

Audio-enabled interlinear template – setup guide

This document explains how to use the audio-enabled interlinear template with your own audio files and transcriptions.

Pre-requisites

This guide assumes you have the following¹:

- A WAV or MP3 file with the text you want to display
- A Toolbox file containing an ELAN time-aligned interlinearised annotation of this text²
- A basic understanding of XML

1 Copying the Interlinear Template files

- Copy the contents of the `interlinear_viewer` folder onto your local computer
- Copy your own audio file into the setup directory

2 Convert audio file to 44.1 kHz MP3

This interlinear template uses Flash technology to play sound. This is because most computers already have Flash plug-ins for browsers installed. Even when they don't, it is easy to download them. Flash also seems to be more reliable than other plug-ins such as QuickTime.

A downside of using Flash is that it cannot play WAV files at all. Also it does not play MP3 files sampled at 48 kHz cleanly. This means that any WAV files you have need to be converted to MP3, and if they are 48 kHz they need to be resampled to 44.1 kHz.

If you have a 44.1 kHz WAV file then there are many tools which can convert this to an MP3 (e.g. Audacity, Creative Wave Studio).

If you have a 48 kHz WAV or MP3 then it is a bit more complicated. Audacity and Creative Wave Studio offer the functionality to resample files to 44.1 kHz but it does not seem to work very well. Instead you can use the freely-available BeSweet utility (<http://besweet.notrace.dk/>), which is included with the files distributed with this exercise.

To perform the resampling do the following:

- Find `Resample.bat` in the setup folder
- Open it in a text editor (do **not** double-click it) – you should see the following:

```
besweet\besweet.exe -core( -input "sample.wav" -output
"..\audio\sample.mp3" -logfilea log.txt ) -lame( --preset
medium --resample 44.1 )
```
- Change `sample.wav` to the name of your original file (e.g. `MyFile.wav`).
- Change `..\audio\sample.mp3` to the name you want to call your MP3 file (e.g. `..\audio\MyFile.mp3`)

¹ Note that the software is still of some use for displaying standard Toolbox files on the web, even if there is no audio file and the texts are not aligned.

² By combining Toolbox and ELAN it is possible to create a time-aligned interlinear text. See `sample.txt` for an example, and `ELANToolbox export exercise.pdf` for guidelines on how to do this.

- *Do not* change anything else in this file, including the spaces.
- Save and exit
- Double-click on `Resample.bat`. The process may take a few minutes depending on how long your recording is.
- Check that your MP3 file has been created in the audio directory.

N.B. If you are working with files stored on your computer (which you probably will be if you are testing) then you need to make sure that the Flash security settings are configured to trust the folder where you have installed `interlinear_viewer`. To do this carry out the following:

- Browse to http://www.macromedia.com/support/documentation/en/flashplayer/help/settings_manager04.html
- Click on 'Edit locations' and then 'Add location'
- Add the `interlinear_viewer` folder

This step is *not* required when accessing files over the Internet.

3 Exporting the Toolbox file to XML

For display in a browser, the Toolbox file needs to be converted to XML. The first time you do this you need to set up the XML export process – do this as follows:

- Open Toolbox and make sure the file you want to convert is the active window
- Select 'Export...' from the File menu
- Click 'Add...'
- Choose XML
- Choose a sensible name e.g. 'Browser view'
- If there are fields you don't want to display, uncheck 'All Fields...' and use 'Select Fields...' to choose the correct ones (note by default *all* fields are included)
- Click OK

Now you are ready to export your file:

- You should be back in the File->Export dialog entitled 'Create an exported copy of the file'
- Choose the export process you have just created
- Make sure 'Current record only' is selected
- Click OK
- You will be presented with the defaults from the process you have just created
- For the output file, choose the setup directory under `interlinear_viewer`. All you need to do is specify the location name of the Output File e.g. `MyText.xml`.
- Click OK

4 Massaging the Toolbox export

The XML file exported by Toolbox is not quite as we want it for display in a browser. A batch file

`transform_toolbox.bat` is provided to perform this step³:

- Open `transform_toolbox.bat` in a text editor (do **not** double-click it).
- You should see the following text:
`msxsl sample.xml transform_toolbox.xsl -o ..\texts\sample.xml`
- Change `sample.xml` to the name of your own XML file in both places
- Save and exit
- Double-click on `transform_toolbox.bat`
- Go to the `texts` folder and check the new XML file is there. Open it in a text editor and check the following:
 - Near the top of the file there should be the following line:
`<?xml-stylesheet href="interlinear.xsl" type="text/xsl"?>`
 - `<m_wrapper>` tags should have been placed around each morpheme⁴.
 - the `<ELANMediaURL>` tag at the end should have the MP3 extension

5 Start playing

The hard work is done now, all you need to do is browse to the XML file using an Internet browser⁵ and enjoy!

³This batch file only works on Windows computers. See Appendix 2 for alternative methods.

⁴See Appendix 1 for the reasons for this step.

⁵Right-click the file and choose 'Open with Internet Explorer' or 'Open with Firefox'.

Appendix 1 – XML structure of Toolbox export

The batch file in step 4 changed the structure of the Toolbox XML for multi-morphemic words. The XML exported by Toolbox is less than ideal in that morpheme tags are simply concatenated together e.g.

```
<txGroup>
  <tx>kàràkúmí</tx>
  <mb>kà-</mb>
  <ge>NC1-</ge>
  <ps>afx-</ps>
  <mb>ràkúmí</mb>
  <ge>camel</ge>
  <ps>n</ps>
</txGroup>
```

To allow the interlinear viewer to recognise individual morphemes, we need to put a wrapper around each morpheme:

```
<txGroup>
  <tx>kàràkúmí</tx>
  <m_wrapper>
    <mb>kà-</mb>
    <ge>NC1-</ge>
    <ps>afx-</ps>
  </m_wrapper>
  <m_wrapper>
    <mb>ràkúmí</mb>
    <ge>camel</ge>
    <ps>n</ps>
  </m_wrapper>
</txGroup>
```

Appendix 2 – manual changes for non-Windows platforms

Step 4 involved running the `transform_toolbox.bat` file which is only supported by Windows. This batch file uses the Windows tool `MSXSL.exe` to apply `transform_toolbox.xsl` to the output of the Toolbox export.

If you have access to an alternative XSL parser for your platform then simply use it to transform the Toolbox export with `transform_toolbox.xsl`.

This appendix shows how to make the changes without using XSL.

Adding/Editing the stylesheet processing instruction

- Open the XML file in a text editor
- Near the start, immediately after the line

```
<?xml version="1.0" encoding="UTF-8"?>
```

add

```
<?xml-stylesheet href="interlinear.xsl" type="text/xsl"?>
```

The start of your file should look like:

```
<?xml version="1.0" encoding="UTF-8"?>
```

```
<?xml-stylesheet href="interlinear.xsl" type="text/xsl"?>
```

```
<database>
```

```
<idGroup>
```

...

Changing the filename

- Open the XML file in a text editor. Find the tag

```
<ELANMediaURL>MyFile.wav</ELANMediaURL>
```

- Change as appropriate e.g.

```
<ELANMediaURL>MyFile.mp3</ELANMediaURL>
```

Don't worry about file path here – it is ignored.

Adding the morpheme wrapper

This change can be done using a Regular Expressions editor. The 'Search' expression (using Edit Pad Pro syntax) is:

```
<mb>(.*</mb>(\r\n|\n)<ge>(.*</ge>(\r\n|\n)<ps>(.*</ps>
```

The 'Replace' expression is:

```
<m_wrapper><mb>\1</mb><ge>\2</ge><ps>\3</ps></m_wrapper>
```