Step 1

1. Tokenization and correction of transcriptions and lexicon

2. Lemmatization and lexicon enrichment and correction of transcriptions

3. Lexicon: check for enclitic co-occurrence (for ASR)

4. Get everything into ELAN

Import:

File>Import> Transcriber file

Asks how many tiers:

<ANNOTATION>

<ALIGNABLE\_ANNOTATION ANNOTATION\_ID="a42" TIME\_SLOT\_REF1="ts120" TIME\_SLOT\_REF2="ts123">

<ANNOTATION\_VALUE>**ka3gah3 nej3 chrun3 nej3 sij3 nej3 ni2yaj3 ta3 chi2hi2 taj3**</ANNOTATION\_VALUE>

</ALIGNABLE\_ANNOTATION>

</ANNOTATION>