Transcribing with Praat

If you already have Praat TextGrids that are transcripts of recordings, you may be able to upload them directly into LaBB-CAT, instead of converting them to Transcriber files.

Prerequisites

There are several conditions that your TextGrids must meet in order to use them as source transcripts for Praat:

1. They must be transcripts - i.e. each must contain at least one tier that contains intervals containing the words uttered by the speaker(s) in ordinary orthographic spelling.
In particular, if the TextGrid contains only sparse, isolated intervals that code individual words or phonemes, but no transcription of whole utterances, then it cannot be used as a transcript for uploading directly into LaBB-CAT (but once you have a full transcript for the recording, you *may* be able to upload the coded annotations in your original TextGrid)
2. The speaker of each utterance must be systematically identifiable - possible ways of doing this are illustrated below.
3. In addition to utterance interval tiers, the TextGrid may contain other tiers that have intervals marking noises, comments, and other annotations. If you want to import these into LaBB-CAT, then you must first ensure that, for each such non-transcription tier, there is a layer in LaBB-CAT already created, ready to receive the annotations. It's best to make sure that the TextGrid tier name and the LaBB-CAT layer's 'name' are the same.
(there are already LaBB-CAT layers for 'noise' and 'comment', so you don't need to create those)

TextGrid Transcript Structures

As mentioned above, the speaker of each utterance in the TextGrid must be systematically identifiable. There are two possible structures for your TextGrid to facilitate this:

- One tier per speaker, tier name = speaker name
- Turns tier and utterances tier, turns tier containing speaker names

One tier per speaker, tier name = speaker name

You may have one tier for each speaker. In this case, each interval marks the start/end times of each utterance (equivalent to each line in Transcriber transcripts), and the name of the tier is the name of the speaker, as shown in Figure 1.

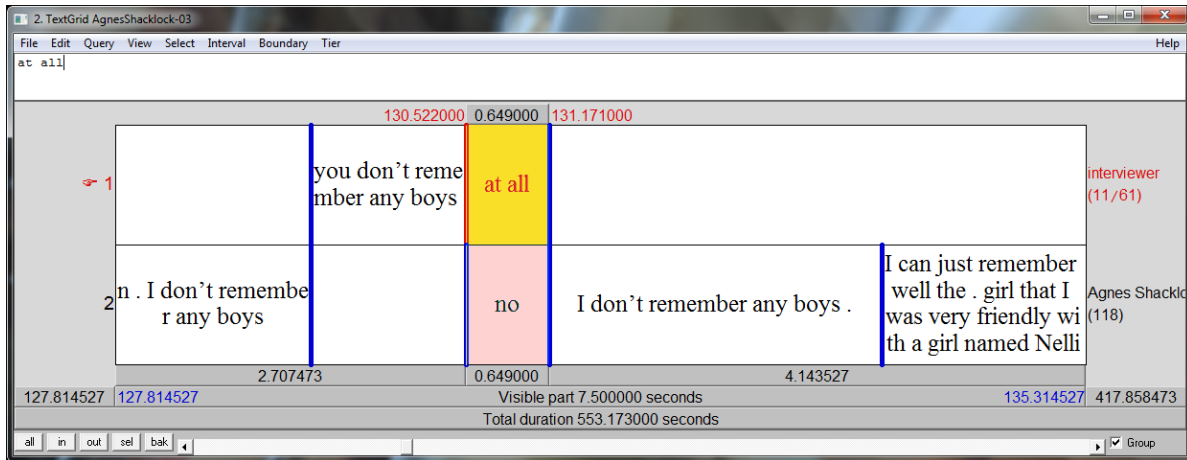With this structure, you can annotate simultaneous speech.

Figure 1: Each tier corresponds to the transcript of one speaker

Turns tier and utterances tier, turns tier containing speaker names

Alternatively, your TextGrid may have two tiers, one called `turns` and the other called `utterances`. The `turns` tier marks the start/end times of speaker turns, and contains the names of the speakers speaking, and the `utterances` tier contains the actual words spoken by the speaker speaking at that time, as shown in Figure 2.

You can have multiple utterances per turn, but you must ensure that the first utterance starts at the same time as the start of its turn, and that the last utterance ends at the same time as the end of its turn.

With this structure, you cannot annotate simultaneous speech.

Uploading TextGrid Transcripts

When uploading transcripts that are TextGrid, use exactly the same upload page as for all other formats - i.e. select *upload* on the menu, and then select *upload transcripts.*
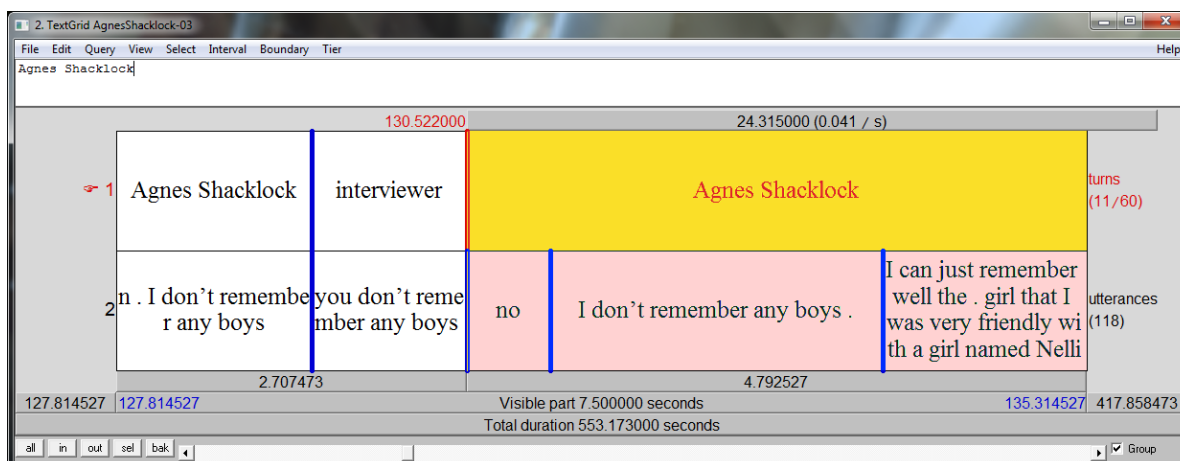
Figure 2: The turns tier labels the current speaker, and the utterances tier contains utterance transcription

(Make sure you **don't** select the *upload textgrid annotations* option on the *upload* page - this is for uploading annotations into transcripts that already exist in the LaBB-CAT database)

If your TextGrids contain tiers that are annotations rather than utterances or turns, and you want to import those annotations, then you must first ensure that matching layers exist in LaBB-CAT. Most likely you will want to create *span layers,* so select that option on the menu, and when you create new layers, set the alignment of the layer to 'Time Stamps' for point tiers, and 'Time Intervals' for interval tiers.

Once you're sure all the layers you need exist, the steps are:

(1) Select *upload* on the LaBB-CAT menu.
(2) Select *upload transcripts.*
(3) Press *Choose File* on the left and select the transcript *.TextGrid* file.
   On the right the text "Praat TextGrid" should appear.
(4) Press the *Choose File* to the right of this to select the media file(s) that correspond to the transcript.
(5) Press *Upload.*
(6) Now you must specify how LaBB-CAT should interpret the tiers. On the left are listed the tiers in the TextGrid, and for each you can specify a LaBB-CAT layer for it.

   - For tiers that contain transcriptions, select the *utterances* option, and if you're using the "turns tier and utterances tier" structure described above, ensure that the `turns` tier has the *turns* layer selected.
   - For tiers that contain annotations for noises or other events, select the appropriate layer.
   - For tiers that you don't want to import, select *[none]

Figure 3: Mapping TextGrid tiers to LaBB-CAT layers

In Figure 3, the `crosscheck` tier is not being imported, the annotations on the `noise` tier will be inserted into the *noise* layer in LaBB-CAT, and the `interviewer` and `Agnes Shacklock` tiers contain the transcriptions of the speech of those two speakers, so they are mapped to the *utterances* layer in LaBB-CAT.

(7) Click Set Mappings, and the data from the TextGrids will be imported into LaBB-CAT.

Conventions for non-speech annotations

Praat has no direct mechanism for marking non-speech annotations in their position within the transcript text. However, LaBB-CAT supports the use of textual conventions in various ways to make certain annotations:

- To tag a word with its pronunciation, enter the pronunciation (with no spaces) in square brackets, directly following the word (i.e. with no intervening space), e.g.:
  `…this was at Wingatui[wIN@tui]…`
- To tag a word with its full orthography (if the transcript doesn't include it), enter the orthography (with no spaces) in round parentheses, directly following the word (i.e. with no intervening space), e.g.:
  `…I can't remem~(remember)…`
- To insert a noise annotation within the text, enclose it in square brackets (surrounded by spaces so it's not taken as a pronunciation annotation), e.g.:
  `…sometimes me [laughs] not always but sometimes…`
- To insert a comment annotation within the text, enclose it in curly braces (surrounded by spaces), e.g.:
  `…beautifully warm {softly} but its…`

During upload, these annotations will be extracted from the transcript text and inserted into corresponding LaBB-CAT layers.