze, oder komplexere Strukturen, wie z.B. ineinander geschachtelte Unterlisten.

Tabelle

Die Tabelle ist eine besondere Darstellungsform für Angaben und soll vorsichtig verwendet werden. Man muss immer die Frage stellen, ob eine Tabelle notwendig ist (oder ob eine Liste genügt oder sogar lediglich einige einzelne Absätze). Ein empfohlener Test: Eine Tabelle liegt dann vor, wenn Spalten über ein Etikett verfügen; das Element kann eine Tabelle sein, wenn alle Spalten klar benannt werden können; anderenfalls ist es wahrscheinlich keine Tabelle.

table

TEI-Modul	figures – Tables, Formulæ, Graphics and Notated Music
Unterelement von	div
Überelement für	head , row (pb , note)
Beschreibung	14.1.1 TEI Tables ➤ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/FT.html#FTTAB1

Umfasst eine Tabelle oder tabellarische Angaben.

row

TEI-Modul	figures – Tables, Formulæ, Graphics and Notated Music
Unterelement von	table
Überelement für	cell
Beschreibung	14.1.1 TEI Tables ➤ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/FT.html#FTTAB1

Umfasst eine Zeile / ganze Reihe der Tabelle.

row/@role

TEI-Modul	tei – The TEI Infrastructure
Beschreibung	att.tableDecoration ➤ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-

att.tableDecoration.html#tei_att.role

Die genutzten Werte: *label* (für den Tabellenkopf).

cell

TEI-Modul	figures – Tables, Formulæ, Graphics and Notated Music
Unterelement von	row
Überelement für	#PCDATA , pc , w
Beschreibung	14.1.1 TEI Tables ➤ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/FT.html#FTTAB1

Umfasst eine einzelne Tabellenzelle.

cell/@cols

TEI-Modul	tei – The TEI Infrastructure
Beschreibung	att.tableDecoration ➤ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.tableDecoration.html#tei_att.cols

Zeigt, wie viel Spalten eine Zeile umspannt.

cell/@rows

TEI-Modul	tei – The TEI Infrastructure
Beschreibung	att.tableDecoration ➤ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.tableDecoration.html#tei_att.rows

Zeigt, wie viel Reihen eine Zeile umspannt.

Zitat

cit

TEI-Modul	core – Elements Available in All TEI Documents
Unterelement von	p
Überelement für	bibl , quote (pb , note)
Beschreibung	3.3.3 Quotation ➤ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html#COHQQ

Umfasst ein Zitat (nur wenn es im Original deutlich als Zitat ausgegliedert ist).

quote

TEI-Modul	core – Elements Available in All TEI Documents
Unterelement von	cit
Überelement für	p
Beschreibung	3.3.3 Quotation ➤ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html#COHQQ

Umfasst den Inhalt des Zitats.

bibl

TEI-Modul	core – Elements Available in All TEI Documents
Unterelement von	cit
Überelement für	#PCDATA , pc , w
Beschreibung	3.3.3 Quotation ➤ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html#COHQQ

Umfasst den Autor, die Autorin oder die Quelle des Zitats.

Unstrukturierter Text

In besonderen Fällen wird die Modellierung des Textes übersprungen. Auf diese Weise darf z.B. bei Tabellen oder Anzeigen verfahren werden. Die zugelassene Struktur dafür ist

ein `p`-Element mit einem `note`-Element am Anfang und dem darauf folgenden laufenden unstrukturierten Text:

```
<p><note type="unstructured" resp="MSI">unstrukturierte Tabelle</note>kupki jednotliwce
Gubin/Baršć 137% 60% Běža Woda 114% 104% Worjejcje 113% 72% Chošebuz 95% 97% Budyšyn 81%
61% Kamjeńc 45% 35% Grodk 43% 37% Kalawa/Lubin 40% 90% „Jan Skala“ 33% 88% Niska 29%
228%</p>
```

Bemerkung: Der unstrukturierte Textinhalt umfasst nur Zeichen und Leerzeichen – die Gliederung oder alle Zeilenumbrüche werden ausgelassen.

Elemente der Textformatierung

Nur die wesentlichen (aussagekräftigen) Aspekte der Dokument- und Textstruktur sollen modelliert werden. Die rein graphischen (typographischen) Phänomene werden ausgelassen. Die folgenden Elemente liegen an der Grenze zwischen dem Semantischen und dem Graphischen.

hi

TEI-Modul	core – Elements Available in All TEI Documents
Unterelement von	item , p , (w)
Überelement für	#PCDATA , pc , w
Beschreibung	3.3.1 What Is Highlighting? ➤ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html#COHQW

Umfasst ein (kurzes) Fragment des fortlaufenden Textes, das graphisch hervorgehoben (fett, kursiv, unterstrichen) ist – verwendet nur in dem Fall, wenn es inhaltlich sinnvoll ist, diese Hervorhebung beizubehalten.

hi/@rend

TEI-Modul	tei – The TEI Infrastructure
Beschreibung	3.3.1 What Is Highlighting? ➤ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html#COHQW

Präzisiert die graphische Art der Hervorhebung, z.B. *bold*, *italic*, *underlined*, *stretched*. Dieses Attribut ist fakultativ, soll aber nicht leer bleiben.

lb

TEI-Modul	core – Elements Available in All TEI Documents
Unterelement von	byline , docImprint , foreign , head , item , p
Überelement für	—
Beschreibung	3.10.3 Milestone Elements ➤ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html#CORS5

Markiert einen Zeilenumbruch – wenn er inhaltlich aussagekräftig ist.

Eigentlich soll das `lb` Element als unsichtbares Interpunktionszeichen verstanden werden.

Dieses Element soll selten genutzt werden. Besser ist es durch ein Leerzeichen, eine Liste o.ä. zu vertreten.

pb

TEI-Modul	core – Elements Available in All TEI Documents
Unterelement von	back , body , div , foreign , front , item , lg , list , p , titlePage
Überelement für	—
Beschreibung	3.10.3 Milestone Elements ➤ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html#CORS5

Markiert den Anfang einer neuen Seite: die nachfolgenden Elemente stehen auf der bezeichneten Seite.

Die Funktion dieses Elements besteht nicht darin, einen physischen Bruch zu modellieren, sondern eine Information zu geben, wo die folgenden Elemente positioniert werden, z.B. zum Zweck der Zitierung.

pb/@n

TEI-Modul	tei – The TEI Infrastructure
Beschreibung	3.10.3 Milestone Elements • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html#CORS5

Umfasst die Seitenzahl, d.h. die Nummer der Seite, die an dieser Stelle anfängt.

Für die Umschlagseiten werden die Werte wie *outside front cover*, *inside front cover*, *inside back cover* und *outside back cover* benutzt.

Vertikaler Abstand oder (Trenn-)Ornament

Ein deutlicher Abstand oder eine andere besondere Trennung zwischen zwei Blockelementen wird nicht modelliert. TEI bietet keine elegante Lösung dafür. Es wäre auch häufig unklar, wo es angewandt werden soll.

Sonstige besondere Elemente

note

TEI-Modul	core – Elements Available in All TEI Documents
Unterelement von	div , figure , item , p
Überelement für	#PCDATA , p , pc , w
Beschreibung	3.8.1 Notes and Simple Annotation • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html#CONONO

Umfasst eine Fußnote.

Außerdem erfüllt das `note`-Element leider mehrere andere und sehr verschiedenartige Funktionen. In gewissem Maße werden durch die Anwendung dieses Elements einige fehlende Optionen der TEI-Kodierung ausgeglichen.

Das `@type`-Attribut zeigt die Funktion an:

- *footnote* = eine Fußnote im Text
- *encoders-comment* = eine Anmerkung der XML-Bearbeiterin oder des XML-Bearbeiters, z.B. »In der Druckversion eine Linie an falscher Stelle gedruckt.«
- *editors-note* = ein vom Verfasser oder von der Redaktion eingefügter Text, z.B. »(něco skrotcone)«
- *continuation* = ein Fortsetzungshinweis, z.B. »(dalej na 2. boce)«
- *text-source* = eine Bestimmung der Textquelle, z.B. »(z kniglickow „Na Kałpjeńcu“)«
- *marginalia* = ein Text ist in einer Randspalte positioniert und mit dem Haupttext verbunden, aber nicht wie eine Fußnote verankert
- *unstructured* = kennzeichnet einen unstrukturierten Inhalt

Dass kein deutlicher Unterschied zwischen einer Fußnote und Bearbeitungsanmerkungen

gemacht wird, ist zwar unvorteilhaft, aber vielleicht von TEI beabsichtigt.

Die drei Funktionen *editors-note*, *continuation* und *text-source* werden künstlich als Notiz modelliert, weil TEI (wohl fehlerhaft) keine Möglichkeit gibt, ein *p*-Element hinter dem *byline*-Element und vor dem Ende der Sektion zu benutzen.

Konvention: wenn ein Fortsetzungshinweis (*note*[@*typ*='continuation']) genauso in einem Zeitungsartikel (*div*[@*type*='article']) wie in einer Rubrik (*div*[@*type*='section']) stehen könnte, wird der Artikel bevorzugt.

note/@type

TEI-Modul	tei – The TEI Infrastructure
Beschreibung	att.typed ➤ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.typed.html#tei_att.type

Die genutzten Werte: *footnote*, *encoders-comment*, *editors-note*, *continuation*, *text-source*, *marginalia*, *unstructured*.

note/@subtype

TEI-Modul	tei – The TEI Infrastructure
Beschreibung	att.typed ➤ http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.typed.html#tei_att.subtype

Das *@subtype*-Attribut klassifiziert die Bearbeiter-Anmerkung (*note*[@*type*='encoders-comment']). Diese Klassifikation ermöglicht die automatische Nachbearbeitung und befreit den Bearbeiter von sehr präzisen Formulierungen.

In Verbindung mit *@type*='encoders-comment' ist das *@subtype*-Attribut obligatorisch. Es kann mit dem XML-Schema leider nicht erzwungen werden, und die Bearbeiter müssen selbständig darauf achten.

Die Klassifizierungswerte sind:

- *missing-text* = Die Anmerkung beschreibt ein fehlendes Textfragment (z.B. »Die Seiten 131-134 fehlen.«).
- *description* = Die Anmerkung beschreibt ein besonderes Merkmal des Textes, das dokumentiert werden soll (z.B. »Das Bild umfasst dt. Text.«).
- *adjustment* = Die Anmerkung beschreibt eine Korrektur oder andere Änderung des Textes (z.B. »«).
- *technical* = Die Anmerkung bezieht sich direkt auf die XML-Struktur (z.B. »TEI does not allow p elements here, so we have to use a note.«).

note/@rend

TEI-Modul	tei – The TEI Infrastructure
Beschreibung	Tagging the anchor mark ➤ http://www.wwp.northeastern.edu/research/publications/guide/html/referencing_noteLinking.html

Bei Fußnoten wird hier das Zeichen oder die Nummer angegeben, mit welcher die Fußnote bezeichnet oder nummeriert ist.

Z.B. um zu kodieren, dass im Druck ein Sternchen gesetzt wird, benutzt man den Wert *siglum(*)*.

note/@resp

TEI-Modul	tei – The TEI Infrastructure
-----------	------------------------------

Beschreibung	3.8.1 Notes and Simple Annotation • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html#CONONO
--------------	---

Für eine Bearbeitungsanmerkung ist es notwendig, den Bearbeiter oder die Bearbeiterin zu nennen.

Die bisher genutzten Werte: *JSA*, *MSI*, *TML*.

stage

TEI-Modul	core — Elements Available in All TEI Documents
Unterelement von	div , l , lg , p
Überelement für	hi , note , pb
Beschreibung	3.12 Passages of Verse or Drama → https://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html#CODV 7.2.4 Stage Directions → https://www.tei-c.org/release/doc/tei-p5-doc/en/html/DR.html#DRSTA

Umfasst Didaskalien – sowohl die Regieanweisung als auch die Zuteilung der Rolle bzw. die Bezeichnung der sprechenden/singenden Person. Wird benutzt vor allem für die Modellierung der dramatischen Texte und der Lieder, in denen die Verse zwischen unterschiedlichen Sängern verteilt sind.

Obwohl das TEI ein reiches System zur Annotation des Dramas bietet, wurde hier mit Absicht nur *stage* ausgewählt, um die Annotation nicht unnötig zu verkomplizieren.

ref

TEI-Modul	core – Elements Available in All TEI Documents
Unterelement von	item
Überelement für	#PCDATA , pc , w
Beschreibung	3.6 Simple Links and Cross-References • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html#COXR

Umfasst einen im Text gegebenen Verweis auf einen anderen Ort im Text.

Derzeit wird der Verweis nur markiert, ohne weitere Modellierung des Ziels und anderer Merkmale.

ref/@type

TEI-Modul	tei – The TEI Infrastructure
Beschreibung	att.typed • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.typed.html#tei_att.type

Bezeichnet den Zieltyp.

Der bisher genutzte Wert: *page*.

foreign

TEI-Modul	core – Elements Available in All TEI Documents
Unterelement von	item , l , p
Überelement für	#PCDATA , lb , pb , pc , w
Beschreibung	3.3.2.1 Foreign Words or Expressions • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html#COHQHF

Umfasst ein (kurzes) Fragment des fortlaufenden Textes, dessen Sprache der Sprache seines Überelements nicht gleicht (wobei dieses Fragment nicht als Zitat modelliert werden soll).

Die Hauptfunktion dieser Markierung ist, das Fragment bei der späteren linguistischen Analyse aus- bzw. einzuschließen.

Wahrscheinlich wird es immer fraglich sein, ob ein Fragment fremd oder nur fremder Herkunft ist (Grad der sprachlichen Integration).

foreign/@xml:lang

TEI-Modul	tei – The TEI Infrastructure
Beschreibung	3.3.2.1 Foreign Words or Expressions • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html#COHQHF ISO 639 Code Tables • http://www-01.sil.org/iso639-3/codes.asp

Sprachcode nach ISO-639-3. Das Attribut ist für dieses Element obligatorisch.

Siehe auch Kapitel »Annotation der Sprache«.

space

TEI-Modul	transcr — Representation of Primary Sources
Unterelement von	item , p
Überelement für	#PCDATA
Beschreibung	11.5.1 Space • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/PH.html#PHSP

Markiert einen im Original frei gelassenen Textbereich (z.B. Lücken zum Ausfüllen in einem Formular).

Annotation der Sprache

Die sorbischen Korpora umfassen Dokumente in zwei Hauptsprachen (d.h. Ober- und Niedersorbisch); häufig enthalten diese Dokumente aber Textfragmente in einer anderen als der Hauptsprache (Niedersorbisch, Obersorbisch, Deutsch u.a.).

Für Annotation der Sprache dient das Attribut `@xml:lang`.

TEI-Modul	tei – The TEI Infrastructure
Beschreibung	att.global • http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.global.html#tei_att.xml-lang

Die Fremdsprache wird nach ISO-639-3 kodiert (Beispielwerte: *ces*, *deu*, *dsb*, *eng*, *fra*, *hsb*, *ita*, *lat*, *lit*, *pol*, *slk*, *spa*, *srp*). Bemerkung: ISO-639-3 hat auch besondere Codes z.B. für eine unerkannte Sprache (*und*) oder für eine Mischung aus mehreren Sprachen (*mul*). Siehe https://en.wikipedia.org/wiki/ISO_639-3#Special_codes.

Dieses Attribut ist ein besonderes, weil es fast jedem strukturellen Element zugewiesen werden kann. Ausnahmen sind Elemente, die keinen Text darstellen, z.B. `lb` und `pb`.

Da die Annotationselemente eine hierarchische Struktur bilden, hat auch die Sprachinformation einen hierarchischen Charakter. Das bedeutet, dass die Unterelemente die Sprache von ihren Überelementen übernehmen und, solange ihre Sprache gleich der Sprache ihres Überelementes ist, kein explizites `@xml:lang`-Attribut haben müssen. Beispielweise: eine Rubrik ist niedersorbisch, alle Artikel in dieser Rubrik sind automatisch auch niedersorbisch; nur die nicht-niedersorbischen Artikel (in dieser Rubrik) brauchen dann ein `@xml:lang`-Attribut; die niedersorbischen Textfragmente in diesen nicht-niedersorbischen Artikeln, müssen jedoch das Attribut `@xml:lang='dsb'` bekommen.

```
<div type="section" xml:lang="dsb">
  <head>Zasadnje dolnoserbiska rubrika</head>
  <div type="article">
    <head>Dolnoserbiski nastawk</head>
    <p>Dolnoserbiski tekst.</p>
    <cit><quote xml:lang="hsb">görnoserbski citat</quote></cit>
  </div>
  <div type="article" xml:lang="deu">
    <head>Artikel auf Deutsch</head>
    <p>Text auf Deutsch.</p>
  </div>
</div>
```

```
<div>
  <p xml:lang="dsb">Zasajžony dolnoserbški wótstawk.</p>
</div>
```

Jedes Dokument soll seine Haupt- bzw. Standardsprache deklarieren. Dafür dient das `@xml:lang`-Attribut, das obligatorisch beim `/TEI/text`-Element (also beim Hauptelement der Textstruktur) steht.

Bemerkung: Die Sprachdeklaration unter `/TEI/teiHeader/profileDesc/langUsage` bestimmt eine oder mehrere Hauptsprachen des Dokumentes. Da diese Metaangabe nicht immer präzise genug ist, wird sie von der automatischen Auswertung nicht beachtet.

Textanalyse

Die Textanalyse ist (hier und nach den TEI-Richtlinien) eine linguistische bzw. philologische Annotation des Textes. Vorerst werden nur die rudimentären Elemente genutzt, die die Annotation von Tokens ermöglichen. Die linguistische Annotation wird grundsätzlich automatisch durchgeführt, kann aber manuell angepasst werden.

W

TEI-Modul	analysis — Simple Analytic Mechanisms
Unterelement von	author , bibl , byline , cell , change , dateline , distributor , docAuthor , docDate , docEdition , docImprint , foreign , head , hi , item , l , label , note , p , quote , ref , title , titlePart
Überelement für	#PCDATA , figure , hi , lb , note , pb , space
Beschreibung	17.4.2 Lightweight Linguistic Annotation • https://www.tei-c.org/release/doc/tei-p5-doc/en/html/AI.html#AILALW

Umfasst ein Text-Token. Die geltende, genaue Deklaration dieses Begriffs wird anderswo gegeben. Grob formuliert ist ein Token eine ununterbrochene Reihe von Buchstaben, Ziffern und eventuell gewissen Interpunktionszeichen.

w/@type

TEI-Modul	tei – The TEI Infrastructure
Beschreibung	17.1.1.1 Words and Above • https://www.tei-c.org/release/doc/tei-p5-doc/en/html/AI.html#AILCW

Dieses Attribut dient zum Speichern von verschiedenen zusätzlichen Merkmalen und Verarbeitungshinweisen des Tokens.

Die genutzten Werte:

- `name` das Token bildet einen Eigennamen
- `abbr` das Token bildet eine Abkürzung mit finalem Punkt
- `ignore` das Token wurde (vorerst) von der weiteren Verarbeitung ausgenommen, Gründe hierfür können in `w/@subtype` angegeben werden

Es können mehrere mit Whitespace getrennte Werte im Attribut gespeichert werden.

w/@subtype

TEI-Modul	tei — The TEI Infrastructure
Beschreibung	17.1.1.1 Words and Above • https://www.tei-c.org/release/doc/tei-p5-doc/en/html/AI.html#AILCW

Wenn das Token von der weiteren Verarbeitung ausgeschlossen wurde (`w/@type` den Wert `ignore` annimmt), kann in diesem Attribut ein Grund für den Ausschluss angegeben werden.

Die genutzten Werte:

- *number* eine mit Ziffern (Zahlzeichen) geschriebene Zahl
- *symbol* ein oder mehrere Symbolzeichen, z. B. »€«.
- *dialect* ein Token mit dialektalen (nichtstandardsprachlichen) Merkmalen
- *old* ein Token mit veralteten (nichtstandardsprachlichen) Merkmalen
- *appendedToPrev* ein historisch korrektes Token, das nach gegenwärtigem Orthografieverständnis mit dem vorangegangenen Token vereint werden muss

w/@orig

TEI-Modul	analysis — Simple Analytic Mechanisms
Beschreibung	17.4.2 Lightweight Linguistic Annotation • https://www.tei-c.org/release/doc/tei-p5-doc/en/html/AI.html#AILALW

Falls das Token korrigiert bzw. anders geändert wurde, umfasst das `@orig`-Attribut seine originale Form. Die von den TEI-Richtlinien dafür bestimmte `choice`-Struktur wird vorerst nicht benutzt, da sie zu komplex ist.

w/@norm

TEI-Modul	analysis — Simple Analytic Mechanisms
Beschreibung	17.4.2 Lightweight Linguistic Annotation • https://www.tei-c.org/release/doc/tei-p5-doc/en/html/AI.html#AILALW

Ordnet dem Token eine oder mehrere normierte (standardisierte) Formen zu. Innerhalb der Attributwerte können folgende Metazeichen(folgen) auftreten, aufgelistet in aufsteigender Operatorenpräzedenz:

- `|` (senkrechter Strich, Pipe) steht zwischen zwei Ausdrücken und markiert sie als alternative Lesarten
- `!` (einzelnes Ausrufezeichen) steht am Anfang eines Ausdrucks und markiert ihn als manuell festgelegt
- `!!` (zwei aufeinanderfolgende Ausrufezeichen) steht am Anfang eines Ausdrucks und markiert ihn als (manuell festgelegte) Metanormalisierung

Ein Ausdruck ohne vorangestelltes Ausrufezeichen gilt als über automatisierte Verfahren ermittelt. Beispiele für die Verwendung obiger Zeichen:

```
<w norm="k njomu" lemma="K WÓN|K WÓNO">kñomu</w>
<w norm="!móžo|!m'žo" lemma="MÓC">mžo</w>
<w norm="pśiś|pśiž|!pśiś" lemma="PŚIŚ">přchiřch</w>
```

w/@lemma

TEI-Modul	analysis — Simple Analytic Mechanisms
Beschreibung	17.4.2 Lightweight Linguistic Annotation • https://www.tei-c.org/release/doc/tei-p5-doc/en/html/AI.html#AILALW

Ordnet dem Token einen oder mehrere Flexemidentifikatoren (\approx Lemmata) zu. Innerhalb der Attributwerte können folgende Metazeichen(folgen) auftreten, aufgelistet in aufsteigender Operatorenpräzedenz:

- `|` (senkrechter Strich, Pipe) steht zwischen zwei Ausdrücken und markiert sie als alternative Lesarten
- `!` (einzelnes Ausrufezeichen) steht am Anfang eines Ausdrucks und markiert ihn als manuell festgelegt

- **!!** (zwei aufeinanderfolgende Ausrufezeichen) steht am Anfang eines Ausdrucks und markiert ihn als (manuell festgelegtes) Metalemma

Ein Ausdruck ohne vorangestelltes Ausrufezeichen gilt als über automatisierte Verfahren ermittelt. Für ein Beispiel vgl. [w/@norm](#).

pc

TEI-Modul	analysis — Simple Analytic Mechanisms
Unterelement von	author , bibl , byline , cell , change , dateline , distributor , docAuthor , docDate , docEdition , docImprint , foreign , head , hi , item , l , label , note , p , quote , ref , title , titlePart
Überelement für	#PCDATA
Beschreibung	17.4.2 Lightweight Linguistic Annotation • https://www.tei-c.org/release/doc/tei-p5-doc/en/html/AI.html#AILALW

Umfasst ein Interpunktionszeichen. Die Leerzeichen werden nicht annotiert, müssen aber unbedingt im Text an ihren Positionen (d.h. zwischen den [w](#)- bzw. [pc](#)-Elementen) bleiben.

Änderungsprotokoll

- 2016-02-26 Marcin verfasst die Dokumentation.
- 2016-02-29 Marcin ergänzt die Schemaübersicht.
- 2016-03-08 Marcin korrigiert kleine Fehler.
- 2016-03-18 Neues Attribut: `hi/@rend`; neue Elemente: `dateline` und `space`; neue Wert für `div/@type`: `slogan`.
- 2016-03-18 Marcin erweitert die Bemerkungen für `lb`, `note`, Gedicht und vertikaler Abstand.
- 2016-03-31 Marcin erweitert das Kapitel »Ziel« und fügt das Kapitel »Nicht kodierte Angaben« hinzu.
- 2016-04-01 Neues Element: `epigraph`; neuer Wert für `div/@type`: `work`.
- 2016-04-04 Neue Werte für `div/@type`: `text-frame` und `dummy`.
- 2016-04-05 Marcin erweitert das Kapitel »Textaufteilung«.
- 2016-04-08 Marcin erweitert die Kapitel »`titlePage`«, »`front`«, »`body`«, »`back`«.
- 2016-04-27 Neue Attribute: `head/@xml:lang` und `p/@xml:lang`.
- 2016-04-27 Marcin erweitert die Kapitel »Fehler: Verweis nicht gefunden«, »`slogan`« unter »`div/@type`«, »`p`« und »Tabelle«.
- 2016-05-26 Nach der Beratung der Abteilung Sprachwissenschaft: Das Attribut `hi/@rend` ist nicht mehr obligatorisch; unstrukturierter Text wird zugelassen (Marcin fügt das Kapitel »Unstrukturierter Text« und den neuen Wert für `note/@type`: `unstructured` hinzu).
- 2016-05-30 Marcin stellt das Attribut `hi/@rend` als obligatorisch wieder her.
- 2016-06-30 Neuer Wert für `lg/@type`: `part` und für `note/@type`: `marginalia`.
- 2016-07-12 Die Werte für das Attribut `@xml:lang` wurden auf ISO-639-3 beschränkt.
- 2016-08-01 Thomas liest Korrektur.
- 2016-09-15 Marcin fügt das Kapitel »Stufen der strukturellen Kodierung« hinzu.
- 2016-09-15 Marcin erweitert das Kapitel »`revisionDesc`«.
- 2016-09-15 Neue Attribute: `revisionDesc/@status`, `change/@status` und `note/@subtype`.
- 2016-11-15 Neue Attribute: `cell/@cols`, `cell/@rows`.

- 2016-11-30 Neues Attribut: [lg/@xml:lang](#).
- 2016-12-01 Das Attribut [hi/@rend](#) ist nicht mehr obligatorisch.
- 2016-12-09 Neues Attribut: [quote/@xml:lang](#).
- 2017-06-01 Joanna ändert die Richtlinien zu den Umschlagtexten und -bildern (siehe Kap. »Nicht kodierte Angaben«, »Stufe 2: Text mit minimaler TEI-Struktur«, »[front](#)«, »[back](#)« und »[pb/@n](#)«).
- 2018-04-11 Thomas und Marcin erweitern das Kapitel »[change/@status](#)«.
- 2019-07-15 Marcin fügt das Kapitel »Textanalyse« hinzu. Neue Elemente: [w](#) und [pc](#).
- 2019-07-19 Marcin ändert den Titel – das Dokument bezieht sich nicht mehr nur auf Niedersorbisch.
- 2019-07-19 Marcin entfernt die verstreuten Deklarationen vom Attribut [@xml:lang](#) und fügt das Kapitel »Annotation der Sprache« hinzu.
- 2020-01-27 Marcin präzisiert die Anwendung von [hi](#), [foreign](#) und [ref](#) im Kapitel »Stufen der strukturellen Kodierung«.
- 2020-06-22 Joanna entfernt die fehlerhafte Anwendung des Elements [div\[@type='poem'\]](#) im Kapitel »Stufe 2: Text mit minimaler TEI-Struktur«.
- 2020-08-14 Neues Element: [stage](#).
- 2020-12-03 Neue Attribute: [w/@norm](#) und [w/@orig](#) im Kapitel »Textanalyse« (Entwurf).
- 2022-01-26 Marek definiert die Attribute [w/@type](#), [w/@subtype](#) (neu), [w/@lemma](#) und [w/@norm](#) um.
- 2022-01-26 Marcin erweitert das Kapitel »revisionDesc« mit neuen Werten für [@status](#).
- 2022-10-10 Marcin erweitert das Attribut [lg/@type](#) mit dem neuen Wert [fragment](#).
- 2022-10-25 Hauke nimmt ein paar sprachliche Korrekturen vor.
- 2023-05-02 Marcin ändert das Zeitformat für [change/@when](#).