Specifying Language in ELAN transcripts

For multilingual corpora, it's important that each transcript specifies the language of the speech that has been transcribed.

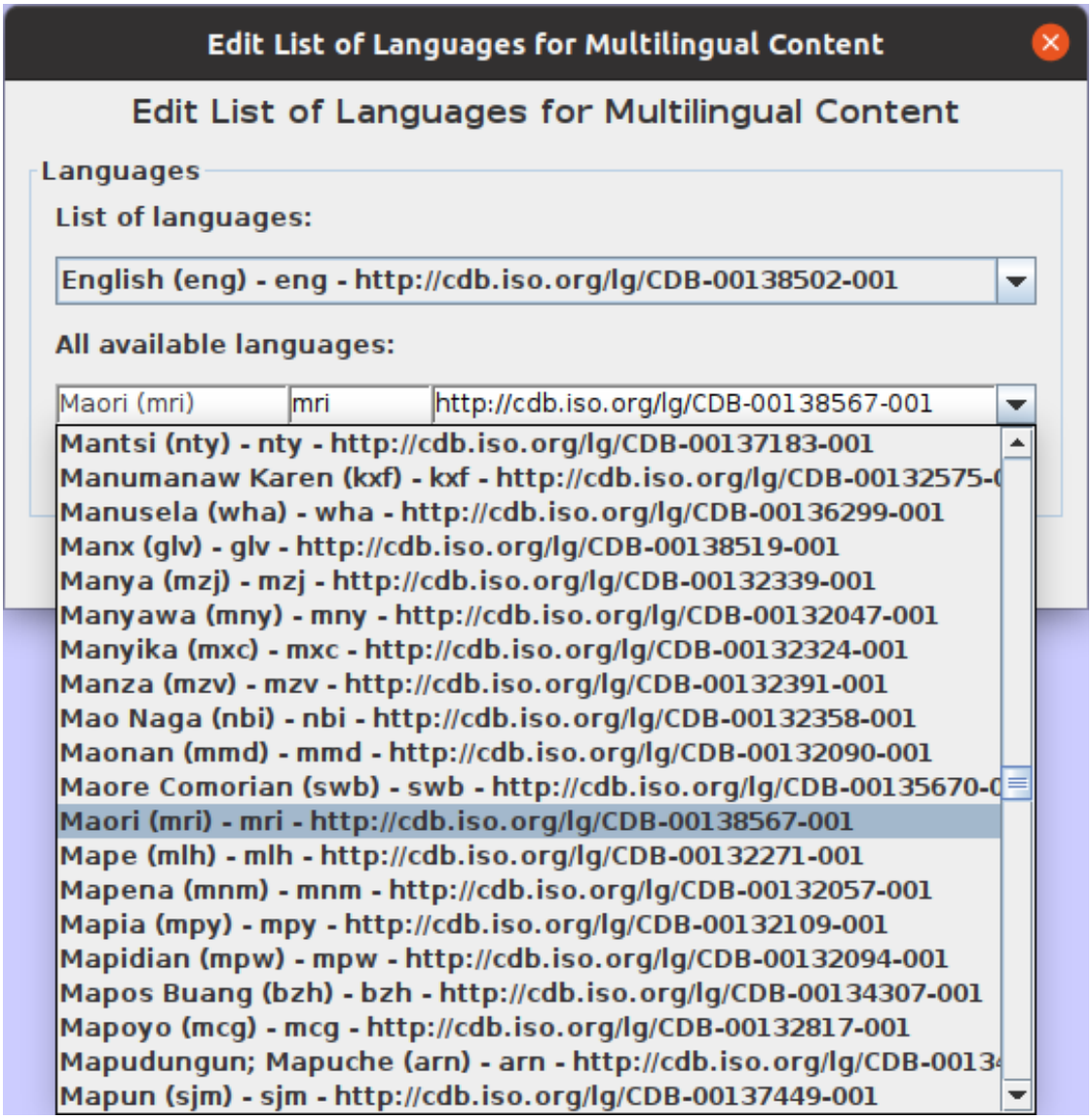
In ELAN, each tier can have a language specified by setting the *Content Language* attributed of the tier. This is set using a dropdown box, which unfortunately is often populated with a single item: *English (eng)*.

If the speech is in a language other than English, you will first have to add the desired language to the list of options.

Adding a new language option

1. Click the *Edit* menu and select the *Edit List of Languages…* option.

2. Click the lower dropdown box (directly above the buttons) and select the language you want.



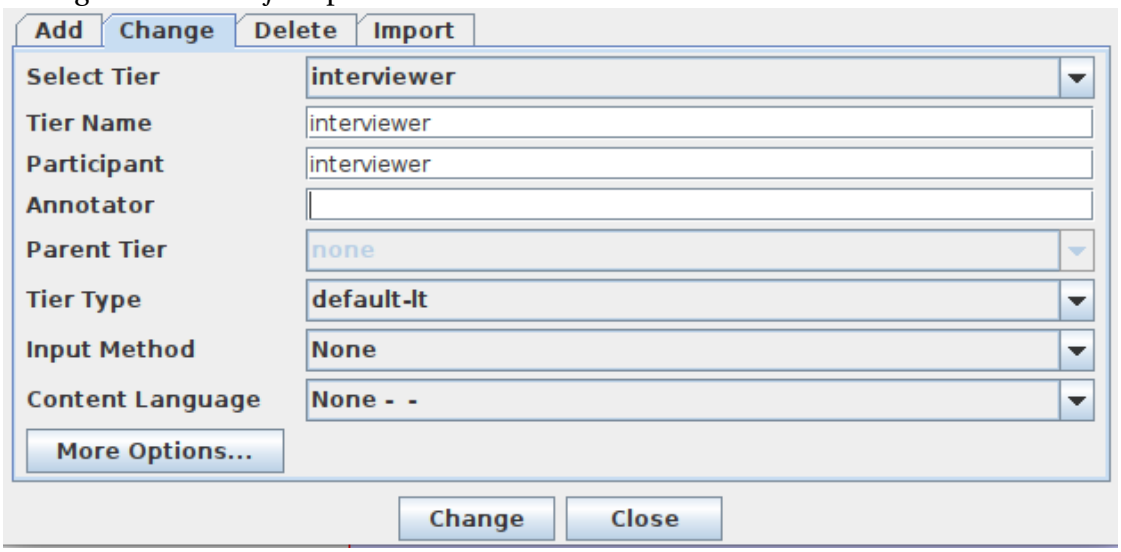
3. Click *Add*.
   The selected language should appear in the upper dropdown box.

4. Click *Close*

Defining the language for the transcript

Your transcript tiers now need to have this language selected.

1. In the transcript, right-click the name of a transcript tier on the left and select the *Change Attributes Of…* option.



2. Change the *Content Language* option by selecting the language from the dropdown list.

3. Click Change.
   The tier attributes window will close.
4. Repeat the previous three steps for each transcript tier.

Speech in a Different Language

Sometime the speaker may include words or phrases in a different language from the rest of the transcript. It may be important to tag these words with the language they're in (e.g. to ensure that after upload to LaBB-CAT, the pronunciation is correctly inferred).

LaBB-CAT recognises a transcription convention for tagging 'code-switches' (CS) into another language. Immediately following the word (i.e. with no intevening space) in square brackes the code "CS:" is added with the ISO 639 3-letter code for the language. e.g.

> …me mudé de Christchurch[CS:eng] en 2004…

For longer phrases, the code-switch tag can be placed immediately before the first word and immediately after the last word, to mark those and all intervening words as being in a different language. e.g.

> …it has a certain [CS:fre]je ne sais quoi[CS:fre] I think…