5. Manual Annotation

Now we're going to create our own layer for manual annotations, and explore ways of populating it. Let's say we're interested in the pronunciation of the vowel in the word "the" when the following word starts with a vowel. We're going to:

- Create a layer for annotations on tokens of the word "the".
- Search for tokens using word orthography, and identify 'false positives' (e.g. cases like "…the one…" where the spelling of the following word starts with a vowel but it's not pronounced as a vowel).
- For the 'true positives', perform some auditory analysis (i.e. listen to them) and tag each token accordingly.

---

(1) To embark on this mini project, we're first going to create a 'project' in LaBB-CAT to categorize our annotations.
Select the *projects* link on the menu.
(2) Add a project called "the" with a description something like *Pronunciation of the vowel in the when followed by a word-initial vowel*, by filling in the form and pressing the *New* button.

Now we're going to create a layer to store our annotations…

(3) Select on the *word layers* option on the menu.
You will see a list of existing word layers, including the *orthography* layer, the *lexical* layer, etc.
The row of column headings at the top is also a form for adding a new layer.
(4) Fill in the top row with the following details:

- **Layer ID:** `the`
- **Type:** *Text*
- **Alignment:** *None* (our annotations are simply tags on words, inheriting their start/end times from the word token they tag)
- **Manager:** don't select any manager, as we'll be adding manual annotations, rather than automatically generated ones
- **Generate:** don't select any option (this setting is only relevant for managed layers, so it doesn't actually matter what you select here)
- **Project:** *the*
- **Description:** `"the" followed by a word-initial vowel`

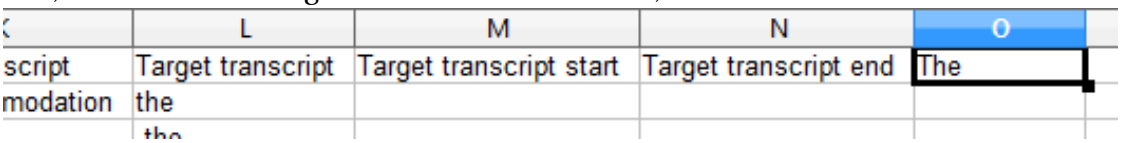
(5) Press *New* to add the layer.

Now we're ready to find some tokens…

(6) Select *participants* on the menu.
(7) We're going to search all male monolinguals, so filter by the appropriate attribute values if they're not already filtered, and then click *Layered Search*.
(8) Search for instances of the word "the" followed immediately by a word starting with a vowel, on the *orthography* layer.
(9) Export the results to a CSV file and open it.

Now we're going to annotate the CSV file to identify false positives.

(10) Add a column to the right-hand side of the spreadsheet, called "The" - i.e. on the first line, in the cell to the right the last column header, enter the word *The*



	L	M	N	O
script	Target transcript	Target transcript start	Target transcript end	The
modation	the			
	the			

(11) For each row in the spreadsheet check the contents of the "Match transcript" column, and decide whether the match is a 'false positive' or not. False positives are cases like "the one", where the second word actually starts with a non-vowel sound (i.e. the word "one" actually starts with a /w/ phoneme).
For false positives, enter `FP` in your new "The" column. For all the others, enter `TP`.
(12) Save the CSV file.
You may be asked if you want to change the format of the file. Resist the temptation to do this - we are going to upload this file into LaBB-CAT, and it can only understand CSV files.

Now that we've annotated our results, we're going to to load our annotations into the new layer we created in LaBB-CAT…

(13) In LaBB-CAT, select the *upload* menu option.
(14) Select the *upload csv annotations* option.
(15) Press *Choose File* and select the annotated CSV file you just saved.
(16) Press *Upload*.
(17) On the form that appears, you can leave the default choices for the options. Just ensure that the *Tag Words* option is selected at the top, and at the bottom the *The* column in the spreadsheet is mapped to the *the* layer in LaBB-CAT.
(18) Click *Insert Annotations*.
You will see a message about how many annotations were added. Now, within LaBB-CAT, each token mentioned in your CSV file has been tagged with either "TP" or "FP"

Now that we've seen one way to add annotations to the database, using CSV files, we will try another way - editing word annotations directly from the interactive transcript.

We're going to find our 'true positive' tokens of the word "the", and annotate each depending on how the speaker pronounces it in the recording.

(19) In LaBB-CAT, select the *search* menu option, which by default searches utterances of all participants.
This time we're going to search for the true-positive annotations we just inserted.
(20) Under the "Pattern" heading, there's now a "Projects" column that includes the "the" project we added at the start. Tick that project, so that the layer associated with it is displayed in the list of *Word* layers to the right.
(21) Tick your custom layer (called "the") in the *Word* column.
Your search matrix is now two layers high by one word wide.
(22) Search for `TP` on the *the* layer.
The results page should show you all the words you annotated with `TP` in your CSV file above.
(23) Click on the first match.
This will open the interactive transcript for the match. You'll be able to see not only the transcript text, but also the *TP*/*FP* tags you have added.
(24) Click on the first match in the transcript, and select the *Play* option to play the line.
(25) Listen carefully to see whether the speaker pronounces the word "the" like "thee" or not. If they do, we're going to annotate the word with the code `i`. Otherwise we're going to annotate it with the code `@`.
(26) Click the match word again, and select the last option on the menu: *Edit*. A window will appear, with a list of layers. Each line has the token's annotation on the given row. So at the bottom, the value on the *word* layer is probably "the" or maybe "The" or "the ." or something similar. On the *orthography* layer, the annotation will be "the". Other layers may be blank, except for your *the* layer, whose annotation will be "TP".
(27) We're going to change the "TP" annotation depending on the pronunciation of the token. So replace "TP" with `i` or `@` as appropriate.
(28) Click the *Save* button  to save your annotation to the database.
(29) Close the "edit word" window.
(30) Back in the interactive transcript, find the next match result - it will be highlighted.
(31) Annotate the next match in exactly the same way - play the utterance, listen to the pronunciation, and change the `TP` to an appropriate code.
(32) Similarly annotate the rest of the matches in the transcript.

> **Note**
>
> You may notice that, although you're changing labels to `i` or `@`, the tags still appear as `TP` in the transcript page; this is simply because that's what the tag was when you opened the transcript. If you refresh the page, you'll see your new tags instead of the old ones.

(33) Once you've annotated the last match in the transcript, close the browser's tab.
This will take you back to the search results page.
(34) You've already annotated all the matches in the first transcript, so move to the next

transcript in the results list, and click the first match.

(35) Annotate all the “TP” tokens in the transcript.

(36) Repeat the above steps until you’ve annotated all the matches.

You’ve now used two methods for annotating words. Although this is a small, toy example, you can hopefully see that you could manage a larger annotation project involving much more tokens, possibly multiple annotators, and working either offline (with a CSV file and maybe extracted WAV files) or online (directly in the interactive transcript page), as preferred.

There are other ways to add manual annotations, which relate to concrete points or intervals in time during the recording. We will see how to do this later…