

Preparing the files

Before starting ELAN, create separate .wav and .eaf files for FW, EOQ, Long questionnaire (= IV) and any other conversation. You can easily cut up or paste together files using Audacity (<http://www.audacityteam.org/download/>). Either select the relevant portion and choose “Export selection as .wav” from the File menu *or* copy and paste relevant sections, one following another, into a new Audacity file and then be sure to “Export as .wav” as the final step for any new file you create. In the interview_catalog.xlsx, record the starting time-point in the original long .wav for FW and EOQ, so we can find them again if necessary.

If the digital recording of a particular interview is divided up on lots of different .wav files you should **paste them together before linking to ELAN**. Use Audacity.

Note that we won't transcribe the responses to the ethnic orientation questionnaire for now, but if there's a good story or other information elicited during it, then create a .wav and .eaf file with “_EOQ” as part of the name.

Match .wav and .eaf filenames.

Using ELAN

- 1) Start [ELAN](#).
- 2) Choose “New” from the “File” menu. Select the .wav you are going to transcribe and “Add New Media.”
- 3) Immediately save your file with the appropriate filename for the speaker you are starting to annotate. I'll call it **I2F35A_IV.eaf**. Save it in the folder with your name on it.
- 4) Choose “AUTO BACKUP > 1 minute” from the File menu.
- 5) **Edit>Preferences>Edit Preferences > Editing > $\sqrt{\quad}$ Enter key commits change in the inline box.** (This lets you save an annotation just by hitting Enter before you leave the annotation box.)
- 6) You should now see a soundwave in the center of the **I2F35A_IV.eaf** window. If you don't, Command-Click (\mathbb{A} +Mouse click) on the place where the soundwave should be (it will be showing as a flat horizontal line) and then choose a number to magnify/Vertical Zoom in by. Keep trying until you see the soundwave enough to be able to distinguish speech from silence. [If it is still too quiet to hear quickly, use the “Amplify” function to edit the (whole) .wav in Audacity.]
- 7) Test volume. (Apple menu > System Preferences > Sound > Output or the \blacktriangleleft icon on the upper right of your monitor.)
- 8) Segment the file into clauses, on the Default tier, using the Segmentation mode in the Options menu.

- 9) Switch back to Annotation mode in the Options menu.
- 10) Create a new Type to associate the tiers for each speaker to the segmented Default tier. (Type menu > Add New Ling. Type. Name it “samesize” and choose “Symbolic Association” for the type.)
- 11) Create separate tiers in ELAN for each person who speaks in the .wav. You can add more tiers at any point. Make each one the daughter of the Default tier, using “samesize” as the Type to associate them.
- 12) Name the tiers as follows:

<u>Tier name</u>	<u>Actual speaker ID</u>
<i>Interviewer</i>	NN (Interviewer initials)
<i>Main Speaker</i>	I2F35A (Speaker code)
<i>Other Speaker2, etc.</i>	I2F64B (Speaker code)
<i>Notes</i>	Any questions/comments you want to share.

- 13) Switch to Transcription mode (in Options menu) to transcribe the text. Transcribe one speaker at a time.
- 14) In the Interviewer tier, simply type “Q1” as the annotation when they ask the first question of the long questionnaire, etc. Unless they diverge significantly from the script, you don’t need to transcribe the questions. BUT TRANSCRIBE ANY OTHER SPEECH BY THE INTERVIEWER, OTHER THAN A SCRIPTED QUESTION. (You can download an English version of the questionnaire from the website or the HLVC/Documents folder on iMac1. Other language versions should be available in the relevant server/folder.)
- 15) If you have any questions or are not sure about something, type a note in the Notes tier, at the relevant timepoint. If you are not sure how to type something, put “XX”. (Don’t use “?” because that is hard to search for!)
- 16) Clicking on an annotation in “Grid” or “Text” view (top half of window) will take you to that part of the wave. We (the analysts) will then be able to search for answers to each question in the questionnaire by just typing the question number. You (the transcribers) will have “bookmarks” to help see where you are in the transcription.

General rules for transcribing for this project.

Break up the recording into Annotation segments. Normally make 1 Annotation per sentence or turn, but that can be modified later, so don’t worry about it a lot.

You may find the ELAN Shortcut Keys helpful.

Transcribe word for word what is said in the orthography specified for your LANGUAGE.

LANGUAGE	Orthographic convention
Cantonese	Matthews & Yip Romanization ¹ (see Cantonese Transcribing Tips.doc)
Faetar	IPA
Hungarian	standard spelling for Hungarian
Italian	standard spelling for Italian / Calabrese
Korean	Hangul
Portuguese	standard spelling for Portuguese (Use 1990 'Novo Acordo'): http://www.portaldalinguaportuguesa.org/novoacordo.php?action=novoacordo&act=list&version=al
Russian	Comrie & Corbett's Linguistic Transliteration ²
Tagalog	standard spelling for Tagalog
Ukrainian	Comrie & Corbett's Linguistic Transliteration ²

Use standard spellings, not “eye-spelling.”

Do not change the word order or morpheme order, of what was said.

Make your best guess when you come to something unclear or hard to understand. A proofreader will check your work. (If you want, make a note in the *Notes* tier.)

Use periods, question marks, and exclamation points to mark the end of a sentence. Don't use a dash (or anything) to indicate an abrupt stop. Use a dash to indicate incomplete words.

Hyphenate multi-word proper nouns like “UNIVERSITY-OF-TORONTO.”

Do not type in any last names of personal acquaintances or relatives that might get mentioned during the interview. Just put an initial, e.g. “Maria G.” For any other proper nouns that might identify a participant, type “[REDACTED]”.

Use ALL CAPITAL LETTERS to transcribe ENGLISH WORDS.

¹ Matthews, Stephen and Virginia Yip. 1994. *Cantonese: a comprehensive grammar*. London; New York: Routledge. pp. 13-22. (In Robarts PL1733 .M38 1994) Also (partially) available through books.google.ca.

² Comrie, Bernard & Greville Corbett. 2002. *The Slavonic Languages*. London & New York: Routledge. 827, 832-833. **Exceptions to this system** that we will follow: the symbols with hachek (use zh, sh, ch, and shh instead), the 'hard sign' (#) and 'open e' (je). With this system, our annotations can be automatically transliterated to Cyrillic using the engine on <http://www.translit.ru/>. You can consult www.gramota.ru, sponsored by the Russian government. It has official rules for standard Russian spelling and punctuation, and a spellchecker.

Spell out any numbers the speaker says (like years).

Don't worry about transcribing pauses, etc. They are easy to find in ELAN.

Don't transcribe stutters or false starts. (You can exclude them from the annotation field.)

Use a consistent spelling for fillers like *aha*, *mmm*, etc.

Special notes for Russian:

Particles *-to*, *-libo*, *-nibud'*, *-ka*, all require a dash. Ex: *Chto-libo*, *kto-nibud'*, *chto-to*, *davaj-ka*, etc.

Particles *koe-*, *koj-* require a dash. Ex: *koe-kto*.

Po- requires a dash only if the following word ends in *-ki*, *-i*, *'omu*, *'emu*. Ex: *po-russki*, *po-nashemu*, *po-sobach'i* (swimming)

-taki only requires a dash in *vsjo-taki*. No dash in other cases.

Capitalize *Vy* for the polite form of 2nd person singular.

Wear headphones to transcribe. You'll hear better and get distracted less (and distract others less). There are some in the lab.

Take a short break every 10 minutes or so, so you don't burn out.

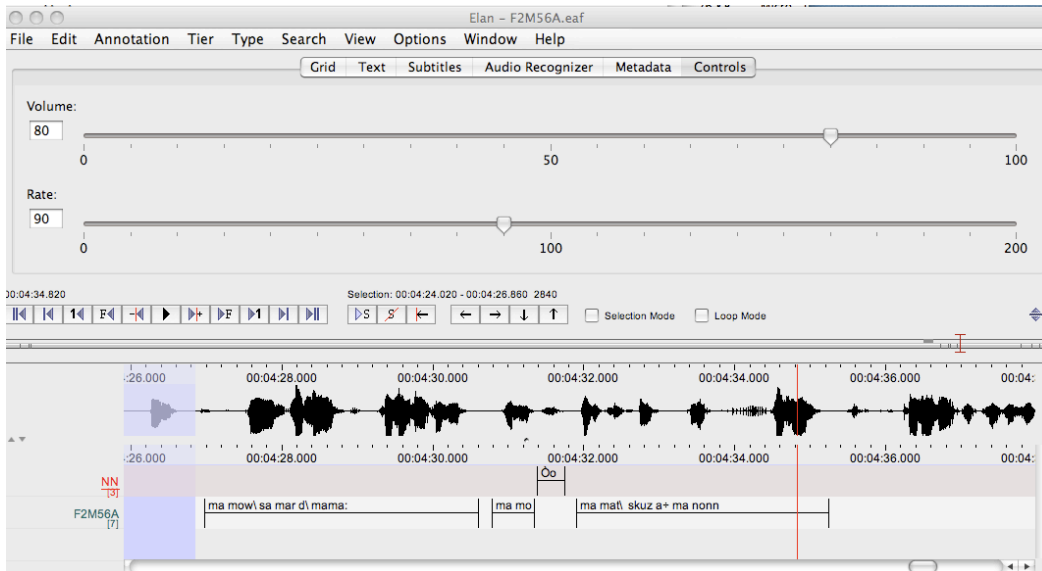
On the computer where you are working, make a folder with your name, inside your LANGUAGE folder and save all your files.

Back-up everything to the appropriate server and folder at the end of the workday. Put it in the appropriate LANGUAGE folder inside HLVC. These get backed up hourly.

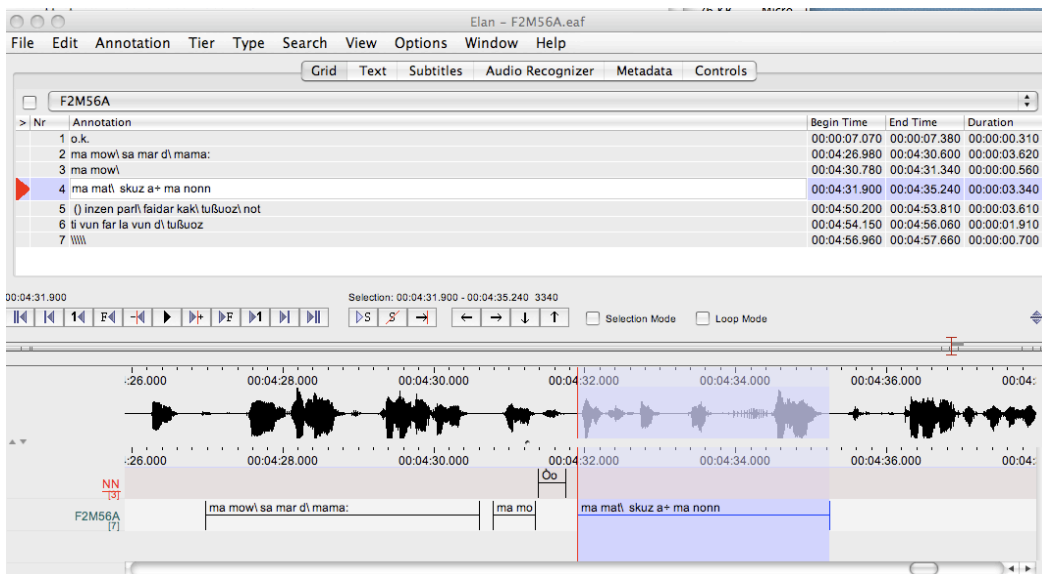
When you finish a file, move it to the "Finished Transcriptions" folder. Update the IV catalog with the completion date, your initials, and how long it took.

What ELAN can look like

Top window with "Controls" selected (adjust volume and playback speed here):



Top window with "Grid" selected (see a list of annotations in the selected tier):



Top window with “Text” selected (see transcriptions in paragraph format, for selected tier):

