

Automatically adding text to blank ELAN annotations

March 2014

Sky Onosson

onosson@gmail.com

I am annotating a number of audio files which are recordings of speakers reading from a wordlist. The wordlist itself is saved as a text file with spaces between the words rather than tabs.

I wanted to segment out the words as annotations without having to re-type the entire wordlist, instead loading it in to the blank annotations directly from the wordlist file. I kind of achieved that. I'm still not sure which way is more efficient, but I will be trying it out for a bit longer to see if it's helpful.

=====

The following method is based on Jesse Stewart's recommendations combined with my own attempts, and specified for Mac users. I am using the following software for this process:

ELAN v.4.6.2

TextEdit v.1.9 / Notepad

Microsoft Excel 2011 v.14.0.0.0

The following instructions are fairly detailed and specific - modify as necessary for your purposes.

=====

1. Open up a new file in ELAN -> ⌘N (Win Ctrl+N)
2. Click -> Add Media File
3. Select the audio file for the current speaker
4. Click -> Options>Segmentation Mode
5. Begin playing through your file and use the key command to mark the begin and end of the annotation area for each word in your wordlist. For my purposes, working with a wordlist

task, I am completely ignoring all other speech, background noise, etc. It doesn't get marked. The default commands are CTRL-Space for playback, and ENTER for marking an annotation, and these can be changed under Edit>Preferences>Edit Shortcuts... I prefer using just the spacebar for playback control, as that is standard for most non-linguistic digital audio recording software such as Logic, ProTools etc. which is what I am used to.

6. When you are finished choose File>Export As>Tab-Delimited Text... There will be several options here, so you need to know what your ultimate needs are and choose accordingly. Since I am ultimately uploading these files to FAVE online, I am following the requirements for that which are to only select the Begin Time, End Time, and ss.msec boxes, but you may choose whichever are appropriate for your purposes. This will be relevant in later steps.

7. Save your file somewhere.

8. Open Excel, and select File>Import... then choose Text File and locate your file from step 7.

9. The default import options should work, so just click on Finish.

10. Leave the Excel file for a moment, and open your wordlist file in TextEdit.

11. If your file is, like mine, a list of words with spaces between, you need to convert the spaces to Tabs. The easiest way to do this is go to the very beginning of your file and press Tab, then select that Tab and ⌘C (Win Ctrl+C) to copy it, then delete it. Now press ⌘F (Win Ctrl+F) and check the Replace box. Put a space in the top Find box, then ⌘V (Win Ctrl+V) to paste your Tab in the lower Replace box. Then click All to replace all the spaces with Tabs. Done.

12. ⌘A (Win Ctrl+A) to copy all of the text.

¹13. The following few steps are not very efficient - if you know a better way to do this please let me know!

13a. Go back to the Excel file and ⌘V (Win Ctrl+V) to paste the text into a blank cell.

13b. Now select all the words again by dragging across and ⌘C (Win Ctrl+C)

13c. Click on the top cell in the column immediately to the right of your time data from ELAN. If you used the same choices as me in step 6, this will be column E

13d. Select -> Edit>Paste Special... In the dialogue that comes up, check off the Transpose box.

13e. Delete the text you pasted in 13a which isn't in column E (or wherever you pasted it in 13c-d).

¹ Note from Jesse Stewart: As far as step 13 is concerned, If you have tons of data that take up all the rows of an excel sheet, I might suggest pasting the data into Word. Then instead of replacing the 'space' with a 'tab', replace it with '^p' (paragraph marker). This will put everything into a column. Then paste it into excel.

14. Shift-⌘S (Win Ctrl+S) to save your file and select Tab-Delimited Text

15. Go back to ELAN and select -> Import>CSV/Tab-delimited Text File... locate the Excel file from 14.

16. Under each column you need to specify its data type. For me, column 1 is Tier, column 2 is blank (uncheck the box), 3 is Begin Time, 4 is End Time, and 5 is Annotation. Check your information and select accordingly, then click OK.

17. This will open a *new* ELAN file. Select -> Edit>Linked Files and click Add... then select your audio file from way back at step 3, then click Apply.

You should now have annotations in the positions where you marked them in Step 5, filled in with the words from your prepared wordlist file, without having to type any of them in!

I recommend checking a few random spots throughout the annotations to make sure everything looks good, then save your ELAN file somewhere. If you are uploading to FAVE like me, you can Export it as a Tab-Delimited file for those purposes.

=====

I hope this works for everyone! And thanks again to Jesse Stewart for the crucial information in steps 6 through 16. I've merely applied this to my particular scenario and filled in the details.