Upload and Force-align Transcripts

This workflow details the steps required to upload new transcripts and then run forced alignment on them.

> **Assumptions**
>
> These instructions assume that
>
> - you've got a LaBB-CAT server set up
> - it's already configured for forced alignment with HTK
> - the HTK configuration is designed for the per-speaker train/align procedure

There are two sub-tasks:

1. upload the transcripts for a given speaker, and then
2. run forced alignment for all of that speaker's utterances.

These two broad steps will be repeated for each speaker.

1. Upload Transcripts

As forced alignment is done on a per-speaker database, it's best to upload all the transcripts that a speaker appears in before running forced alignment.

(1) Choose a speaker/participant to upload.
(2) Identify all the transcripts/recordings that they appear in. This may be only one transcript, which is fine. But if they're in more than one recording, they should all be uploaded before forced alignment, to maximise the amount of speech available for training acoustic models for alignment.
(3) In LaBB-CAT, select the *upload* option in the menu.
(4) Select the first option, *upload transcripts*.
(5) Press the left-hand *Choose File* button and select the first transcript file (it may be a *.eaf* ELAN file, or a Praat *.TextGrid* file, etc.).

When you select a file, a new row of *Choose File* buttons will appear below the first. This is for adding more transcripts in the 'episode'. An 'episode' is a set of transcripts that belong together because they were recorded during the same session. Unless your recordings for a speaker wer all recorded on the same day, each recording session has only one recording.

(6) Next to *Media* on the first row, press *Choose File*
(7) Select the *.wav* file that corresponds to the transcript.
(8) Ensure the *Corpus* option is correct for the transcript you're uploading.

(9) Ensure the *Type* option is correct (e.g. *interview* for an interview, *word-list* for a word list reading, etc.)
(10) Leave the other options as-is, and press *Upload*

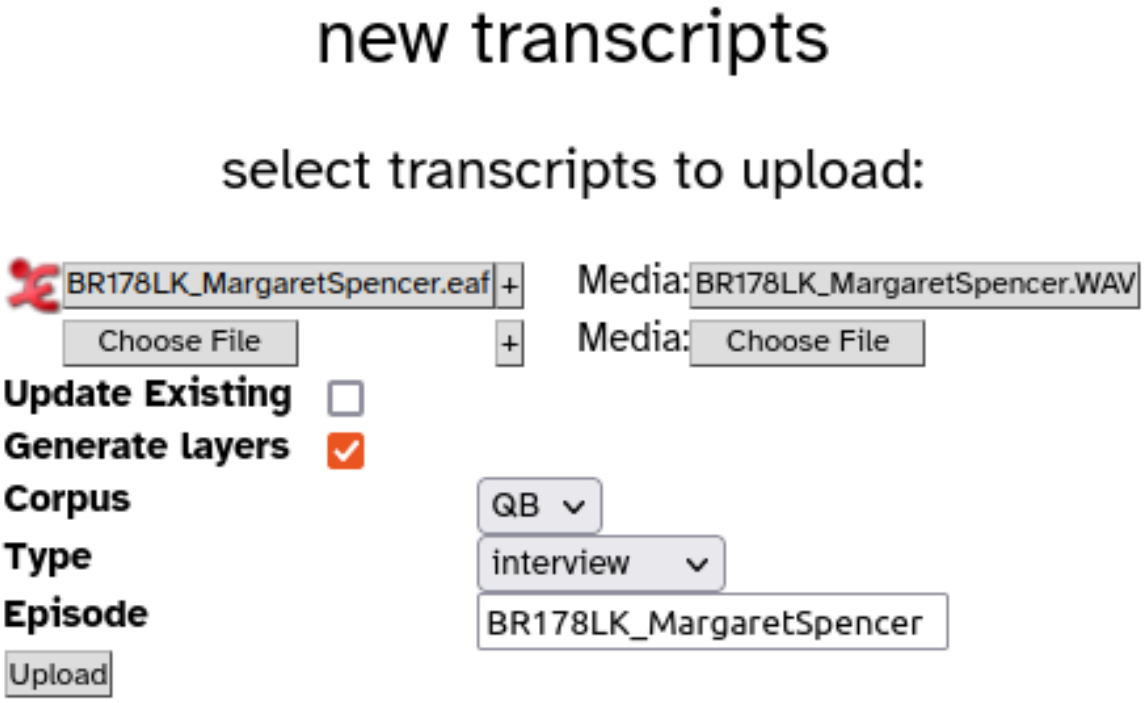

Figure 1: Select the transcript on the left and the media on the right

ELAN/Praat transcripts have a number of Tiers defined in them, e.g.:

- one for the participant's utterances,
- another for an 'interviewer' if there is one,
- one for noise annotations,
- one for transcriber comments, and
- one for topic annotations.

Each tier must be mapped to a LaBB-CAT annotation layer.

Now that you've uploaded the file, LaBB-CAT has analysed the structure of the transcript file and pre-selected some default options for layer mappings. These defaults are correct under most circumstances, but it's a good idea to double-check that

- tiers that contain transcripts of speech are mapped to *utterance* and
- tiers that contain other information aren't mapped to *utterance*.

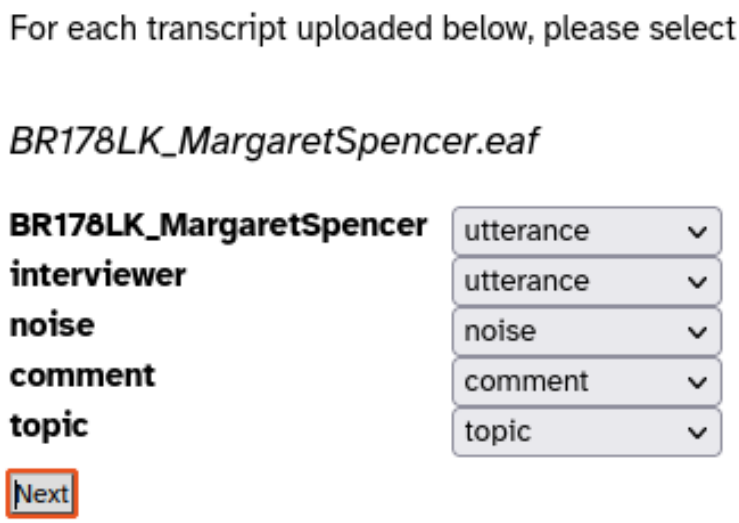

Figure 2: Check mappings of tiers to layers

(11) Click *Next* to continue.
This will display a page listing all the speakers in the transcript, so you can select which one is the 'main participant', which is the speaker selected by default for searches and other processing.

(12) Ensure that the target speaker is ticked, and others like the interviewer are not ticked, and click *Set Main Participants*.
This will display a page with the name of transcript you uploaded, with an *edit meta-data* link, and a progress bar (which may have already finished). The progress bar indicates progress with processing of automatically generated annotations.

(13) Ensure the progress bar finishes, and there are no errors.

If there are more transcripts to upload for this speaker, you can repeat from step 5. above, using the form below the heading 'select transcripts to upload'.

(14) Click the name of the (last) transcript you uploaded.
This opens LaBB-CAT's interactive transcript page.

(15) Double check that the transcript text appears correct and that the media playback control appears on the top right.

(16) Click on an utterance and select the *Play* option from the resulting menu.
Ensure that the audio corresponding to the given utterance is played.

At this point, the transcripts have been uploaded and are ready for forced alignment.

2. Forced Alignment

To start a forced-alignment process per-speaker, you need to first select a speaker who will be aligned. Then you will fill in any pronunciations that are missing from the dictionart. After that, HTK will automatically force-align their utterances, producing start/end times for each speech sound.

(17) In LaBB-CAT, select the *participants* option on the menu
(18) Find and tick the speaker.
(19) Press the *All Utterances* button
(20) Click *List*
(21) Once the paginated list of utterances appears, press the *Htk* button below.

Basically you need to fill in the boxes with the pronunciations and click *Save Pronunciations*.

> **Note**
>
> - You don't have to fill them all in at once, you can do a few, and click *Save*, which will save your work and list what's left.
> - You don't have to fill them all in, you can leave some empty and continue with the HTK forced-alignment by clicking *Start* (HTK will ignore any lines where the remaining unknown words appear, but the ones you filled in will be included).
> - Some of the boxes will be initially filled in with a suggestion from the lexicon layer manager - these may or may not be correct, and aren't saved until you save them.
> - The pronunciations have to be in the 'DISC' format - i.e. one character per phoneme, with no spaces. There's a 'helper' link on the right of each pronunciation box - if you click it, it expands into a list of clickable phonemes - just the ones that aren't ordinary letters, and diphthongs etc.
> - The *search* button lets you look up the lexicon for similar words - this probably won't help for place names, but for words like "tarseal", you can click the *lookup* button, enter "tar seal" in the box as two separate words, and you'll get back the DISC pronunciation of each word, with clickable buttons to copy the given pronunciation into the box. This is useful for digits and numbers too, which may not be in the lexicon - so for "1", search for "one" and copy the pronunciation.
> - If you click on the word itself, the transcript for the first instance of that word is opened, in case you want to listen to it, or in case it's actually just a typo and you want to correct the transcript.
> - If you're using CELEX, when you specify the pronunciations, it's recommended to put syllable separators (-) and primary stress markers (') too - e.g. for "tarseal" you can put *t#sil* but it would actually be better to put *t#-'sil*. These markers are entered into the dictionary even though they're stripped out for HTK, and they may come in handy later (e.g. the syllable separators are used by the CELEX layer manager to count syllables).

When you add pronunciations this way, they're added to the dictionary and all the instances of those words in LaBB-CAT are updated with the pronunciations - not just the participant you're looking at, but all participants in the database. So you only have to come up with a pronunciation for each word once.

(22) Once you've filled in all the missing pronunciations, forced alignment will start automatically. If you want to start forced alignment before you've entered all pronunciations, click the *Start* button at the bottom of the page.

     You should see a progress bar while the forced alignment is running. It will take a few minutes to complete.

Once HTK has produced the word and segment alignments, it:

- sets the start/end times of the words on the transcript layer accordingly,
- adds new phone annotations to the *segments* layer with the alignments of the phones, and
- saves a timestamp in the *htk* layer.

When the layer manager has finished, you'll see a message saying "Complete - words and phones from selected utterances are now aligned."