

Elements, attributes and tokens

This page is still incomplete but already useful.

Last update #MB181029

Please use the page to check which attributes and values you can use to annotate your texts. If the options given here does not suffice your need, please leave a comment here: [questions re attributes and values](#)

#AZ172906 values added del rend overwritten, add place inline

#MB181029 for <lb>, added attribute 'rend' with the value "indent"

#AZ201515 added table syntax

Formal annotations (div, ab, pb, lb, w, s, space, table, row, cell, add, del, gap, unclear)

Element	explanation	possible attributes	possibel values	note
<div>	text division, subdivision of <body> used by us for abstract, edition, translation and commentary	xml:id	abs ed et commentary	
		type	abstract edition English_translation synopsis commentary	
		xml:lang	eng (for English) nep (for Nepali) new (for Newari, Nepal Bhasa) san (for Sanskrit) tib (for Tibetan)	follow ISO 639-2 codes, click here
<ab>	anonymous block, used in <div> edition for text blocks	type	invocatio seal space addition margin maintext signature ...	
		corresp	# ... (points to the xml:id within <zone>)	
<pb>	page break, beginning of a new page	n	1r 1v ... 1r-part1 1r-part2 ... 1 2 ...	
<lb/>	beginning of a new line	n	1 2 3 ...	
		break	no	to be used if line break splits a word
		rend	indent	to be used to record indents at the beginning

				of lines, or to indicate the first-line indent at the beginning of a paragraph
<w>	marks a word			
<s>	marks a sentence like unit	xml:lang	see under <div>	Use only if language of a sentence differs from main language. Use iso 639-2 codes for the value of the attribute. A list of these codes can be found at: https://en.wikipedia.org/wiki/List_of_ISO_639-2_codes#N
<space>	marks space between words or lines	quantity	1 2 ...	
		unit	chars lines	

Table Syntax				
<table>	outer frame of table	n	1 2 ... 1-part1 1-part2 ...	
		cols	1, 2, ...	gives the number of columns of a table
		rows	1, 2, ...	gives the number of rows of a table
<row>	marks row inside <table>	n	1 2 ...	count from top to bottom for each table individually use only in edition, not in translation
		role	label	use to mark the row that gives labels for columns
<cell>	marks cell inside <row>	cols	1 2 ...	use if a cell extents over more than one column
		rows	1 2 ...	use if a cell extents over more than one rows
<lb/>	line break in a cell			use unnumbered to split lines within cells only use if a cell has more than one line

Other characteristics of the original to be annotated				
<add>	Scribal addition	place	above the line below the line in the left margin in the right margin in the upper margin in the lower margin inline in the middle	

		hand	first second uncertain	note: OH is asked to make @hand non-mandatory, but at present even first hand needs to be given
	Scribal deletion	rend	adapted crossed out erasure overdotted overstrike overwritten understrike underdotted	
<gap>	marks <u>lost</u> text (if text can be restored, use <supplied>)	reason	lost	
		agent	breakage folding ...	
		extent	1 character 2 characters ... 1 word 2 words ... ca. 1 word ca. 2 words ... 1 line 2 lines ... ca 1 line ca. 2 lines ... unknown	
<gap>	marks <u>illegible</u> text (if text can be restored, use <supplied>)	reason	illegible	
		agent	water seal ink photo quality handwriting ...	
		extent	see under <gap> for <u>lost</u> text	
<unclear>	marks an doubtful reading	cert	high medium low	cert is optional

If the document contains a table use <table>, <row> and <cell>, if not needed, without any attributes. For examples, see here ####.

Editorial changes (supplied, surplus, sic, corr, orig, reg)

Element	explanation	possible attributes	possibel values	note
<supplied>	editor's restoration of lost text	reason	lost	
<supplied>	editor's restoration of	reason	illegible	

	illegible text			
<supplied>	editor's addition of omitted text	reason	omitted	
<surplus>	marks a word or phrase of the original which the editor considers to be superfluous or redundant			for superfluous letters in a word use <sic> and <corr>
<sic> and <corr> in <choice>	corrections by the editor (mark original word by <sic> and corrected one by <corr>, nest both in <choice>)			In Nepali text, apply wordwise, i.e. inside <w>. In long Sanskrit composita correct syllable wise
<orig> and <reg> in <choice>	standardization by the editor (mark original word by <orig> and standardized one by <reg>, nest both in <choice>)			used for standardizing ॐ/ॐ and ॐ/ॐ (apply wordwise, i.e. inside <w>) and graphical features
<orig> (alone)	standardization of middle dots and nuktas			use only if middle dot separates words or sentences, if middle dot marks an abbreviation: no <orig>

Content annotation (persName, placeName, geogName)

Element	explanation	possible attributes	possibel values	note
<persName>	marks a personal name	type (not necessary for a human name)	god goddess elephant ...	
		corresp	"pointer_to_the_persona lnames_repository"	
<placeName>	marks a place name	corresp	"pointer_to_the_persona lnames_repository"	
<geogName>	marks a a name associated with some geographical feature such as mountains, river etc.	corresp	"pointer_to_the_persona lnames_repository"	

Tokens for Translation and Commentary (p, note, term, bibl, foreign)

Element	explanation	attributes	values	note
<p>	paragraph			
<note>	footnote	place	end	
<term>	links a word to the wordlist database	ref	ID in wordlist database	term is displayed in italics
<bibl>	links a bibliographical reference to the bibliography database	corresp	ID in bibliography database	
<foreign>	italizes a word or phrase not annotated as <term>			