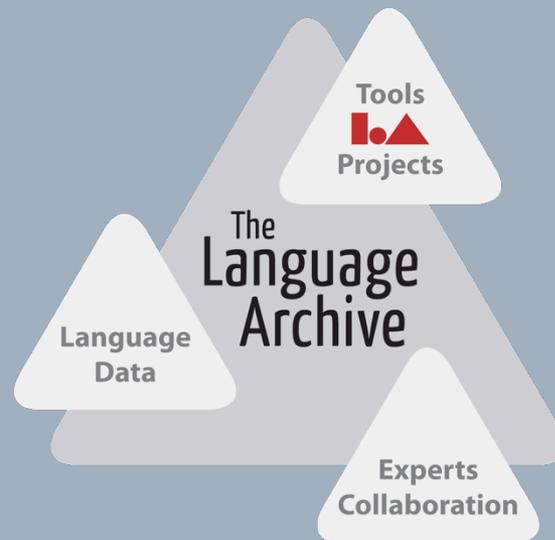




MAX-PLANCK-GESELLSCHAFT



ELAN - novel developments

Han Sloetjes, Aarthu Somasundaram, Herman Stehouwer, Sebastian Drude

The Language Archive - Max Planck Institute for Psycholinguistics
Nijmegen, The Netherlands



Overview, current state of ELAN



- ELAN main window

The screenshot displays the ELAN main window for the file '010_autobiogr.eaf'. The interface includes a video player on the left showing a woman speaking, a grid of annotations on the right, and a timeline at the bottom with audio and text tracks.

The grid shows the following annotations:

Nr	Annotation	Begin Time	End Time	Duration
1	it was like that, son	00:00:00.000	00:00:02.317	00:00:02.317
2	here my mother had a love relationship, with her man, with my f...	00:00:02.317	00:00:07.824	00:00:05.507
3	they were still lovers	00:00:09.876	00:00:14.832	00:00:04.956
4	one day people saw them	00:00:14.832	00:00:18.268	00:00:03.436
5	father got married	00:00:25.137	00:00:28.420	00:00:03.283
6	mother got married	00:00:28.420	00:00:31.055	00:00:02.635
7	mother had a child	00:00:36.794	00:00:38.606	00:00:01.812
8	I grew in my mother's belly	00:00:38.606	00:00:41.515	00:00:02.909
9	ready, the belly grew very big	00:00:46.297	00:00:54.369	00:00:08.072
10	she brought me there	00:00:54.369	00:00:57.479	00:00:03.110
11	to Jaibiaia, they brought me to the place	00:00:57.479	00:01:02.581	00:00:05.102

The timeline at the bottom shows the audio waveform and the corresponding text tracks for the selected annotation (Nr 1). The text tracks are:

- ref@010 [276]: akyj jê, pi'a: | ujape ange 'ajo'atu a'y'n, omen ta me, apai ta me. | ujapywo a'y'n ne.
- Sort@010 [276]: it was like that, son | here my mother had a love relationship, with her man, with my father. |
- SmngE@010 [226]: foi assim, filho | aqui minha mãe namorou com o marido, com meu pai | aqui neste lugar
- ref@NOBODY [2]:
- Sort@NOBOD [2]:



Overview, current state of ELAN



MAX-PLANCK-GESELLSCHAFT

- Multi-tier, multi-speaker, time-linked annotation of audio and video recordings
- Available for Windows, Mac OS and Linux
- Open source, (mainly) written in the Java programming language
- Current version 4.5.1
- <http://tla.mpi.nl/tools/tla-tools/elan/>





- General overview, current state of ELAN
- Working modes - Transcription mode
- Multiple file editing
- Tier based operations
- Interoperability
 - Toolbox and Praat interoperability
 - FLEx interoperability
- Interlinearization mode - Lexan - Suggestion modules towards (semi-)automatic annotation
 - analyzers/recognizers - “annotyzers”
 - web services



- Different working modes (“perspectives”) for different tasks
- Current available working modes
 - Annotation mode
 - Synchronization mode
 - Segmentation mode
 - Transcription mode
 - Interlinearization mode



Segmentation mode



MAX-PLANCK-GESellschaft

- Creation of segments while the media is playing
- Switching between tiers while playing

ELAN 4.5.0 File Edit Annotation Tier Type Search View Options Window Help

ELAN - Household_1_small.eaf

Segmentation Controls

- Two keystrokes per annotation (non-adjacent annotations)
- One keystroke per annotation (adjacent annotations)
- One keystroke per annotation, fixed duration

Duration (ms.): 500

- Stroke marks begin of annotation
- Stroke marks end of annotation

Delayed mode: compensate for keystroke delay (ms.) 225

00:03:09.660 Selection: 00:00:00.000 - 00:00:00.000 0

03:03.000 00:03:04.000 00:03:05.000 00:03:06.000 00:03:07.000 00:03:08.000 00:03:09.000 00:03:10.000 00:03:11.00

A_po

A_po

A_vb

B_po | su iroi irèrè tà mc. | Bòò su iroi irèrè tà |

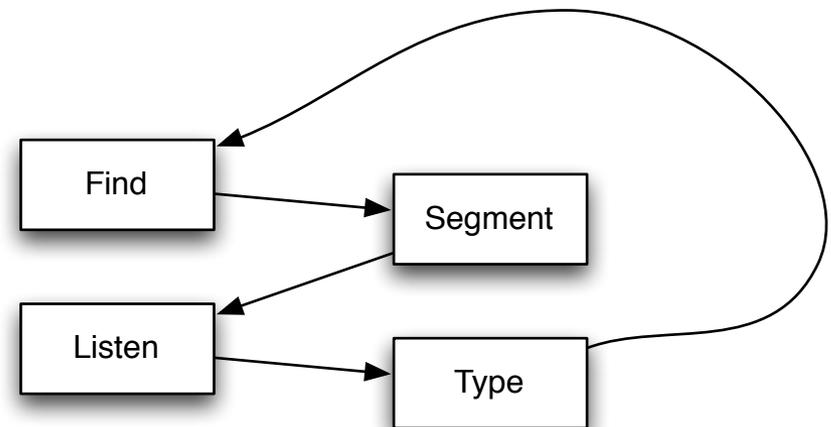
B_vb

C_po

QIR_sample



- Optimized for listening (or watching!) and typing text
- Based on the workflow of *segment first, label later*
- Keyboard oriented





Transcription mode



MAX-PLANCK-GESellschaft

- Configuration of the table based on tier types
- Hiding/disabling of individual tiers
- Cross-column or within-column traversal

No.	Type1 : p (practical orthography)	Type2 : t (literal translation)	Type3 : f (free translation)
258	lãa yã ásarã.	I will buy snuff.	I want to buy snuff [tobacco]
259	onyũ gũ lo tsuaraã fufu ɔwetè ikayc ga ngbe amc nẽ.	the person that last pounded fufu in this mortar.	Whoever last pounded fufu in this mortar
260	áasarã na bó.	we do not have snuff.	We don't have snuff.
261	() a- a bara ãra.	() you- you've done something.	() y- you've done something
262	lò na kura lãa yere onè gũ mmò.	I have nothing to tell that person.	I have nothing to tell that person.
263	so be?	that what?	What?
264	onyũ gũ lo tsuara fufu ɔwetè ikayc ga ngbe amc lo too.	the person that finally pounded fufu in this mortar I am referring to.	Whoever last pounded fufu in this mortar I mean.
265	a nyã masamũra i kube.	did you find the tortoise.	
266	oòf masamũra ɔyeɛ nẽt.	oof tortoise are many --.	
267	mĩ nyũ nɔũ i kãpòt amc.	you should watch water in the bowl.	
268	Eɛ nyũ fo socks nɔũ.	Edi look at your socks water.	
269	be ʒãó ásu fo Dãdɛ socks asé áa pia ma sòcks futzãá má mmò ánc ána.	why did you take your father's socks you went and put into those white socks again.	
270	hal.	hey.	
271	(kòka)	(song)	
272	(ɔ su kapòt ma ùr ikãrũ)	(she hit the aluminium bowl on the ground)	
273	rere masocks q - i masocks amc - bé ʒãó abá ʒãé ngbo ní.	pick socks out - in those socks - why are you difficult like that.	
274	oòo kí nyũ.	no ohi! look.	
275	(kòka)	(song)	
276	su áa la (i kpaá)	take and dry (unclear)	
277	(i kpaá) nɔũ ámẽ mare ncc - rjma sc ma fe mmo ncc -	(unclear) those in the water habitat - those that have been passing there - -	
278	nɔũ amc mare - nɔũ ámẽ mare ʒ nẽc.	those in water habitat - as for those in aquatic habitat.	
279	(to oca kocko)	(she is sacking the chicken)	
280	(i kpaá) ñĩ sc áye ngbe?	(unclear) but you know that right?	
281	nɔũ amc mare ncc - bo - boó ma ùsã ngbe nẽ boó -	those in aquatic habitat - we - we those who are sitting here we too -	
282	boó nɔũ ámẽ (i kpaá)	we too aquatic inside (unclear)	
283	hè -	(calling)	
284	ʒo nya gũ Bóáb kã ʒto ʒ nyũ kpɛkpɛtĩ ngbe.	have you see that Boab is looking at the puppy here.	



Transcription mode



MAX-PLANCK-GESellschaft

- Configuration of the table based on tier types
- Hiding/disabling of individual tiers
- Cross-column or within-column traversal

Transcription mode settings

Font size: 10

Number of columns: 5

Column	Select type for column
1	po (practical orthography)
2	
3	dt (detailed transcript)
4	tl (literal translation)
5	tf (free translation)
	tn (translation in lingua franca)

Select tiers... Apply Cancel

No	Type1: po (practical orthography)	Type2: tl (literal translation)	Type3: tf (free translation)
258	lāa yā āsarā.	I will buy snuff.	I want to buy snuff [obacco]
259	onyū gō lo tsuarā fufu owetē ikayc ga ngbe amc nē.	the person that last pounded fufu in this mortar.	Whoever last pounded fufu in this mortar
260	āsarā na bō.	we do not have snuff.	We don't have snuff.
261	() a- a bara ira.	() you- you've done something.	() y- you've done something
262	lō na kuira lāa yere onō go mmō.	I have nothing to tell that person.	I have nothing to tell that person.
263	sc be?	that what?	What?
264	onyū gō lo tsuara fufu owetē ikayc ga ngbe amc lo too.	the person that finally pounded fufu in this mortar I am referring to.	Whoever last pounded fufu in this mortar I mean.
265	a nyā masamūra i kube.	did you find the tortoise.	
266	oōl masamūra zyeđe nē.	oo! tortoise are many --.	
267	mī nyō ngju i kápōi amc.	you should watch water in the bowl.	
268	Eđ nyō fo socks ngju.	Eđ! look at your socks water.	
269	be šsō āsu fo Dāđ socks asē āa pia ma šsōcks fudzaā má mmō amc āna.	why did you take your father's socks you went and put into those white socks again.	
270	hai.	hey,	
271	(kōka)	(song)	
272	() su kapōi ma ū ikāŕū)	(she hit the aluminium bowl on the ground)	
273	rzc masocks d- i masocks amc - bé šsō abā šlē ngbo nē.	pick socks out - in those socks - why are you difficult like that.	
274	oōo lōl nyō.	no oh! look.	
275	(kōka)	(song)	
276	su āa la (l kpaā)	take and dry (unclear)	
277	(l kpaā) ngju āmē mare ncc - rma sc ma fe mmo nē -	(unclear) those in the water habitat - those that have been passing there --	
278	ngju amc mare - ngju āmē mare ō nē.	those in water habitat - as for those in aquatic habitat.	
279	(to zsa kōka)	(she is sacking the chicken)	
280	(l kpaā) inī so āye ngbe?	(unclear) but you know that right?	
281	ngju amc mare nē - bō - bō ma ūsi ngbe nē bō -	those in aquatic habitat - we - we those who are sitting here we too -	
282	boō ngju āmē (l kpaā)	we too aquatic inside (unclear)	
283	hē -	(calling)	
284	lo nya gō Bōab kā ūo ō nyō ikpīkpībi ngbe.	have you see that Boab is looking at the puppy here.	



Transcription mode



MAX-PLANCK-GESELLSCHAFT

- Configuration of the table based on tier types
- Hiding/disabling of individual tiers
- **Cross-column or within-column traversal**

No	Type1 : po (practical orthography)	Type2 : tl (literal translation)	Type3 : tf (free translation)
258	làa yá àsarà.	I will buy snuff.	I want to buy snuff [tobacco]
259	onyũḡ gó lo tsuará fufu owetè ikaye ga ngbe amè nê.	the person that last pounded fufu in this mortar.	Whoever last pounded fufu in this mortar
260	àsarà na bó.	we do not have snuff.	we don't have snuff.
261	() a- a bara ìra.	() you- you've done something	() y- you've done something
262	lò na kuira làa yere onìḡ gò mmò.	I have nothing to tell that person.	I have nothing to tell that person.
263	so be?	that what?	What?
264	onyũḡ gó lo tsuará fufu owetè ikaye ga ngbe amè lo toò.	the person that finally pounded fufu in this mortar I am referring to.	Whoever last pounded fufu in this mortar I mean.
265	a nyá masamùrà í kùlè.	did you find the tortoise.	
266	oò! masamùrà oyede nê.	oo! tortoise are many ~.	
267	mì nyḡ nḡu í kápòì amè.	you should watch water in the bowl.	
268	Edí nyḡ fò socks nḡu.	Edi look at your socks water.	
269	be ósó ásu fò Dádí socks asè àa pia ma sócks fudzaá má mmò amè àna.	why did you take your father's socks you went and put into those white socks again.	



Multiple file processing



MAX-PLANCK-GESELLSCHAFT

- Creation of eaf files for a series of recordings
- Changing tier and type properties
- Tier-based operations; overlaps, subtraction, merging of annotations
- Find and Replace
- Import
 - Toolbox, Praat
- Export
 - Toolbox, Praat, Tab-delimited, Statistics, Wordlist ...

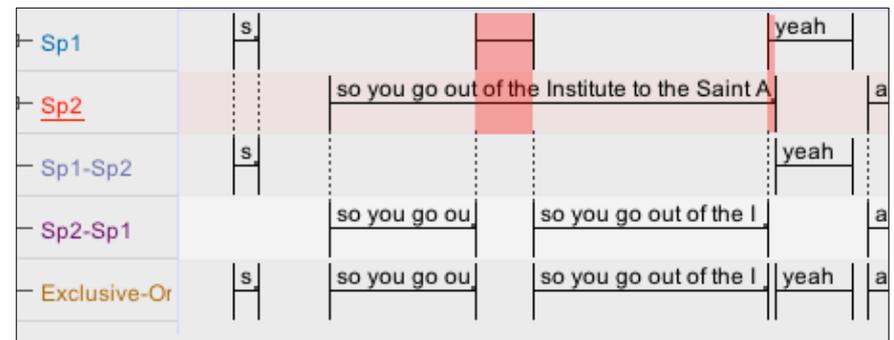
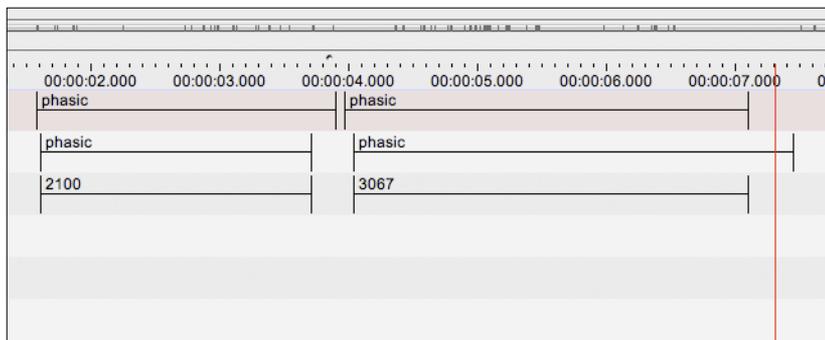


Tier-based operations



MAX-PLANCK-GESELLSCHAFT

- Creation of tiers with annotations resulting from applying logic to existing annotations
 - Annotations from overlaps
 - Annotations by subtracting, XOR logic
 - Annotations by merging
 - Annotations from gaps





- **Toolbox** (import and export)
 - remains problematic (easy to break)
 - but, stabilized somehow
- **Praat** (import and export)
 - TextGrid files with Interval and Point tiers
 - short notation files now supported
 - ELAN doesn't support single point annotations yet
- **Transcriber** (import only)
 - support for the new file format (.tag) to be added



- FLEx import
 - introduced in ELAN version 3.8.0 (Aug 2009)
 - time and speaker information not available, problem with retaining time alignment
 - Upcoming changes in import function
 - support for attributes begin-time-offset, end-time-offset and speaker
 - multiple top level tiers for multiple speakers
 - removed morph element as possible time-alignable level
 - retain punctuation information (by linking to an ISOcat data category), language to font mapping



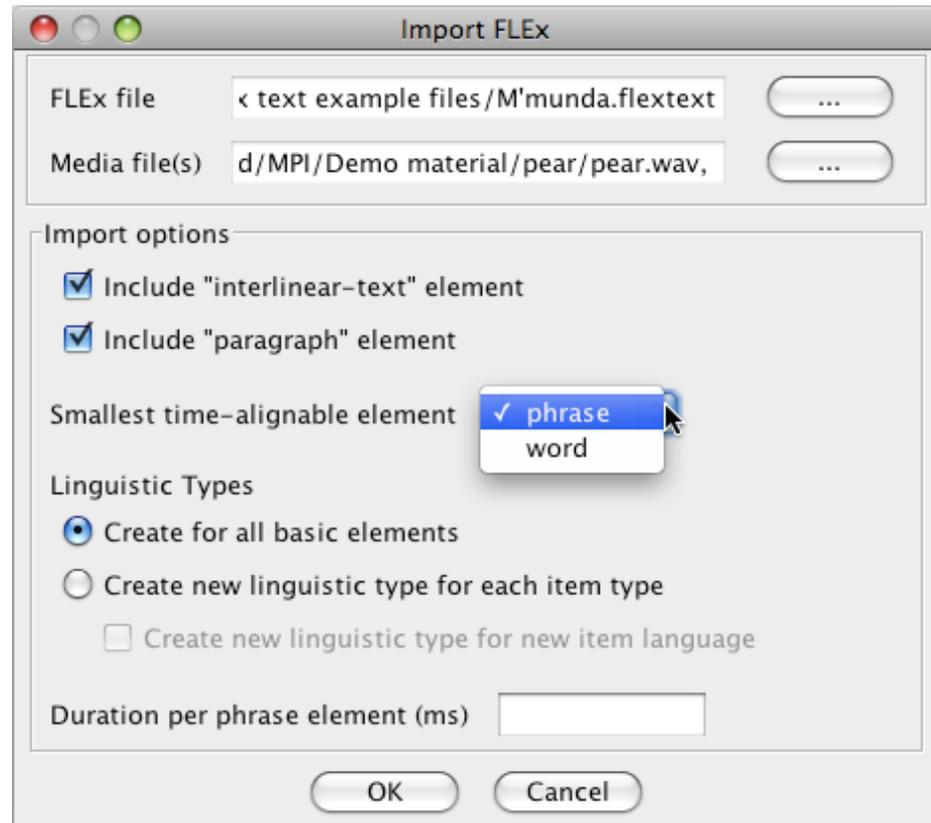
- FLEx import
 - element to tier name mapping convention:
speaker_<major_element>-<item_type>-<item_lang>

e.g. A_phrase-txt-en

```
<phrase begin-time-offset="2000" end-time-offset="4000" speaker="BBB">  
  <item lang="en" type="segnum">2</item>  
  <item lang="en" type="gls">There he has many trees.</item>  
  <words>  
    <word>  
      <item lang="seh" type="txt">Mwenemo</item>  
      <morphemes>  
        <morph type="stem">  
          <item lang="seh" type="txt">mwenemo</item>
```



- FLEx import
 - optional import of
 - interlinear-text element
 - paragraph element
 - morph no longer time-alignable
 - word time-alignment will be lost on export
 - linguistic types for
 - basic elements (phrase, word, morph, paragraph)
 - basic element + item type
 - basic element + item type + language





- FLEx import

Texts & Words

- Interlinear Texts
- Concordance
- Word List Concordance
- Word Analyses
- Bulk Edit Wordforms
- Statistics

Texts

Title

Show All

M'munda

Phatani ntima

Thika na Ntsikana

Wapakila m'mwadia

Text

Title Sen M'munda
Por Na Machamba

Info Baseline Gloss Analyze Tagging Print View Text Chart

1 Word Babanga ana munda ukulu

Morphemes ^0- baba nga a- n -a mu- nda u- kulu

Lex. Entries mu-₁ baba nga a-₅ n₂ -a₁ mu-₂ nda u-₃ kulu

Lex. Gloss 1 pai 1S 3S+1 ter IND 3 machamba 3 grande

Lex. Gram. Info. N:ncl N 1/2 Poss V:subj Irreg-na V:vf N:ncl N 3/4 Poss:possncl Adj

Word Gloss pai meu machamba grande

Word Cat. N Irreg-na N Adj

Free My father has a large garden.

2 Word Mwenemo muna mitti

Morphemes mwenemo mu- n -a mi- ti

Lex. Entries mwenemo mu-₃ n₂ -a₁ mi- ti

Lex. Gloss aquele mesmo 3 ter IND 4 arvora

Lex. Gram. Info. dem6 V:subj Irreg-na V:vf N:ncl N 3/4

Word Gloss lá tem árvoras

Word Cat. dem6 Irreg-na N

Free There he has many trees.

3 Word Muna milaranja , mifigu , mindimu na mimanga

Morphemes mu- n -a mi- laranja , mi- figu , mi- ndimu na mi- manga

Lex. Entries mu-₃ n₂ -a₁ mi- laranja mi- pfigu mi- ndimu₂ na₁ mi- manga₂

Lex. Gloss 3 ter IND 4 laranjeira 4 bananeira 4 limoneira e 4 mangeira

Lex. Gram. Info. V:subj Irreg-na V:vf N:ncl N 3/4 N:ncl N 3/4 N:ncl N 3/4 Conj N:ncl N 3/4

Word Gloss laranjeiras e mangeira

Word Cat. Irreg-na N N N Conj N

Free He has orange trees, banana trees, lemon trees and mango trees.

4 Word Ikkalipo minga m'munda

Morphemes i- kha- li po mi- nga m'- mu- nda

Lex. Entries i- kha- li ipo mi- nga₂ mu-₁ mu-₂ nda

Lex. Gloss 4+5+9 IMP ser ele 4 espinho 18 3 machamba

Lex. Gram. Info. V:(obj)/sbj V:(mba tam)/tam1/neg/rel tam1 Irreg-li Pro N:ncl N 3/4 N:dblnc N:ncl N 3/4

Word Gloss esoinho na machamba

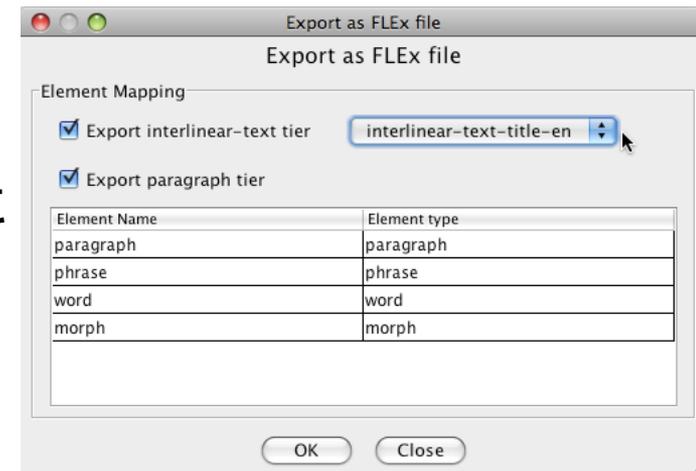


- FLEX import

	00.000	00:00:00.500	00:00:01.000	00:00:01.500	00:00:02.000	0						
interlinear-text-title-en [1]	In the Garden											
interlinear-text-comment-en [1]	"M'munda" ("In the Garden") a language learning text, (c) Luis Nyazeze, used by permission.											
interlinear-text-comment-pt [1]	"M'munda" ("In the Garden") a language learning text, (c) Luis Nyazeze, used by permission.											
interlinear-text-title-pt [1]	Na Machamba											
interlinear-text-title-seh [1]	M'munda											
paragraph [7]	1											
A_phrase-segnum-en [7]	My father has a large garden.					2						
A_phrase-gls-en [7]	My father has a large garden.					There he has many trees						
A_word-txt-seh [39]	Babanga	ana	munda	ukulu	.	Mwenemo						
A_morph-txt-seh [66]	^0-	baba	nga	a-	n	-a	mu-	nda	u-	kulu		mwenemo
A_morph-cf-seh [66]	mu-	baba	nga	a-	n	-a	mu-	nda	u-	kulu		mwenemo
A_morph-gls-pt [66]	1	pai	1S	3S+1	ter	IND	3	machamb	3	grande		aquele mesmo
A_morph-hn-en [33]	2		5	2	1	2		3				
A_morph-msa-pt [66]	N:ncl	N 1/2	Poss	V:subj	Irreg-n	V:vf	N:ncl	N 3/4	Poss:poss	Adj		dem6
A_morph-type [66]	prefix	root	root	prefix	root	suffix	prefix	root	prefix	stem		stem
A_word-gls-pt [29]	pai meu					machamba		grande				lá
A_word-pos-pt [30]	N		Irreg-na			N		Adj				dem6



- FLEx export
 - export to .flextext format
 - (currently) assumes a certain tier naming convention (as produced by the import function)
 - includes time and speaker information
 - retains FLEx id's (guid's) where possible
 - aims at minimal loss round-trip
 - to do: provide means for specifying tier(type)-to-element mapping





- Plug in mechanism for audio and video recognizer components (-> AVATeCH)
- Metadata in CMDI
- Recognizers apply algorithms to produce annotations

- Some available as a web service
- Call web services from ELAN
 - WebLicht (in progress)
 - WebMAUS
 - TypeCraft



- In preparation
- Text oriented, for (morphological) parsing, glossing, tagging
- Interface to connect with modules for (semi-automatic) processing of text - Lexan suggestion modules:
 - analyzers - recognizers - annotyzers
- Extensible, API for suggestion modules