

Phonemic Tagging using the Character Mapper

The orthography of some languages is closely related to its phonology.

For example, in the case of Te Reo Māori, it's possible to devise a relatively simple mapping from spelling to phonemes; most letters can represent themselves, with some letter clusters mapping to specific phonemes (e.g. “ng” → /N/, “wh” → /f/, etc.).

LaBB-CAT's Character Mapper layer manager can be used to define the details of such a mapping, in order to generate the phonemic transcription annotation layer.

Creating a Phonemes Layer

The Character Mapper maps characters strings in the source layer to new characters strings to generate in the target layer. e.g. for mapping spelling to phonemic transcription.

To create a new layer with annotations from your dictionary:

1. Select the *word layers* option from the menu - this will display a list of all the word layers you already have in the database.
2. At the top of the list, there's a blank form for creating a new layer - fill this form in:
 - *Layer ID* - enter a one- or two-word description - e.g. **phonemes**
 - *Type* - select *Phonological*
 - *Manager* - select *Character Mapper*
 - *Alignment* - select *None* (as these are simply tags on the orthographic words)
 - *Generate* - select *Always*
3. Press the *New* button to create the layer
4. You will see the layer configuration page. Check the online help for explanations of all options.
 - *Source Layer* - the layer from which annotations will have their characters matched.
 - *Language* - a regular expression identifying the ISO 639 code of the language that this configuration should annotate, e.g. en-NZ for New Zealand English, de.* for any variety of German, or blank for any language at all.
 - *Mappings* - a list of mappings from source characters to destination characters. If a character (or sequence of characters) in the left column is found in the source layer, then the corresponding character(s) in the right column is saved in the destination layer. Mapping rows are processed in order, and once a match is found, the rest of the rows are ignored for that character.
 - Rows can be added using the “add” button on the right, or removed by selecting a mapping (by clicking in the source or destination box) and using the “remove” button.

- The order of mappings can be changed by selecting a mapping and using the “up” button to move it up, or the “down” button to move it down.
- For characters in the source layer that don’t match any character in the left column, you can either Copy them into the destination layer, or Ignore them (so they are not copied).

Source layer: orthography Language: mi

Mappings

ai	→	W
ae	→	2
ao	→	6
au	→	5
ei	→	1
ou	→	O
oi	→	4
oe	→	X
ā	→	a
ē	→	e
ī	→	i
ō	→	o
ū	→	u
wh	→	f
ng	→	N



Copy ▼ characters with no mapping

Save

5. Press *Save*
6. Press *Regenerate*.

You will see a progress bar while the layer manager annotates all the transcripts that have already been uploaded.

LaBB-CAT will then generate annotations for all the transcripts you already have in your database. If you have a lot of data, this may take a while.

From now on, when you upload a new transcript, annotations will automatically be generated by using the mapping rules you specified.