

Improving the exploitation of linguistic annotations in ELAN



Onno Crasborn

Centre for Language Studies Radboud University Nijmegen

Han Slöetjes

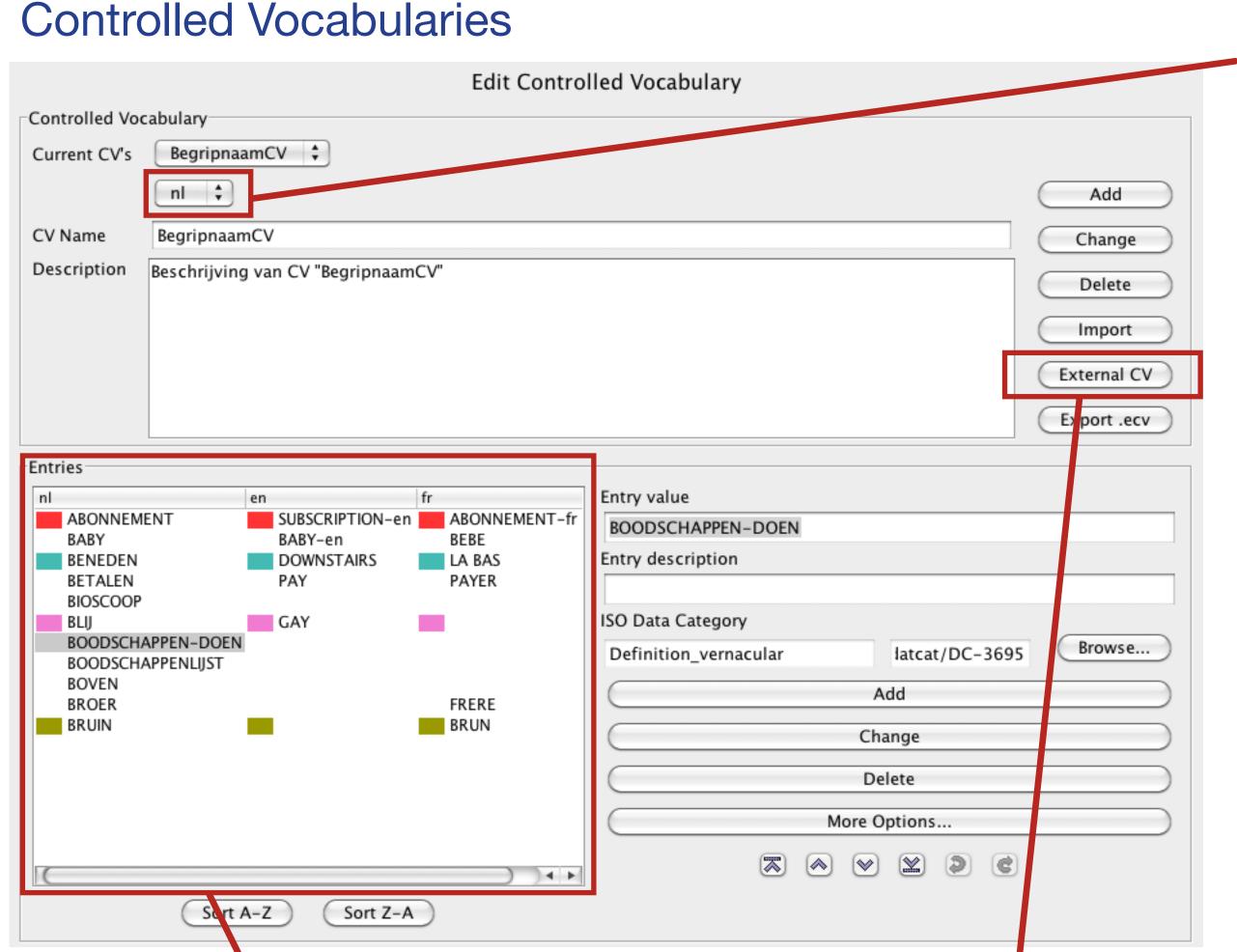
Max Planck Institute for Psycholinguistics Nijmegen





Two major enhancements to the multimodal annotation tool ELAN

- Support for multilingual content to enhance research in multilingual contexts. Starting with multilingual controlled vocabularies; metadata and data categories will follow Autumn 2014 release
 - [•] Multi-layered concordances; improved query formulation; more intuitive visualisation of search results and tier



Languages can be added to a vocabulary.

Languages can be selected from the ISO 639-3 list of languages (where possible).

Autumn 2014 release A display language for multilingual content can be selected in ELAN. Annotations, tooltips and metadata are automatically updated when the display language is changed.

Edit languages of Controlled Vocabulary
Languages
Languages available in CV:
Dutch Sign Language (dse) – dse – http://cdb.iso.org/lg/CDB-00134
Edit the language identification:
tch Sign Language (dse) – dse – http://cdb.iso.org/lg/CDB-00134158-001
Dusner (dsn) – dsn – http://cdb.iso.org/lg/CDB-00134163-001
Dusun Deyah (dun) - dun - http://cdb.iso.org/lg/CDB-00134130-001
Dusun Malang (duq) - duq - http://cdb.iso.org/lg/CDB-00134127-001
Dusun Witu (duw) – duw – http://cdb.iso.org/lg/CDB-00134131-001
Dutch (nld) – nld – http://cdb.iso.org/lg/CDB-00138580-001
Dutch Sign Language (dse) - dse - http://cdb.iso.org/lg/CDB-00134158-0
Dutton World Speedwords (dws) – dws – http://cdb.iso.org/lg/CDB-00138
Duungooma (dux) – dux – http://cdb.iso.org/lg/CDB-00134121-001

Language Resource Ontology Lexicography	articulatoryOrientation	http://www.isocat. org/datcat/DC-28 03	Sign Language	Description
Language Codes Terminology Multilingual Information Manager Lexical Resources Lexical Semantics Translation	audienceKnownToInterpre ter	http://www.isocat. org/datcat/DC-47 45	Sign Language	Rotation of the active articlator in terms of the rotation of the forearm (on
	audienceNotPresent	http://www.isocat. org/datcat/DC-47 44	Sign Language	dimension). Sometimes us as a synonym for palm up (supine) and palm down
Sign Language Dialogue Acts	audiencePartlyKnownToInt erpreter	http://www.isocat. org/datcat/DC-47 46	Sign Language	(prone).
	auxiliarySignLanguage	http://www.isocat. org/datcat/DC-46 74	Sign Language	
	bilingual	http://www.isocat. org/datcat/DC-47 68	Sign Language	
	bilingualHomeProgramme	http://www.isocat. org/datcat/DC-47 57	Sign Language	
	boardingSchool	http://www.isocat.	Sign Language	

Entries in a vocabulary can contain values for multiple languages...

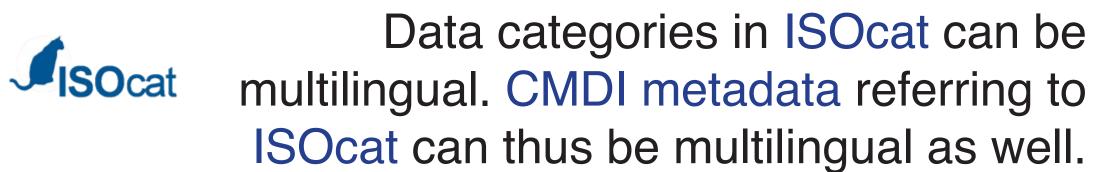
...also in external vocabularies.

Structured Search Multiple eaf: variable match

000	Search eaf files	
	Substring Search Single Layer Search Multiple La	ayer Search
Domain: 99 eaf files		Define Domain
Query History: <	> New Query	Save query Load query
Mode: case insensitive	variable match	\$ Clear
Minimal Duration	Maximal Duration Begin After End Before)
X	= 0 ann. • Y	All Tiers 🔻
Overlap 🔻	▼	Must be in same file 🔹
X		Tier Type: phonetic_transcription
Find	le query Fewer Columns More Colu	Imns Fewer Layers More Layers

Construct a query using variables: find identical or different annotations on the same or on different tiers without specifying a particular value.

Current release Various new options in the Multiple Layer Search allow for queries using variables. New presentation options provide a



Alignment view

	● ● ● Search	n eaf files
The query	Substring Search Single Lay	er Search Multiple Layer Search
	Domain: 292 eaf files	Define Domain
can be	Query History: < > N	New Query Save query Load query
hidden to		Save query Load query
dienlav	Mode: case insensitive \$ var	iable match Clear
display	Minimal Duration Maximal Duration	Begin After End Before
more		
search	X	Tier Type: gloss 🔻
	Overlap 💌	Must be in same file 👻
hits.	Х	Tier Type: gloss 👻
	Find Hide query Fewer Columns	More Columns Fewer Layers More La
	Found 65510 hits in 65510 annotations (of 132802)	Ready Cancel
	< > Hit 105 - 117 of 6551	.0 Save hits Time scale: 5 sec.
	00:05:28.789 00:00:05.000 CNGT0004_en	00:05:33.789 Lingui Partic Begin Time Durati
ne	GlossL S1 LAUGH-A	gloss S003 05:28.899 0.922
w for	GlossR S1 LAUGH-A	gloss S003 05:28.789 1.058
	CNGT0004_en	00:07:34.919
lew	GlossL S1 PALM-UP	gloss S003 05:29.919 0.344
de a	GlossR S1 PALM-UP 00:05:30.383 00:00:05.000	Show Concordance view Show Frequency view (by annotation)

better grip on the data.

Alignment view

<	> hit 7 - 9 of 11118	Time scale:	1 sec.	\$	
	00:01:04.960	00:00:01.000		00:01:05.960	
CNGT0001.eaf					Lingui An Partic Begin Time End Time Duratio
GlossL S1	PO		PO		gloss SG S003 01:05.040 01:07.200 2.160
GlossR S1	PO				gloss SG S003 01:04.960 01:05.560 0.600
	00:01:05.680	00:00:01.000		00:01:06.680	
CNGT0001.eaf	-				
GlossL S1	PO		KLAAR-A		gloss SG S003 01:05.680 01:08.818 3.138
GlossR S2	PO				gloss SG S004 01:06.060 01:06.339 0.279
	00:01:05.640	00:00:01.000		00:01:06.640	
CNGT0001.eaf					
GlossL S1	РО		KLAAR-A		gloss SG S003 01:05.680 01:08.818 3.138
GlossR S1	РО				gloss SG S003 01:05.640 01:07.240 1.600

Show Frequency view (by frequency) CNGT0004_en GlossL S1 COSY Show hit in transcription GlossR S1 COSY Hide info balloons 00:05:30.677 CNGT0004_en Linguistic Type Visible columns GlossL S1 Hide alignment times Annotator GlossR S1 ✓ Participant Save hits ✓ Begin Time 00:00:10.253 Save hit statistics CNGT0004_en End Time gloss S004 10.253 0.962 gloss S004 10.253 0.962 GlossL S2 ✓ Duration

Columns with tier and annotation properties can be individually hidden or displayed.

Blue bars show the alignment of annotations in the query result with respect to each other, as well as their duration.

Acknowledgments

GlossR S2

The work presented here was funded by CLARIN-NL, projects MultiCon and EXILSEA.

