

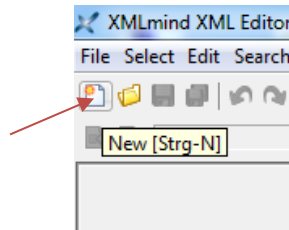
SWIKO: Transcriptions with XML Mind

The transcription process with XML Mind has five basic steps:

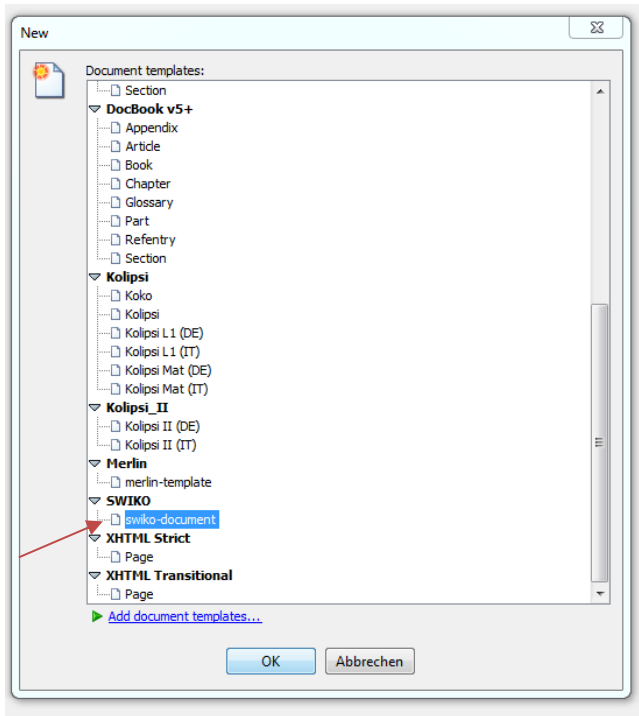
1. Open the SWIKO template to create a new document.
2. Write metadata information (transcriber, author, TaskID).
3. Write original text as closely related to the author's intentions.
4. Copy original text and tag.
5. Save as a new file.

1. Create a new document

1. File → New ... or Ctrl+N or use the “New” button:



2. Choose “swiko-document”¹ in the dialogue box (you may have to scroll down):



3. Click OK.

¹ If it is not there, the XML mind installation may not be complete. Refer to the installation instruction and check contents of the installation folders. Also, if you have several XML mind installations, choose the one using the add-on folder with the SWIKO installation.

2. Write metadata

The following metadata has to be written:

Tag	Description	Example	Permitted values	Comment
Transcriber <transcriber>	Transcriber's initials		three capital letters: first letter of first name and two first letters of last name	Identifier of the person who transcribed the text
Author ID <author_id>	The ID of the author as given in the text		Letters and numbers. SWIKO: Two first letters of school location + a 3-digit- number. HarmoS: 7 digits, starting with 6 or 9.	Identifier of the person who wrote the text
Checked by ID <checked_by>	The ID of the person who double-checked the transcription	<p>Copyright 2012 project</p> <p>Transcriber: KKA</p> <p>Author ID: Es175</p> <p>Task ID: SWI02_fE</p> <p>Original Text:</p> <p>■</p> <p>Tagged Text:</p> <p>■</p>	As Transcriber	Identifier of the person who double- checked the text. May be empty.
Task ID <task_ID>	The identifier of the task as given by the project. Languages are adapted to UGK standard: first letter of L1 as lower letter, followed by first letter of L2 as capital letter (if applicable)		Three capital letters (SWI or HAR), two numbers (01-12) and the language identifier (a lower letter for L1 texts or a lower and a capital letter for L2 texts).	Identifier of the task the student solved, usually with project name (i.e. SWI or HAR) and language identifier. Currently, transcripts with task ids that do not conform to this pattern will output an error message.


2.1. Original text

The text must be transcribed as closely as possible to what the author (probably intended). In the field “Original Text”, the text is transcribed as the final version the student intended:

- Spelling errors are transcribed as the student wrote them.
- Deleted parts are not transcribed.
- Inserted parts are transcribed where the author intended them to be.
- Single or a few words from other languages are transcribed as they are. Longer sequences written in another language (i.e. whole sentences) are not transcribed.
- Only five tags may be used:
 - o Number
 - o Abbreviation
 - o Par(agraph)
 - o Period
 - o Unreadable

Tag	Description	Example	Permitted values	Comment
number	<p>A number. Including any characters related to it (1.; 1'500; 3:30 ...).</p> <p>A number that is part of a word or name (e.g. “PS4” or “iPhone8” or “1990er”) does not have to be tagged as a number!</p>	<p><u>Original Text:</u> this is a good idée because i don't like wake up at #06:30#, this is me. Because i would a school finish at #16:00# because i have h #1.#verry job to the home[.] mittime is for #2#h#30# of time wake up at #06:30#, this</p>	Numbers and related characters. When in doubt, put in more.	For technical reasons, numbers have to be excluded from the POS tagging. In order to do that, they need to be tagged in both versions of the text as numbers. This is especially important if there are other characters, such as dots, apostrophes or colons. If there is a letter between numbers (e.g. 2h30), they can also be tagged separately with

				the letter as text in between.
abbreviation	The student used an abbreviation instead of writing the complete words.	<p><u>Original Text:</u> Hello BBC...!</p> <p>I think it's good for us children, when you told us how dangerous handys are. I'm \</p> <p>#1.# I think Handys are good, because the parents haven't angst. They can always</p> <p>#2.# When you are a girl and a men want vergewaltigen you, you can call anywhe</p> <p>On the other hand:</p> <p>#1.# I think handys aren't good, because they sent lacer, who aren't good for you.</p> <p>#2.# Most the children need handys for a bad zweck, <<z. B.>> "Happy sleeping".</p> <p>I think handys mustn't be forbidden, then most they are good. They only must be f</p>	Abbreviations that contain periods, spaces or other characters. Words we would also say out loud such as BBC, USA, TV or SMS don't have to be tagged as abbreviations. When in doubt, tag as abbreviation.	For technical reasons, abbreviations have to be excluded from the POS tagging if they contain other characters, such as dots, apostrophes or colons. Therefore, they are tagged in both the originalText and the taggedText. Abbreviations that only consist of letters can also be left untagged. Errors are never corrected.
unreadable	An unreadable part of the text.	<p><u>Original Text:</u> this is a good idée because i don't like \ for me. Because i would a school finish riding and i have a very homework and a [??] because i like this.</p>	This tag does not contain information.	This tag is used when a word or a number (or more than one) cannot be deciphered. In case the transcriber has an idea or a hypothesis, a comment-Tag can be used after this tag.
par	A new paragraph is started.	<p><u>Original Text:</u> this is a good idée because i don't like wake up for me. Because i would a school finish at 16:00 riding and i have a very homework and i sleep t</p> <p>I don't would changer a [??] because i like this.</p>	This tag does not contain information.	A new paragraph can be entered by pressing Esc and then p. It is only meant to be tagged when a student obviously started a new paragraph. When the author just arrived at the end of a line and

				continued on another, no par-Tag is entered.
period	A student forgot to put a period at the end of a sentence.	Original Text: this is a good idée because i don't like wake up at 01 for me. Because i would a school finish at 16:00 bec riding and i have a very homework and i sleep tôt. I don't would changer a [??] because i like this[.] 	This tag does not contain information.	For technical reasons, a sentence has to be ended by a period. If the students did not write one, this tag is used.

Any other information must be put into the tagged text.

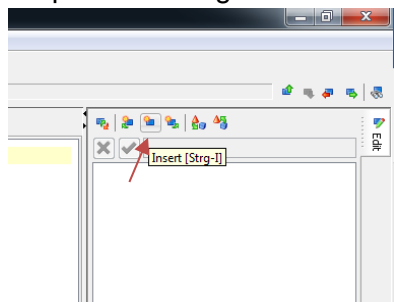
All tags mentioned above (including all paragraph breaks) must appear in both versions of the text!

2.2. Tagged text

Once the original text has been transcribed, it can be copied into the field for the tagged text. Usually, all tags (numbers, unreadables, pars etc.) from the original will be copied along with the text and can be left alone. Sometimes, the last tag at the end has to be put back in.

Step 1: Check whether the text was correctly copied from the original and re-insert tags if necessary. All tags in the originalText must also appear in the taggedText!

Step 2: Insert tags. To insert a tag, type Ctrl+i or click the “Insert tag”-button (right side of the screen):





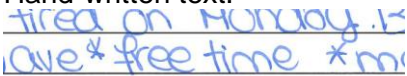

The following tags can be inserted in the tagged text. Usually, it will be necessary to delete the word(s) meant to be put into a tag. Then, the tag is inserted and the word (and, possibly, its correction or translation) are re-typed.

Please note that a “word” in the following descriptions is delimited by empty spaces. Apostrophes and hyphens are not considered to delimit a word: *S-Bahn*, *orang-utan*, *c'est* and *qu'est-ce* are all “single words”. Thus, any tag, even if it concerns only a part of the word, must contain the whole “word” (e.g. “*l'amitié*” must be tagged as `<error><originalForm>l'amitié</originalForm>...`). In other words: no apostrophe (') must be directly in front or directly behind a tag.

Tag	Description	Example	Permitted values	Comment
ambiguous → alternative (2x)	If the transcriber is unsure what the author meant, (s)he may give two alternatives.	<p>like /wake/wack/ up :</p> <p>change a [!]planing </p> <p>In the original text, the <u>first</u> word must be transcribed without tags.</p>	Two words (or numbers), one for each alternative.	Only two alternatives are possible, two alternatives must be given. In the original text, the first alternative is transcribed.


Tag	Description	Example	Permitted values	Comment
anonymized	<p>An information which could identify the author was changed.</p> <p>No identifiable information must remain in either text.</p>	<p>Original Text: this is a good idée because i don't like wake up at 06:30, this is a c for me. Because i would a school finish at 16:00 because i have hc riding and i have a very homework and i sleep tôt. </p> <p>I don't would changer a [??] because i like this[,] Love, =Maria=</p> <p>The anonymized names must also appear in the original text (but without the tag)!</p>	Names, places	The anonymization table contains a list of names and places that will be used instead of what the students write. No information that may identify the students must remain in the text!
comment	A comment by the transcriber (!).	<p>!Maria=</p> <p>this is a [[Etwas wurde weggestrichen, unlesbar]]diff</p>	<p>Any characters or numbers.</p> <p><i>No tags must be inside this tag.</i></p>	<p>Please keep as short as possible.</p> <p>Comments only appear in the tagged version of the text. They will be transferred to EXMARaLDA on a comment line. Do not use any other tags within comment.</p>

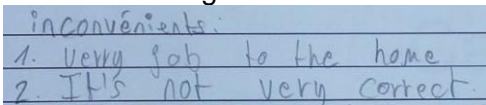
Tag	Description	Example	Permitted values	Comment
insertion	The author (!) corrected his/her own text by inserting letters or words into the text.	<p>Original text: and have not time to do.</p> <p>Tagged text: and ^{^have^} not time to do.</p> 	one or more words or numbers	<p>These are corrections done by the author! Mark as insertion only what was obviously inserted as a correction to the text. If some words were moved within the text, the final position is marked as an insertion.</p> <p>Inserted text must be at the same position as in the original text version. There, it is <u>untagged</u>.</p>
deletion	The author (!) corrected his/her own text by erasing a part.	<p><u>Original Text:</u> I can't ist too difficult[.]</p> <p><u>Tagged Text:</u> I [can't can't] [ist is] too schwer difficult[.]</p> 	<p>one or more words or numbers</p> <p><i>No tags must be inside this tag!</i></p>	<p>These are corrections by the author! If the deletion is unreadable, it is not inserted as a tag. If some words were moved within the text, the original position is inserted in the tagged text as a deletion. A deletion must <u>not</u> be part of the original text.</p>

Tag	Description	Example	Permitted values	Comment
emphasis	The author emphasized something.	<p>Hand-written text: </p> <p>Original text in XMLmind: But we don't have free time most on</p> <p>Tagged text: But we don't have <u>free time</u> most on</p>	One or more words or numbers	When the author underlined something, wrote something in capitals, used another color or otherwise emphasized something. It is possible (but not necessary) to write a comment to better describe the emphasis. Any symbols used for the emphasis are <u>not</u> transcribed (see the asterisks in the example).
emoticon	The author used an emoticon, either as a picture -☺- or in punctuation :-)-.	<p>Original Text: I can't ist too difficult[.]</p> <p>Tagged Text: I [can't can't] [ist is] too difficult [[a crying smiley]][.]</p> 	Anything, as short as possible.	Insert a short description of the smiley or the smiley itself, if written in symbols. Content within this tag will appear on the comment line. The smiley must <u>not</u> be part of the original text.
error	The author made a spelling error.	<p>Original Text: es gibt ein bischen Bier</p> <p>Tagged Text: es gibt ein [bischen bisschen] Bier</p>	2 words or word forms: originalForm: the word written by the author	Error correction is meant to facilitate tagging: a word of the correct word class must be created that can be recognized


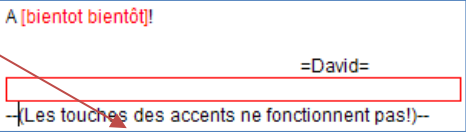
Tag	Description	Example	Permitted values	Comment
		<p>Original Text: En question d'amitié on les</p> <p>Tagged Text: En question [d'amitié d'amitié] on les</p>	targetForm: the corrected form of the word	<p>by the tagger. No further corrections are necessary!</p> <p>Either the originalForm or the targetForm <u>must be one word</u>. The other may vary, i.e. in cases like "freetime" → "free time" or "Jugend Herberge" → "Jugendherberge". The originalForm must be identical to the one in the original text.</p>
foreign_Deu, foreign_Fra, foreign_Eng or foreign_oth	The author wrote a word, which obviously is from another language than the language of the text <u>but not</u> in the language of schooling.	<p>Original Text: can start at neun hour</p> <p>Tagged Text: can start at [neun nine] hour</p> <hr/> <p>Original Text: many hours for schul .</p> <p>Tagged Text: many hours for [schul school]</p>	<p>2 words or word forms:</p> <p>originalForm: the word written by the author (incl. spelling mistakes)</p> <p>targetForm: the word translated to the language of the text (correctly spelled)</p>	<p>Translation is meant to facilitate tagging: a translation as close as possible to the original word, in the correct word form. Either the originalForm or the targetForm <u>must be one word</u>. The other may vary, i.e. in cases like "free time" → "Freizeit" or "Jugendherberge" → "camp de jeunesse". The originalForm must be identical to the one in the original text.</p>

Tag	Description	Example	Permitted values	Comment
foreign_text	The author wrote a longer sentence or text in another language than the text should be.	This tag and the text in it only appear in the Tagged Text!	Any text that would be bothersome to tag as single foreign words.	If it would be too much to tag whole passages in the foreign language, this tag can be used. It is treated like a comment tag: It does not appear in the Original Text and it will not be tagged.
schlg_item	The author wrote a word in the language of schooling, which is not the language of the text.	<u>Original Text:</u> Die camps de jeunesse finde ich gut <u>Tagged Text:</u> Die [camp de jeunesse Jugendlager] finde ich gut	2 words or word forms: originalForm: the word written by the author (incl. spelling mistakes) targetForm: the word translated to the language of the text (no errors)	Translation is meant to facilitate tagging: a translation as close as possible to the original word, in the correct word form. Either the originalForm or the targetForm must be one word. The other may vary, i.e. in cases like "camp de jeunesse" → "Jugendherberge" or vice-versa.

Tag	Description	Example	Permitted values	Comment
schlg_text	The author wrote at least an entire sentence in the language of schooling	<p>This tag and the text in it only appear in the Tagged Text!</p> <p>Original Text: #2 #It's not very correct.</p> <p>#3# </p> <p>Tagged Text: #START#2.#It's not very correct.#END</p> <p>#START#3.#[[On apprend mieu à faire les devoirs en classe qu'à l'école.]]#END</p> <p>[[Ne pas faire les devoirs à la maison mais plutôt tous faire à l'école.]]</p> <p>Note the par-Tag – the red box resp. the empty line after #3# – that appears in both versions!</p>	Any character. Tags can be inserted, but they will be shown as simple text in the Exmaralda file.	If it would be too much to tag whole passages in the language of schooling, this tag can be used. It is treated like a comment tag: It does <u>not</u> appear in the Original Text and it will not be tagged.
image	The author inserted a picture or a graph of some kind.	<p>Original Text: ich mag Skifahren sehr.</p> <p>Tagged Text: ich mag Skifahren sehr. [[Bild von einem Herz.]]</p>	A description of the image. Any character.	This tag works like comment. It must not be in the original text, should be a succinct description and will be put into the comment line.

Tag	Description	Example	Permitted values	Comment
list	The author made a list of some kind.	<p>Handwritten original:</p>  <p>Original Text:</p> <p>inconvenients:</p> <p>#1.#very job to the home[.]</p> <p>#2.#It's not very correct.</p> <p>Tagged Text:</p> <p>[inconvenients disadvantages]:</p> <p>#START#1.#[very Very] job to the home#END[.]</p> <p>#START#2.#It's not very correct.#END</p> <p>Limits of each listing: #START and #END. Note that the numbers were written by the students. If there were none, no numbers would appear.</p>	Everything the author wrote into one point in the list.	The author made a list, with bullet points, numbering or other. Each point in the list goes into one list tag to determine the beginning and the end. Whatever is used to denote the list (the bullet points, number etc.) can be put in the first place after the list START.
number	A number. Including any characters related to it (1.; 1'500; 3:30 ...).	<p>Original Text:</p> <p>this is a good idée because i don't like wake up at #06:30#, this is me. Because i would a school finish at #16:00# because i have h</p> <p>#1.#very job to the home[.]</p> <p>mittime is for #2#h#30# of time</p> <p>wake up at #06:30#, this</p> <p>This tag also needs to be in the original text version!</p>	Numbers and related characters. When in doubt, put in more.	For technical reasons, numbers have to be excluded from the POS tagging. In order to do that, they need to be tagged in both versions of the text as numbers. This is especially important if there are other characters, such as dots, apostrophes or colons.

Tag	Description	Example	Permitted values	Comment
par	A new paragraph is started.	<p>this is a good idée because i don't like w for me. Because i would a school finish : riding and i have a very homework and i</p> <p>I don't would changer a [??] because i ill</p> <p>This tag also needs to be in the original text version! Par-Tags at the end of the text (i.e. when a comment is added in the tagged text) must also appear in both versions.</p>	This tag does not contain information.	A new paragraph can be entered by pressing Esc and then p. It is only meant to be tagged when a student obviously started a new paragraph. When the author just arrived at the end of a line and continued on another, no par-Tag is entered.
period	A student forgot to put a period at the end of a sentence.	<p>I don't would changer a [??] because i like this.]</p> <p>This tag also needs to be in the original text version!</p>	This tag does not contain information.	For technical reasons, a sentence has to be ended by a period. If the students did not write one, this tag is used.
symbol	The student used a symbol that cannot be easily typed as such.	<p>Original Text: auch das Howerboard in</p> <p>Tagged Text: auch das [Howerboard Hoverboard] [(R) registered trademark] in</p>	A description of the symbol.	If the student wrote a symbol that is difficult to type, use this tag and describe it. It must not appear in the Original Text.
entity	The student spoke of something specific that cannot be translated.		A proper name.	Proper names, such as names of institutions or things are marked with entity. Spelling mistakes can be kept. Independent of whether it's written in target language or not.

Tag	Description	Example	Permitted values	Comment
off_topic	The student wrote something that has nothing to do with the text itself.	<p>Original Text:</p>  <p>Tagged Text:</p>  <p>(Off-topic text must not appear in the original text. Note the par-Tag – the red box denoting an empty line – that appears in both versions.)</p>	Any text, not corrected or changed in any way.	<p>If a student wrote a comment or a text that is not part of the task, use this tag. The text must <u>not</u> appear in the original text.</p> <p>There must be no other tag within this tag!</p>
Incomplete	The text has obviously not been finished in time.		Nothing	This tag appears at the end of a text when a student obviously did not finish the text they were writing (e.g. for time reasons). This tag must <u>not</u> appear in the original text.