Intonation Units and "Sentences"
in ELAN and Toolbox
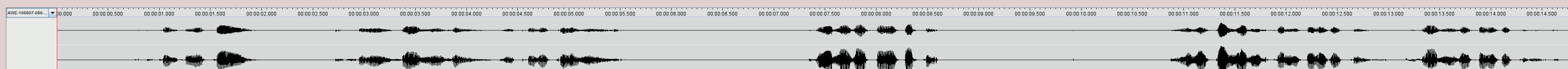
Workflow:

Intonation units:

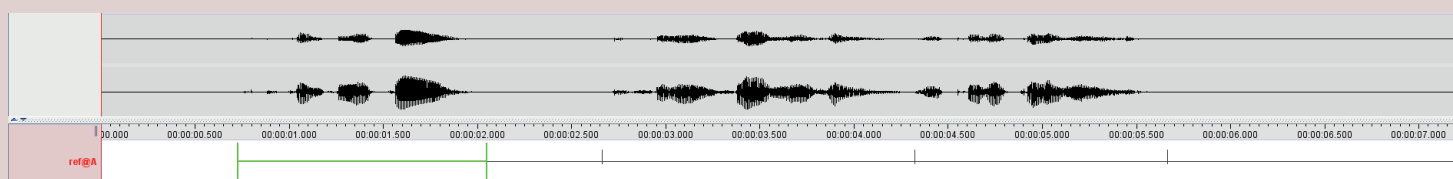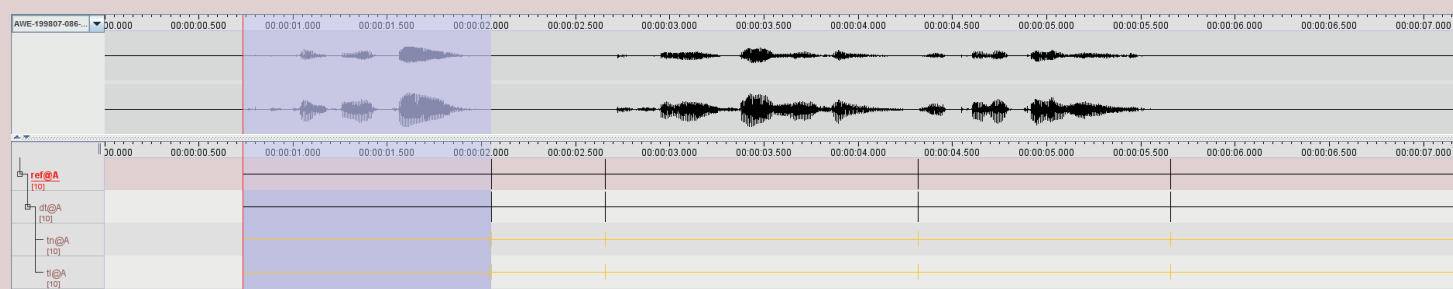*Syntactic units:*

Open the media file(s) in ELAN using the template file *ELAN-Template-IU+Sent.etf*

In <u>Segmentation mode</u>, use the root tier **ref@A** for segmentation into intonation units

Create (empty) annotations on dependent tiers **dt@A**, **tn@A** and **tl@A**

In <u>Transcription mode</u> with **dt@A** and **tl@A**, use **dt@A** for narrow transcription (with pauses, false starts, etc.), and use **tl@A** and **tn@A** for a literal translation (close to word-to-word)

In <u>Annotation mode</u>, merge those **ref@A-segments** that belong to the same syntactic unit

Copy **dt@A** to a tier **po@A** with linguistic type po under ref (rename **dt@A-cp** to **po@A**), and copy **tl@A** to a tier **tf@A** with linguistic type **tf** under **ref** (rename **tl@A-cp** to **tf@A**), merging adjacent transcriptions and translations belonging to the same sentences

In <u>Transcription mode</u> (with **po@A** and **tf@A**), clean the transcription (elimination hesitation, false starts, length markers etc.), and change the translation in **tf@A** into a free translation (idiomatic in the target language)
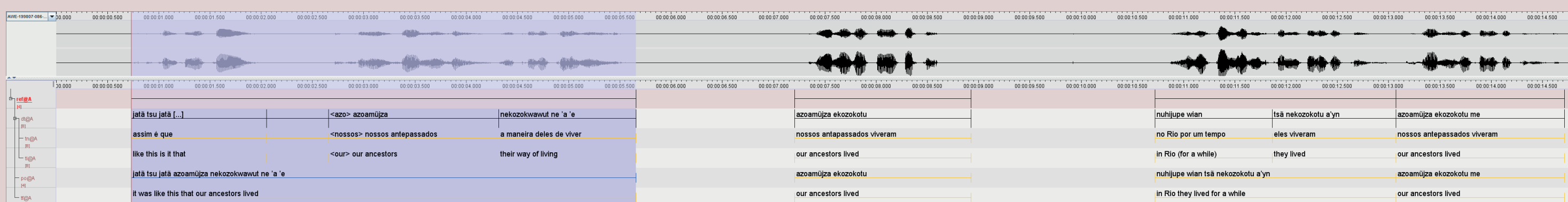
Copy **po@A** to a tier **tx@A** with linguistic type **tx** under **ref@A** (rename **po@A-cp** to **tx@A**)
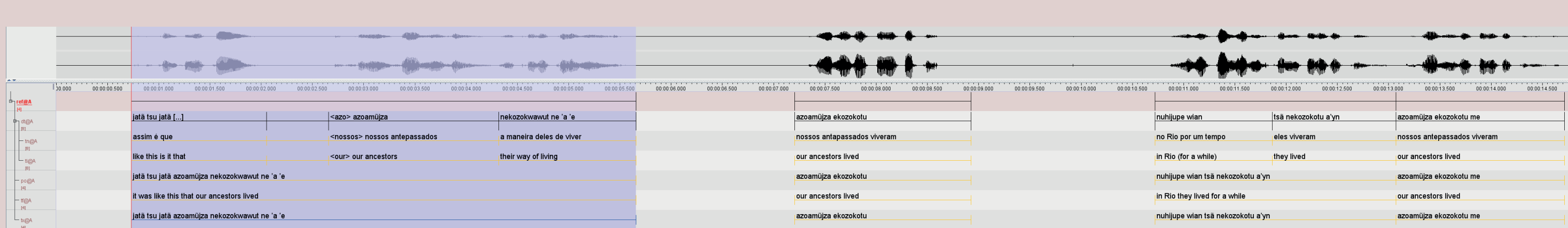
Export tiers **\tx** and **\tf** to Toolbox using Toolbox Database Type *sentences.typ*, creating tier **\ref** (for automatic numbering) and the tiers **\ELANbegin**, **\ELANend**, **\ELANspeaker**.

| \ref Reference	001
\po Practical Orthography	jatã tsu jatã azoamũjza nekozokwawut ne 'a 'e
\tf Free Translation	it was like this that our ancestors lived
\tx Text	jatã tsu jatã azoamũjza nekozokwawut ne 'a 'e

In Toolbox, interlinearize **\tx**, creating **\mb**, **\ge**, and **\ps**

\ref Reference	001						
\po Practical Orthography	jatã tsu jatã azoamũjza nekozokwawut ne 'a 'e						
\tf Free Translation	it was like this that our ancestors lived						
\tx Text	jatã tsu jatã azoamũjza		nekozokwawut		ne 'a 'e		
\mb Morphemes	jatã tsu jatã azo-	amũjza n-	ekozoko	-ap	-put me	'a	'e
\ge Gloss	this! like topic 13-	ancestors 3-	live	-instnr -pst	def	empa	emph
\ps Part of Speech	dprn posp ptl ps.pref- n	ps.pref- v	v>n	-n>n ptl	ptl	ptl	

Import the result back to ELAN, using the same Toolbox Database Type *sentences.typ*
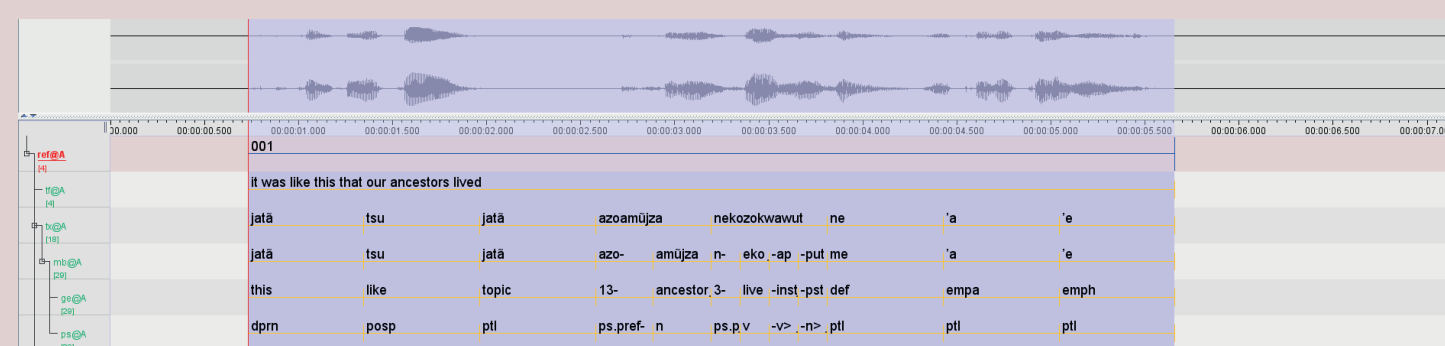
In ELAN, merge the result with the original file, taking only **tx@A** and its dependent tier from the *Toolbox-re-import*
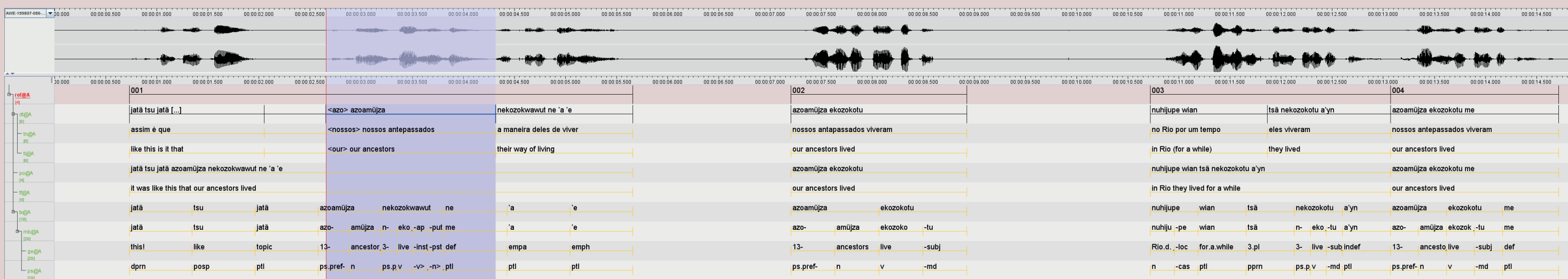
A segmentation into intonation units (IU) is useful for studying prosody and other speech phenomena. For the grammatical analysis, syntactic units (SU, "sentences") are more appropriate, which also apply in spoken language. We propose to annotate both in language documentation.

IU stand in a many-to-one-relation to SU, and are therefore the better starting point. We present an efficient workflow with recommended settings for ELAN and Toolbox that (a) avoids doubled work (segmenting and annotating) in ELAN, and that (b) includes a solid round-trip-configuration for exporting basic annotation from ELAN to Toolbox, where basic glossing is done, and importing the result back from Toolbox into ELAN.

The result is a documentation format that provides a good basis for both discourse and grammatical analysis.

Sebastian Drude
(CLARIN ERIC, Utrecht University,
Radboud Univ. Nijmegen,
Goethe-Universität Frankfurt)
drude@xs4all.nl

Han Sloetjes
(Max-Planck-Institute for
Psycholinguistics, Nijmegen)
han.sloetjes@mpi.nl