2. Setting Up

In this exercise you will:

- Define corpora
- Define transcript types
- Define speaker meta data

After this you will have an empty LaBB-CAT database set up ready to upload transcripts into.

Now that the software is installed, we will set up a basic structure for receiving data:

(1) Open LaBB-CAT.

> **Note**
>
> If you will be running LaBB-CAT on your own computer, you will need to start LaBB-CAT, and this will open your browser on the LaBB-CAT start page.
> If you are using a LaBB-CAT server that's already been installed for you elsewhere, you will have been given a link or URL to use; open your browser on that page now.

(2) The start page has a link on it called *Where do I start?* - you may like to click on this link and read the first section, which explains a little about how to navigate around LaBB-CAT and where to find online help and hints.

(3) Click back on the start page of LaBB-CAT (the page with the *Where do I start?* link).

Now we will set up some corpus names...

(4) On the menu at the top, click the *corpora* link.
This page shows a list of current corpora, which only contains one corpus, called *corpus*.

(5) Above the *corpus* corpus, there's a form that you can fill in to add a new corpus. Fill in the following information:

- **Name** : `QB`
- **Language** : *English*
- **Description** : `Quakebox recordings`

(6) Click the *New* button to add the "QB" corpus.
You should see a message at the top of the page saying *Record created* and now the *QB* corpus is in the list, under the "corpus" corpus.

(7) Add another corpus called *UC* with the description *Campus recordings*.

(8) We won't actually be using the corpus called *corpus*, so we want to delete it. To do this, click the *Delete* button to the right of the *corpus* corpus in the list.

(9) You will be asked *Are you sure you want to delete corpus?* You are sure, so click *OK*.
The row will be deleted from the list.

Now you have some corpora set up with the names you've provided.

The data we are using is a collection of stories about peoples' experiences during the devastating earthquakes that hit the Canterbury region of New Zealand in 2010 and 2011. Some recordings are interviews, where an interviewer asks the participant questions, and others are monologues. Now we're going to set up these two transcript types…

(10) Click on the *transcript types* menu option.
You will see a list of transcript types, although there's currently only one type in the list, called *interview*.
(11) Above this, fill in the empty *Type* box with the word: `monologue`
(12) Click the *New* button.
You will notice that now the list has two transcript types, *interview* and *monologue*. A *Save* button has appeared, because your changes aren't yet saved to LaBB-CAT.
(13) Click the *Save* button.
You will see a message at the top saying *Layer saved: transcript_type*.

Now that we have both corpora and transcript types, we're going to set up meta-data options for the participants in our corpus…

(14) Click on the *participant attributes* menu option.
You will see a list of fields or attributes that participants (or speakers) can have. There are currently only three attributes:

- *Gender* - i.e. whether the participant is female or male or something else
- *Birth Year* - i.e. the year the participant was born
- *Notes* - general arbitrary notes

For our corpus data, we don't have the exact year of birth. Instead we have the age of the participant, defined by various age group categories.

(15) As with previous pages, the headings at the top are also a form you can fill in to add a new row, which we will now fill in with the following information:

- **Attribute ID** : `age_category`
- **Type** : *Select* (this is because we want to be able to select from a list of possible values)
- **Layer Label** : `Age`
- **Access** : *Public*
- **Searchability** : *Searchable* (so that it appears on the search page)

- **Category** : *General*
- **Description** : `Age Category`

(16) Press the *New* button to add the attribute.
You will see a message saying *Layer added: participant_age_category* and the new attribute will now appear at the bottom of the list. Now we need to define the options for it…

> **Tip**
>
> If you want more information about what each of these are for, check the online help for this page.

(17) To the right of the *Age* attribute's *Category* there's an icon like a tag .
Hover your mouse over this icon to see what it does.
(18) Press the *Valid labels* icon.
This shows a (currently empty) list of options for the *participant_age_category* attribute.
(19) In the blank *Label* box, enter: `18-25`
(20) Press the *New* button to add the option.
You will see that "18-25" appears twice on the row that's added:

- On the left is the value that is saved in LaBB-CAT's database.
- On the right is an editable description that is displayed in various places in LaBB-CAT, which can be used to provide a little further explanation about the value.

(21) Change the *Description* to `18-25 years`.
(22) In the *Label* row at the top, enter: `26-35`
(23) Press the *New* button to add the option.
(24) Change the *Description* to `26-35 years`
(25) Similarly, add the following age categories:

- `36-45`
- `46-55`
- `56-65`
- `66-75`
- `76-85`

(26) Add a final option:

- **Label** : `85+`
- **Description** : `85 years or more`

(27) Lastly, press the *New* button without filling in a *Label* to add a 'default' option for participants with missing data.
(28) Press the *Save* to save all your changes to LaBB-CAT.
You will see a message saying *Layer saved: participant_age_category*

(29) Now select the *participant attributes* option on the menu to return the list of all attributes.

We are going to add a few more attributes, but they will be 'free text' fields without predefined options.

(30) Add another attribute, called `ethnicity`. For 'Type' select *String*, and make it *Public* and *Searchable*.

(31) Similarly, add the following *Public Searchable string* attribute:

- `languages_spoken` - a list of languages they speak

  ... and the following *Not Searchable* attributes:
- `grew_up` - what country they grew up in
- `grew_up_region` - what region of New Zealand they grew up in
- `grew_up_town` - what town or city they grew up in

(32) Lastly, as we will not be using it, delete the *Birth Year* attribute.

Now you have an empty database for which you've:

- created two corpora, *QB* and *UC*,
- created a new transcript type, so that we can have monologues as well as interviews, and
- created some new attributes for participants, so we can record the ages of our speakers, their place of origin, and the languages they speak, in addition to their genders.