# Tokens used in xml and possible pseudocodes in Word

## Rules

* We use following tag symbols. # @ { } and :
  + # begins and end a markup
  + @ begins an attribute
  + {} contains xml tag content
  + : if the attribute is unclear from the order,we use : after attribute name @attribute:value
* If reserved symbols are necessary for the text, they are escaped by a \ e.g. \@
* @# is used if a cetain attribute begins with #
* Within hashtags upper-small case is respected.
* Order of the arttributes is important and will be read from let to right.
* Content for XML or attributes is written in {}, always at the end of the markup.

## Questions ? 1. I took a generic example for word (w) from tei doc , is it ok so ? 2. Did I forget to include any use-case in DEL ? 3. For the sentence, I suggested s 4. When ENTER key is typed, we being new text division, I feel may be users will add more than necessary enters or it is not easy to count. May be #LB# is better and they can copy and paste. 5. with + we accept a gap with illegible text., if we need agent there, we add it as an attribute and : 6. Representation for addition by second hand ? 7. Unclear was modified a little by me. 8.

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

|  |  |  |  |
| --- | --- | --- | --- |
| Markup | Markup Example | TEI Example | Remarks |
| #-# | #-# | <lb @break=no> | ?? [line beginning without break](https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-lb.html) |
| #++++@agent:value# | #++++@agent: test # | <gap @reason=“illegible“ extent=“4 characters“> | ?? [gap](https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-gap.html) |
| #///# | #///# | <gap @reason=“lost extent=“3 characters“> | [gap](https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-gap.html) |
| #&@place: place content# | #&@place: this is a place&# | <add place="this is a place"/> | [add: text additions](https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-add.html) |
| #?{text}# | #?{text unclear}# | <unclear @reason=“lost> </unclear> | ?7? [unclear](https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-unclear.html) |
| #&&@place:place content# | #&&@place: this is a place&# | <add place="this is a place"/> | ?6? [add: text additions](https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-add.html) |
| #ab@type@correspond# | #@ab@addition@#addition1# | <ab type="addition" corresp="# addition1"> | [Annonymous Block](https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-ab.html) |
| #del@rend{text}# | #DEL@rend:overstrike{deleted text}# | <del @rend="overstrike">"deleted text"></del> | ?2? [deletion](https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-del.html) |
| #div@id@type@lang# | #div@abs@abstract@eng# #div@ed@edition@nep-san@ | <div xml:id="abs" type="abstract" xml:lang="eng"> | [Text division](https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-div.html) |
| Enter Key | `|| ?4? [line beginning](https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-lb.html) | | #pb@p:page-number@facs# |#@pb@p=12@#surface1#|| [page beginning](https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-pb.html) | | #s{Content}# |#S{A short affair}#|A short affair| ?3? [s-unit: sentence like division](https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-s.html) | | #sp@quantity@unit# |#SP@7@chars#|| [space](https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-space.html) | | #w# |#w#|` | ?? [Word](https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-w.html) |  |

#w# | #-# | <w lemma="play" pos="vvz" xml:id="A19883-003-a-0180"> | ?? [Word](https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-w.html) |

php docx2tei.php ~/projects/nhdp/input/Template.docx ~/projects/nhdp/input/output.xml