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 FLEx 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 FLEx



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 FLEx 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 FLEx or just when we bring it back out of FLEx (the latter would entail that we don’t use FLEx to gloss things in this tier)



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.



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.