

NB: with these questions, feel free to tell me to figure out a solution. I'm asking these questions in case there's a way you want it done

Do we want to preserve these annotations? If so, we need to settle on a filler for FLE_x so that we can have the [Spanish] annotation (probably would need another field for that? Idk, each -sp is its own tier in the ELAN file so if we want to have -sp tiers at all we'll need to make a call on what's appropriate in FLE_x

Category	Start	End
KGW-tns (99)	0	100
KGW-notes (88)	0	100
Child1-sp (25)	20	30
Child2-sp (15)	35	45
Child3-sp (2)	50	60
Other-tca (13)	0	100
Other-tns (13)	0	100
Other-notes (11)	0	100

Below we see Ticuna and Spanish in the same annotation but on the spanish tier for the child—do we want to make sure this comes up in FLE_x so it can be part of the parsing work? If we keep this tier on the downlow and don't treat it like it's there while we're annotating, I guess we'll miss some things like this! Seems like something we wouldn't want to miss, so what is the preferred strategy for making sure it's included, either in FLE_x or just when we bring it back out of FLE_x (the latter would entail that we don't use FLE_x to gloss things in this tier)

TAA-tca [83]	ngel7ta5 wa317i5 p	
KGW-tca [100]		
TAA-tns [83]	l'onde van a calentar	En
TAA-notes [68]	et this - congrats. en	yo
KGW-tns [99]		
KGW-notes [88]		
Child1-sp [23]		
Child2-sp [19]	na31pal y tu cabeza	
Child3-sp [2]		
Other-tca [13]		
Other-tns [13]		
Other-notes [11]		
Child4-tca [26]		
Child4-tns [26]		

Title		Tca		tca_20170527_disc_004_en	
Eng		tca_20170527_disc_004_en			
Info	Baseline	Gloss	Analyze	Tagging	Print View Text Chart
1	Word	Other-tca	t_a17a4k+4	k+2a4na4	?
	Word Gloss	***	whatever	Intj:tag	
	Word Cat.	***	pro	interj	
Free Spn Qu'e es?					
Eng What is it?					
2.1	Word	TAA-tca	nge3ma2wa5+1k+2	i5ta4mu4I3	nu5a2A4ma4
	Word Gloss	***	***	***	***
	Word Cat.	***	***	***	***
Free Spn Vengan por ac'a a mirar.					
Eng Y'all come and watch it here (spkr-prox)!					
Note i hear no r+1 - nope					
2.2	Word	nu5a2	pe3lr+3	grabacion	na4ng_o1
	Word Gloss	Dloc:SpkrProx	2Pl.Alposs	***	***
	Word Cat.	n	n	***	***
Free Spn Esta sus grabaciones se han prendido.					
Eng Y'all's recording is appearing here (spkr-prox)!					

So is this what we are looking for? What I've done here is added the tier name ("Other-tca" and "TAA-tca") as the first thing in the paragraph, to try to achieve the goal of each paragraph being a different speaker's turn. An alternative with similar results would be to add a sentence delimiter so it appears on a new line (ie "Other-tca" would be 1.1 and t_a17a4k+4 k+2a4na4? would be 1.2) Below are the corresponding ELAN and FLEx baseline representations

An alternative is to make *everything* from one speaker appear in the same paragraph, which would really make reading the text difficult while trying to gloss it, since instead of being sort of off, the timing will be completely nonsensical. Benefit is that everything for each tier stays together.

	00:02:20.000	00:02:22.000	00:02:24.000	00:02:26.000
TAA-tca	t_a17a4k+4 k+2a4na4? n+317#3 pe3r+3dau2n+4!			
KGW-tca				
TAA-tns	Vengan por ac'a a mirar.			
TAA-notes	Esta sus grabaciones se ha			
KGW-tns	i hear no r+1 - nope			
KGW-notes				
Child1-sp				
Child2-sp				
Child3-sp				
Other-tca	t_a17a4k+4			
Other-tns	Qu'e es?			
Other-notes				
Child4-tca				
Child4-tns				
Child4-notes				
TAA-tns-en	Y'all come and watch it here (s			
KGW-tns-en	Y'all's recording is appearing			

Eng	tca_20170527_disc_004_en				
Info	Baseline	Gloss	Analyze	Tagging	Print View Text Chart
Other-tca t_a17a4k+4 k+2a4na4?					
TAA-tca nge3ma2wa5+1k+2 i5ta4mu4I3 nu5a2A4ma4 n+317#3 pe3r+3dau2n+4! nu5a2 pe3lr+3 grabacion na4ng_o1.					

Should I be preserving the written would-be deadkey accent marks <'>, or can/should I replace them with the full composed character (ie should I keep <qu'e> or make it <qué>)?

Given that there are timestamps in the flextext for tca_20170825_abs_ahs_haldi, could you put the corresponding ELAN file up on the Box/Drive so I can cross-check it and see where things might go wrong in the translation process (I can find it online). I really think having some level of gold standard, even if just localized to particular milestones, would be incredibly helpful, to make sure I'm not totally off base. What that would mean is, for example, a partially filled text in FLEx (NOT the raw flextext file only) with a corresponding ELAN file that was imported according to the standards defined by MPI/SIL for that transfer. This would be to check where data can get lost, and how certain fields get translated. Ideally, it would also be amazing to have a sample text showing how the text should look in FLEx. This way I actually have an idea as to what I'm shooting for. I'm fine with giving mock-ups like I do above, the issue will just be that I'll have to await your approval/feedback for each step.