

Using strategic annotations to add value to, and extract value from, a signed language corpus

Trevor Johnston
Centre for Language Sciences
Department of Linguistics
Macquarie University
Sydney, Australia

CISL Summer School, Charles University, Prague
25-27 August 2014

Abstract

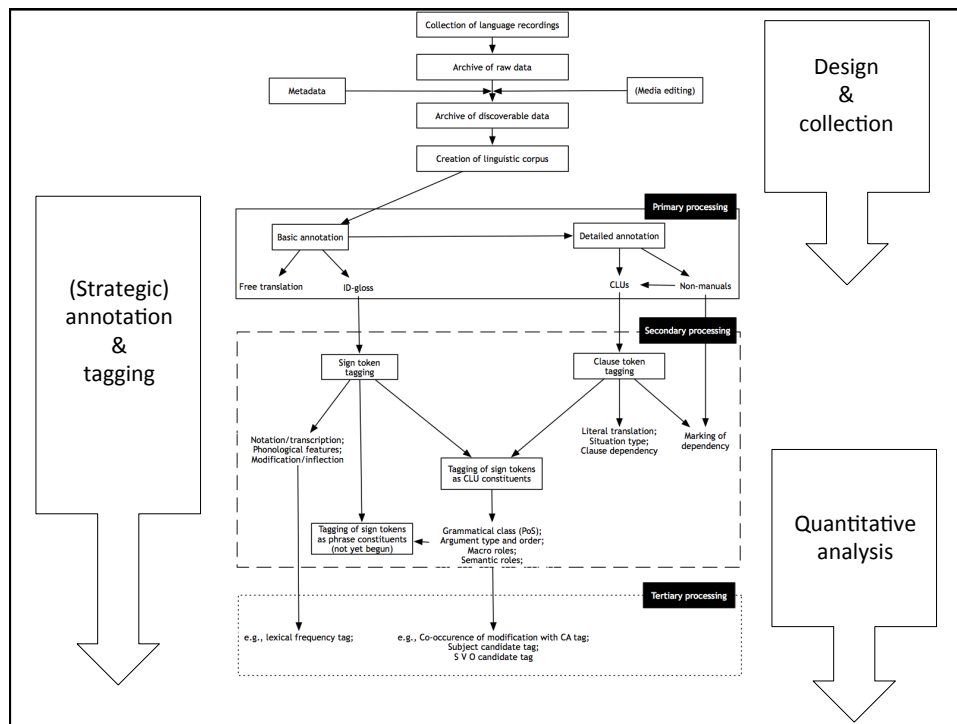
This lecture I explain the ways in which multimedia annotation software is being used to transform an archive of Auslan recordings into a true machine-readable linguistic corpus. After the basic structure of the annotation files in the Auslan corpus is described and the exercise differentiated from transcription, then the glossing and annotation conventions are explained. Following this the searching and pattern-matching at different levels of linguistic organization these annotations make possible is exemplified.

I explain why it is important to be clear about the difference between transcription and annotation otherwise at the end of the exercise—despite time consuming and expensive processing of the video recordings—we may not actually be able to discern the types of patterns in our corpora that we had hoped we could. The conventions are designed to ensure that the annotations really do enable researchers to identify regularities at different levels of linguistic organization in the corpus and thus to test, or build on, existing descriptions of the language.

Outline

- Contextualization
 - SL corpora at CISE Summer School
 - the case for SL corpus linguistics
 - same basic concepts and principles
 - the Auslan Corpus
- Using strategic annotations
 - Primary processing
 - basic and detailed annotation
 - Secondary processing
 - sign tokens
 - clause tokens
 - Tertiary processing
 - adding corpus-derived information to the corpus
 - processing annotated data outside of ELAN
- Research using an annotated SL corpus
 - Single sign searches
 - single tier
 - multi-tiers
 - Multiple sign searches
 - single tier
 - multi-tiers
 - Other operations
 - Annotations from overlaps
 - Annotation overlap information
 - Preparation for quantitative and multivariate analysis (e.g., Rbrul)
- Conclusion

Contextualization



What do we want to do?

- empirically ground SL description
- validate previous research
- generate new observations
- document linguistic community
- create teaching/learning resources

Why do we want to do it?

The case for SL corpus linguistics

Characteristics

- SL using communities
 - minority communities
 - no real location
 - interrupted inter-generational transmission
 - few native signers
 - no written form
 - high number/% of adult L2 users
 - cf notion of 'socio-linguistic typology'

Consequences

- intuitions less useful
 - high degree of variation
 - norms 'less-established'
- ad-hoc glossing
 - idiosyncratic
 - token/type confusion
 - inaccessible primary data
- corpora needed not just to compensate, but to enable!

Why do we want to do it?

- no easily or commonly used written form
- lack of language documentation
 - cf. preservation
- language endangerment
 - cf. maintenance, revitalization
- limits to intuitions and introspection
- unique usage/acquisition environments
- difficult for learners to gain exposure

How do we do it?

- create language archives
 - i.e., documentary linguistics
- adopt a corpus-based approach
- value-add to language archives using
 - multi-media annotation software
 - annotation, not necessarily transcription
 - controlled/consistent glossing (ID-glosses)
 - systematic linguistic tagging
- open access for researchers and community
 - learners and teachers
 - peer review

The Auslan Corpus

As at mid-2014

- 256 signers (approx. 200 hours digital video)
- native or near-native (before age 7)
- >1100 clips of which 459 clips with some basic annotations
 - ID glossing, free translation
- 200 (7 hours) with more extensive annotations

Types of detailed annotations

- clause-like units
- grammatical class
- spatial and directional modification
- aspectual modification
- event and situation type (Aktionsart)
- lexical frequency
- constituency
 - argument position and macro- and semantic roles
- clause relationships
 - dependency, embedding

Some basic concepts & principles

Notation: more or less a symbol system

Transcription: more or less a writing system

Annotation: something appended to text

Tagging: codes appended to text

Transcription ≠ Annotation

- basic annotation vs detailed representation
 - does detailed representation only mean phonetic/phonological detail (“transcription”)?
 - “generation” (i.e. reproduction) not the aim
- morpho-syntax & conventional signs/constructions
 - type/token relationship
 - form/meaning relationship
 - annotation as tagging for assessing the systematic co-occurrence of certain potential/assumed morpho-syntactic markers with certain types of meanings

Glossing & annotation conventions

- The Auslan Corpus Annotation Guidelines
 - downloadable from www.auslan.org.au/about/corpus/
 - only mentioned here in broad and general terms
- This afternoon’s workshop
 - annotation discussed in more detail
 - examples of glossing
 - experimentation with a Czech SL sample
 - discussion of issues
- More information/justification
 - following references

Some previous work on SL transcription, SL lexicalization, and SL corpora

- Johnston, T. (2014). The reluctant oracle: Adding value to, and extracting value from, a signed signed language corpus through strategic annotations. *Corpora*, 9(2), pages not yet available.
- Johnston, T. (2013). Towards a comparative semiotics of pointing actions in signed and spoken languages. *Gesture*, 13(2), 109-142.
- Johnston, T., & Schembri, A. (2013). Corpus Analysis of Deaf Sign Languages. In C. Chapelle (Ed.), *Encyclopedia of Applied Linguistics*. London: Wiley-Blackwell.
- Johnston, T. (2013). Formational and functional characteristics of pointing signs in a corpus of Auslan (Australian sign language): are the data sufficient to posit a grammatical class of 'pronouns' in Auslan? *Corpus Linguistics and Linguistic Theory*, 9(1), 109-15.
- Johnston, T. (2012). Lexical frequency in signed languages. *Journal of Deaf Studies and Deaf Education*, 17(2), 163-193.
- Johnston, T., & Schembri, A. (2010). Variation, lexicalization and grammaticalization in signed languages. *Langage et Société*, 131(mars 2010), 19-35.
- Johnston, T. (2010). From archive to corpus: transcription and annotation in the creation of signed language corpora. *International Journal of Corpus Linguistics*, 15(1), 104-129.
- Johnston, T. (2008). Corpus linguistics and signed languages: no lemmata, no corpus. In O. Crasborn, E. Efthimiou, T. Hanke, E. D. Thoutenhoofd & I. Zwitserlood (Eds.), *Proceedings of the Sixth International Language Representation and Evaluation Conference (3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Signed Language Corpora)* (pp. 82-87). Marrakech, Morocco, May 26-June 1.
- Johnston, T., & Schembri, A. (2006). Issues in the creation of a digital archive of a signed language. In L. Barwick & N. Thieberger (Eds.), *Sustainable data from digital fieldwork: Proceedings of the conference held at the University of Sydney, 4-6 December 2006* (pp. 7-16). Sydney: Sydney University Press.
- Johnston, T. (2001). The lexical database of Auslan (Australian Sign Language). *Sign Language & Linguistics*, 4(1/2), 145-169.
- Johnston, T., & Schembri, A. (1999). On defining lexeme in a sign language. *Sign Language & Linguistics*, 2(1), 115-185.
- Johnston, T. (1991). Transcription and glossing of sign language texts: Examples from Auslan (Australian Sign Language). *International Journal of Sign Linguistics*, 2(1), 3-28.

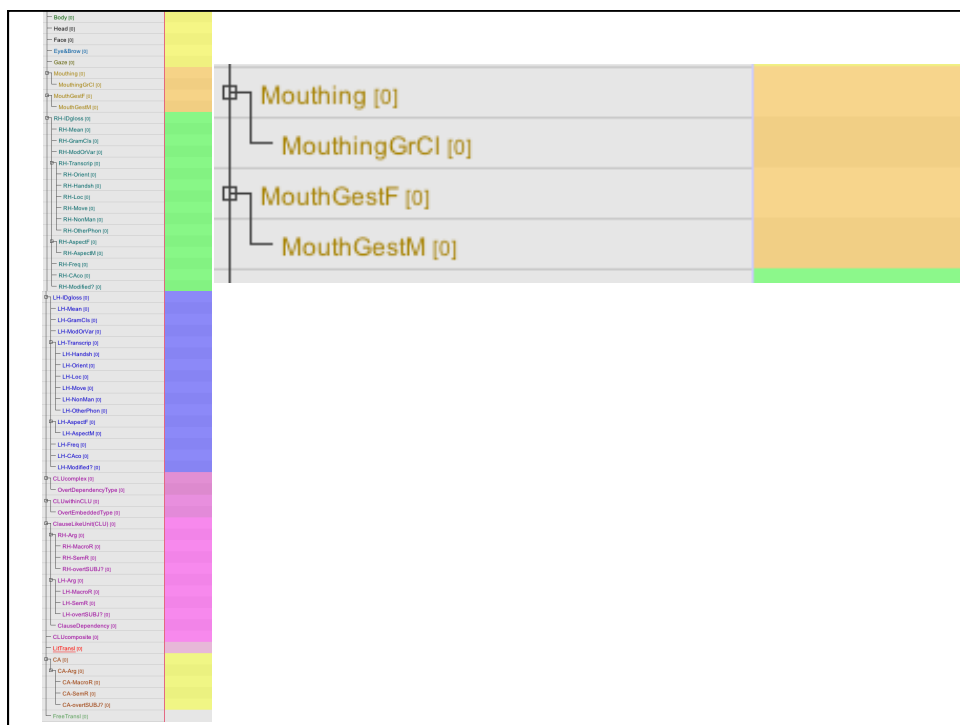
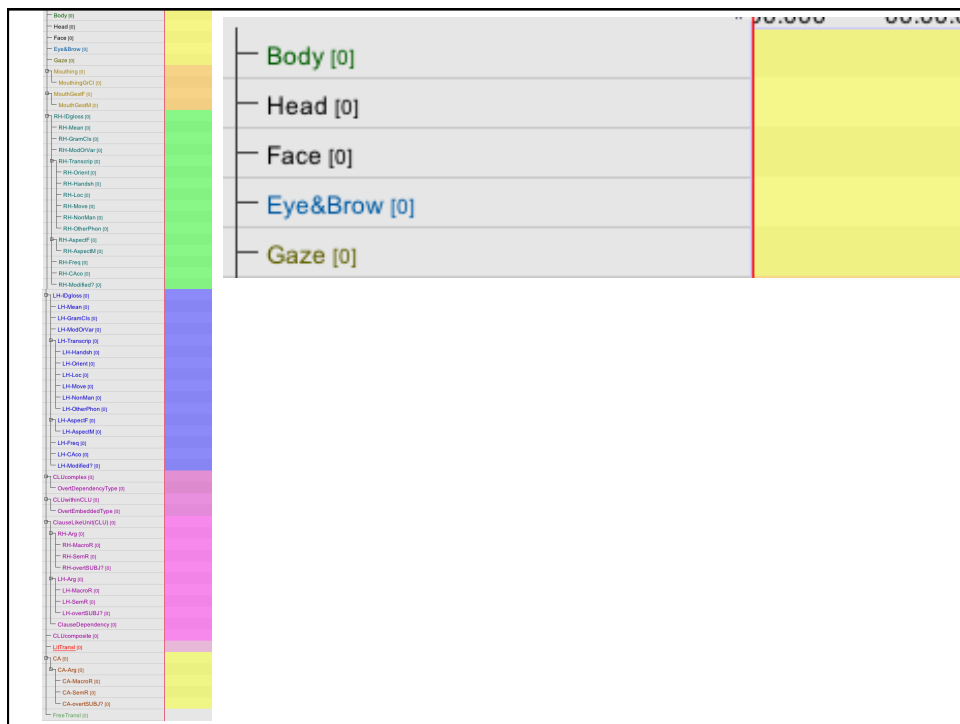
The screenshot shows the Elan software interface with a video of a signer on the left and a list of glosses on the right. The glosses are organized into a table with columns for different linguistic levels. The selected time range is 00:00:00.945 - 00:00:01.380.

Level	Gloss
LOOK	G-GO-AWAY · DS(B):TURTLE-MOVING · RABBIT · DS(B):HARE-RUNNING · TIRED(BENT5) · SLEEPY · TIRED(BENT5) · SLEEPY · LOOK · BIG(BENT7) · TREE · AREA · GOOD · PT:LOC · GOOD · DS(H):HARE-RUNNING · DS(2):HARE-LIE-DOWN · SLEEP-2H · FAR · SLEEP-2H · FINISH-FIVE · TURTLE · DS(B):TURTLE-MOVING · DS(1):PATH-OF-TURTLE · PT:DET · RABBIT · SLEEP-2H · AGES · PT:PRO3SG · INSTANT · NOTHING · AGES · PT:DET · TURTLE · GOOD · DS(B):TURTLE-MOVING · ARRIVE · APPROACH · PT:DET · RABBIT · SLEEP-2H · WAKE · G(CA):STRETCHING · THINK · RIGHT · G-HOLD-ON-THERE · RIGHT · PT:PRO1SG(7) · LONG · SLEEP-2H · RIGHT · THINK · FRIEND · WHERE · TURTLE · WHERE · GET-UP · LOOK ·

The timeline at the bottom shows the following glosses for the selected time range:

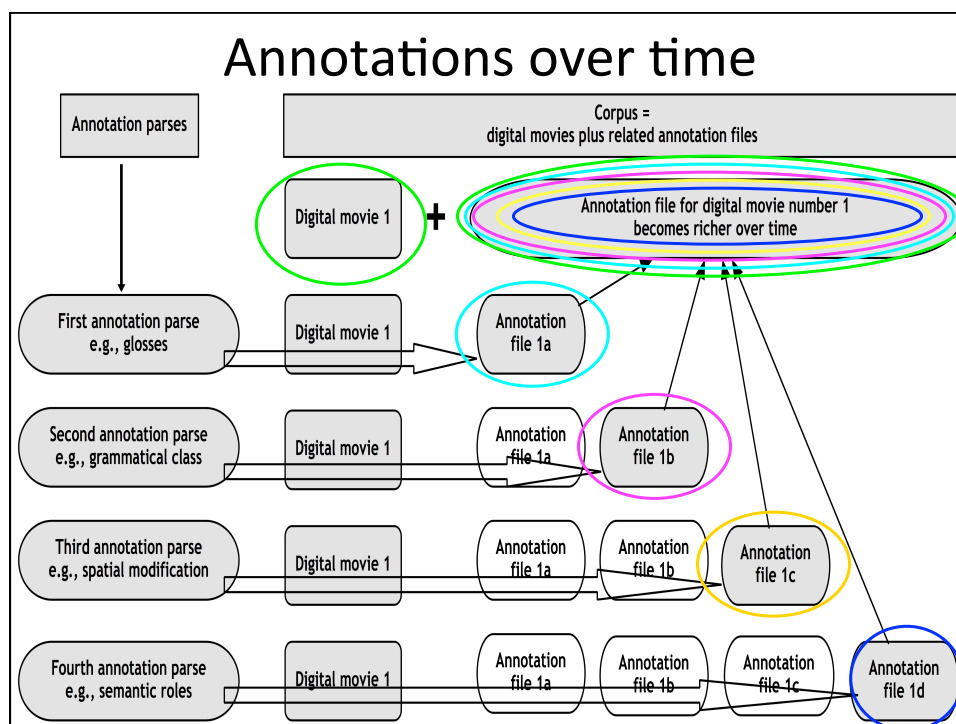
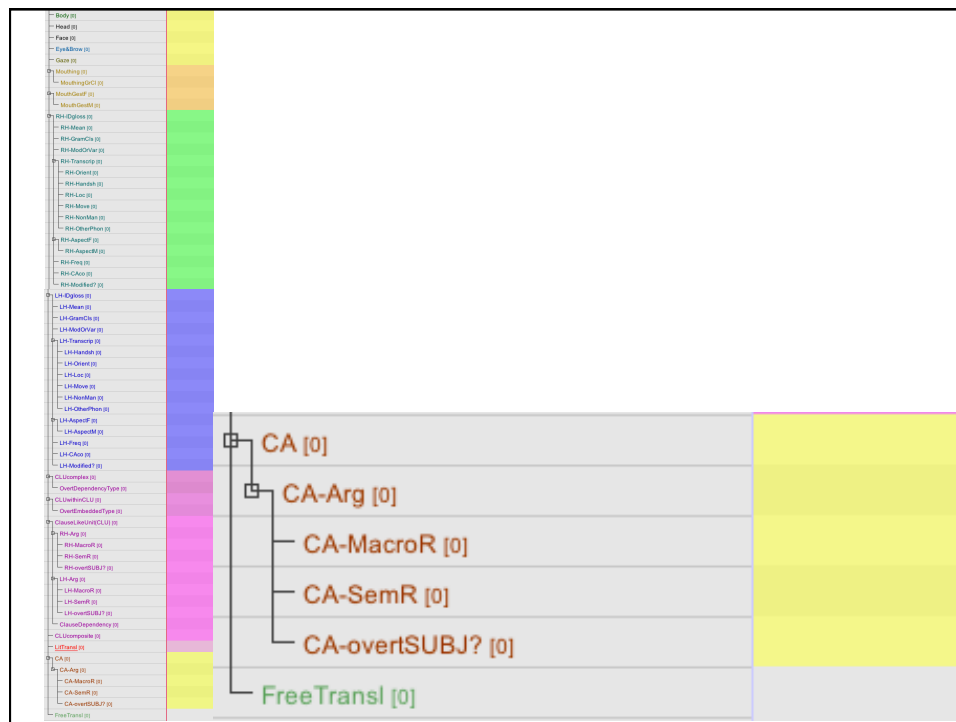
Time	Gloss
00:01:56.500	PT:DET
00:01:57.000	RABBIT
00:01:57.500	SLEEP-2H
00:01:58.000	AGES
00:01:58.500	PT:PRO
00:01:59.000	INST

The FreeTransl track shows the following text: "The rabbit intended on having only a short nap but it turned into a very long deep sleep."



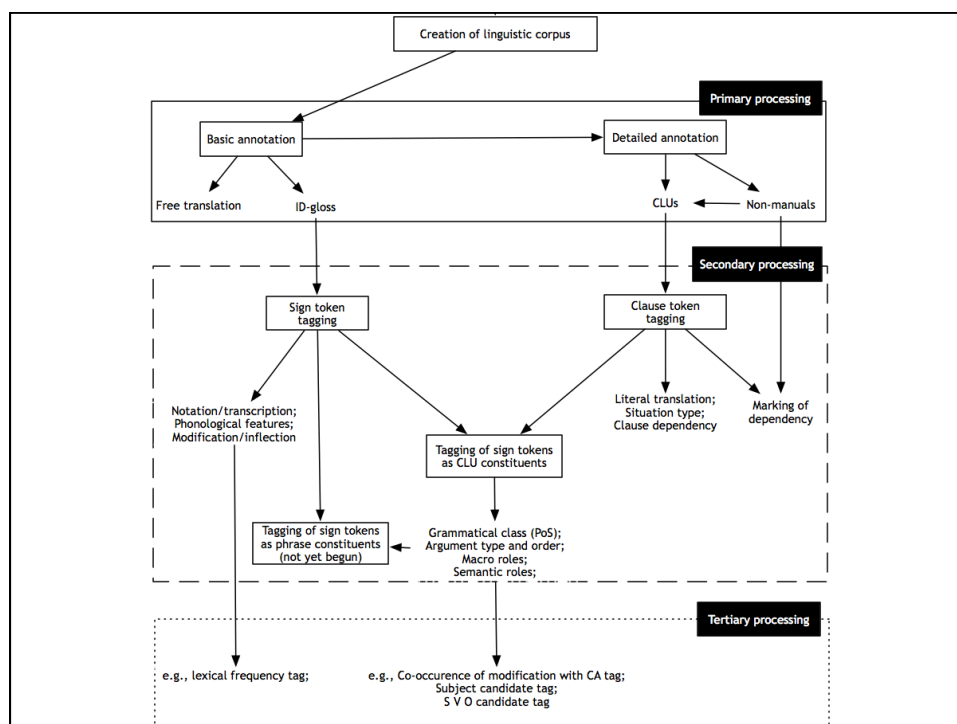
Body [0]		
Head [0]		
Face [0]		
Equilibr [0]		
Quar [0]		
Mouth [0]		
MouthGloss [0]		
MouthGloss [0]		
MouthGloss [0]		
RH-IDgloss [0]		
RH-Mean [0]		
RH-GramCls [0]		
RH-ModOrVar [0]		
RH-Transcrip [0]		
RH-Orient [0]		
RH-Handsh [0]		
RH-Loc [0]		
RH-Move [0]		
RH-NonMan [0]		
RH-OtherPhon [0]		
RH-AspectF [0]		
RH-AspectM [0]		
RH-Freq [0]		
RH-Caco [0]		
RH-Modified? [0]		
LH-IDgloss [0]		
LH-Mean [0]		
LH-GramCls [0]		
LH-ModOrVar [0]		
LH-Transcrip [0]		
LH-Orient [0]		
LH-Handsh [0]		
LH-Loc [0]		
LH-Move [0]		
LH-NonMan [0]		
LH-OtherPhon [0]		
LH-AspectF [0]		
LH-AspectM [0]		
LH-Freq [0]		
LH-Caco [0]		
LH-Modified? [0]		
CLUcomplex [0]		
OvertDependencyType [0]		
CLUwithinCLU [0]		
OvertEmbeddedType [0]		
ClauseLikeUnit(CLU) [0]		
RH-Arg [0]		
RH-MacroR [0]		
RH-SemR [0]		
RH-overtSUBJ? [0]		
LH-Arg [0]		
LH-MacroR [0]		
LH-SemR [0]		
LH-overtSUBJ? [0]		
ClauseDependency [0]		
CLUcomposite [0]		
LiTransl [0]		
CA [0]		
CA-Arg [0]		
CA-MacroR [0]		
CA-SemR [0]		
CA-overtSUBJ? [0]		
FinalTransl [0]		

Body [0]		
Head [0]		
Face [0]		
Equilibr [0]		
Quar [0]		
Mouth [0]		
MouthGloss [0]		
MouthGloss [0]		
MouthGloss [0]		
RH-IDgloss [0]		
RH-Mean [0]		
RH-GramCls [0]		
RH-ModOrVar [0]		
RH-Transcrip [0]		
RH-Orient [0]		
RH-Handsh [0]		
RH-Loc [0]		
RH-Move [0]		
RH-NonMan [0]		
RH-OtherPhon [0]		
RH-AspectF [0]		
RH-AspectM [0]		
RH-Freq [0]		
RH-Caco [0]		
RH-Modified? [0]		
LH-IDgloss [0]		
LH-Mean [0]		
LH-GramCls [0]		
LH-ModOrVar [0]		
LH-Transcrip [0]		
LH-Orient [0]		
LH-Handsh [0]		
LH-Loc [0]		
LH-Move [0]		
LH-NonMan [0]		
LH-OtherPhon [0]		
LH-AspectF [0]		
LH-AspectM [0]		
LH-Freq [0]		
LH-Caco [0]		
LH-Modified? [0]		
CLUcomplex [0]		
OvertDependencyType [0]		
CLUwithinCLU [0]		
OvertEmbeddedType [0]		
ClauseLikeUnit(CLU) [0]		
RH-Arg [0]		
RH-MacroR [0]		
RH-SemR [0]		
RH-overtSUBJ? [0]		
LH-Arg [0]		
LH-MacroR [0]		
LH-SemR [0]		
LH-overtSUBJ? [0]		
ClauseDependency [0]		
CLUcomposite [0]		
LiTransl [0]		
CA [0]		
CA-Arg [0]		
CA-MacroR [0]		
CA-SemR [0]		
CA-overtSUBJ? [0]		
FinalTransl [0]		



Using strategic annotations

in the creation of a linguistic corpus



Primary processing

Basic

- Free translation
 - written (preferable as minimum)
 - spoken is a bonus (potentially quicker), but
 - written is immediately searchable
- Tokenization of the signing stream
 - identify and gloss
 - discriminate types of signs

Additional possible detailed annotations

- Units larger than the individual sign
 - clause-like units (CLUs)
- Identified (provisionally) through
 - delivery (articulatory units)
 - meaning (coherent proposition or 'move' in exchange)
 - form/structure (constructional schema)
- Non-manual features (*relates to Delivery*)
 - body posture, head movements, gaze, facial expressions, mouthing, mouth gestures

Primary processing

IDglosses & free translation

The screenshot shows the Elan software interface for processing signing data. The window title is "Elan - SLRB1c2b.eaf". The interface includes a video player on the left showing a signer, a list of IDglosses on the right, and a timeline at the bottom.

IDglosses:

- LOOK · G:GO-AWAY · DS(B):TURTLE-MOVING · RABBIT ·
- DS(B):HARE-RUNNING · TIRED(BENT5) · SLEEPY · TIRED(BENT5) · SLEEPY
- LOOK · BIG(BENT7) · TREE · AREA · GOOD · PT:LOC · GOOD ·
- DS(H):HARE-RUNNING · DS(2):HARE-LIE-DOWN · SLEEP-2H · FAR ·
- SLEEP-2H · FINISH-FIVE · TURTLE · DS(B):TURTLE-MOVING ·
- DS(1):PATH-OF-TURTLE · PT:DET · RABBIT · SLEEP-2H · AGES
- PT:PRO3SG · INSTANT · NOTHING · AGES · PT:DET · TURTLE · GOOD ·
- DS(B):TURTLE-MOVING · ARRIVE · APPROACH · PT:DET · RABBIT ·
- SLEEP-2H · WAKE · G(CA):STRETCHING · THINK · RIGHT ·
- G:HOLD-ON-THERE · RIGHT · PT:PRO1SG(7) · LONG · SLEEP-2H · RIGHT ·
- THINK · FRIEND · WHERE · TURTLE · WHERE · GET-UP · LOOK ·

Timeline:

Time	RH-IDgloss [250]	LH-IDgloss [232]	FreeTransl [40]
00:01:56.500	PT:DET	RABBIT	The rabbit intended on having only a short nap but it turned into a very long deep sleep.
00:01:57.000	SLEEP-2H	SLEEP-2H	
00:01:57.500	AGES	AGES	
00:01:58.000	PT:PRO		
00:01:59.500	INST		

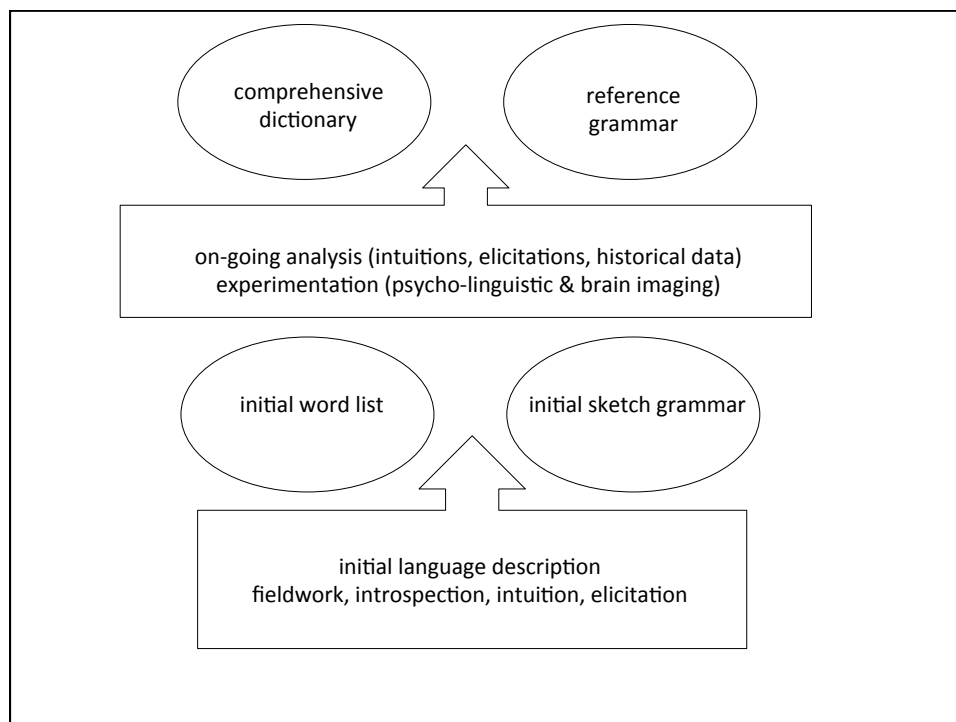
Basic glossing conventions for different types of signs

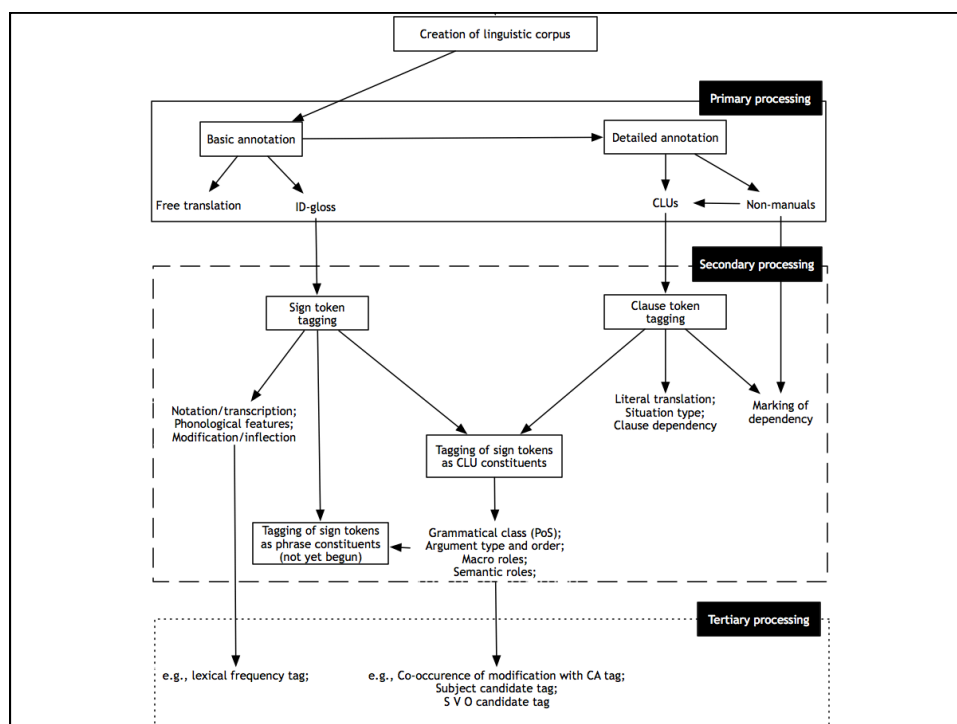
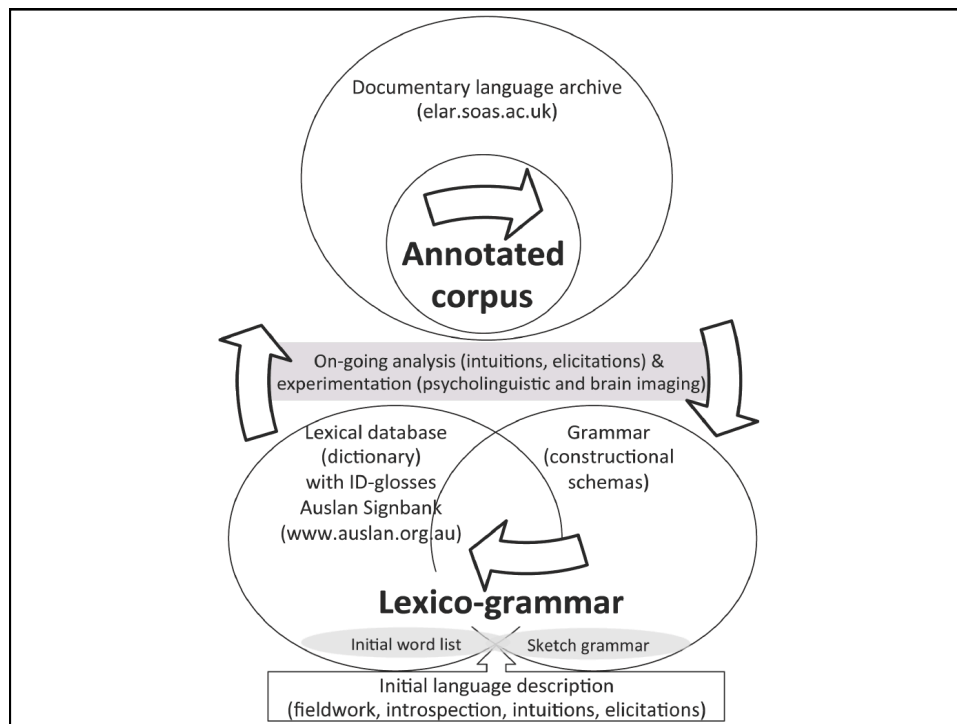
Sign type	Example Gloss	Type component	Type-like component	Token-like component
Fully lexical	WATER	WATER	n/a	n/a
Partly lexical	PT:PRO1SG	PT:	PRO1SG	n/a
	DSM(1-VERT):PERSON-GO-DOWN-HILL	DS	M(1-VERT):	PERSON-GO-DOWN-HILL
Non-lexical	G(5-UPWARDS):WELL	G	(5-UPWARDS):	WELL
	G:FLAILING-ABOUT-IN-WATER	G:	n/a	FLAILING-ABOUT-IN-WATER

Annotation conventions for different types of signs and for different purposes will be discussed this afternoon in the workshop talk, but for now....

How does one gloss fully lexical signs? By using the ID gloss.

How does one know what ID gloss to use? By consulting a dictionary of the language.





Secondary processing

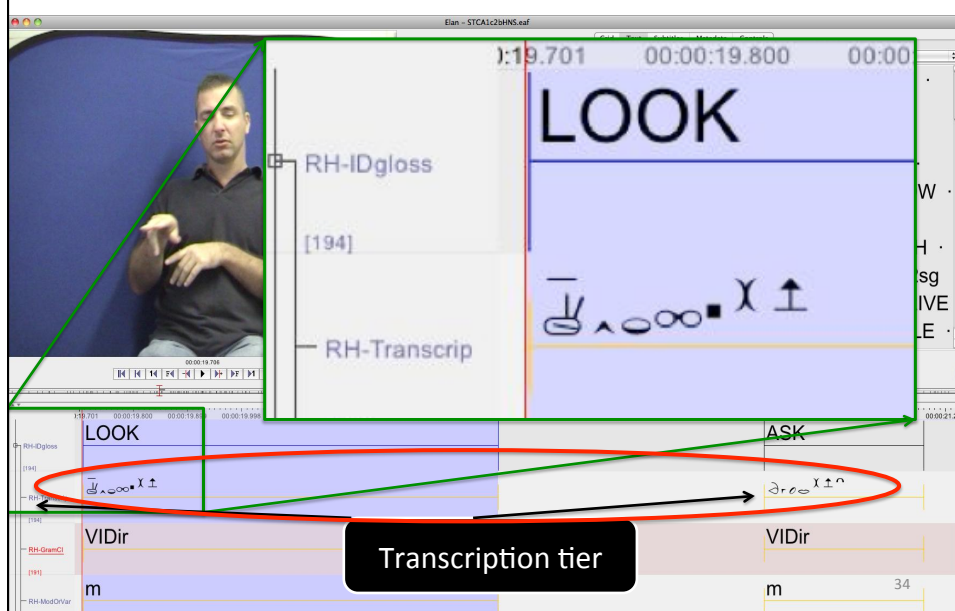
Sign token tagging

- transcription (with or without dedication notation system)
- phonological features (standard parameter model)
- morphological features
 - direction, location etc.

CLU token tagging

- Literal translation
- Situation type
 - state, accomplishment, achievement, activity
- CLU dependency status
 - independent, dependent, embedded
 - manner of marking of dependency/relationship
 - lexis, intonation contour, none (juxtaposition & context)

Phonological information (e.g., form notation)



Grammatical modification tag

The screenshot shows the ELAN software interface. At the top, there's a menu bar (File, Edit, Annotation, Tier, Type, Search, View, Options, Window, Help) and a toolbar. Below the menu bar, there's a video window showing a man speaking. To the right of the video, there's a list of tiers: 'RH ID gloss' (LOOK), 'RH mod' (m), 'RH-gram dis' (VIDir), and 'free t/ation' (The hare continued to laugh until suddenly the tortoise turned to look at him with distaste.). The 'RH mod' tier is highlighted with a red circle. Below the video, there's a timeline with various tiers. A black box labeled 'Modification tier' points to the 'm' tag in the 'RH mod' tier. The timeline also shows other tiers like 'RH ID gloss', 'RH-gram dis', 'LH ID gloss', 'LH-gram dis', and 'CAtrolshift'.

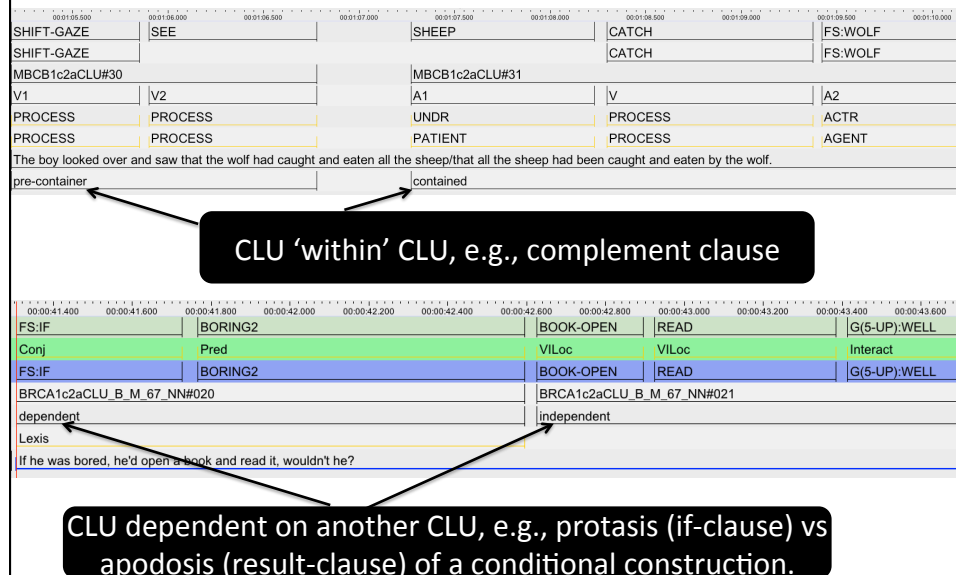
Literal and free translations

Literal translation ('CLU by CLU')

	00:00:29.500	00:00:30.000	00:00:30.500	00:00:31.000	00:00:31.500	00:00:32.000
RH ID gloss	FS:V(V)	SPRINT	DSM(5-HORI):MANY-HU	WHAT	PT:P	LAUGH
LH ID gloss	FS:V(V)	SPRINT	DSM(5-HORI):MANY-HU	FBUOY:DSM(5-HORI):MAN		LAUGH
CA	CA:VILLAGERS			CD:VILLAGERS		CA:BOY
Causal-likeUnit(CLU)	BAOBB1c2aCLU#1			BAOBB1c2aCLU#2		BAOBB1c2aCLU#3
Transl	(the) village sprint move-uphill			(villagers) whatted		he laugh (at them)
FreeTransl	The villagers raced up the hill.			When the villagers arrived and wondered what the fuss was about, the boy just laughed at them.		

Free translation (English-like chunks)

CLU relationships



Secondary processing (cont.)

- Tagging of sign tokens as CLU constituents
 - Grammatical class (PoS)
 - Argument type and order
 - Macro roles
 - Semantic roles
- Tagging of sign tokens as phrase constituents
 - Not attempted yet

Grammatical class tag

Grammatical class tag

The screenshot displays the ELAN software interface for linguistic analysis. The top window shows a video of a woman speaking. Below the video is a list of time-coded utterances with their corresponding time intervals. The bottom window shows a detailed grammatical analysis of the video, with a timeline and a list of grammatical tags. A red box highlights the 'VIDir' tag in the 'LH-Idgloss' list. Below the video, a timeline shows the same utterances with grammatical tags like 'AUSLAN', 'PERHA|GET-BETTER', 'CHECK-2H', 'FINISH FIVE', 'CHECK-2H', 'WANT', 'PT-PR', 'CHECK-2H'. A red box highlights the 'VIDir' tag in the 'LH-Idgloss' list. Below the video, a timeline shows the same utterances with grammatical tags like 'AUSLAN', 'PERHA|GET-BETTER', 'CHECK-2H', 'FINISH FIVE', 'CHECK-2H', 'WANT', 'PT-PR', 'CHECK-2H'. A red box highlights the 'VIDir' tag in the 'LH-Idgloss' list.

Argument order & role

Argument order & role

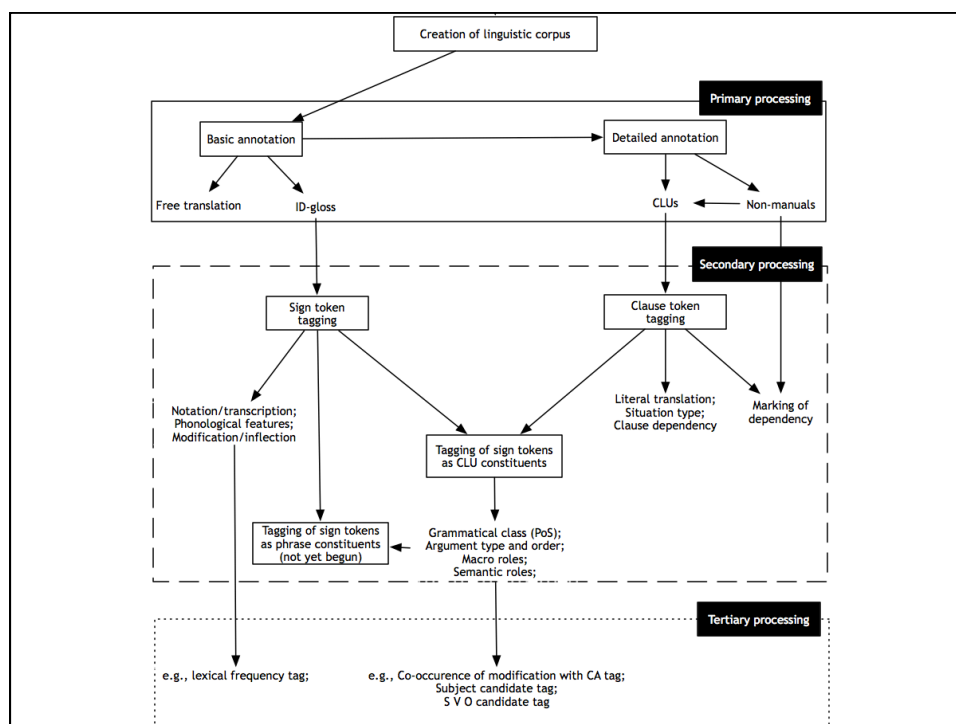
The screenshot displays the RH-Idgloss software interface. On the left, a small video window shows a woman speaking. The main area features a timeline from 00:00:01.695 to 00:08:00.000. Above the timeline, a text bar contains the sentence: "ONE · BOY · DSL(1)PERSON-THERE · HAVE · FS-VILLAGE · DSL(5)HOUSES-LOCATED-IN-A-GROUP · BOY · PT-DET · ALWAYS1 · MORNING · ALWAYS1 · NIGHT2 · DSM(C)MOVE-GROUP · SHEEP2 · DSH(C)MOVE-GROUP · FS:PASTURE · PT.LOC(B) · DINNER · FS:GRASS · DINNER · ALWAYS1 · NIGHT2 · DSH(5)MOVE-GROUP · DSH(5)MOVE-GROUP · DSH(5)MOVE-GROUP · DSH(5)MOVE-GROUP · BOY · PT.LOC · BORED · WHY · DINNER · ALL-DAY · MORNING · EARLY · UNTIL · LATE · NIGHT2 · DINNER · BOY · BORED · G-WHAT-CAN-I-DO · THINK · PLAY · FS:GAME · GOOD · G(NMS):CA · DINNER · BOY · PT.LOC · WHOLE · FS:WOLF · WOLF · PEOPLE/FALSE START · FS-VILLAGE". Below the timeline, several tiers are visible, including "RH-Idgloss [133]", "RH-GramCls [133]", "Clause [53]", "RH-Arg [130]", "RH-MacroR [84]", "RH-SemR [83]", "Clause-cp [53]", "ClauseAndRHArg [53]", "RHHargAndRHSemRoles [130]", and "CIRHargAndSemRoles [53]". A red circle highlights the "Clause 3" tier, which contains the roles "A1", "nonA", "nonA", and "nonA". An arrow points from the bottom left corner to this circle.

Argument order & role

cover the combined tiers

Tertiary processing

- creating new annotations based on existing annotations
- using the corpus to enrich the corpus by
 - incorporating results of searches, sorting and quantification based on primary and secondary processing into the corpus itself as new annotations
- both enable one to conduct subsequent searches using the newly incorporated annotations as given information or as new constraints



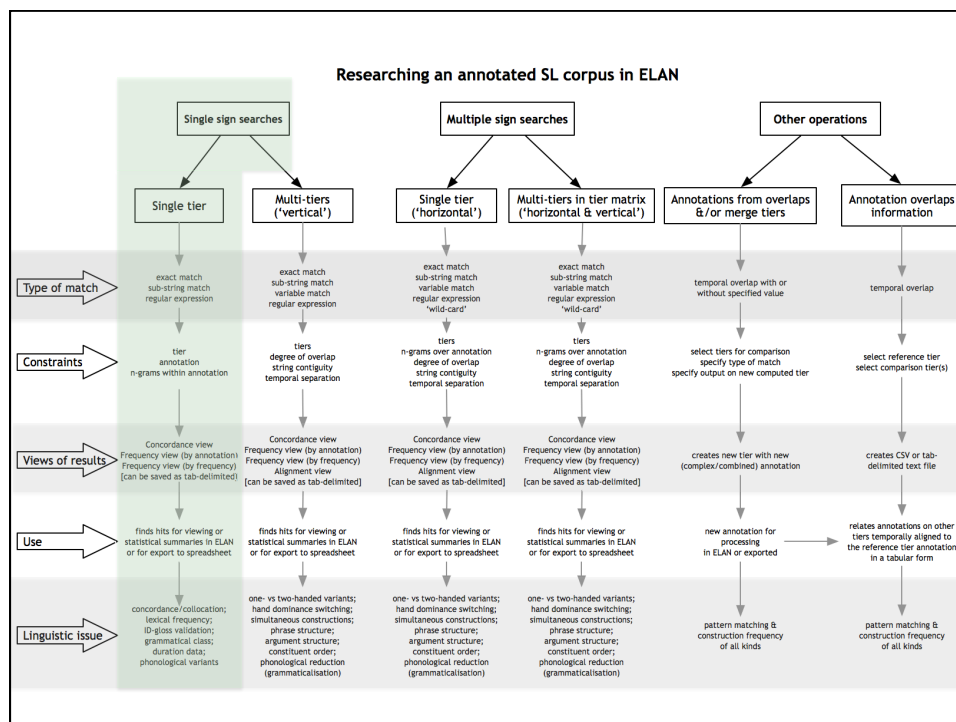
New CLU annotation containing core arguments and verb

The screenshot displays the RH-Arg gloss tool interface. On the left, a video of a woman speaking is shown. The main window displays the sentence: "ONE · BOY · DSL(1):PERSON-THERE · HAVE · FS:VILLAGE · DSL(5):HOUSES-LOCATED-IN-A-GROUP · BOY · PT:DET · ALWAYS1 · MORNING · ALWAYS1 · NIGHT2 · DSM(C):MOVE-GROUP · SHEEP2 · DSH(C):MOVE-GROUP · FS:PASTURE · PT:LOC(B) · DINNER · FS:GRASS · DINNER · ALWAYS1 · NIGHT2 · DSH(5):MOVE-GROUP · DSH(5):MOVE-GROUP · DSH(5):MOVE-GROUP · DSH(5):MOVE-GROUP · BOY · PT:LOC · BORED · WHY · DINNER · ALL-DAY · MORNING · EARLY · UNTIL · LATE · NIGHT2 · DINNER · BOY · BORED · G:WHAT-CAN-I-DO · THINK · PLAY · FS:GAME · GOOD · G(NM):CASE · DINNER · BOY · PT:LOC · YELLOW · PEOPLE:FALSE-START · NONSENSE". Below the sentence, a timeline shows the temporal structure of the utterance. The right panel displays the linguistic analysis, including the sentence structure, the temporal structure, and the semantic roles. A red circle highlights the "Clause-cp (copy) + RH-Arg" section, which shows the semantic roles of the clause: "A1 V A2", "A1 AGENT", "nonA", "nonA", and "nonA".

New CLU constituent annotation argument & semantic role

The screenshot displays the RH-Idgloss software interface. On the left, a video window shows a woman speaking. The main area features a timeline with various linguistic annotations. A black callout box with the text "Argument + Semantic Role" has arrows pointing to the "A1 AGENT" label in the timeline. The timeline includes labels such as "RH-IDgloss (133)", "RH-GramCls (133)", "Clause (53)", "RH-Arg (130)", "RH-MacroR (84)", "RH-SemR (83)", "Clause-cp (53)", "ClauseAndRHArg (53)", "RHArgAndRHSemRoles (130)", and "CIRHArgAndSemRoles (53)". The timeline itself shows segments for "BOY", "PT-DET", "ALWAYS1", "MORNI", "NP", "Adv", "nonA", "A1", "ACTR", "AGENT", "A1 v A2", "A1 AGENT", "nonA", "nonA", and "nonA".

Researching an annotated corpus



Single sign searches: single tier

Types of matches

- exact match, i.e., “specific string of characters”
- substring matches, i.e., “any part of a string”
- regular expressions, e.g.,
 - .+ = ‘any string of characters’ (i.e., any annotation)
 - ^ = ‘begins with’

Sign type frequency

Exploiting distinctions made possible through the implementation of IDglossing conventions

- search all IDgloss annotations on dominant hand (^, or .+, etc.)
- search for sub-type searches for partly- or non-lexical signs
 - points (^PT), depicting signs (^DS) , gestures (^G:), fingerspellings (^FS)
- view hits in frequency view for statistics
 - export and sort if desired or conduct sub-searches

Three views of search results in ELAN

Concordance	Frequency (annotation, percent, count)	Frequency by frequency (annotation, percent, count)
PAST GO-IN PAST GO-IN PT:DET GO-IN PT:DET COUNTRY PT:DET COUNTRY FAR COUNTRY FAR FS:FARM FAR FS:FARM HAVE FS:FARM HAVE ONE HAVE ONE BOY ONE BOY WHO BOY WHO ALWAYS	ABORIGINE 0.00% 3 ABORIGINE1 0.00% 1 ABORT 0.02% 23 ABOUT 0.00% 2 ABOUT-NUMBER 0.01% 9 ABOUT1 0.01% 6 ABOUT2 0.10% 97 ABOVE-AVERAGE 0.00% 1 ABSENT-MIND 0.00% 1 ABUSE 0.00% 2	PT:PROISG 3.64% 3376 G(S-UP):WELL 2.86% 2653 PT:PRO3SG 2.14% 1983 LOOK 1.09% 1009 SAME 1.06% 980 HAVE 1.05% 976 DEAF-AND-DUMB 0.98% 910 PT:LOC 0.92% 850 PT:DET 0.76% 702 BOY 0.74% 687

Substring Search Single Layer Search Multiple Layer Search

Domain: 459 eaf files Define Domain

Query History: < > New Query

Mode: Annotation case sensitive exact match

Find AUSLAN Tier Name: SH-IDgloss

Found 80 hits in 80 annotations (of 220923) Ready Cancel

hit 1 - 12 of 80 >

MOST SIGN AUSLAN HANDICAPPED REAL
KNOW SIGN AUSLAN SIGN PT:PRO2SG(7)
PT:PRO3SG NOT AUSLAN ENGLAND SIGN
G(5-TOWARDS):UMM TEACHER AUSLAN UNIT AREA
SIGN SAME AUSLAN PT:PRO3SG G(5-UP):WELL(7)
BELIEVE2 PT:PRO3SG AUSLAN CAN HELP1
PT:PRO3PL TALENT AUSLAN NOTHING TEACHER
SPEECH WITH1 AUSLAN G(5-UP):WELL CAN
DEAF HOUSE AUSLAN ALWAYS1 NOISE-2H
SIGN SORE AUSLAN HAVE-NOT-2H MISS
DS(C):TIGHT-GROUP SCHOOL AUSLAN KNOW-2H FS:INDECIPHERABLE
MUM FILL1 AUSLAN SIGN SAME

.+ on RH-IDgloss tier

Substring Search Single Layer Search Multiple Layer Search

Domain: 180 eaf files Define Domain

Query History: < >

Mode: Annotation case sensitive regular expression

Find .+ Tier Name: RH-IDgloss

#hits : 42605
 #annotations with a hit : 42605
 #annotations investigated : 138142 Ready

Frequency 1 - 20 of 9007

Annotation	Percentage	Count
PT:PRO3SG	3.89%	1659
G:WELL	2.09%	890
PT:PRO3SG	1.94%	827
DEAF	1.45%	617
LOOK	1.38%	588
BOY	1.17%	497
SAME	1.10%	467
PT:LOC	1.07%	454
HAVE	1.02%	434
PT:DET	0.88%	377
PT:PRO2SG	0.79%	335
THINK	0.77%	326
NOTHING	0.76%	322
GOOD	0.73%	311
WHAT	0.68%	291
WHY	0.66%	281
REAL	0.65%	278
NOT	0.65%	277
PEOPLE	0.62%	266
ONE	0.62%	263

Substring Search Single Layer Search Multiple Layer Search

Domain: 180 eaf files Define Domain

Query History: < >

Mode: Annotation case sensitive regular expression

Find +. Tier Name: RH-IDgloss

#hits : 42605
 #annotations with a hit : 42605
 #annotations investigated : 138142

Ready

frequency 1 - 20 of 8007

Annotation	Percentage	Count
PT:PRO1SG	3.89%	1659
G:WELL	2.09%	890
PT:PRO3SG	1.94%	827
DEAF	1.45%	617
LOOK	1.38%	588
BOY	1.17%	497
SAME	1.10%	467
PT:LOC	1.07%	454
HAVE	1.02%	434
PT:DET	0.88%	377
PT:PRO2SG	0.79%	335
THINK	0.77%	326
NOTHING	0.76%	322
GOOD	0.73%	311
WHAT	0.68%	291
WHY	0.66%	281
REAL	0.65%	278
NOT	0.65%	277
PEOPLE	0.62%	266
ONE	0.62%	263

Exploiting ID-glossing to generate a frequency list of fully lexical signs

Regular expression: `^[^\QPT\E|^\QDS\E|^\QFS\E|^\QG:\E|^\QG\(\E]`

Concordance view

Frequency by frequency, (i.e., lexical frequency list: annotation, percent, count)

PAST GO-IN
 PAST GO-IN PT:DET
 PT:DET COUNTRY FAR
 COUNTRY FAR FS:FARM
 FS:FARM HAVE ONE
 HAVE ONE BOY
 ONE BOY WHO
 BOY WHO ALWAYS1
 WHO ALWAYS1 TAKE-2H
 ALWAYS1 TAKE-2H PT:DET

LOOK	1.74%	1009
SAME	1.69%	980
HAVE	1.68%	976
DEAF-AND-DUMB	1.57%	910
BOY	1.18%	687
SAY	1.13%	656
THINK	1.12%	648
GOOD	1.10%	636
DEAF	1.09%	633
SIGN	1.07%	624

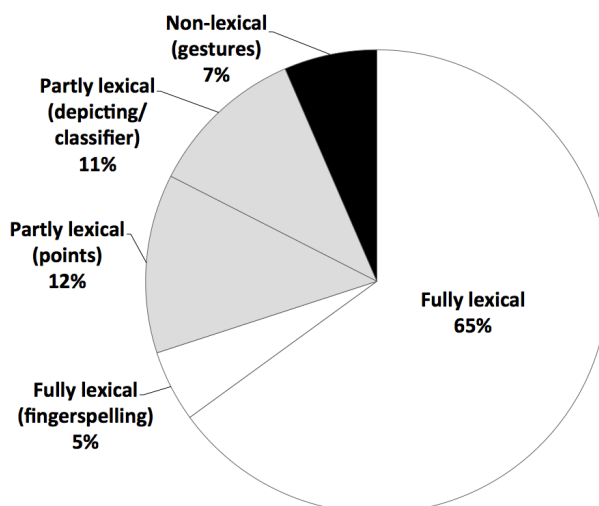
`^[^\QPT\E|^\QDS\E|^\QFS\E|^\QG:\E|^\QG\(\E]`

Exploiting ID-glossing to generate a frequency list of partly lexical signs

Regular expression: ^PT[^DS]		
Concordance view	Frequency list of partly lexical signs	
<i>GO-IN</i> PT:DET COUNTRY	PT:PRO1SG	17.19% 3376
<i>TAKE-2H</i> PT:DET SHEEP2	PT:PRO3SG	10.10% 1983
<i>AREA</i> PT:POSS3PL SHEEP2	PT:LOC	4.33% 850
<i>HOME-2H</i> PT:DET ONE	PT:DET	3.57% 702
<i>FS:DAY</i> PT:PRO3SG GET(SE)	PT:PRO2SG	3.22% 633
<i>FS:SO</i> PT:PRO3SG WANT	PT:PRO1SG(B)	2.54% 499
<i>WOLF</i> PT:PRO3SG FS:SO	PT:PRO1SG(7)	2.06% 404
<i>FS:SO</i> PT:PRO3SG YELL2	PT:PRO3PL	1.78% 349
<i>FROM</i> PT:DET FS:VILLAGE	PT:PRO3SG(7)	1.01% 199
<i>HEAR</i> DSM(G):MANY-HUMANS-RUN ONLY	PT:POSS1SG	0.97% 191

Note: the coding in parentheses at the end of the gloss indicates variant handshapes

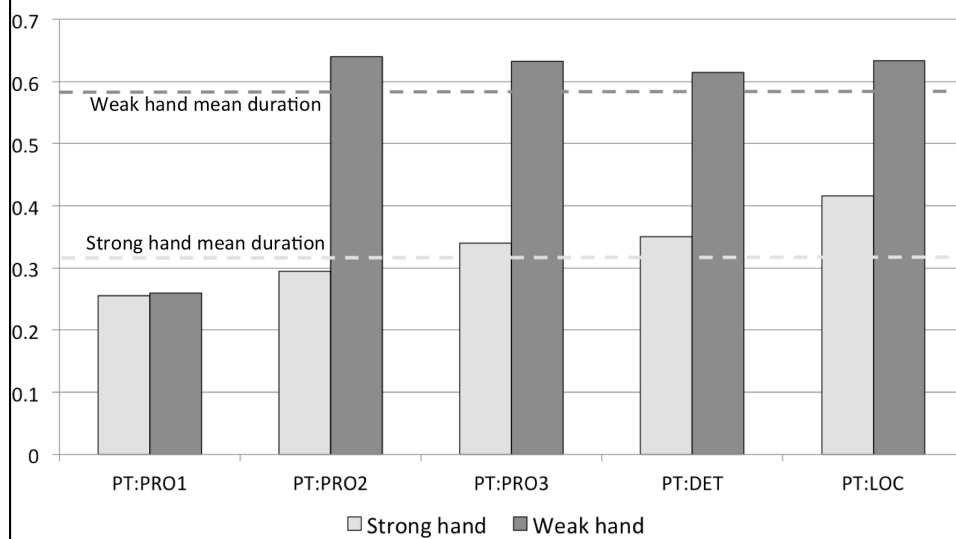
Distribution of sign types in corpus



Distribution of signs by type & text type

Sign type	Narrative	Interview	Retell	All texts
Lexical signs	64%	70%	61%	65%
Fingerspelling	6%	5%	5%	5%
Gestures	6%	9%	5%	7%
Points	16%	15%	7%	14%
Depicting signs	7%	2%	21%	4%

Average duration of pointing signs by strong and weak hand



Handshake variation

Domain: 281 caf files

Query History: < >

Mode: Annotation case insensitive regular expression

Find ^PT:PRO.+?() Tier Name: RH-IDgloss

#hits : 543
#annotations with a hit : 543
#annotations investigated : 160217

Ready

frequency 1 - 21 of 53

Annotation	Percentage	Count
PT:PRO1SG(B)	38.31%	208
PT:PRO1SG(5)	19.34%	105
PT:PRO1SG(7)	10.31%	56
PT:PRO1PL(2)	3.13%	17
PT:PRO3SG(7)	2.95%	16
PT:PRO1SG(BENTB)	2.95%	16
PT:PRO1SG(?)	2.39%	13
PT:PRO1SG(6)	1.84%	10
PT:PRO3SG(B)	1.29%	7
PT:PRO1SG(Y)	1.29%	7
PT:PRO1SG(F)	1.29%	7
PT:PRO3SG(5)	1.10%	6
PT:PRO1SG(B)-2H	0.92%	5
PT:PRO3PL(B)	0.74%	4
PT:PRO3PL(7)	0.74%	4
PT:PRO1SG(7)-2H	0.74%	4
PT:PRO1PL(B)	0.74%	4
PT:PRO3SG(FALSE-START)	0.55%	3
PT:PRO3PL(5)	0.55%	3
PT:PRO1SG(ILY)	0.55%	3
PT:PRO1SG(A)	0.55%	3

Translation equivalents used to locate types of data (1): conditionals

- How is conditional marked?
 - non-manual brow raise?
 - what order the clauses?
 - lexical 'if'
 - what order the clauses?
 - no marking
 - context only

Search for
\\bif\\b

Domain: 281 caf files

Query History: < >

Mode: Annotation case insensitive regular expression

Find \\bif\\b Tier Name: ProuTransl

#hits : 52
#annotations with a hit : 52
#annotations investigated : 160217

Ready

frequency 1 - 21 of 53

All the village people came to the pasture and looked at the boy asking him if he wanted their help.

So we have to remember, if someone is always willing him, you should.

Well really, it means the same as saying "If you go slowly and are patient, then you will make it."

It means that if you have patience, then you will be successful in achieving your goals.

And anyway, there's no point going if it takes you so long.

So he asked the hare if he wanted to race.

The hare asked if the hare would want to join him in a race to the end of the course.

The tortoise said he would be alright and the hare asked if he and the hare could have a bet on the race.

The hare asked if the hare was ready.

It ignored the rabbit and just kept on walking. A thought came to his mind. "If you race him, you could beat him," he told himself.

Off went again, looking OVER his shoulder to see he could see the tortoise. "If the tortoise has passed me then we're in a hurry. I can catch up and overtake him.

on so slow that it would be impossible for you to win," he mused. "But, if that is what you want, then I agree," he replied to the tortoise's invitation.

"It should be fine if I take an afternoon nap, so need to worry," he answered himself.

hate thoughtfully. "Is that what you think of me? Why don't we make a bet. If you want we can have a race to see who arrives first?" he proposed.

The hare looked back down at the tortoise and agreed. "Yes fine, if that's what you want. Great! Adios!" he replied to be good off.

The hare was angry and looked at the tortoise as if to say something. He realized that the tortoise had put him in his place and so

The tortoise then suggested OVER and asked me if I wanted to dance with him.

His friend was excited and came OVER to ask me if I wanted to dance with him.

We argued then he asked if I could give him my phone number.

He walked me to my car and asked if I would give him my phone number.

we number on a piece of paper. I was a bit embarrassed because I didn't know if he would be turned on by the lipstick.

Domain: 281 .caf files

Query History: < >

Mode: Annotation case insensitive regular expression

Find: \bif\b Tier Name: FreeTransl

#hits : 52
#annotations with a hit : 52
#annotations investigated : 160217

Ready

hit 1 - 22 of 52

All the village people came to the pasture and looked at the boy asking him if he wanted their help.
So we have to remember, if someone is always telling lies, you should...
Well really, it means the same as saying " If you go slowly and are patient, then you will make it."
It means that if you have patience, then you will be successful in achieving your goals.
And anyway, there's no point going if it takes you so long."
So he asked the hare if he wanted to race.
The hare asked if you(the tortoise) would want to join him in a race to the end of the course.
The tortoise said he would be alright and the hare asked if the tortoise would be ready.
en the tortoise was sitting, he drew the attention from the hare and asked him if he and the hare could have a bet on the race.
The tortoise asked if the hare was ready.
le ignored the rabbit and just kept on walking. A thought came to his mind, "If you race him, you could beat him." he told himself.
ff racing again, looking OVER his shoulder in case he could see the turtle. " If the turtle has passed me then not to worry. I can catch up and overtake him.
ou are so slow that it would be impossible for you to win." he mocked. "But, if that is what you want, then I agree," he replied to the tortoise's invitation.
"It should be fine if I take an afternoon nap, no need to worry." he convinced himself.
hare thoughtfully. "Is that what you think of me? Why don't we make a bet, if you want we can have a race to see who arrives first?" he proposed.
The hare looked back down at the tortoise and agreed. "Yes fine, if that's what you want. Great! Adios!" he replied as he sped off.
The hare was angry and looked at the tortoise as if to say something. He realised that the tortoise had put him in his place and se
The handsome man swaggered OVER and asked me if I wanted to dance with him.
His friend was excited and came OVER to ask me if I wanted to dance with him.
We argued then he asked if I want a drink
He walked me to my car and asked if I could give him my phone number.
re number on a piece of paper. I was a bit embarrassed because I didn't know if he would be turned on by the lipstick.

Translation equivalents used to locate types of data (2): negation

- Is negation of KNOW
 - suppletive negative incorporation (KNOW-NOT)?
 - lexical NEG KNOW (or KNOW NEG)?
 - non-manual, i.e. with headshake?

Search for
"n't know"

Domain: 281 .caf files

Query History: < >

Mode: N-gram within annotation case insensitive substring match

Find: n't know Tier Name: FreeTransl

#hits : 13
#annotations with a hit : 13
#annotations investigated : 160217

Ready

hit 1 - 13 of 13

The boy didn't know what to do.
The rabbit didn't know what to say and looked down in embarrassment. "You are such a fool!"
He continued to talk about him of different things. I don't know, but it seemed like he was exaggerating. He talked and argued with me.
wrote the number on a piece of paper. I was a bit embarrassed because I didn't know if he would be turned on by the lipstick.
I got into the lift and then the lift stopped. I didn't know what to do because I'm afraid. sorry, the lift stopped.
I didn't know when the baby would be born so I held onto my mobile when I went to
I didn't know that it was time for the birth, we had gone to hospital but it was two to 1
asked (my partner/who) why she had sent me the sms. She said that she didn't know when it was for real. I told her that I understood, no worries.
suddenly. He asked Alla what was going on and Alla persisted that he didn't know. The principal signed intensely at him and Alla just nodded.
it was wonderful to know what people were saying. Without them we wouldn't know what was going on.
I don't know what is going on because I can't hear.
And... the mother frog, well one of the frogs. I don't know whether the boy caught the number on the father frog, one of the frogs g.
Then I don't know what you call it there was an animal and they got a big check.

Substring Search **Single Layer Search** Multiple Layer Search

Domain: 281 eaf files Define Domain

Query History: < >

Mode: N-gram within annotation | case insensitive | substring match

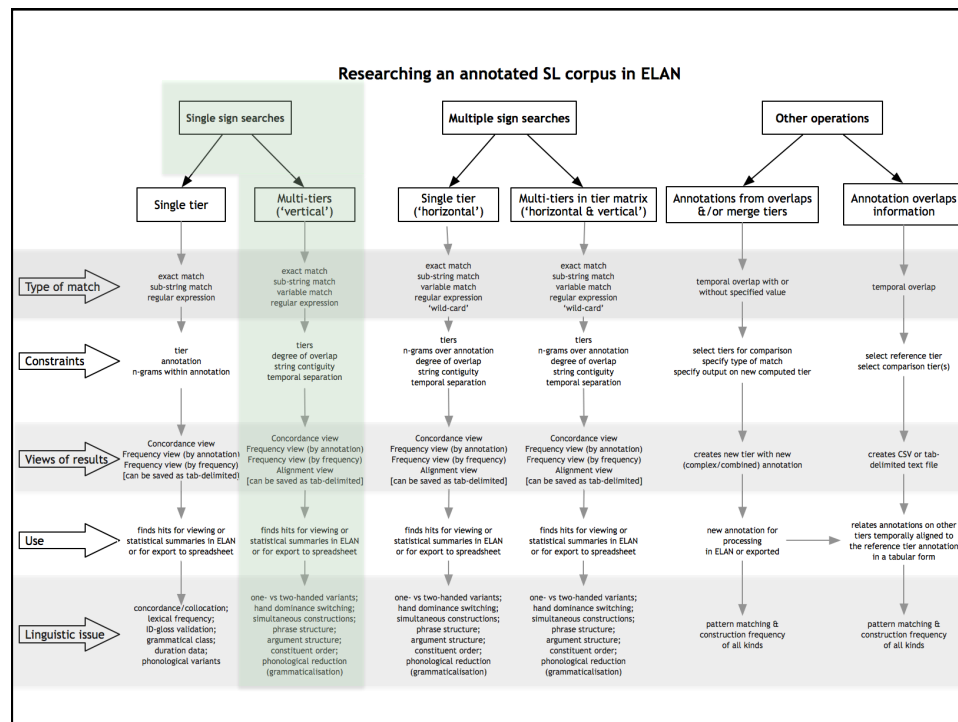
Find Tier Name: FreeTransl

#hits : 13
 #annotations with a hit : 13
 #annotations investigated : 160217 Ready

hit 1 - 13 of 13

The boy **didn't know** what to do.
 The rabbit **didn't know** what to say and looked down in embarrassment. "You are such a fool!"
 He continued to talk about lots of different things. I **don't know**, but it sounded like he was exaggerating. He talked and argued with me.
 wrote the number on a piece of paper. I was a bit embarrassed because I **didn't know** if he would be turned on by the lipstick.
 I got into the lift and then the lift stopped. I **didn't know** what to do because I'm Deaf...sorry, the lift stopped.
 I **didn't know** when the baby would be born so I held onto my mobile when I went to
 I **didn't know** that it was time for the birth, we had gone to hospital but it was two to 1
 asked (my partner/wife) why she had sent me the sms. She said that she **didn't know** when it was for real. I told her that I understood, no worries.
 suddenly. He asked Alfa what was going on and Alfa pretended that he **didn't know**. The principal signed furiously at him and Alfa just nodded.
 ; it was wonderful to know what people were saying. Without them we **wouldn't know** what was going on.
 I **don't know** what is going on because I can't hear.
 And. . . the mother frog, well one of the frogs, I **don't know** whether the boy caught the mother or the father frog, one of the frogs g!
 Then I **don't know** what you call it there was an animal and they got a big shock.

Single sign searches: multi-tier



Examples of secondary searches

- Lexical frequency by grammatical class
- Points (PT)
 - palm orientation and PT-type
 - eye-gaze and PT-type
- Clause types
 - number and order of overt arguments
 - semantic role of argument in position

An ID-gloss searched by grammatical class

Found 87 hits in 87 annotations (of 220923)

frequency 1 - 12 of 12

Annotation	Percentage	Count
#1 FINISH.FIVE-2H #2 Aux	31.03%	27
#1 FINISH.FIVE-2H #2 Conj	21.84%	19
#1 FINISH.FIVE-2H #2 VLoc	18.39%	16
#1 FINISH.FIVE-1H #2 Aux	5.75%	5
#1 FINISH.FIVE-2H #2 Adv	5.75%	5
#1 FINISH.FIVE-2H #2 DM	5.75%	5
#1 FINISH.FIVE-2H #2 Interact	4.60%	4
#1 FINISH.FIVE-2H #2 NLoc	2.30%	2
#1 FINISH.FIVE-1H #2 Conj	1.15%	1
#1 FINISH.FIVE-1H #2 Interact	1.15%	1
#1 FINISH.FIVE-1H #2 VLoc	1.15%	1
#1 FINISH.FIVE-2H #2 VDir	1.15%	1

An ID-gloss searched by grammatical class and modification

Found 122 hits in 122 annotations (of 220923)

frequency 1 - 5 of 5

Annotation	Percentage	Count
#1 LOOK #2 VDir #3 m	86.07%	105
#1 LOOK #2 VDir #3 cg	8.20%	10
#1 LOOK #2 VDir #3 n	4.10%	5
#1 LOOK #2 NLoc #3 m	0.82%	1
#1 LOOK #2 NorV #3 m	0.82%	1

Key: Grammatical class: VDir = directional verb, NLoc = locatable noun, NorV = either noun or verb. Modification: m = modified, cg = congruent, n = not modified)

Palm orientation and PT type

- Is there a relationship with palm orientation and type of point?
- Search for all pointing signs and determine the distribution of their palm orientations
 - IDglosses being with “PT” (^PT)
 - IDglosses overlap an orientation annotation (.+)
 - Export hits into a database, sort and extract statistics
- Answer: strong apparent relationship between palm orientation and point type

The screenshot shows the 'Multiple Layer Search' interface. At the top, there are tabs for 'Substring Search', 'Single Layer Search', and 'Multiple Layer Search'. Below the tabs, the 'Domain' is set to '180 eaf files'. The 'Query History' section has navigation buttons. The 'Mode' is set to 'case sensitive' and 'regular expression'. There are buttons for 'Minimal Duration', 'Maximal Duration', 'Begin After', and 'End Before'. The search criteria are entered in a table-like structure with columns for 'Tier Name' and 'Search Term'. The first row has 'Tier Name: RH-IDgloss' and '^PT'. The second row has 'Tier Name: RH-Orient' and '.*'. The 'Find' button is highlighted. Below the search criteria, the results are displayed as a list of hits. The first hit is '#1 || PT:DET || #2 || Id || #3 || ||'. The second hit is '#1 || PT:LOC || #2 || Id || #3 || ||'. The third hit is '#1 || PT:LOC-2H || #2 || Id || #3 || ||'. The fourth hit is '#1 || PT:LOC || #2 || Id || #3 || ||'. The fifth hit is '#1 || PT:DET || #2 || Id || #3 || ||'. The sixth hit is '#1 || PT:DET-2H || #2 || Id || #3 || ||'. The seventh hit is '#1 || PT:DET || #2 || Id || #3 || ||'. The eighth hit is '#1 || PT:PROSGI || #2 || Id || #3 || ||'. The 'Ready' button is highlighted. The status bar shows 'hit 1 - 8 of 654'.

Domain: 180 eaf files

Query History: < >

Mode: case sensitive regular expression Clear

Minimal Duration Maximal Duration Begin After End Before

Tier Name	Search Term
RH-IDgloss	^PT
RH-Orient	.*
All Tiers	

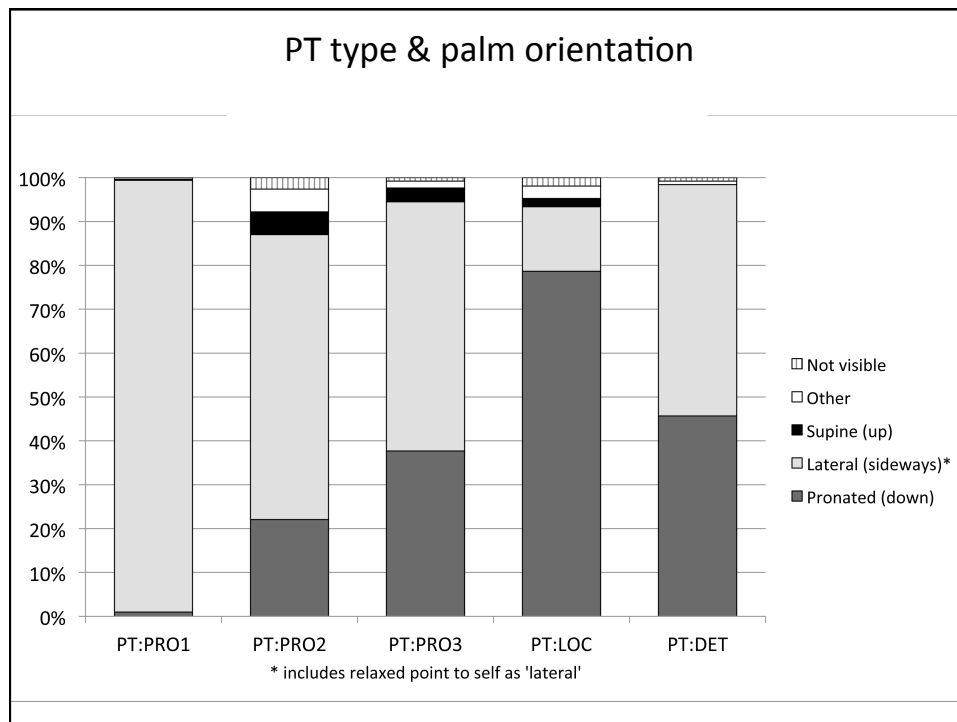
Find

#hits : 654
#annotations with a hit : 654
#annotations investigated : 139485

Ready

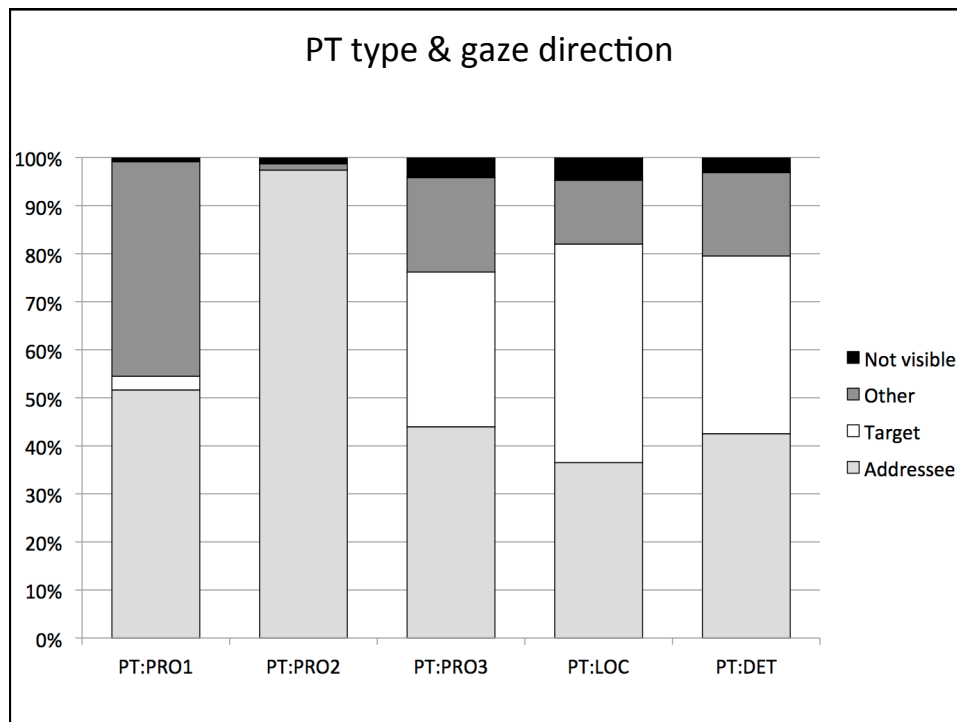
hit 1 - 8 of 654

```
#1 || PT:DET || #2 || Id || #3 || ||
#1 || PT:LOC || #2 || Id || #3 || ||
#1 || PT:LOC-2H || #2 || Id || #3 || ||
#1 || PT:LOC || #2 || Id || #3 || ||
#1 || PT:DET || #2 || Id || #3 || ||
#1 || PT:DET-2H || #2 || Id || #3 || ||
#1 || PT:DET || #2 || Id || #3 || ||
#1 || PT:PROSGI || #2 || Id || #3 || ||
```



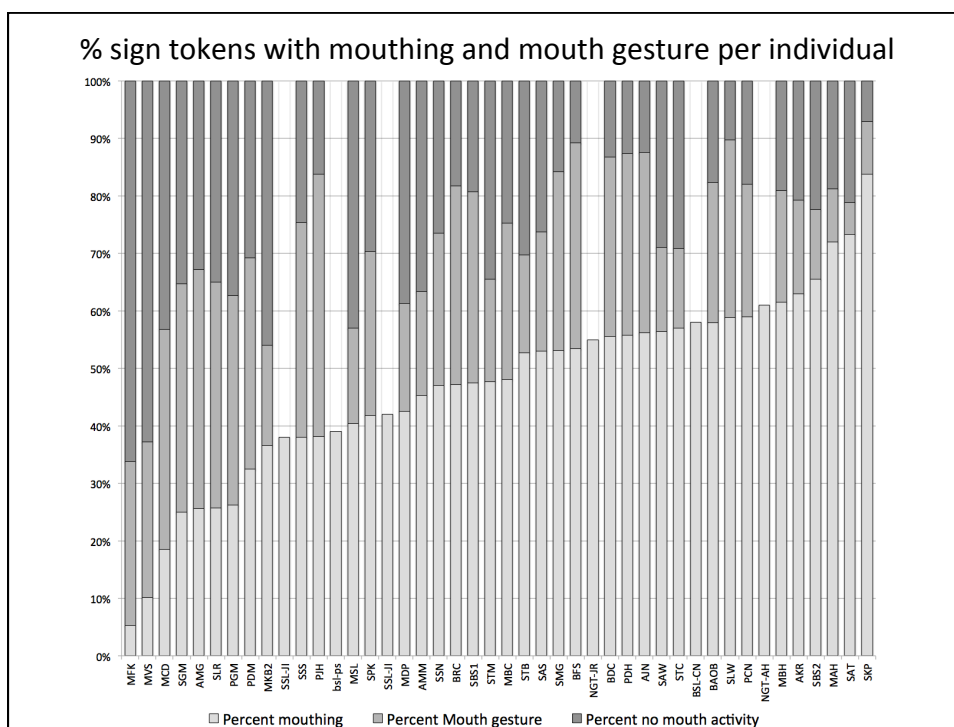
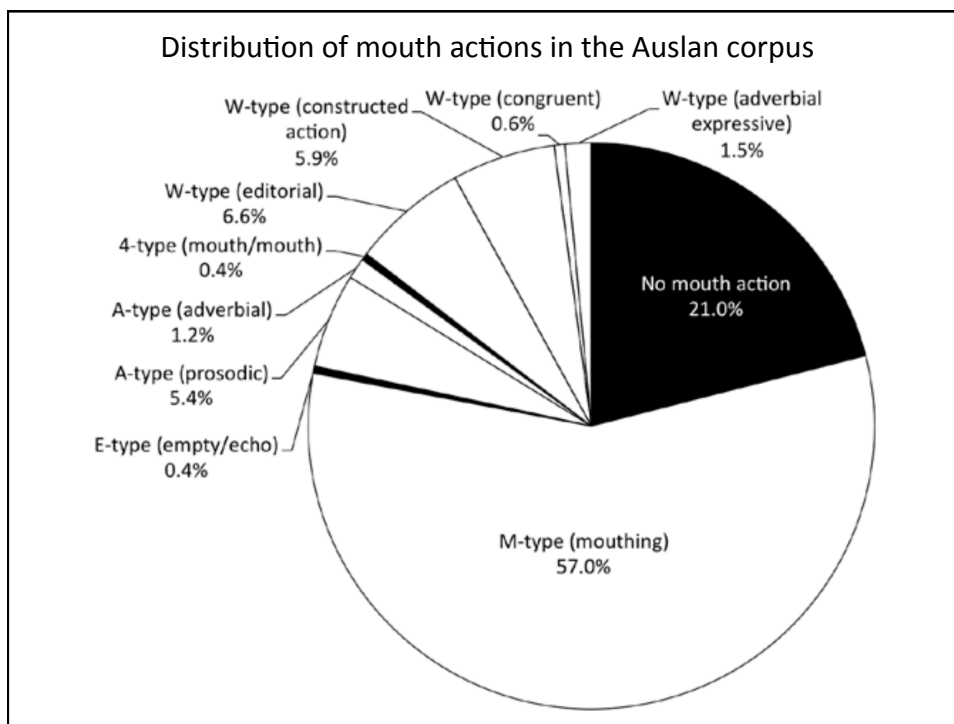
Eye gaze and PT type

- Is there a relationship with direction of eye gaze and type of point?
- Search for all pointing signs and determine the distribution of the co-temporal eye gaze
 - IDglosses being with “PT” (^PT)
 - IDglosses overlap an eye gaze annotation (.+)
 - Get annotation/hit statistics or export hits into a database, sort and extract statistics
- Answer: strong apparent relationship between eye gaze and point type

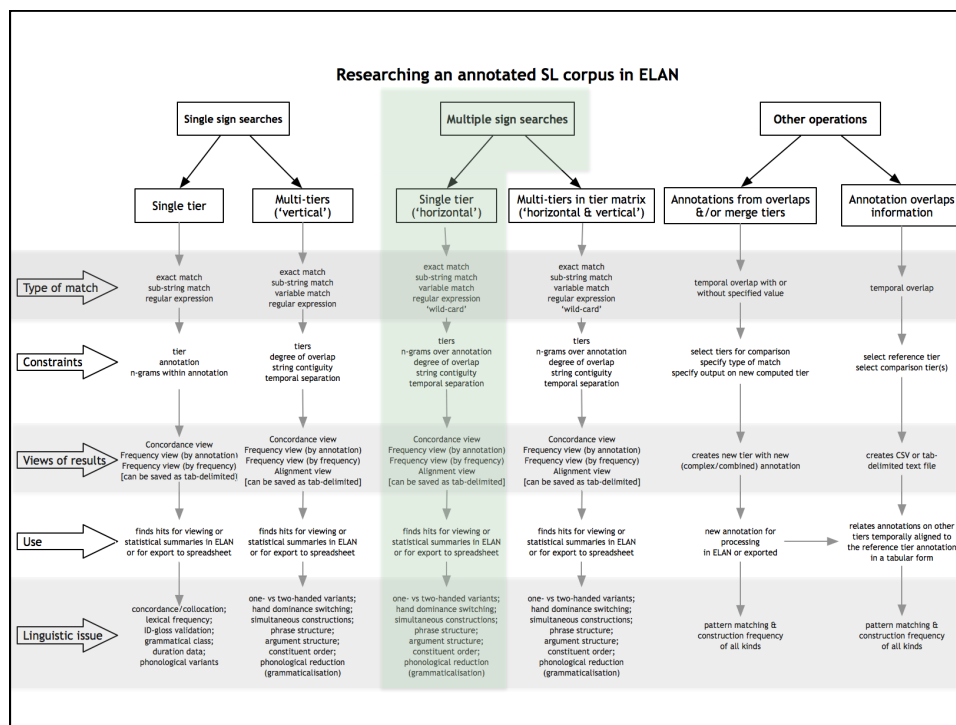


Mouth actions in Auslan

- What is the rate of mouth actions per signs
 - by kind (mouthing, mouth gestures, none)
 - by sign type
 - by person
 - by text type
- Search for all sign tokens
 - any IDglosses (.+)
 - overlaps with any mouth action annotation (.+)
 - Get annotation/hit statistics or export hits into a database, sort and extract statistics



Multiple sign searches: single tier ('horizontal')



A search for a sequence of two specific annotations separated by an unspecified annotations

Search query: Tier Name: SH-IDgloss

Buttons: Find, Fewer Columns, More Columns, Fewer Layers, More Layers

Status: Found 0 hits in 0 annotations (of 220923) Ready Cancel

- ID-gloss validation & compounds/blends
- Established compounds given their own ID-gloss in Auslan Corpus
 - ‘tomato’ is TOMATO, not RED+BALL
 - compounding elements information entered in dictionary
 - criteria for compound/blend status (semantic shift, reduction in form, different mouthing, inseparability of elements)

Collocation or compound: DEAF CLUB?

Search query: Tier Name: RH-IDgloss

Buttons: Find, Fewer Columns, More Columns, Fewer Layers, More Layers

Status: Found 0 hits in 0 annotations (of 220923) Ready Cancel

Search results (left):

```

341 - 343
DISAPPOINTED WITH DEAF CLUB NOT PROVIDE
PRESENT FROM DEAF CLUB PT:PROSG REAL
OBJECT PT:LOC DEAF FS CLUB OBJECT PT:PROI
FSTO PT:DET DEAF FS CLUB PT:LOC WHEN
DEAF DEAF DEAF CLUB DEAF COMMUNITY
DIFFERENT PT:LOC DEAF CLUB PT:PROSG:SH FINISH-FIVE
HAVE PT:POSSIBLE DEAF FS CLUB PT:PROPL HAVE
PT:LOC G:YOU-KNOW DEAF FS CLUB WHAT G:YOU-KNOW
PT:PROSG GO DEAF FS CLUB FIVE FS-ELIZ
  
```

Search results (right):

```

341 - 343
LOOK-SOMEONE G:come-here DEAF CLUB LEADER DS(S) many-people-go
INTEGRATE KNOW-NOT-SH DEAF CLUB FUTURE G:WELL
PT:PROSG GO DEAF CLUB PT:PROPL G-GO-AWAY
DEAF THINK DEAF CLUB BIG PART(S)
PT:PROPL WANT DEAF CLUB ENCOURAGE-SH MAKE
AVAILABLE FOR DEAF CLUB PEOPLE NEED
F:DO NOT DEAF CLUB PT:POSSING FAULT
G:CA HAVE-NOT DEAF CLUB FISS-OFF-SH G:YOU-SEE
  
```

Substring Search **Single Layer Search** Multiple Layer Search

Domain: 281 eaf files Define Domain

Query History: < >

Mode: N-gram over annotations case insensitive substring match

Find Tier Name: RH-IDgloss

#hits : 9
 #annotations with a hit : 9
 #annotations investigated : 160217 Ready

hit 1 - 9 of 9

DISAPPOINTED WITH **DEAF CLUB** NOT PROVIDE
 PRESENT FROM **DEAF CLUB** PT:PRO1SG REAL
 OBJECT PT:LOC **DEAF FS:CLUB** OBJECT PT:PRO3
 FS:TO PT:DET **DEAF FS:CLUB** PT:LOC WHEN
 DEAF DEAF **DEAF CLUB1** DEAF COMMUNITY
 DIFFERENT PT:LOCPL **DEAF CLUB1** PT:PRO3SG-2H FINISH-FIVE
 HAVE PT:POSS3PL **DEAF FS:CLUB** PT:PRO3PL HAVE
 PT:LOC G:YOU-KNOW **DEAF FS:CLUB** WHAT G:YOU-KNOW
 PT:PRO1SG GO **DEAF FS:CLUB** FIVE FS:ELIZ

Substring Search **Single Layer Search** Multiple Layer Search

Domain: 281 eaf files Define Domain

Query History: < >

Mode: N-gram within annotation case insensitive substring match

Find Tier Name: RH-IDgloss

#hits : 8
 #annotations with a hit : 8
 #annotations investigated : 160217 Ready

hit 1 - 8 of 8

LOOK-SOMEONE G:come-here **DEAF-CLUB** LEADER DS(5):many-people-go
 INTEGRATE KNOW-NOT-2H **DEAF-CLUB** FUTURE G:WELL
 PT:PRO1SG GO **DEAF-CLUB** PT:PRO3PL G:GO-AWAY
 DEAF THINK **DEAF-CLUB** BIG PART(SE)
 PT:PRO3PL WANT **DEAF-CLUB** ENCOURAGE-1H MAKE
 AVAILABLE FOR **DEAF-CLUB** PEOPLE NEED
 FS:DO NOT **DEAF-CLUB** PT:POSS2SG FAULT
 G:CA HAVE-NOT **DEAF-CLUB** PISS-OFF-2H G:YOU-SEE

Substring Search **Single Layer Search** Multiple Layer Search

Domain: 281 caf files Define Domain

Query History: < >

Mode: Annotation case insensitive regular expression

Find Tier Name: RH-IDgloss

#hits : 238
 #annotations with a hit : 238
 #annotations investigated : 160217

Ready

hit 23 - 44 of 238

PT:PRO1SG(7) WANT PT:PRO1SG(B) IN G:THERE-YOU-GO
 TORTOISE GET-ATTENTION PT:PRO1SG(B) KNOW WANT
 STUPID KNOW-NOT-2H PT:PRO1SG(B) GET-LOST2 G:WELL
 WOLF G:WELL PT:PRO1SG(B) ENOUGH
 EAT GRASS PT:PRO1SG(B) HERD DS(T):USING-CROOK-TO-HERD
 SEE G:WELL PT:PRO1SG(B) WIN RABBIT
 SLEEP-2H G:don't-worry-about-it PT:PRO1SG(B) GOOD SLEEP-2H
 DS(5):TURTLE-STAND KNOW PT:PRO1SG(B) ALWAYS1 SLOW
 ALWAYS1 SLOW PT:PRO1SG(B) KNOW G:WELL
 AUSTRALIA NOT-YET PT:PRO1SG(B) OVERSEAS PT:PRO1SG
 PT:PRO1SG PAST PT:PRO1SG(B) HAVE CAR
 NICE CAR PT:PRO1SG(B) DS(b):parked-car ALRIGHT
 PT:PRO1SG(5) DEAF PT:PRO1SG(B) G:UMM DOORBELL
 SCREAM G:WELL PT:PRO1SG(B) DEAF M:CANT
 PANIC WHAT PT:PRO1SG(B) PATIENCE STAND
 TELEPHONE G:WELL PT:PRO1SG(B) G:DEAF G:WELL(1)
 ALWAYS1 CHRISTMAS PT:PRO1SG(B) REAL DISAPPOINTED
 STILL WORK PT:PRO1SG(B) CONTINUE STOP
 DEPEND-ON SIXTY-YEARS-OLD PT:PRO1SG(B) DONT-KNOW TIME-GOES-BY
 AMERICA HELP PT:PRO1SG(B) PT:PRO3PL PT:PRO1SG
 PT:PRO1SG(B) PAST ONE
 WHY NOT PT:PRO1SG(B) DRIVE WANT

PT:PRO1SG(7) WANT PT:PRO1SG(B) IN G:THERE-YOU-GO
 TORTOISE GET-ATTENTION PT:PRO1SG(B) KNOW WANT
 STUPID KNOW-NOT-2H PT:PRO1SG(B) GET-LOST2 G:WELL
 WOLF G:WELL PT:PRO1SG(B) ENOUGH
 EAT GRASS PT:PRO1SG(B) HERD DS(T):USING-CROOK-TO-HERD
 SEE G:WELL PT:PRO1SG(B) WIN RABBIT
 SLEEP-2H G:don't-worry-about-it PT:PRO1SG(B) GOOD SLEEP-2H
 DS(5):TURTLE-STAND KNOW PT:PRO1SG(B) ALWAYS1 SLOW
 ALWAYS1 SLOW PT:PRO1SG(B) KNOW G:WELL
 AUSTRALIA NOT-YET PT:PRO1SG(B) OVERSEAS PT:PRO1SG
 PT:PRO1SG PAST PT:PRO1SG(B) HAVE CAR
 NICE CAR PT:PRO1SG(B) DS(b):parked-car ALRIGHT
 PT:PRO1SG(5) DEAF PT:PRO1SG(B) G:UMM DOORBELL
 SCREAM G:WELL PT:PRO1SG(B) DEAF M:CANT
 PANIC WHAT PT:PRO1SG(B) PATIENCE STAND
 TELEPHONE G:WELL PT:PRO1SG(B) G:DEAF G:WELL(1)
 ALWAYS1 CHRISTMAS PT:PRO1SG(B) REAL DISAPPOINTED
 STILL WORK PT:PRO1SG(B) CONTINUE STOP
 DEPEND-ON SIXTY-YEARS-OLD PT:PRO1SG(B) DONT-KNOW TIME-GOES-BY
 AMERICA HELP PT:PRO1SG(B) PT:PRO3PL PT:PRO1SG
 PT:PRO1SG(B) PAST ONE
 WHY NOT PT:PRO1SG(B) DRIVE WANT

What can occur between “I” and “THINK”?

PT:PRO1SG | = 0 ann. | .+ | = 0 ann. | ^THINK\$ | Tier Name: SH-ID

Fewer Columns | More Columns | Fewer Layers | More Layers

ts in 50 annotations (of 238929) | Ready

frequency 1 - 7 of 30 | >

Annotation	Percentage	Count
#1 PT:PRO1SG NOT THINK	24.00%	12
1 PT:PRO1SG NEVER THINK	8.00%	4
PT:PRO1SG ALWAYS THINK	6.00%	3
PT:PRO1SG CAN-NOT THINK	4.00%	2
#1 PT:PRO1SG LIKE THINK	4.00%	2
#1 PT:PRO1SG ONE THINK	4.00%	2
PT:PRO1SG PT:PRO1SG THINK	4.00%	2

Is the NEGATOR position fixed?

^PT:PRO1SG\$ | = 0 ann. | THINK | = 0 ann. | ^NOT|^NOTHING | Tier Name: SH-IDgloss

Find | Fewer Columns | More Columns | Fewer Layers | More Layers

Found 11 hits in 11 annotations (of 238929) | Ready | Cancel

frequency 1 - 2 of 2

Annotation	Percentage	Count
#1 PT:PRO1SG THINK NOTHING	54.55%	6
#1 PT:PRO1SG THINK NOT	45.45%	5

^NOT|^NOTHING | = 0 a... | THINK | = 0 a... | ^PT:PRO1SG | Tier Name: SH-ID...

Find | Fewer Columns | More Columns | Fewer Layers | More Layers

Found 2 hits in 2 annotations (of 238929) | Ready | Cancel

frequency 1 - 1 of 1

Annotation	Percentage	Count
#1 NOT THINK PT:PRO1SG	100.00%	2

Is “I think” becoming a sentence adverbial?*

00:16:20.512	00:00:01.000	00:16:21.512
PT:PRO1SG	THINK	
00:16:27.732	00:00:01.000	00:16:28.732
PT:PRO1SG	THINK	
00:00:48.510	00:00:01.000	00:00:49.510
PT:PRO1SG	THINK	
00:00:41.425	00:00:01.000	00:00:42.425
PT:PRO1SG	THINK	
00:00:27.915	00:00:01.000	00:00:28.915
PT:PRO1SG	THINK	
00:13:26.720	00:00:01.000	00:13:27.720
PT:PRO1SG	THINK	

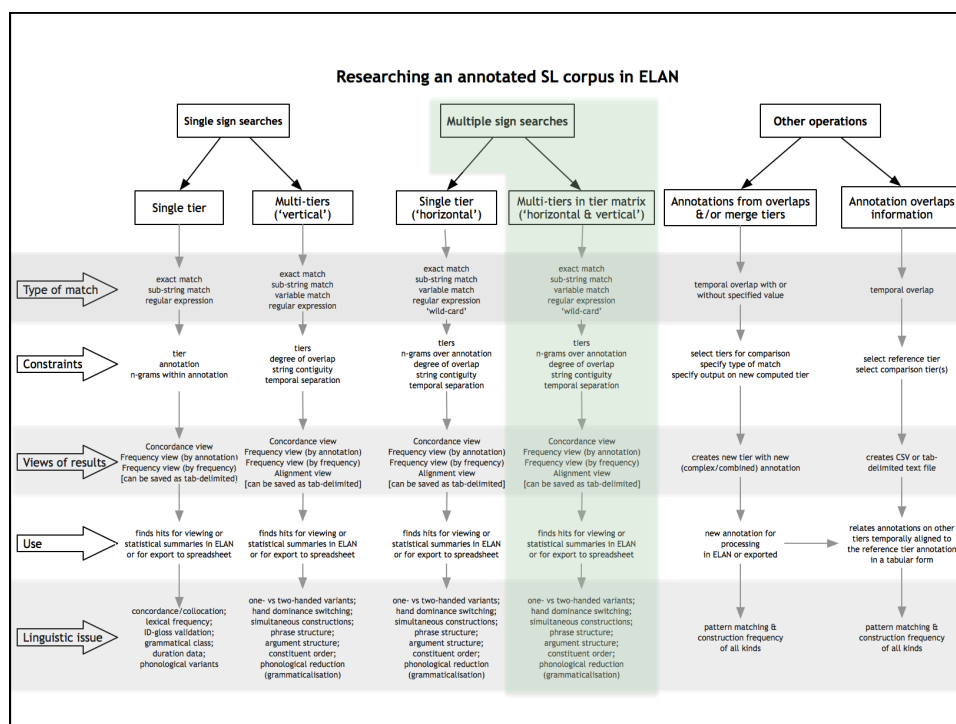
*Just 6 examples out of 203 hits

Simultaneity

- e.g., when points co-occur are they double articulations, co-articulations, or something else?

ADCB1c4a.eaf	00:23:34.818	00:00:02.000	00:23:36.818
SH-IDgloss		PT:PRO1SG(7)	
WH-IDgloss		PT:PRO3SG	
ADCB1c4a.eaf	00:26:27.820	00:00:02.000	00:26:29.820
SH-IDgloss		PT:PRO1SG	
WH-IDgloss		PT:PRO3SG	
ADCB1c4a.eaf	00:28:29.855	00:00:02.000	00:28:31.855
SH-IDgloss		PT:PRO1SG	
WH-IDgloss		PT:PRO3SG	
ADCB1c4a.eaf	00:28:34.380	00:00:02.000	00:28:36.380
SH-IDgloss		PT:PRO1SG	
WH-IDgloss		PT:PRO3SG	

Multiple sign searches: multi-tier ('horizontal & vertical')



Substring Search | Single Layer Search | Multiple Layer Search

Domain: 416 eaf files Define Domain

Query History: < > New Query Save query Load query

Mode: case insensitive | regular expression Clear

Minimal Duration Maximal Duration Begin After End Before

= 0 ann. Tier Name: RH-IDgloss
 Must be in same file
 = 0 ann. Tier Name: RH-Arg
 Must be in same file
 = 0 ann. Tier Name: RH-MacroR
 Must be in same file
 Tier Name: RH-ModOrVar

Find Hide query Fewer Columns More Columns Fewer Layers More Layers

Found 17 hits in 17 annotations (of 331188) Ready Cancel

< > Frequency 1 – 13 of 13 Save frequencies

Percentage	Count	Annotation
17.65%	3	#1 IRABBIT ILOOKI #2 IA I IVI #3 IACTRI IPROCESSI #4 II lml
17.65%	3	#1 ITORTOISE ILOOKI #2 IA I IVI #3 IACTRI IPROCESSI #4 II lml
5.88%	1	#1 IFRIENDI IDISPERSEI #2 IA I IVI #3 IACTRI IPROCESSI #4 II lml
5.88%	1	#1 INS:ALFA IGET-ATTENTIONI #2 IA I IVI #3 IACTRI IPROCESSI #4 II lml
5.88%	1	#1 IPEOPLE IBLANKI #2 IA I IVI #3 IUNDR IPROCESSI #4 II lml
5.88%	1	#1 IPT:PROIPL(2) IOPPOSEI #2 IA I IVI #3 IACTRI IPROCESSI #4 II lml
5.88%	1	#1 IPT:PROISGI ISAYI #2 IA I IVI #3 IACTRI IPROCESSI #4 II lml
5.88%	1	#1 IPT:PROISGI ISTOP I #2 IA I IVI #3 IACTRI IPROCESSI #4 II lml
5.88%	1	#1 IPT:PRO2SG(2) IARRIVEI #2 IA I IVI #3 IACTRI IPROCESSI #4 II lml
5.88%	1	#1 IPT:PRO2SG IGO-DOWNI #2 IA I IVI #3 IACTRI IPROCESSI #4 II lml
5.88%	1	#1 IPT:PRO2SG ISEND-TOI #2 IA I IVI #3 IACTRI IPROCESSI #4 II lml
5.88%	1	#1 IRABBIT ILOOKI #2 IA I IVI #3 IACTRI IPROCESSI #4 II lml
5.88%	1	#1 ISHEEP I CATCHI #2 IA I IVI #3 IUNDR IPROCESSI #4 II lml

Substring Search | Single Layer Search | Multiple Layer Search

Domain: 416 eaf files Define Domain

Query History: < > New Query Save query Load query

Mode: case insensitive | regular expression Clear

Minimal Duration Maximal Duration Begin After End Before

= 0 ann. Tier Name: RH-IDgloss
 Must be in same file
 = 0 ann. Tier Name: RH-Arg
 Must be in same file
 = 0 ann. Tier Name: RH-MacroR
 Must be in same file
 Tier Name: RH-ModOrVar

Find Hide query Fewer Columns More Columns Fewer Layers More Layers

Found 24 hits in 24 annotations (of 331188) Ready Cancel

< > Frequency 1 – 14 of 22 Save frequencies

Percentage	Count	Annotation
8.33%	2	#1 IFS:LIFT ISTOP-2HI #2 IA I IVI #3 IUNDR IPROCESSI #4 II lml
8.33%	2	#1 ITORTOISE ISAYI #2 IA I IVI #3 IACTRI IPROCESSI #4 II lml
4.17%	1	#1 ICANEI IBANI #2 IA I IVI #3 IUNDR IPROCESSI #4 II lml
4.17%	1	#1 IDOORI IHAVEI #2 IA2 I IVI #3 IUNDR IPROCESSI #4 II lml
4.17%	1	#1 IDSS:LEVEL IWINI #2 IA I IVI #3 IACTRI IPROCESSI #4 II lml
4.17%	1	#1 IFS:MORAL I TITLEI #2 IA I IVI #3 IACTRI IPROCESSI #4 II lml
4.17%	1	#1 IFS:TRADITION ICONTINUEI #2 IA I IVI #3 IACTRI IPROCESSI #4 II lml
4.17%	1	#1 IHELP IFAILI #2 IA2 I IVI #3 IUNDR IPROCESSI #4 II lml
4.17%	1	#1 INEWSPAPER IREADI #2 IA I IVI #3 IUNDR IPROCESSI #4 II lml
4.17%	1	#1 INS:ALFA IGUFFAWI #2 IA I IVI #3 IACTRI IPROCESSI #4 II lml
4.17%	1	#1 IPEOPLE IHAVEI #2 IA I IVI #3 IUNDR IPROCESSI #4 II lml
4.17%	1	#1 IPT:FBUOY(RELAXED) ISAYI #2 IA I IVI #3 IACTRI IPROCESSI #4 II lml
4.17%	1	#1 IPT:PROIPL(2) IDECIDEI #2 IA I IVI #3 IACTRI IPROCESSI #4 II lml
4.17%	1	#1 IPT:PROISG(5) IHAVEI #2 IA I IVI #3 IACTRI IPROCESSI #4 II lml

Substring Search | Single Layer Search | Multiple Layer Search

Domain: 416 eaf files Define Domain

Query History: < > New Query Save query Load query

Mode: case insensitive regular expression Clear

Minimal Duration Maximal Duration Begin After End Before

<input type="text" value="."/>	= 0 ann.	<input type="text" value="."/>	Tier Name: RH-IDgloss
Overlap		Overlap	Must be in same file
<input type="text" value="^A ^A1 ^A2"/>	= 0 ann.	<input type="text" value="^V"/>	Tier Name: RH-Arg
Overlap		Overlap	Must be in same file
<input type="text" value="^ACTR ^U"/>	= 0 ann.	<input type="text" value="."/>	Tier Name: RH-MacroR
Overlap		Overlap	Must be in same file
<input type="text" value=""/>		<input type="text" value="cg"/>	Tier Name: RH-ModOrVar

Find Hide query Fewer Columns More Columns Fewer Layers More Layers

Found 6 hits in 6 annotations (of 331188) Ready Cancel

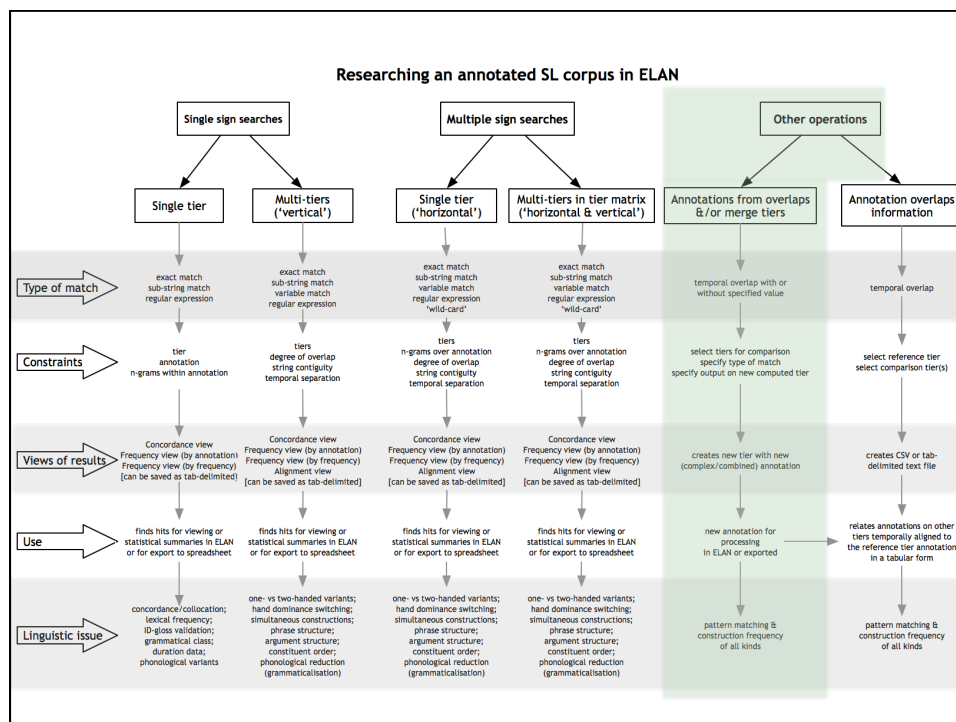
< > Frequency 1 – 6 of 6 Save frequencies

Percentage	Count	Annotation
16.67%	1	#1 IFS:RABBIT SIT-ON #2 A IV #3 ACTR IPROCESS #4 lcg
16.67%	1	#1 INS:ALFA IGUFFAW #2 A IV #3 ACTR IPROCESS #4 lcg
16.67%	1	#1 IPEOPLE IG(5-DOWN):PHOOEY #2 A IV #3 ACTR IPROCESS #4 lcg
16.67%	1	#1 IPT:PRO3SG(D) IDEMAND #2 A IV #3 ACTR IPROCESS #4 lcg
16.67%	1	#1 IRABBIT IGUFFAW #2 A IV #3 ACTR IPROCESS #4 lcg
16.67%	1	#1 ITORTOISE ILOOK #2 A IV #3 IUNDR IPROCESS #4 lcg

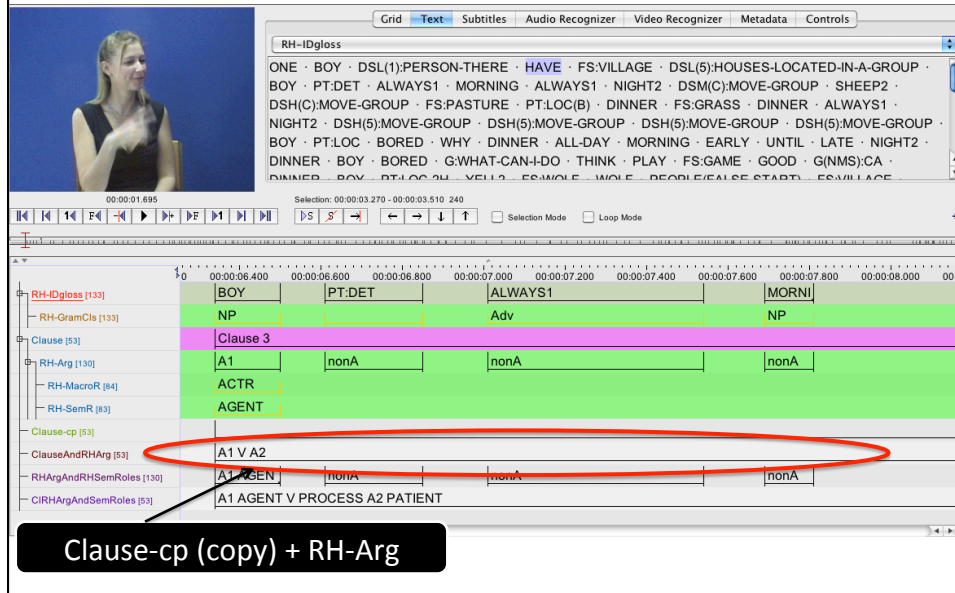
What's missing?

- The ability to combine data dispersed across several tiers into a single complex annotation that can then be the target of a search constrained by other factors
- Any indication that the sequences identified are CLU-specific
 - i.e., they occur within CLU constructions (and are thus potential schemas), rather than being unrelated juxtapositions of annotations (they span clause boundaries)

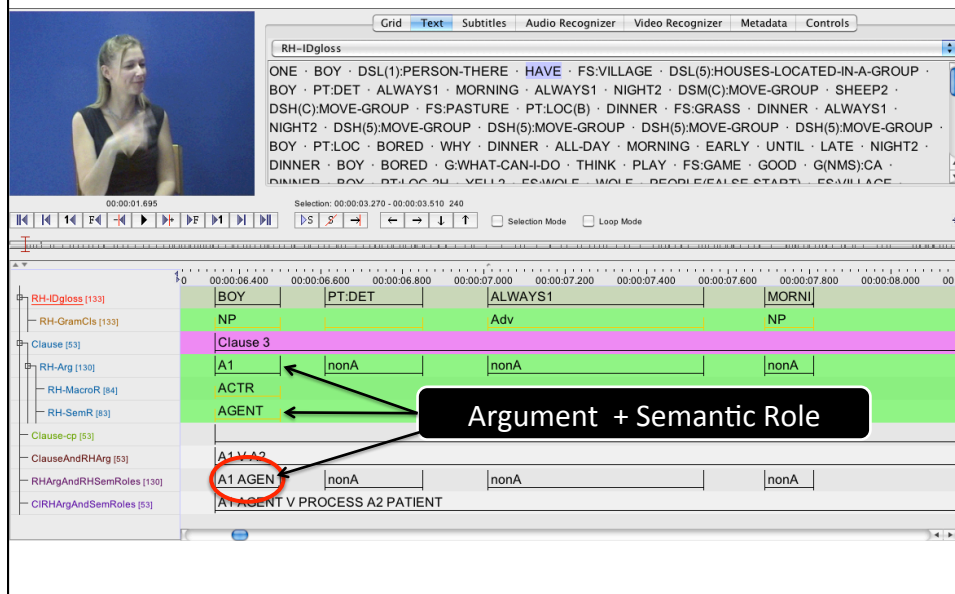
Other operations: merge & annotation from overlaps



New CLU annotation containing core arguments and verb
(Merge tiers)



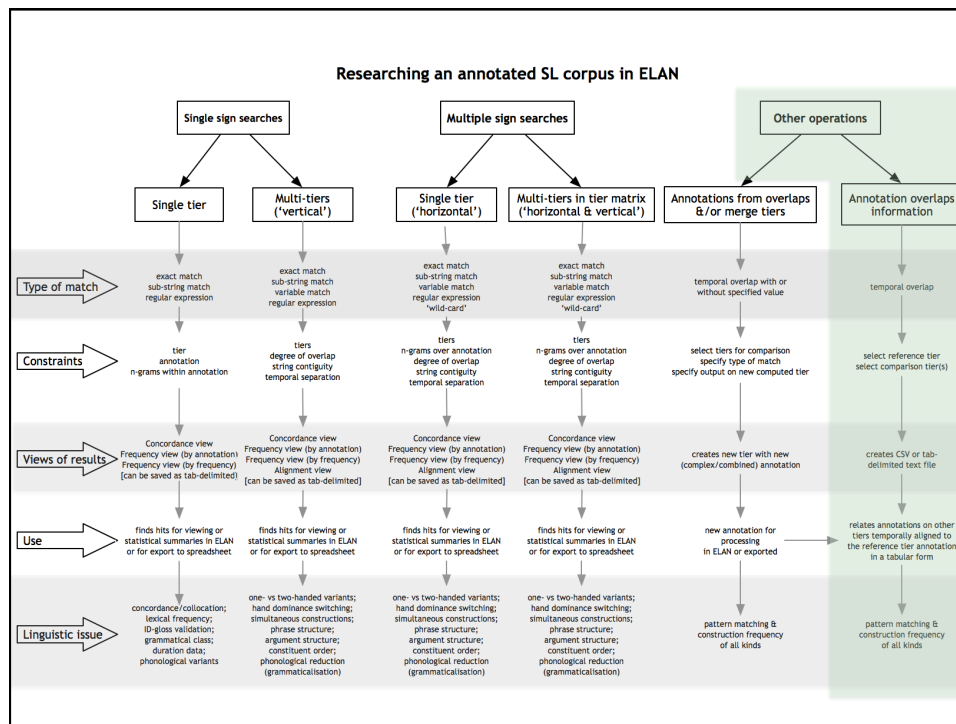
New CLU constituent annotation argument & semantic role (Annotation from overlaps)



Results can be tagged rather than merged

- tag as 'potential subject'
- tag as 'potential passive construction'
- whatever the research interest is...

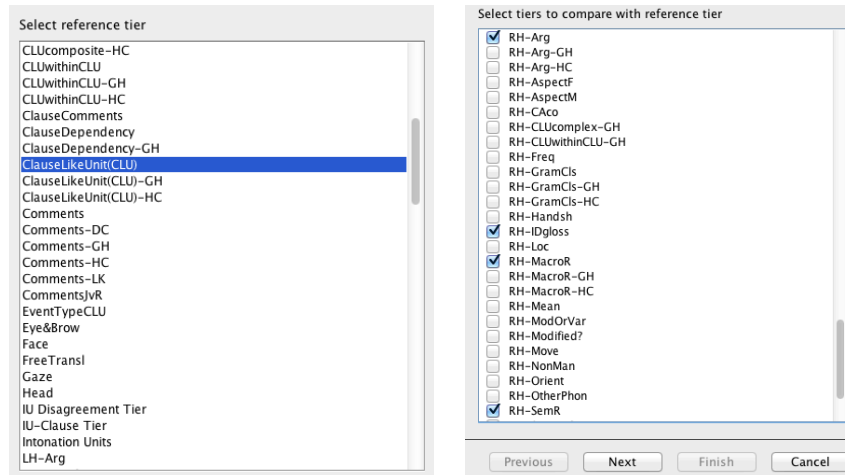
Other operations
Annotation overlaps information
(export only)



Annotation overlaps information

- Like “annotations from overlaps” except that
 - the output does not alter the ELAN annotation file itself (by creating a new tier with the merged information)
 - the results are automatically exported as CSV files which can be opened and further processed in databases, such as EXCEL (the results cannot be viewed in ELAN)
- Advantages of “Annotation overlaps information”
 - better at processing overlaps that involve a temporally larger domain (e.g., CLUs) overlapping multiple temporally smaller domains (e.g., sign tokens)
 - thus excellent for analyzing constituency
 - each reference tier annotation is exported with file identifying information and token identifying information
 - exported data can be subjected to multiple sorting, searching, and filtering operations (especially recursively on found sets) than is possible within ELAN
 - exported data can be easily tagged with codes for multivariate analysis (e.g., in Rbrul)

Annotation from overlaps



AOI data in Excel

The image shows two side-by-side screenshots of an Excel spreadsheet titled "ForPugExample.xlsx". The spreadsheet displays AOI data for a specific tier, "ClauseLikeUnit(CLU)".

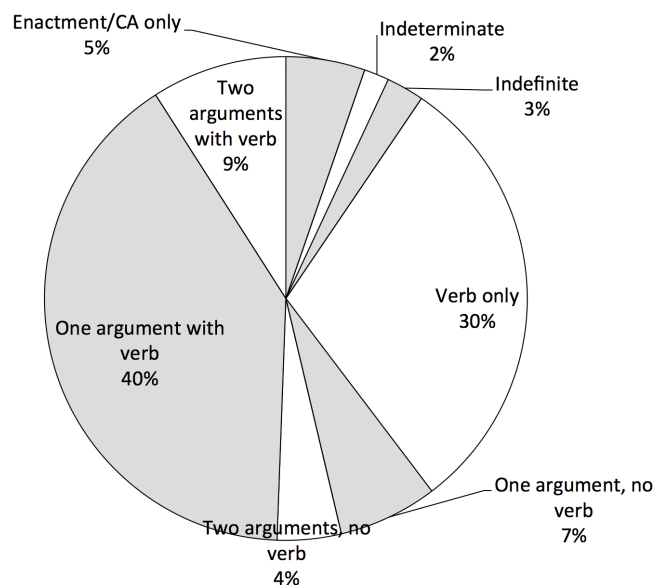
The left screenshot shows the first 23 rows of data. The columns are labeled "CLU", "SH-ARG", "WH-ARG", and "CA-ARG". The data is organized into a table with 23 rows and 4 columns. The first row is the header, and the subsequent rows contain data for each tier.

The right screenshot shows the same spreadsheet, but with the "Next" button highlighted, indicating that the data is being processed or compared.

Exported data can be tagged for additional processing, e.g., in Rbrul

	A	B	D	E	G	H	L	N	P	Q	R	S
1	ClauseLikeUnit(CLU)	Form	PoS	Modifier position	CLU_aktionsart	CompletiveVsAnterior	TrstType	TokenInCLU	Age	Dialect	City	Sex
2	AAM1c4CLU_A_M_34_N#01	GOOD	Aux	postmodifier	act	Completive	Interview	SingleToken	34	Southern	Adelaide	Male
3	AAM1c4CLU_A_M_34_N#03	GOOD	Aux	postmodifier	act	Anterior	Interview	SingleToken	34	Southern	Adelaide	Male
4	AAM1c4CLU_A_M_34_N#06	GOOD	Aux	premodifier	act	Anterior	Interview	SingleToken	34	Southern	Adelaide	Male
5	AAM2b4CLU_A_F_42_N#01	GOOD	Adj	head	state	n/a	Retell	SingleToken	42	Southern	Adelaide	Female
6	AAM2b4CLU_A_F_42_N#04	GOOD	Aux	premodifier	ach	Completive	Retell	SingleToken	42	Southern	Adelaide	Female
7	AAP2b4CLU_A_F_51_N#97	FINALLY	Aux	postmodifier	ach	Completive	Retell	SingleToken	51	Southern	Adelaide	Female
8	AAP2b4CLU_A_F_51_N#108	FINALLY	Adv	postmodifier	ach	n/a	Retell	SingleToken	51	Southern	Adelaide	Female
9	AAP2b4CLU_A_F_51_N#112	FINALLY	Adv	postmodifier	ach	n/a	Retell	SingleToken	51	Southern	Adelaide	Female
10	AAP3c4CLU_A_F_51_N#04	GOOD	Verb	head	ach	n/a	Narrative	SingleToken	51	Southern	Adelaide	Female
11	AAP4c4CLU_A_F_51_N#03	GOOD	Verb	head	ach	Completive	Interview	SingleToken	51	Southern	Adelaide	Female
12	AAP4c4CLU_A_F_51_N#04	FINALLY	Verb	head	ach	n/a	Interview	SingleToken	51	Southern	Adelaide	Female
13	AAP4c4CLU_A_F_51_N#05	FINALLY	Adv	premodifier	act	n/a	Interview	SingleToken	51	Southern	Adelaide	Female
14	AAP4c4CLU_A_F_51_N#06	GOOD	Adv	premodifier	acc	n/a	Interview	SingleToken	51	Southern	Adelaide	Female
15	AAP4c4CLU_A_F_51_N#07	GOOD	Aux	premodifier	acc	Anterior	Interview	SingleToken	51	Southern	Adelaide	Female
16	AAP4c4CLU_A_F_51_N#08	GOOD	Aux	premodifier	acc	n/a	Interview	SingleToken	51	Southern	Adelaide	Female
17	AAP4c4CLU_A_F_51_N#09	GOOD	Aux	premodifier	act	Anterior	Interview	SingleToken	51	Southern	Adelaide	Female
18	AAS81c2a4CLU_A_M_64_N#35	GOOD	Verb	head	ach	n/a	Retell	SingleToken	64	Southern	Adelaide	Male
19	AASc4c4CLU_A_M_64_N#01a	FIVE	Interact	n/a	n/a	n/a	Interview	TwoTokens	64	Southern	Adelaide	Male
20	AASc4c4CLU_A_M_64_N#01b	GOOD	Interact	n/a	n/a	n/a	Interview	TwoTokens	64	Southern	Adelaide	Male
21	AASc4c4CLU_A_M_64_N#06	GOOD	Aux	n/a	ach	Anterior	Interview	SingleToken	64	Southern	Adelaide	Male
22	ACAc4c4CLU_A_M_73_N#01	FIVE	Aux	postmodifier	act	Completive	Interview	SingleToken	73	Southern	Adelaide	Male
23	ACAc6c4CLU_A_M_73_N#01	FIVE	DM	n/a	n/a	n/a	Recount	SingleToken	73	Southern	Adelaide	Male
24	ADCc4c4CLU_A_M_49_N#03	FINALLY	Verb	head	ach	n/a	Interview	SingleToken	49	Southern	Adelaide	Male
25	ADCc4c4CLU_A_M_49_N#07	GOOD	Aux	premodifier	acc	Anterior	Interview	SingleToken	49	Southern	Adelaide	Male
26	ADCc4c4CLU_A_M_49_N#12	FINALLY	Adv	postmodifier	act	n/a	Interview	SingleToken	49	Southern	Adelaide	Male
27	ADCc4c4CLU_A_M_49_N#13	FINALLY	Verb	head	ach	n/a	Interview	SingleToken	49	Southern	Adelaide	Male
28	ADCc4c4CLU_A_M_49_N#15	FINALLY	Verb	head	ach	n/a	Interview	SingleToken	49	Southern	Adelaide	Male
29	ADCc4c4CLU_A_M_49_N#16	FINALLY	Verb	head	ach	n/a	Interview	SingleToken	49	Southern	Adelaide	Male
30	ADCc6c4CLU_A_M_49_N#01	FIVE	DM	n/a	n/a	n/a	Recount	SingleToken	49	Southern	Adelaide	Male
31	ADPc2a4CLU_A_M_71_N#17	FIVE	Aux	postmodifier	ach	Completive	Retell	SingleToken	71	Southern	Adelaide	Male
32	ADPc3c4CLU_A_M_71_N#01	FIVE	Aux	premodifier	ach	Anterior	Narrative	SingleToken	71	Southern	Adelaide	Male
33	ADPc6c4CLU_A_M_71_N#02a	FINALLY	Verb	head	ach	n/a	Recount	TwoTokens	71	Southern	Adelaide	Male
34	ADPc6c4CLU_A_M_71_N#02b	FINALLY	Verb	head	ach	n/a	Recount	TwoTokens	71	Southern	Adelaide	Male

Distribution of CLU types



(Recent presentation, ISGS San Diego, July 2014)

Structure of CLUs

CLU by argument type	% of each type with a constituent which is				
	enacted	depicting	enacted or depicting	pointing	enacted, depicting or pointing
Constructed action only	100%	n/a	100%	n/a	100%
Indeterminate	75%	tbc	>75% (tbc)	tbc	+>75% (tbc)
Indefinite	50%	tbc	>50% (tbc)	tbc	+>50% (tbc)
Verb only	19%	12%	28%	6.5%	35%
One argument, no verb	4%	1%	5%	23%	23%
Two arguments, no verb	4%	12%	16%	58%	58%
One argument with verb	17%	8%	23%	40%	49%
Two arguments with verb	16%	20%	20%	60%	75%

(Recent presentation, ISGS San Diego, July 2014)

Argument patterns

(Recent presentation, ISGS San Diego, July 2014)

Phenomenon	Comment
One argument CLUs	
A V = 70%	pre-verbal argument twice as frequent as post-verbal
V A = 30%	
Two argument CLUs	
A1 V A2 = 83%	A1 V A2 by far the most frequent construction
V A1 A2 = 17%	
A1 A2 V = 0%	A1 A2 V only unattested because of task/genre Johnston et al 2007 has multiple examples from 3 SLs (Volterra picture task, spatial)
*V A2 A1 etc. n/a	cannot exist by definition (1 and 2 refer to 'order of appearance', not syntactic role)

(Recent presentation, ISGS San Diego, July 2014)

Conclusion

