

BLIP Lab's ELAN Tips

Talk Together Study Corpus

If you encounter a language you do not know, or a language that you do know and do not understand.

A: Generate an empty cell in the Utterance and Chunk tier (i.e., do not annotate it). Tag the Language as `#!#langname` or `#!#?`.

If you encounter information you think should be kept private, or you are unsure of how to anonymize certain information.

A: Refer to Privacy Matters under the ELAN handbook or the "Privacy Matters" video.

If you are unsure of how to tag a specific non linguistic sound.

A: Refer to the lists below, or check the more complete list under "List of all Special Codes" under the ELAN handbook/wiki.

If you find that a large amount of data on your recording is inaudible, muffled, or unintelligible.

A: Tag it in the Utterance and Chunk Tier as `###`, and "Vocal Sounds" in the Language Tier. If it is especially inaudible or muffled, make a comment stating this.

If too many tiers getting in the way

A: Hide them!

- Hover over the specific tier > right click > Hide `[participant_name](Tier Name)`.
- To reveal the hidden tier: hover over the tier sidebar > Visible Tiers > check the checkbox of the hidden tiers to reveal them

If you find that the speaker misspeaks in another language but you yourself know what they actually mean

A: Translate the utterance ad verbatim. Subsequently, add `<'intendedspeech'>` to indicate the implied/intended word by the speaker.

If you encounter stuttering or speech that is cut off

A: Indicate stuttering and speech that is cut off with an ellipsis i.e., three dots with no spaces in between. For example "b...b...banana" or "but how..."

Do the utterances need to have proper capitalisation and punctuation?

A: Utterances do not need capitalisation and punctuation. There are a few exceptions that we allow. These include censoring personal information (e.g., BABYNAME, QUALTRICSID), question marks (?), ellipsis to indicate interrupted speech (...), tildes (~) to indicate long interjections (e.g., ohh~), colon/equal/plus signs used within the special codes below, three pound signs (###) to indicate unintelligible/inaudible speech, apostrophes (') in contractions (e.g., don't), and hyphens (-) for word connecting forms (e.g., jalan-jalan)

Quick guide to Baby Language Tier Tags

Description	Marking
Baby Laughing	:v:laughter
Baby Crying	:v:crying
Baby babbling, cooing, includes screaming, whining, exclamations	:v:vocalizations
Baby burping, singing, etc.	:v:airstream

Quick Guide to Special Codes for Adults

All these markings will be done in the Utterance Level Tier.

- {open class} categories can take the characters [a-z 0-9 ? +]
- {closed class} categories can only take from the list of given examples.
- Basically open class categories are free and easy and closed class categories are restricted to a given list.

Marking	Description	Examples
:v:{closed class}	:v: is for Vocal Sounds that are not considered speech. All spaces in these codes must be replaced with an underscore.	:v:laughter :v:shushing_sound :v:gasp :v:tsk :v:animal_sound :v:coughing :v:whistle :v:sigh :v:train_sound :v:kissing_sound
:v:r:{open class}	A subgroup of the above. For when the adult individual makes a repeated rhythmic sound. Code it first as :v:r, then add another colon at the back and transcribe the contents of the utterance.	:v:r:dadadada.
:v:u:{open class}	:v:u: is used to designate Vocal Sounds that cannot be classified. When transcribing these, type v:u then a colon, then transcribe the actual words after the second colon.	
:s:{closed class}	:s: is for Non Vocal Sounds. Tag these articles as "Non Vocal Sound" under the Language Tier.	:s:clapping :s:sneeze
:si:{open class}	:si: is just for singing. All space between lyrics must be replaced by "+". Tag this in the Language Tier as whatever Language tag is applicable. This means if the individual is singing in English, tag it in the Language Tier as English, but if the individual is singing nonsense syllables, tag is as "Vocal Sounds".	:si:mary+had+a+little+lamb :si:humming
:m:{open class}	:m: is for mimicking. All space between lyrics must be replaced by "+". Tag this in the Language Tier as whatever	:m:gagagaga

	Language tag is applicable. This means if the individual is mimicking English words, tag it in the Language Tier as English, but if the individual is mimicking nonsense syllables, tag is as "Vocal Sounds".	
:d:{closed class}	:d: is for discourse markers. Discourse markers are sentence endings like "la", "lor", "leh", or "hor". Unlike the rest of the codes, discourse markers do not need to be put in a separate cell in the Utterance Tier.	:d:lah
:l:langname:{open class}	This is for Loanwords.	:l:malay:jalan
= {open}	This is for Bound Morphemes. So if the Bound morpheme is =le, it will be Mainword =le. If the bound morpheme comes from a different language tag the language accordingly.	=le
#!#langname and #!#?	Unclear language will be tagged as #!#langname on the language tier. Completely unknown language will be tagged as #!#? on the language tier.	#!#Malay

PLEASE SEE ELAN HANDBOOK/WIKI FOR MORE DETAILS AND EXAMPLES