

Exercise 4*

Creating a two-language transcript with interlinearized glossed text (IGT)

Goals:

In this exercise, we will learn to make an ELAN file for IGT. This is the most complicated structure we've seen so far, but it's only a small conceptual leap away from the two-language basic annotation we created yesterday. We will use a recording of the Kannada language with the following structure for our IGT (structures may vary for your own project):

Tier 1: Kannada sentences (normative sentences)

Tier 2: Kannada intonation units (prosodic phrases)

Tier 3: Kannada words by morphemes (words with hyphens to show morpheme boundaries)

Tier 4: English morpheme glosses

Tier 5: English free translation of the entire Kannada sentence.

Today's files are found in the folder called ``kan-pear-story'`.

Step one: Create a new ELAN file.

1. Follow the directions from the last exercise for creating a new ELAN file, but use today's audio file: ``kan-pear-story.wav'`.

Before beginning to define linguistic types and tiers, let's take a moment to think about what types we'll need for the tiers listed above.

Tier 1: This will be a top-level parent tier, linked directly to the audio stream. It will contain a transcription of the actual utterances found in the recording, and they will be the size of a sentence.

Tier 2: This tier will also contain transcriptions of the actual utterances in the recording, but they will be in smaller units than the first tier (i.e., subdivisions of the first tier). This tier is a child of Tier 1, because we want to show that the intonation units are subparts of the sentences. However, it will also be a parent tier to the tiers below it (because we will want to show that the items in the next tier, words and morphemes, are subparts of the intonation unit). This tier is also linked to the timeline in the sense that you can click on a single intonation unit and hear that part of the audio stream.

*This worksheet is closely follows one made by Andrea Berez & Christopher Cox. They have kindly given me permission to recreate it with my own (minor) changes.

Tier 3: This tier is not linked to the timeline (you couldn't click and hear just a single word or morpheme)¹; rather, it's a child of Tier 2. Its contents are subdivisions of Tier 2.

Tier 4: This tier is a child of Tier 3, and contains a one-to-one translation of the items in its parent (a morpheme-by-morpheme gloss).

Tier 5: This tier is a translation of, and thus a child of, Tier 1.

Step two: Define linguistic types.

For this file we need four linguistic types. You are already familiar with the first two: a type called 'text' for the top-level parent tier, the one containing the Kannada sentences; and a type called 'translation' for the two tiers that contain translations (both morpheme glosses and the sentence-level free translation). Notice that we can use the same type for two different tiers.

1. Follow the instructions in the previous exercises to create these two types. You will also need two more types, and they are conceptually similar—they are both for tiers that are subdivisions of their parents. However, ELAN handles them a little differently because one of these types is linked to the audio stream (so you can click- and-play), but the other is not.
2. Go to 'Type' > 'Add New Linguistic Type'
3. The first of these two types we will make will be called 'timed chunk' (go ahead and put that in the 'Type name' box). In the 'Stereotype' menu, select 'Time Subdivision'. This means that the type will be used for tiers that are a subdivision of some other tier, but are still linked to the timeline. Notice that the 'Time-alignable' check box is checked. This is correct.
4. Click 'Add' to add this type.
5. The second of these two types we will make is for tiers that are subdivisions of some other tier but are not linked to the timeline. Call this type 'chunk'.² In the 'Stereotype' menu, select 'Symbolic subdivision'. This is because we are not making timeline subdivisions, we are just creating a space to separate words and morpheme boundaries. Notice that the 'Time-alignable' checkbox is now unchecked. This is correct. Click 'Add' and then 'Close'.

¹It's not that you could never do this with words or morphemes, it's just that here we're going to set it up so that you can't just hear a single morpheme by clicking on the annotation. You could certainly set it up that way if you wanted to! and a type called 'translation' for the two tiers that contain translations (both morpheme glosses and the sentence-level free translation). Notice that we can use the same type for two different tiers.

²Remember, these names can be anything you want, but names like 'chunk' and 'timed chunk' are conceptually appropriate for what they'll be used for.

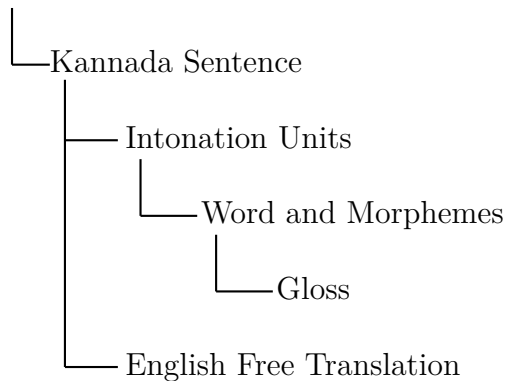
Step three: Define tiers.

For this file we need five tiers.

1. As before, we can use ELAN's default tier as our top-level parent tier, but we should give it a better name, like 'Kannada Sentence'. Please do so.
2. Next, let's work our way down the list of tiers and set each up in turn (you could do this in any order, and the way you specify the parent/child relationships will turn out correct, but it's easier to just go from top to bottom).
3. We need a tier for the intonation units. Add a new tier called 'Intonation Units' (or 'prosodic phrases' or something equally descriptive), fill in the `'Participant'` and `'Annotator'`, and select this tier's parent, 'Kannada Sentence'. Note that ELAN suggests a Linguistic Type in the drop down menu—it doesn't always get it right so be sure to double check. For this tier, we want to set the `'Linguistic Type'` to 'timed chunk.' Click `'Add'` to create the tier.
4. Now we need a tier for the words and morphemes. Create a new tier called 'Words and Morphemes'³, and set its parent tier as 'Intonation Units.' Its proper Linguistic Type is 'chunk' (not 'timed chunk'). Click `'Add'`.
5. Next we need a tier for the glosses of words and morphemes. Make a new tier called 'Gloss', whose parent is 'Words and Morphemes.' Remember that this is one of our two translation tiers, so set the Linguistic Type accordingly. Click `'Add'`.
1. Our final tier can be called something like 'English Free Translation.' This tier will contain a translation of the entire Kannada sentence (not just of words or of intonation units), so its parent is 'Kannada Sentence.' Its linguistic type is, of course, 'translation'. Once you have added all your tiers, it's a good idea to check to be sure that your parent/child relationships are correct. Close the `'Add Tier'` dialog box, and then
2. Go to `'View' > 'Tier Dependencies'`

³Do not use an ampersand (&)! ELAN sometimes has trouble with these as they can be part of the underlying programming language.

You should see a tree structure something like this:



This diagram shows the following information:

- ‘Kannada Sentence’ is linked directly to the audio stream timeline
- ‘Kannada Sentence’ has two immediate children: ‘Intonation Units’ and ‘English Free Translation’
- ‘Intonation Units’ has one immediate child: ‘Words and Morphemes’
- ‘Words and Morphemes’ has one immediate child: ‘Gloss’
- ‘Gloss’ and ‘English Free Translation’ have no children

If you don’t see this, you’ve made a mistake somewhere. Go back and fix your tier relationships.

Step four: add annotations.

Now we are going to add our annotations, working with the transcription in today’s file, `kan-pear-story.txt`.

1. Start with the first Kannada sentence, and add the annotation by listening, selecting, and double clicking just like in the previous exercises.
2. Now we want to divide this sentence into time-based intonation units. First double click in the ‘Intonation Unit’ tier directly below the annotation for the first full sentence. A text box opens up that is just as long as its parent. Go ahead and type the words from just the first intonation unit into that box, and then type **Control-Enter** to make it stick.
3. Now we want to keep adding intonation units—there are five for this first sentence. So here’s what you will do for each new intonation unit:
 - (a) click on the horizontal line containing your first intonation unit so that the line itself turns blue

- (b) right-click on that blue line, and select '**New Annotation After**'. A new text box will open up to the right of the first one.
 - (c) Add the text for the second intonation unit and type **Control-Enter**.
 - (d) Now you want to move the vertical separation line in between the units so that it lines up with the correct point in the wave form where the first unit ends and the second unit begins. Click either of the two intonation units so that the horizontal line turns blue. Now, press the **ALT** key, and while pressing it, use the mouse to pick up the vertical separation bar and drag-and-drop it where you want the new division to be.
1. Repeat this process until you have added all the intonation units for your first Kannada sentence. Notice that by selecting a single intonation unit, you can press Play and hear just that unit.

Next we want to add annotations to the 'Words and Morphemes' tier (for the sake of this exercise, put each word with its morphemes into a single annotation; for example, the entire word haNN-iNA will all go into one annotation box).

2. To add your first word, double click on the 'Words and Morphemes' tier directly below the first intonation unit. A text box that is the same length as the intonation unit above will open up.
3. Type in your first word, and then hit Control-Enter.
4. To add additional words, you will do nearly the same thing you did when you added intonation units (selecting the word so that the horizontal line is blue, clicking '**New Annotation After**', etc.). The difference is that you won't be able to pick up the vertical separation line and move it. Instead, because we used the '**Symbolic Subdivision**' type for this tier, ELAN divides the area into equally-sized units.
5. Continue until you have added all the words for each intonation unit.
6. Next you'll add annotations to the 'Gloss' tier by double clicking directly below each word. ELAN won't allow you to '**Add Annotation After**' in this tier because of the properties of this tier's type. Continue to add glosses to all your words and morphemes.
7. Finally, you can add annotations to the 'English Free Translation' tier. Notice that when you double click, the text box is as long as its parent ('Kannada Sentences'), and not as long as the tier that is physically above it ('Gloss').
8. Now go through the entire file and add all the transcriptions. You'll get faster as you practice!

If you finish early, read in the ELAN manual about the following things:

- there is one linguistic stereotype we haven't used—it's called 'Included In'. What is the difference between this and 'Time Subdivision'?
- read about tokenizing tiers. How can this help you in annotating IGT?

Questions to think about:

You could go even further with your IGT. For example, you could have a tier that puts each underlying morpheme in a separate annotation (good for fusional languages). How would you do this?

What would you do if you had more than one speaker in your recording?

If you want to make a tier that contains metacommentary (notes to yourself, for example), how would you configure it? What kind of linguistic type would you use?