Intonation Units and "Sentences"

in **ELAN** and **Toolbox**

Workflow:

Intonation 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

0.000 00.00.000 00.00.000 00.00.01.500 00.00.02.500 00.00.03.500 00.00.03.500 00.00.04.500 00.00.05.500 00.00.05.500 00.00.05.500 00.00.05.500 00.00.05.500

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

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

In Annotation mode, merge those ref@A-segments that belong to the same syntactic unit

| Section | Part | Part

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

| Supply | S

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)

| Section | Property |

Copy po@A to a tier tx@A with linguistic type tx under ref@A (rename po@A-cp to tx@A)

| Supplementary | Supplementar

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

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

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

Import the result back to ELAN, using the same Toolbox Database Type sentences.typ

				3 4					iliailja	- Marie - Miles - Mile	- spinite	njeres er-ĝa			
				1571	payor-real	- Helita			Man-			Maanners			
	30.000 00:00:0		11.000 00:00:01.500	00:00:02.000	00:00:02:500 0	0:00:03.000	00:	00:03.500	00:00:04.00	00 00:00:04.500	00:00:05.000	00:00:05.500	00:00:06.000	00:00:06.500	00:00:07.000
ref@A		001													
- t@A		it was like this that our ancestors lived													
[4] G-1 tx@A		jatā	tsu	jatā	azoaműja	:a	nekozo	kwawut	ne	'a	'e				
[18] G-] mb@A		jatã	tsu	jatā	azo-	amũjza	n- ek	o -ap -	put me	'a	'e				
[29] — ge@A		this	like	topic	13-	ancestor	3- liv	e -inst-	pst def	empa	emph				
[29]		dprn	posp	pti	ps.pref-	n	ps.p.v	-v> <u> </u> -	n> ptl	ptl	ptl				
□ ps@A [29]															

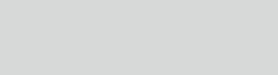
In ELAN, merge the result with the original file, taking only tx@A and its dependent tier from the Toolbox-re-import

AWE-199807-086 ▼ 00.000	00:00:00.500	00:00:01.000	00:00:01.500			10:00:03.000	00:00:03.500		00:00:04.500	00:00:05.000	00:00:05.500	00:00:06.000	00:00:06.500	00:00:07.000	00:00:07.500	00:00:08.000	00:00:08.500	00:00:09.000	00:00:09.500	00:00:10.000	00:00:10.500	00:00:11.000	00:00:11.500	00:00:12.000	00:00:12.500	00:00:13.000	00:00:13.500 0	0:00:14.000 00	0:00:14.500
		- die-		Www.	100	- rengggio				-	Manghas a de .				*	* **	+			·					***		-	***	
_																									Marie Sanda				
30.000	00:00:00.500	00:00:01.000	00:00:01.500	00:00:02.000 0	0:00:02.500	00:00:03.000	00:00:03.500		00:00:04.500	00:00:05.000	00:00:05.500	00:00:06.000	00:00:06.500	00:00:07.000	00:00:07.500	00:00:08.000	00:00:08.500	00:00:09.000	00:00:09.500	00:00:10.000	00:00:10.500	00:00:11.000	00:00:11.500	00:00:12.000	00:00:12.500	00:00:13.000	00:00:13.500 0	10:00:14.000 00	0:00:14.500
[4]																													
₽ ₇ dt@A		jată tsu jată [.1		<azo></azo>	azoaműjza			nekozokwaw	rut ne 'a 'e				8	azoaműjza ekozo	kotu						nuhijupe wian		tsã nekoz	okotu a'yn	azoan	nũjza ekozokotu me	•	
		assim é que			<noss< td=""><td>os> nossos</td><td>antepassado</td><td>os</td><td>a maneira de</td><td>eles de viver</td><td></td><td></td><td></td><td>ļr.</td><td>nossos antapass</td><td>ados viveram</td><td></td><td></td><td></td><td></td><td></td><td>no Rio por um t</td><td>tempo</td><td>eles viver</td><td>am</td><td>nosso</td><td>os antepassados viv</td><td>veram</td><td></td></noss<>	os> nossos	antepassado	os	a maneira de	eles de viver				ļr.	nossos antapass	ados viveram						no Rio por um t	tempo	eles viver	am	nosso	os antepassados viv	veram	
ti@A		like this is it that			<our></our>	<our> our ancestors</our>			their way of living				C	our ancestors lived						in Rio (for a while) they lived			our ar	our ancestors lived					
− po@A		jatā tsu jatā azoamūjza nekozokwawut ne 'a 'e											8	azoamûjza ekozokotu							nuhijupe wian tsā nekozokotu a'yn				azoan	azoamūjza ekozokotu me			
- tr@A		it was like this that our ancestors lived												our ancestors lived						in Rio they lived for a while					our ancestors lived				
tx@A		jatã	tsu	jatã	azoaműj	za n	iekozokwawu	t ne	'a	'e				1	azoaműjza	ekozo	okotu					nuhijupe v	vian ts	ã ne	ekozokotu a'yı	n azoan	nũjza ekozoko	otu me	
mb@A		jatã	tsu	jatã	azo-	amũjza n	- eko -ap	-put me	'a	'e				E	azo- ami	ijza ekozo	oko -tu					nuhiju -pe v	vian ts	ā n-	eko tu a'yı	n azo-	aműjza ekozok	-tu me	
_ ge@A		this!	like	topic	13-	ancestor 3	- live -inst	-pst def	empa	emph	1			ľ	13- and	estors live	-subj					Rio.dloc f	or.a.while 3.	pl 3-	live -sub inde	ef 13-	ancesto live	-subj def	
ps@A [29]		dprn	posp	ptl	ps.pref-	n p	s.p.v -v>	-n> ptl	pti	ptl				ŀ	os.pref- n	V	-md					n -cas p	pti pi	orn ps	s.p.v -md ptl	ps.pre	ef- n v	-md ptl	

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-tripconfiguration 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

5th International Conference on Language Documentation and Conservation (ICLDC) 2-5 March, 2017 University of Hawai'i at Mānoa.