

WWW.MPI.NL

Interlinearization in ELAN

Han Slöetjes

Introduction

ELAN is a manual annotation tool developed by The Language Archive/MPI for **Psycholinguistics**. It supports multi-tier, multispeaker, time-linked annotation of audio and video and is applied in many fields of research, language documentation being one of them.

	2	ELAN 6.0 - r03	_v20_s5.eaf				
	Grid T	ext Subtitles	Lexicon	Comments	Recognizers	Metadata	Controls
	▼ Gloss1						
	> Nr Annotat	on			Begin Time	End Time	Duration
	38 No you c	lrink VK's mbwaa fi	rst		00:04:23.1	66 00:04:24.4	98 00:00:01.332
	39 If I starte	d on this tpile we	o down and et	ay with Mbwokia	00:04:53.80	61 00:04:55.5 07 00:05:11 5	17 00:00:01.656 78 00:00:02 671
	40 Wilen Ki 41 You mus	t come and sing Yi	vémumu, vou i	must come and i	oin 00:05:43.14	41 00:05:44.9	37 00:00:01.796
	42 it's waitin	g in my experience	e has got stuk -	i.e,. I am tired of	f wai 00:05:45.9	16 00:05:49.4	21 00:00:03.505
	43				00:06:00.5	09 00:06:01.4	39 00:00:00.930
	44 Don't be	opening my lime, l	have seen you	ur lime	00:06:01.8	87 00:06:03.8	39 00:00:01.952
	45 Yes they	are waiting for him	1		00:06:26.4	43 00:06:27.5	93 00:00:01.150
	46 What ha	ve they done, we d	ONTKNOW		00:06:30.8	91 00:06:33.0	49 00:00:02.158
r03_v20_s5 🗘	00 00:04:54.00	0 00:04:55.00	0 00:04:5	6.000 00:04	:57.000 00:04:	58.000 00	■ # # # # 1 # 1 # 1 # 1 # 1 # 9:04:59.000 00:
r03_v20_s5 🗘	00 00:04:54.00	0 00:04:55.00	0 00:04:50	6.000 00:04	:57.000 00:04:	58.000 00	• • • • • • • • • • • • • • • • • • •
r03_v20_s5 C	00 00:04:54.00	0 00:04:55.00	0 00:04:50	6.000 00:04	:57.000 00:04:	11 - 5 - 0 - 0 - 0 - 0 - 0 - 0 - 0 - 0 - 0	• H H H H H H H H H H H H H H H H H H H
r03_v20_s5 Gloss1	00 00:04:54.00 00 00:04:54.00 11 11 11 11 11 11 11 11 11 11 11 11 11	0 00:04:55.00 0 00:04:55.00 0 00:04:55.00 ed on this tpile we	0 00:04:50	6.000 00:04	57.000 00:04: 57.000 00:04:	558.000 00	• •
r03_v20_s5 Gloss1 [90] - Gloss2 [90] - Vou are	00 00:04:54.00 00 00:04:54.00 11 start "drinking"/singing	0 00:04:55.00 ed on this tpile we he Tpyîpwî py	0 00:04:50	6.000 00:04	57.000 00:04: 57.000 00:04:	558.000 00	• •
r03_v20_s5 Image: Control of the second secon	00 00:04:54.00 00 00:04:54.00 16 I start 9 "drinking"/singing 1 :aa, kî 1	0 00:04:55.00 0 00:04:55.00 ed on this tpile we he Tpyîpwî py pile p:uu dpî kaa k	0 00:04:50	6.000 00:04	:57.000 00:04: :57.000 00:04: :57.000 00:04: :57.000 00:04: :57.000 00:04:	11 - 11 - 11 - 11 - 11 - 11 - 11 - 11	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
r03_v20_s5 ♥ 00:04:52.000 00:04:53.00 00:04:52.000 00:04:53.00 ● ● <t< td=""><td>00 00:04:54.00 11 11 11 11 11 11 10 00 00:04:54.00 11</td><td>0 00:04:55.00 0 00:04:55.00 ed on this tpile we he Tpyîpwî py pile p:uu dpî kaa ku</td><td>0 00:04:50</td><td>6.000 00:04</td><td>:57.000 00:04: :57.000 00:04: :57.000 00:04: :57.000 00:04: <u>It's the brai</u> mbwaa tpilewe m</td><td>58.000 00 558.000 00 558.000 00</td><td>1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1</td></t<>	00 00:04:54.00 11 11 11 11 11 11 10 00 00:04:54.00 11	0 00:04:55.00 0 00:04:55.00 ed on this tpile we he Tpyîpwî py pile p:uu dpî kaa ku	0 00:04:50	6.000 00:04	:57.000 00:04: :57.000 00:04: :57.000 00:04: :57.000 00:04: <u>It's the brai</u> mbwaa tpilewe m	58.000 00 558.000 00 558.000 00	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

Background and history

- morphological parsing and glossing
- by **Toolbox**³ and **FLEx**⁴
- module, **LEXAN**²



han.sloetjes@mpi.nl

https://archive.mpi.nl/tla/elan



in language documentation the transcription and translation steps are often followed by

• it is possible to perform these steps manually in **ELAN**, but the preferred approach is to apply computer-aided methods as offered

ELAN provides import and export functions for the file formats of these tools, so that users can create a toolchain and move their data from one tool to another for the task at hand

• **TLA** once developed the lexicon tool **LEXUS**¹ and prepared to combine it with **ELAN** by way of a new Natural Language Processing

Interlinearization mode

Configure /	Analyzer Set	tings		▶s	Analyze	/ Interli	nearize	e 🗆	Recurs	ve					4	🕸 F	ont Siz
Analyz	zer & Source	e-Target Co	nfiguration	A phrase-segnum-en	1												
Current nu	mber of co	nfigurations:	: 2	A word-txt-seh	Babanda		ana			mur	nda		ukı	ılu			
					A0 bab	a nga	3	n		mu	nda			kulu			
				A_morph-txt-sen	1 mai		u- 20.	4 4		2	maa		u-	area de			
				A_morpn-gis-pt	i pai	15	35+	1 16	er	3	mac	chamba	3	grande			
• • • • • • • • • • • • • • •			A	A_word-pos-pt	N		Irreg	g-na		N			Adj				
Sena Lexico	n from FLEX		con Actions •	Speaker	A												
Lexical Entr	ies			TC	00:00:00.	.000 - 0	0:00:0	06.00	0								
Lexical	Morph T	Citation	Gloss														
bubu bubu	root	cibubu	[stamme	A_phrase-segnum-en	2												
makuant	stem	makumat	[sixty][se	A_word-txt-seh	Mwenem	0	muna	L		niti		mizinji					
cibese	root	macibese	[morning	A morph-txt-seh	mwenem	0	mu-	n	-a I	ni- ti		mi- z	inji				
dzinza	root		fethnic g	A morph dia pt		esmo	3	tor		1 a	vora	1 n	nuito				
mwenikir	root	mwenikira	[iluminat	A_morph-gis-pt		Como				т и	vora	- II	Tuito				
tsikana	root	ntsikana	[girl][rap	A_word-pos-pt	dem6		Irreg-	na		N		Adj					
nyama y	discontig	nyama y	[fish][pei	Speaker	A												
gw khangala	stem	gwa	play jo	TC	00.00.06	000 - 0	0.00.1		0								
bambaya	root		[sweet p		00.00.00.	.000 - 0	0.00.1	12.00	0								
lekerer	root	lekerera	[forgive][A_phrase-segnum-en	3												
jaia dade	root	njala cidade	fdrving r	A word-txt-seh	Muna		milara	ania		mif	an		n	nindimu		na	min
fik	root	fika	[arrive][c											·		. nex	1
dul	root	dula	[pull out]	A_morph-txt-seh	mu- n	-a	mı-	lalan	ja	mı-	figu		n	ndim	iu	na	mı-
cap	root	capa	[row][re	A morph-gls-pt	3 ter	IND	4	laran	jeira	4	ban	aneira	4	limor	neira	е	4
bondo	root	mbondo	[knee][jo		Irreg.no		N			N			N			Coni	Í N
mbava	root		I thief la	A_word-pos-pt	mey-na		IN			IN						Conj	
panga	root	phanga	[bandit][Speaker	A												
siy	root	siya	[leave so	TC	00.00.12	000 - 0	0.00.1	18 00	0								
lungu	root	Mulungu	God De		00.00.12.		0.00.	10.00	0								
bidzi	root	mbidzi	[zebra][z	A_phrase-segnum-en	4												
munga	stem	munga	[when][q	A word-tyt-seh	Ikhalipo			m	inda		m'mu	nda					
bokho	root		[goat][ca														
ala	root	mwala	[stone][p v	A_morph-txt-seh	- -	kha- li	po	o m	nga		m'-	mu- no	la				
2022	raat	20210	[throw][A_morph-gls-pt	4+5+9 I	MP s	er el	e 4	esp	inho	18	3 m	acha	mba			
	Add D/	amovo	Edit							_							

- new text oriented mode with an Interlinear Glossed Text (IGT) style of user interface
- the **LEXAN** modules became extensions of **ELAN** (after end of support for **LEXUS**)
- combined with a new lexicon component
- supports the process of (machine-assisted) parsing and glossing



ICLDC7, 2021



WWW.MPI.NL

Main characteristics

- text oriented (but still linked to the audio!)
- keyboard driven

displayed at the top.

- optimized for computer assisted morphological parsing and glossing
- text analyzer modules provide 'suggestions' to be disambiguated by the user
- contains a lexicon editor and viewer
- a flexible system, therefore some configuration is required

I																			
• • • • • • • • • • • • • • • • • • • •	Ar	nalyze / I	Interli	nearize	e [Recu	irsiv	e						F	ont Size:	18 –	- +		
A_phrase-segnum-en	1										_							^	
A_word-txt-seh	Bab	anga		ana				mun	da		uk	culu							
A_morph-txt-seh	^0-	°0- baba nga a- n -a mu- nda u- kulu																	
A_morph-gls-pt	1	pai	1S	3S+	1 te	er IN	١D	3	mao	chamb	a 3	grande							
A_word-pos-pt	Ν			Irreg	g-na			Ν			Ac	j							
Speaker	А																		
тс	00:0	0:00.00	0 - 0	0:00:0	06.00	0													
A_phrase-segnum-en	2																		
A_word-txt-seh	Mwe	enemo		muna			m	iti		mizir	ıji								
A_morph-txt-seh	mwe	enemo		mu-	n	-a	m	i- ti		mi-	zinji								
A_morph-gls-pt	aque	ele mes	smo	3	ter	IND	4	ar	vora	4	muito								
A_word-pos-pt	dem	16		Irreg-	na		Ν			Adj									
Speaker	А												milar	anja		mifigu			
ТС	00:0)0:06.00	0 - 0	0:00:1	12.00	0													-
0												E						L	
Interlinear edit	or	with	a p	lay	sel	ecti	ior	ı bı	itto	n		*	Inc	rementa	selection	\checkmark	Remer	nber wi	ndow pos
to play the curr	ren	t acti	ive	phr	ase									S1 P	arse #0 Cł	n Si	2 Parse	#2 Chc	sen 1 tim
	•••			r		•						A_morph-t	xt-se	h mi-	laranj -a	m	i- larar	ij -a	
												A_morph-g	gls-pt	t 4	*** IN	D 4	***	nom	inalizador
														- 	arsa #3 0		Parse #	1	
												A maanala t		- mi-	larani -a	mi	larani	-2	
												A_morph-t	xt-se		***	r 1	***		
												A_morph-g	gis-pi			-			
														S5 Pa	arse #4	S	6 Parse	#5	
												A_morph-t	xt-se	h mi-	laranj -a	m	i- larar	ij -a	
The sugge	stic	ons v	vino	dow	v di	spla	ays	s th	e			A_morph-g	gls-pt	t 4	*** IN	D 4	***	IND	
suggestion	ns i	n the	e sa:	me	lay	out	as	s th	e ec	litoi	r.			S7 Pa	arse #6			S8 Par	se #7
The most	free	quen	tly	sele	ecte	ed p	oar	ses	are			A_morph-t	xt-se	h mi-	laranj -a			mi- la	ranj -a

1. Ringersma, J. and Kemps-Snijders, M. (2007). Creating multimedia dictionaries of endangered languages using LEXUS. Interdpeech 2007

A_morph-gls-pt 4 *** nominalizador 4 ***

2. Stehouwer, H. and Drude, S. (2012). Lexan: A lexical annotation framework for ELAN. Talk presented at LREC 2012

Text analyzers

- are modules that receive text as input (an annotation) and produce output for a single or for multiple annotations
- can be added as extensions through an API can be connected to a lexicon

Analyzer	Source	Target
😵 Parse Analyzer	word-txt-seh	morph-txt-seh
	Applies to tiers: A_word-txt-seh	Applies to tiers: A_morph-txt-seh
🖗 Gloss Analyzer	morph-txt-seh	morph-gls-pt
	Applies to tiers:	Applies to tiers:
	A_morph-txt-seh	A_morph-gls-pt
Show tier mapping		
Configure Parse Analyzer		Remove configuration Edit configurations

- built-in analyzers:
 - whitespace splitt
 - morphological particular
 - gloss analyzer
- the parser and glo
 - require access to lexicon
 - keep record of cl made to improve suggestions • are language independent • the parser is implemented as a finite state

 - machine
- 3. Field Linguist's Toolbox, https://software.sil.org/toolbox/

- 6. http://corpafroas.tge-adonis.fr





Linking analyzers to tier ypes and configuration of an analyzer.

	A	>
1	Configure Analyzer Settings	
ter	☑ Include variants in the parsing process	
	Match longer prefixes/suffixes first	
arser	Exclude aborted parses from results	
	Case sensitive matching	
	Match entry field language against tier content language	
	Only suggest parses with same category constituents	
	Use the citation form of the lexical entry in the output	
ocor:	Maximum number of parse steps: 512	
JS5EL.	Affix marker character:	
	Clitic marker character:	
oa	String for missing values ***	
	"Replace" field in lexicon: replace	
	Apply Settings	
noices		

4. FieldWorks Language Explorer, https://software.sil.org/fieldworks/ 5. Lexicon Interchange Format, https://github.com/sillsdev/lift-standard

Lexicon editor

Lexicon				- Lexical Entry-
Lexicon Name	Sena Lexicon from FLE	x		Levical Unit adidi
Filtor Entries		Posot		ID: 0. 315b8306-207c-4000-b100-f053248d1bdc
		Keset	Columnia All Columnia 🐨	Date Created: 2005-06-22716:20:207
Lexical Entries				
	Marsh Tara	Chatlan	Chara	Morph Type root
Lexical Unit	Morph Type	Citation	GIOSS	Citation
=di	enclitic		[EVID][EVID]	- Variant
=ene	enclitic		[very][muito]	
=mbo	enclitic		[ADD][ADD][INT	
а	root		[ASSOC][ASSOC]	Lunguage riop.
a-	prefix		[3P+2][3P+2][2]	- Note
a-	prefix		[he+PST][ele+PS	Sense S
a-	prefix		[6][6][6][6][6][6	Sense ID: s 802fa71e-e2ch-4e2f-0ca0-d37d77c5acda
a-	prefix		[3S+1][3S+1]	Scher ID. S_SS210/10 CEED 1021 Star 05/07/CS4000
a-	prefix		[PAST][PASSADO]	Gram. Category Adjectivo
a-	prefix		[assocpx][assocpx]	Gloss good
-a	suffix		[-er][nominalizador]	Language Prop. en
-a	suffix		[IND][IND]	Class hom
aa-	prefix		[aa-][n][aa-2]	Language Prop. pt
adia	root	mwadia	[dugout canoe][ca	
adidi	root		[good][bom][righ	Definition
adu	root	cadu	[flea][pulga]	Language Prop.
-aji	suffix		[agent][agent]	Comment
aka	root	caka	[year][ano]	Sense Note Question: m'nhyadidi = ndi + nyadidi Note that N made it become aspirated · Question: Th
aka	stem		[this][este]	Sense Note Question. In physical – nati - physical note and thindae it become aspirated, Question. In
akudya	root	cakudya	[food][comida]	Sense S
akudza	root	wakudza	[foreigner][estran	Sense ID: s_13e26eb7-5809-40d3-a08d-78c704ba8a30
akugawika	root	cakugawika	[dividable][dividivel]	- Cram Category Nome
akutoma	root	cakutoma	[first][premeiro]	Grain. Category Nome
ala	root	mwala	[stone][pedra]	Gloss right side
ala	root	cala	[tinger][dedo]	Language Prop. en
amacadu	root	macadu	[axe][machado]	Gloss direita
ambo	root	mwambo	[law][lei]	Language Prop. pt
ambuk	root	ambuka	[cross][atravessar]	Definition
amyali	root	mwali	[girl][rapariga]	Language Prop.
-an	suffix		[REC][REC]	Commont
ana	1001	mwana		Comment
ana-cinthu	1001	mwana-cinthu	LownerJ[dono]	Im Sense Note
anakati	TOOT	imwanankati	messenger men 🛛 🗸 🛛	

- multiple lexicons can be created and linked
- lexicon entries have a few predefined fields
- custom fields can be added
- sorting of entries according to a custom sort order
- the structure of a lexical entry is similar to the LIFT⁵ format
- import lexicons from **Toolbox**, **FLEx** (**LIFT**) and **CorpAfroAs**⁶ format
- export a lexicon to **LIFT** format

Acknowledgements

The design and implementation of the Interlinearization mode have for a large part been the work of Herman Stehouwer (LEXAN) and Olaf Seibert.



ICLDC7, 2021