# Toolbox/ELAN conversion exercise

Technology and Language Documentation, 19th November 2008

The aim of this exercise is to produce a time-aligned interlinearised transcription of a Cicipu language folktale using both Toolbox and ELAN, making use of the import and export facility in ELAN.

**Preliminary steps**

1. Open the **Cicipu.prj** project in the Toolbox directory under **d:\users\ElanToolboxConversion.**
2. Make sure text **saat002.001** is displayed. Have a quick look at the story.
3. Exit Toolbox.
4. Make a backup copy of **saat002.001.txt** just in case.
5. Find the **Text.typ** file in **d**:**\users\ElanToolboxConversion\Toolbox**. Copy this file into another directory (for example **ElanToolboxConversion**) and rename it to **TextELAN.typ**.
6. Open **TextELAN.typ** in Notepad. Near the top of this file you will see the text **\mkrRecord id**. Change this to **\mkrRecord ref,** and then save the file and exit Notepad[1].

**Importing the Toolbox file into ELAN**

7. Open ELAN
8. Choose **File->Import->Shoebox File...**
9. Tick the **All markers are Unicode** check box.
10. For the **Shoebox file** select the file **saat002.001.txt**
11. In the **Shoebox typ file** box select the **TextELAN.typ** file you created in steps 5-6. This tells ELAN how the Toolbox field markers (\tx, \mb etc...) relate to each other – ELAN needs this information to create the various tiers correctly.
12. A **default block duration** of about 3000 ms is about right for this text. For your own texts you should work out the average duration of your Toolbox references in milliseconds (i.e. file duration divided by the number of references in the text). Doing this will carefully at this stage will make it quicker to time-align the text later.
13. Press **OK**

**Tidying up and linking to the WAV file**

14. To make the display look nicer, right-click on nt@Amos in the lower pane. Choose **Sort Tiers->Sort by Hierarchy**. If you want to, hide the **nt** and **ph** tiers for each speaker.
15. Link the transcription to the WAV file by selecting **Edit->Linked Files...**
16. Click **Add**, and then select **saat002.001.wav** from the **ELANToolboxConversion** directory.
17. Press **Apply** and then 'Cancel' to leave the dialog.
18. Note that the wave form now appears above the transcription
19. **VERY IMPORTANT!!!** Select **Options->Propagate Time Changes->Bulldozer Mode**. If you don't do this you will lose data when aligning transcriptions.
20. Align the transcriptions with the WAV file (if you have not used ELAN yet you will probably need to ask for help here). Just do a few utterances for now as this is time-consuming, but in real-life you would probably do the whole file before exporting back to Toolbox. P.S. Don't forget the children!
21. Save the ELAN file in the **ELANToolboxConversion** directory – choose **Save As...,** and

1 This step is needed because each **id** in the Cicipu Toolbox database relates to a different *text*. In your own text files, if you use a separate **id** for each *utterance* (rather than for each text) then you can omit step 5-6).

call the file **saat002.001.eaf**.

**Exporting the time-aligned ELAN file back to Toolbox**

22. Choose **File->Export As->Toolbox File**.
23. Make sure you are happy with the order of the tiers (you may want to move the **nt** and **ph** tiers below the more standard Toolbox markers **tx**, **mb**, etc... if they are not already there)
24. Uncheck **Wrap blocks**
25. Check the correct TYP file is selected in the **Use Shoebox database type** box.
26. Click OK
27. Choose the same file you imported into ELAN (**saat002.001.txt**), and overwrite it.
28. Exit ELAN

**Opening the (now) time-aligned Toolbox file in Toolbox**

29. *Before opening Toolbox*, open **saat002.001.txt** in Notepad.
    Immediately before the text **`\ref saat002.001.001`** line add the text **`\id saat002.001`**, *on a new line on its own*, save, and exit Notepad. This step needs a bit of care. The top of your file should look something like this:

```
\_sh v3.0  400  Text\_DateStampHasFourDigitYear
\id saat002.001
\ref saat002.001.001\ELANBegin 9.800\ELANEnd
10.800\ELANParticipant Amos\tx mísòoní mísòoní...
```

30. Open Toolbox. You should see your file with the time alignment added.
31. Do some interlinearisation (e.g. three or four lines). There will be a few ambiguity selection boxes – it really doesn't matter which you choose here – this is not a test of your Cicipu!
32. Exit Toolbox

**Finally, repeat steps 7-13 to reimport the file into ELAN and you should see a time-aligned interlinear transcription. Congratulations!**

- Remember this is only one possible 'workflow'. You may do all your interlinear transcriptions in Toolbox first, then convert to ELAN at the end. Conversely you may transcribe directly into ELAN first, and then do the interlinearisation at the end.

Questions to think about:

- Does the \tx field in ELAN look how you would expect?
- For your own project, do you expect to do this process only once or multiple times per text?