

ELAN → Toolbox → ELAN WORKFLOW

Simon Devylder & Lucie Laporte-Devylder May 1st 2020

This document walks you through the steps needed when you want to collect your data for several speakers in ELAN, then gloss it with Toolbox, and transfer the glossed data back into ELAN without losing any time-encoded information.

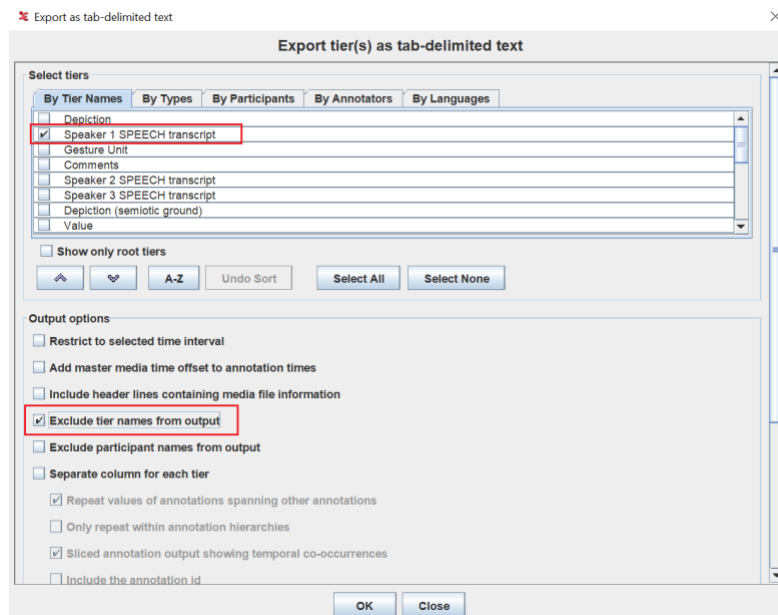
1) Open file in ELAN [e.g. [SAND06.eaf](#)]

→ click “File” > “Export as” > “Tab-delimited Text”

→ select all tiers where the transcripts are recorded, for every speaker: “SPEAKER 1 SPEECH transcript”, “SPEAKER 2 SPEECH transcript”, “SPEAKER 3 SPEECH transcript”, etc.

→ make sure the box “Exclude tier name from output” is **not** ticked

→ save under the name [SAND06_transcript.txt](#)



2) Open [SAND06_transcript.txt](#) in Notepad

→ click “Edit” > “Replace”

- in “Find what”, enter “Speaker 1 SPEECH transcript” (i.e the name of the tier for Speaker 1)
- in “Replace by”, enter “# ” (i.e the # symbol plus a space)
→ click “Replace all”
- repeat the operation, using a different symbol for every speaker
eg. “Speaker 1 SPEECH transcript” replaced by “# ”
“Speaker 2 SPEECH transcript” replaced by “+ ”
“Speaker 3 SPEECH transcript” replaced by “* ”
etc.
- delete the added symbols for every first line (Speaker 1, line 1 ; Speaker 2, line 1 ; Speaker 3, line 1 ; etc.)

Your document should now look like this:

```
SAND06.transcript - Notepad
File Edit Format View Help
Speaker 1 SPEECH transcript eh... mete selus enga
Speaker 1 SPEECH transcript ehisenga ke? tovit six enga
Speaker 1 SPEECH transcript koani ke kei ahite enga tin vatiteka
Speaker 1 SPEECH transcript abowen keilu ke
Speaker 1 SPEECH transcript vatikein ke vatin ke
Speaker 1 SPEECH transcript suval
Speaker 1 SPEECH transcript time uva alesii
Speaker 1 SPEECH transcript umea suk suval ofala kei mulamun

Speaker 2 SPEECH transcript kah kamitel ta koa kur question hema
Speaker 2 SPEECH transcript kosa hehi komaitel vare
Speaker 2 SPEECH transcript six tai tai ar vul te
Speaker 2 SPEECH transcript ehi case es

Speaker 3 SPEECH transcript five en koani
Speaker 3 SPEECH transcript keik kileh iko

Speaker X SPEECH (transcription) mete his tata Mathew mukone ke hel he
Speaker X SPEECH (transcription) keinengani
Speaker X SPEECH (transcription) kei haha kamitel enga ven

Speaker 4 SPEECH (transcription) five ke?hehe
Speaker 4 SPEECH (transcription) kovuli?
Speaker 4 SPEECH (transcription) Bersi ehis ke?
Speaker 4 SPEECH (transcription) start en start en tounah ta enga

Speaker 5 SPEECH (transcription) mulamu arsulul te enga ofala keil amu
Speaker 5 SPEECH (transcription) amutis vareh muko
Speaker 5 SPEECH (transcription) ka kihisi kosa ke sekera vare rovalie keil akilea
Speaker 5 SPEECH (transcription) kei vareh
Speaker 5 SPEECH (transcription) maienga suk houlu umatmat tai heite e
```

BEFORE

```
SAND06.transcript - Notepad
File Edit Format View Help
# eh... mete selus enga
# ehisenga ke? tovit six enga
# koani ke kei ahite enga tin vatiteka
# abowen keilu ke
# vatikein ke vatin ke
# suval
# time uva alesii
# umea suk suval ofala kei mulamun armutis tenga amutisi es tounah keil muk

+ kah kamitel ta koa kur question hema
+ kosa hehi komaitel vare
+ six tai tai ar vul te
+ ehi case es

+ five en koani
+ keik kileh iko

= mete his tata Mathew mukone ke hel helesi
= keinengani
= kei haha kamitel enga ven

five ke?hehe
kovuli?
Bersi ehis ke?
start en start en tounah ta enga

mulamu arsulul te enga ofala keil amumon vare koani
amutis vareh muko
ka kihisi kosa ke sekera vare rovalia vaitnale vensa vare ke keil akilea
kei vareh
maienga suk houlu umatmat tai heite elu enga sesuk
```

AFTER

3) Open Toolbox

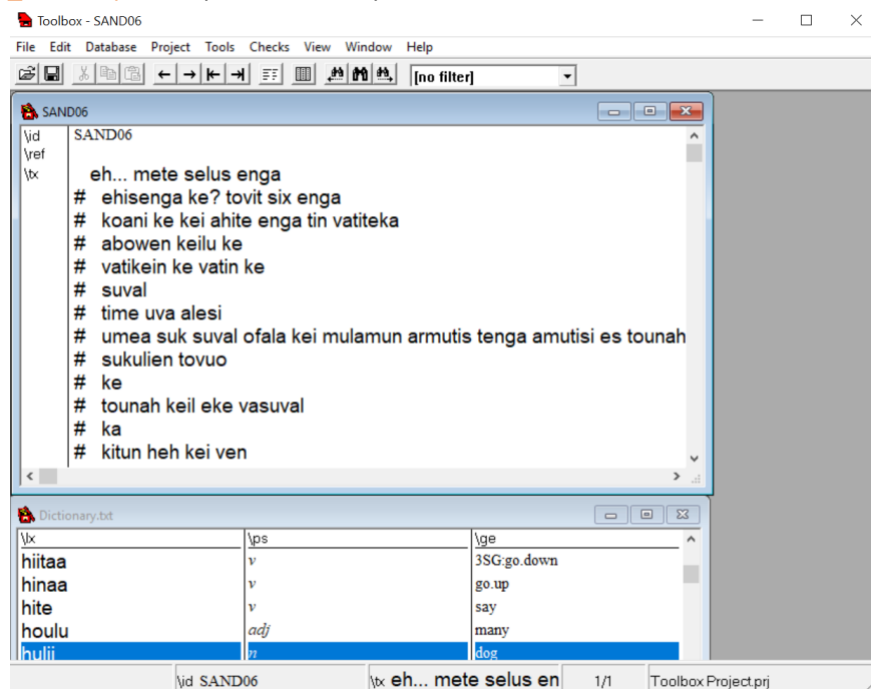
→ click “File” > “New”

Create a new file and call it **SAND06**

Select “Database Type” : “Text”

→ next to \id marker, add “SAND06” (without the quotation marks)

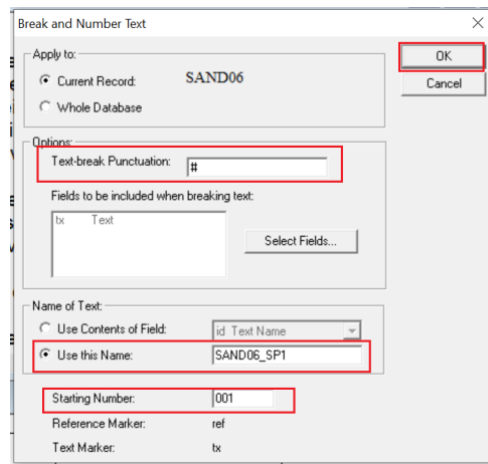
→ copy and paste all the lines from Speaker 1 (the lines that now have a # symbol at their start) from the document **SAND06_transcript.txt** opened in Notepad in front of the \tx marker



→ click “Tools” > “Break/Number text”

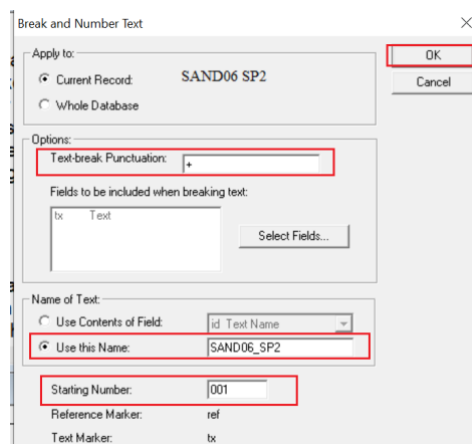
- in “Text-break punctuation”, enter “#”
- click “use this name” and enter “SAND06_SP1”

- in “starting number” check that “001” is entered
→ click OK



- click “Database” > “Insert Record” > name it SAND06 SP2
- copy and paste all the lines from Speaker 2 (the lines that now have a + symbol at their start) from the document *SAND06_transcript.txt* opened in Notepad in front of the \tx marker
- click “Tools” > “Break/Number text”

- in “Text-break punctuation”, enter “+”
- click “use this name” and enter “SAND06_SP2”
- in “starting number” check that “001” is entered
→ click OK



- repeat as many times as necessary, to eventually have 1 sheet per speaker (eg. If your file has 4 speakers you should have 4 sheet: SAND06 ; SAND06 SP2 ; SAND06 SP3 ; SAND06 SP4)

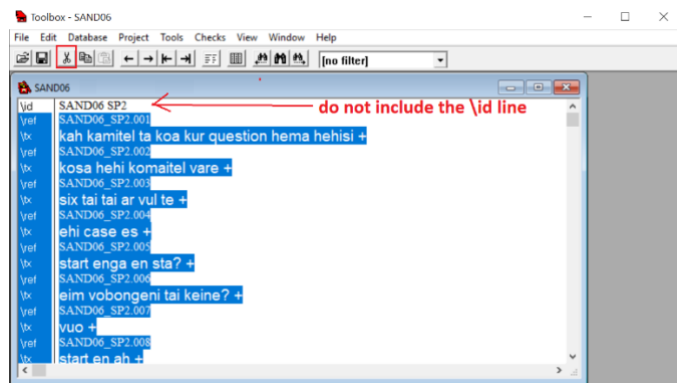
Nb. Speaker 1 should have \ref markers that should look like this: SAND06_SP1.001

Nb. Speaker 2 should have \ref markers that should look like this: SAND06_SP2.001

Nb. Speaker 3 should have \ref markers that should look like this: SAND06_SP3.001

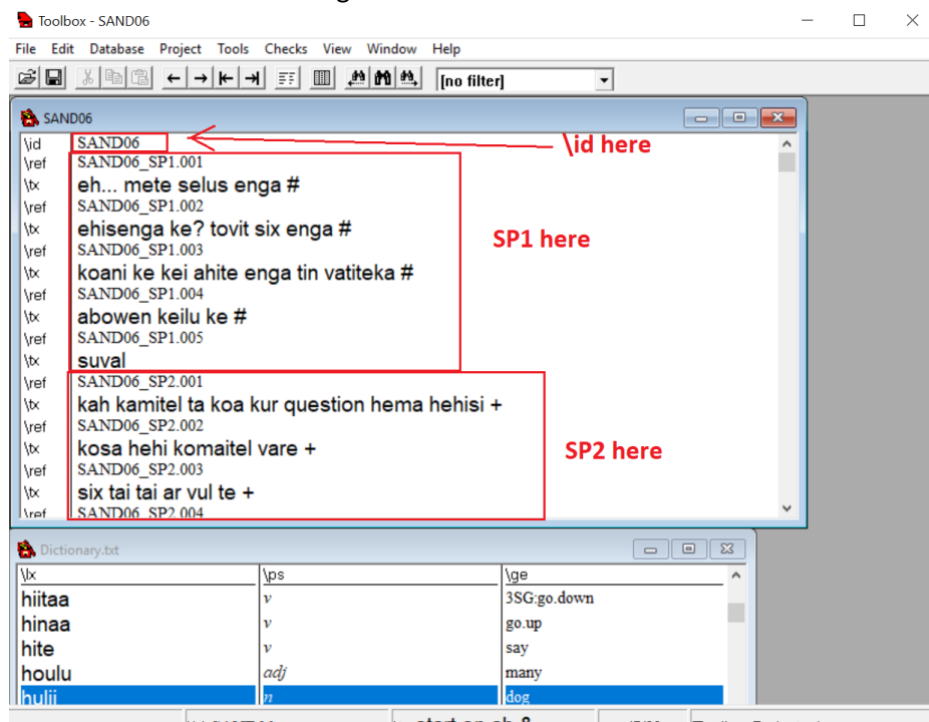
etc.

- cut and paste the text with their \ref form every speaker into the first sheet where speaker 1 is (i.e the sheet with the \id “SAND06”)



→ keep only the sheet with the \id “SAND06”, delete every other sheet (i.e any sheet with an \id that include a speaker name “SAND06 SP2”, “SAND06 SP3”, etc.). To do so, click on “Database” > “Delete record”, once you are on the sheet you want to delete

Your Toolbox file should now look something like:



→ click “File” > “Save” (i.e leave Toolbox open with the SAND06 file you just created)

4) Open again file in ELAN [[SAND06.eaf](#)]

→ create a copy of this file in your working folder

→ click “Save as” > name it “SAND06_toolexport”

→ in ELAN, under the Annotation mode, right click on “SPEAKER 1 SPEECH transcript” > select “change attribute of”

- in “Tier name”, replace “SPEAKER 1 SPEECH transcript” by “ref@SP1”
 - in “Participant”, add “SP1”
- click “Change”

Tier Name	Parent Tier	Tier Type	Participant	Annotator	Input Method	Content La...
Depiction	-	Utterance			English	-
Speaker 1 ...	-	Utterance			English	-
Gesture Unit	-	Utterance			English	-
Comments	-	Utterance			English	-
Speaker 2 ...	-	Utterance			English	-
Speaker 3 ...	-	Utterance			English	-

Add Change Delete Import
 Select Tier: Speaker 1 SPEECH transcript
 Tier Name: ref@SP1
 Participant: SP1
 Annotator:
 Parent Tier: none
 Tier Type: Utterance
 Input Method: English (United Kingdom) (System default)
 Content Language: None - -
 More Options...
 Change Close

> repeat for each speaker

Speaker 2 → Tier name: ref@SP2

→ Participant: SP2

Speaker 3 → Tier name: ref@SP3

→ Participant: SP3

etc.

→ click "Tier" > "Label and Number annotation"

- select the tier now named "ref"
 - click "include label part" and enter "SAND06"
 - click "insert other delimiter" and enter "."
 - Click "min number of integer digit" and enter "3"
 - Check that the label will look like "SAND06_SP1.001" in the blue frame at the bottom of the box
- click "OK" > "Close"

Generate label and number for annotations

Select tier(s)

Single tier
 Multiple tiers

Depiction
 ref@SP1
 Gesture Unit
 Comments
 ref@SP2
 ref@SP3

Options

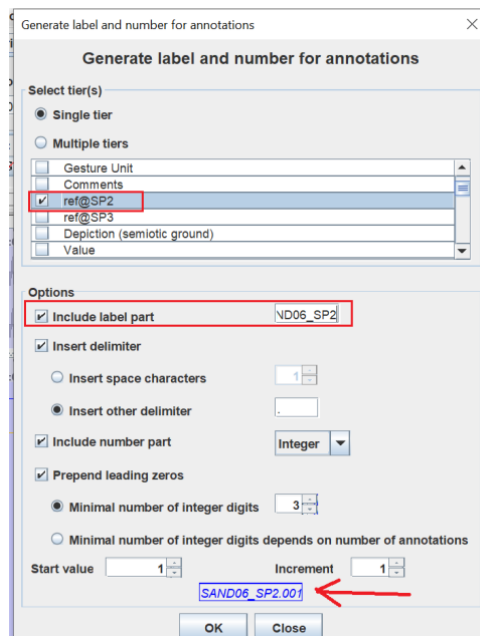
Include label part: SAND06_SP1
 Insert delimiter
 Insert space characters: 1
 Insert other delimiter: .
 Include number part: Integer
 Prepend leading zeros
 Minimal number of integer digits: 3
 Minimal number of integer digits depends on number of annotations

Start value: 1 Increment: 1

SAND06_SP1.001

OK Close

> repeat for every speaker, simply change the number in the case "include label part" : SAND06_SP2



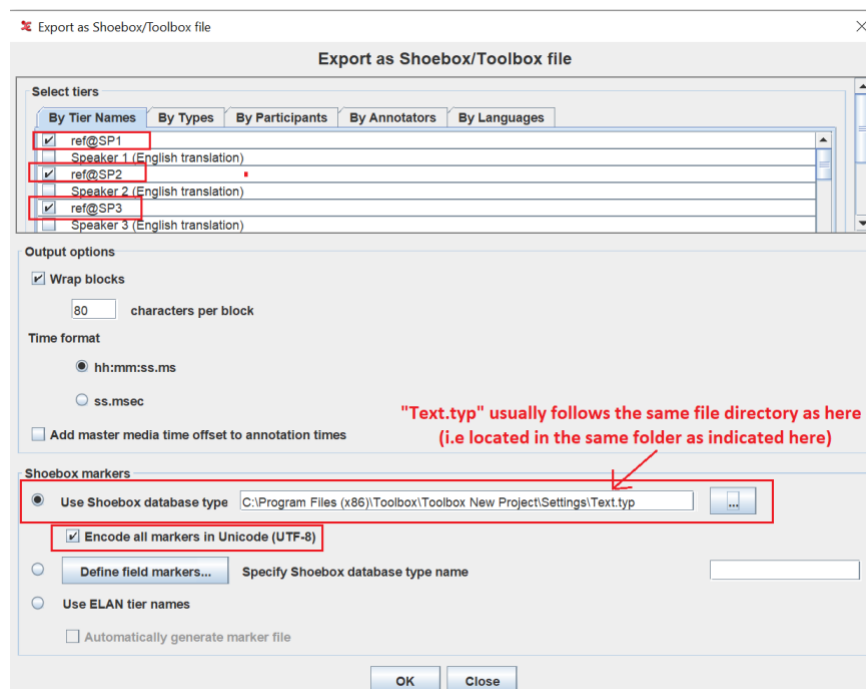
→ click “File” > “Export” > “as Shoebox file”

- in “Select tier” select all speakers: “ref@SP1”, “ref@SP2”, “ref@SP3”, “ref@SP4”, etc.
- in “use Shoebox database type” select the database type “Text.typ”

[n.b. If you are using Windows, this file “Text.typ” is usually located in the folder where Toolbox is installed, in the folder called “Toolbox New Project” > “Setting” > “Text.typ”]

- make sure the box “Encode all markers in Unicode (UTF-8)” is ticked

→ click “OK”



→ save under the name *SAND06_elanexport.txt* (!!) Make sure you select the file format “.txt”

5) Open *SAND06_elanexport.txt* in Notepad

→ add manually “\id SAND06” above the line “\block 001”

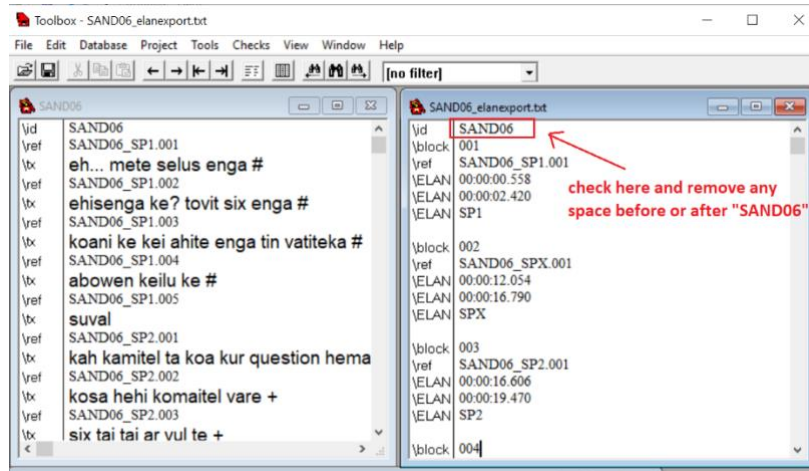
(!! Make sure there is no space after “\id SAND06” (* “\id SAND06 ”))

6) Open *SAND06_elanexport.txt* in Toolbox

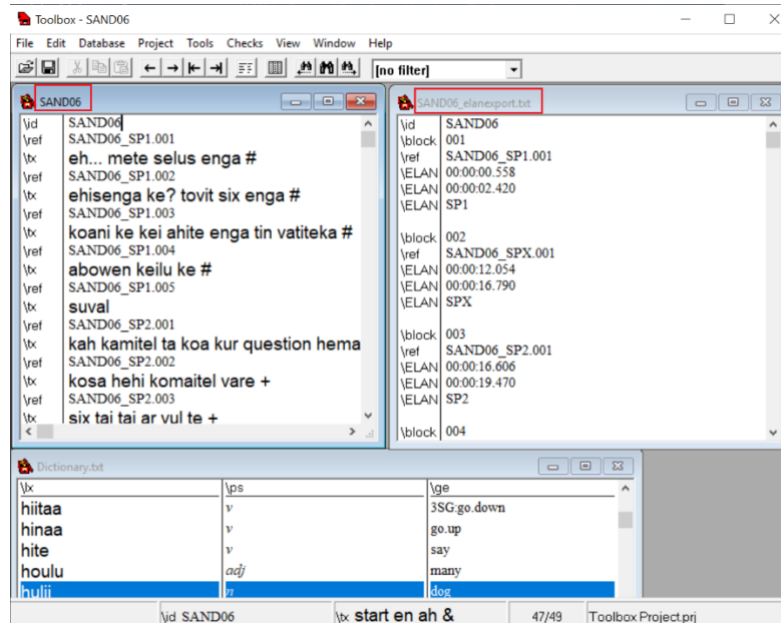
→ click “File” > “Open” > *SAND06_elanexport.txt*

(⚠ Make sure here again that Toolbox did not add any space after “\id SAND06” (~~* “\id SAND06 ”~~)

(⚠ Make sure also that there isn’t any space in the \ref marker (e.g. “\ref SAND06.001” (~~not* “\ref SAND06.001”~~))



You should now have in Toolbox :

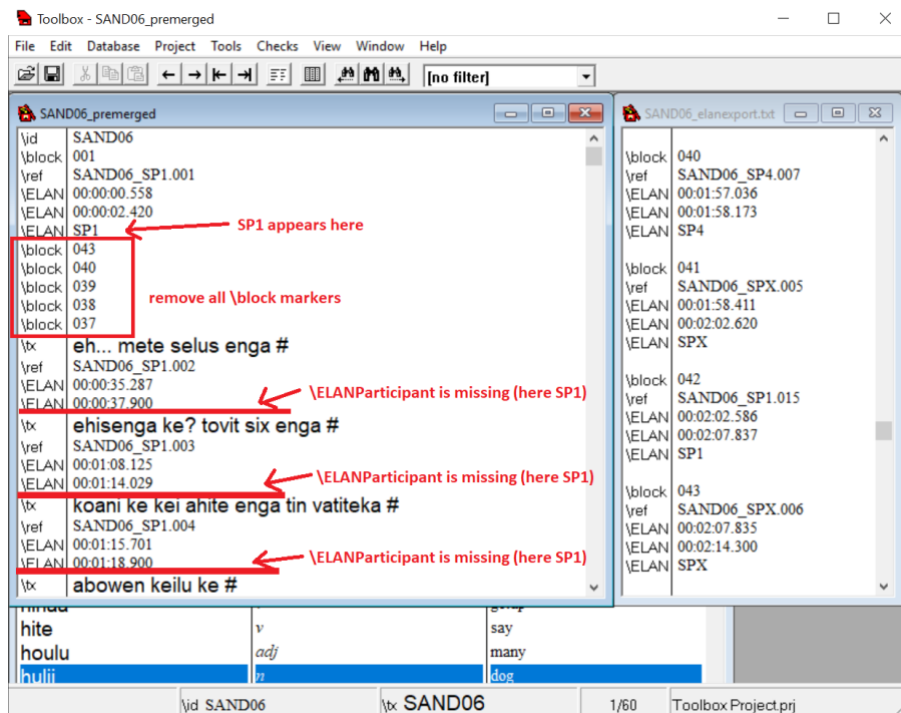


→ click on the window of “SAND06” that you created at step 3 (on the left in the picture above)

→ create a copy of it, click “File” > “Save as” > save under the name *SAND06_premerged.txt*

→ click “Database” > “Merge database” > select “SAND06_elanexport” > “OK”

You should now have in Toolbox :

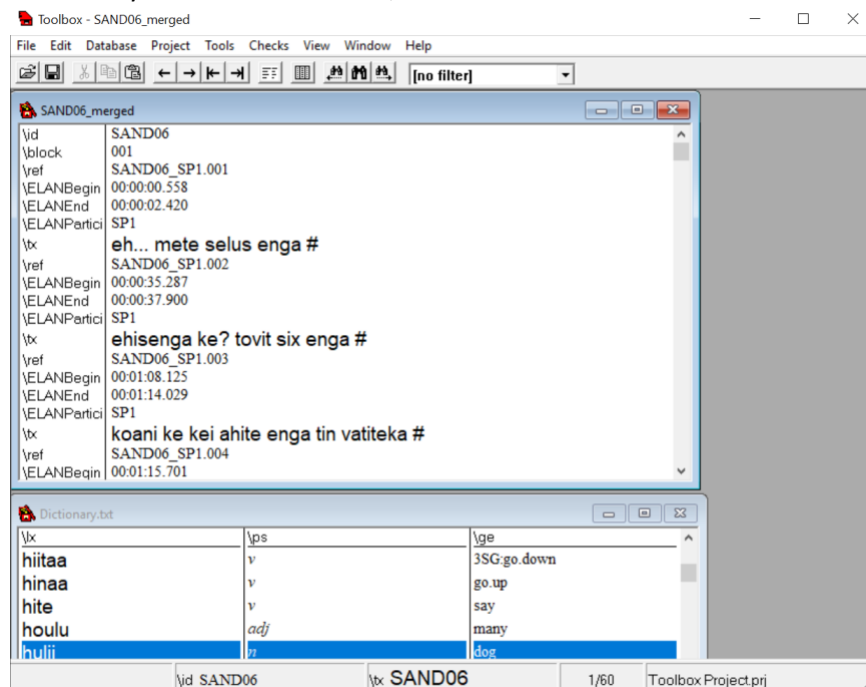


→ close the window of the file “SAND06_elanexport.txt” (on the right in the picture above)

→ in the window of the file “SAND06_premerged”

- delete all “\block” markers (such as those framed in red in the picture above)
- add manually the marker \ELANParticipant with the relevant participant (“SP1”, “SP2”, “SP3”, etc.) everywhere it is missing
- for clarity, move all \ELANMedia markers at the bottom of the file

Your Toolbox file is now ready to be interlinearized, and should look like this:



→ click “File” > “Save as” > save under the name *SAND06_merged.txt*

7) Gloss your transcript in Toolbox

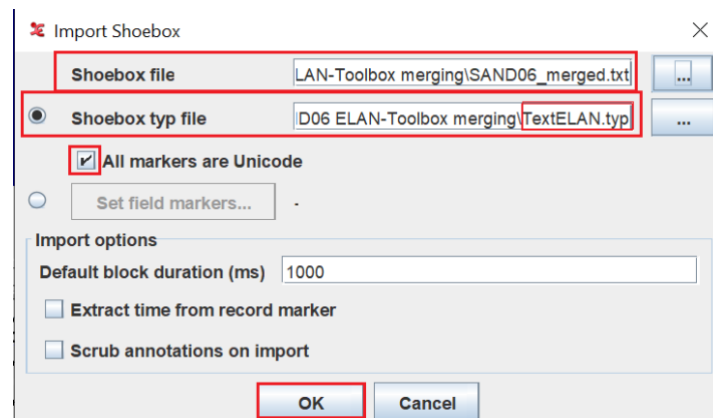
- add the “free translation” (marker “\ft”) after each transcription line
- make sure to save regularly

8) Copy and paste into your working folder the file “Text.typ” from the folder Toolbox > Toolbox New Project > Settings

- open “Text.typ” in Notepad
 - replace the line “\mKrRecord id” by “\mKrRecord ref”
- save under the name “TextELAN”

9) Open ELAN

- click “Import” > “Shoebbox file”
 - in “Shoebbox file” select the file SAND06_merged.txt
 - in “Shoebbox database type” select the modified file from step 7: “TextELAN.typ”
 - make sure “All marker are Unicode” is ticked
 - OK



Example of ELAN file after you have imported the glossed transcript from Toolbox:

File Edit Annotation Tier Type Search View Options Window Help

Grid Text Subtitles Lexicon Comments Recognizers Metadata Controls

ref@unknown

Nr	Annotation	Begin Time	End Time	Duration
1	SAND06_SP1.008	00:01:40.652	00:01:45.610	00:00:04.958

00:01:46.026 Selection: 00:01:40.652 - 00:01:45.610 4958

SAND06_2019062_

00:01:16.000 00:01:17.000 00:01:18.000 00:01:19.000 00:01:20.000 00:01:21.000 00:01:22.000 00:01:23.000 00:01:24.000 00:01:25.000 00:01:26.000 00

tx@SP1 [57] avoue -n keilu eke vatik -n eke vatu -n eke

mb@SP1 [88] wing -3SG 3DL here tail -3S here hea -3S here

ge@SP1 [88] n -poss.aff, pron n n -pos, n n -pos, n

ps@SP1 [88] the two wings are here its tail is here, its head is here

ft@SP1 [15]

ref@SP2 [16] SAND06 SP SAND06 SP2.011

tx@SP2 [69] vatiken vasuval nahisukes_vamuko tounah eke k

mb@SP2 [95] vatike -n va- suva, n, hi, -s, -s, va- muk, tounah eke k

ge@SP2 [95] tail -3SG IMM- rese, 1, a, -s, IMM- do.li, thing here

GOOD LUCK ! 😊