

The reluctant oracle:
annotating a sign language corpus
for answers to questions we can't
ask any other way

Trevor Johnston
Macquarie University
Sydney, Australia

[Sign language corpora: linguistic issues, DCAL, London, 24-25 July 2009]

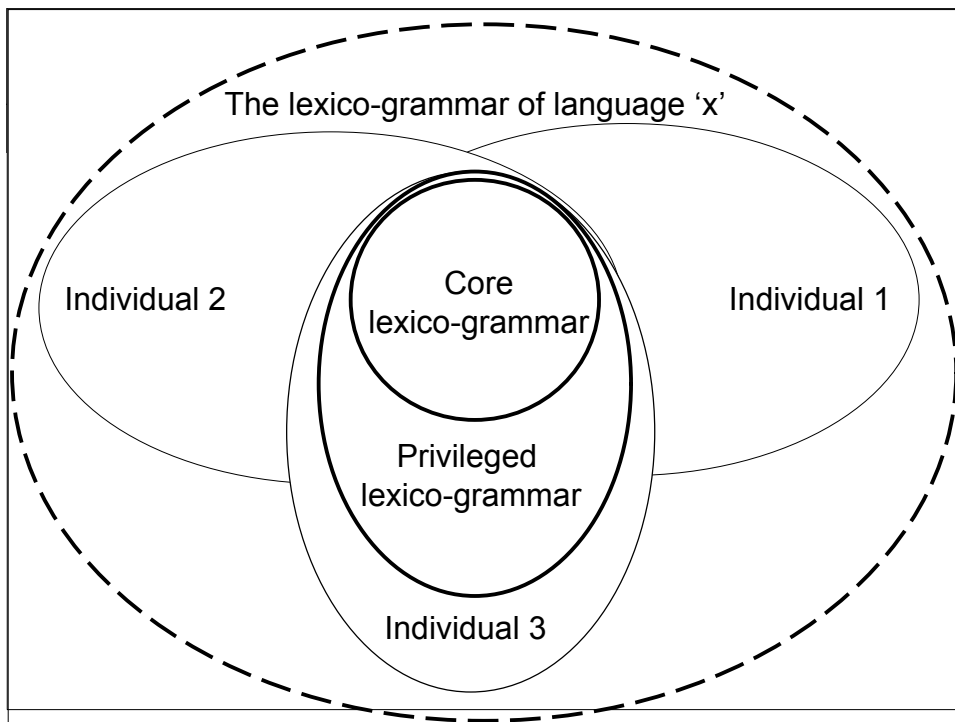
Trevor Johnston Introduction The case for SL corpus linguistics Corpus-based SL research Conclusion	annotating a sign language corpus
---	-----------------------------------

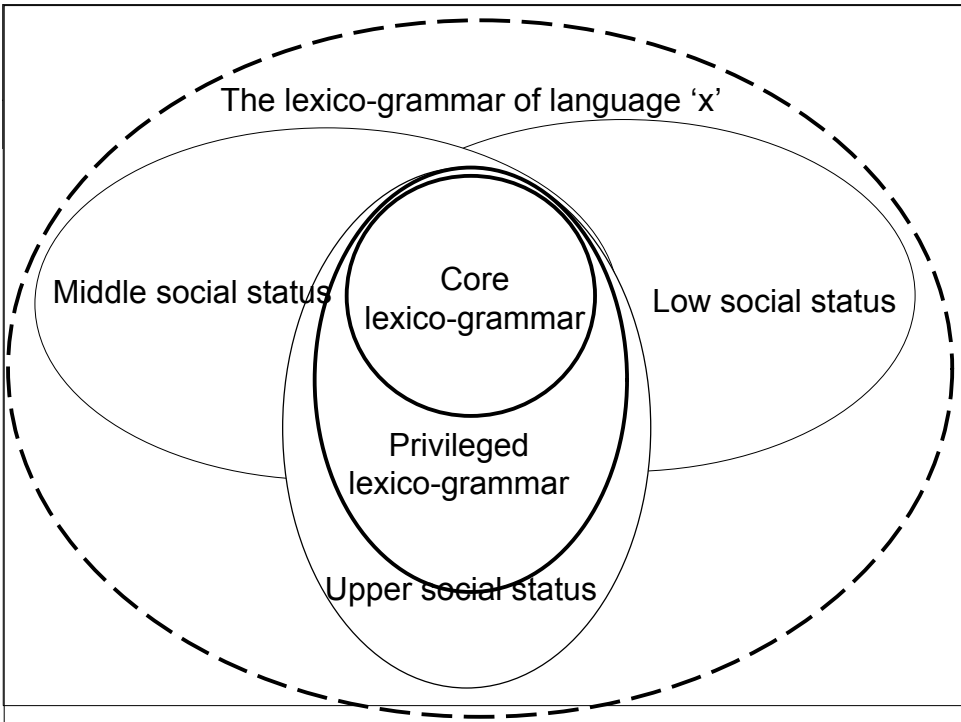
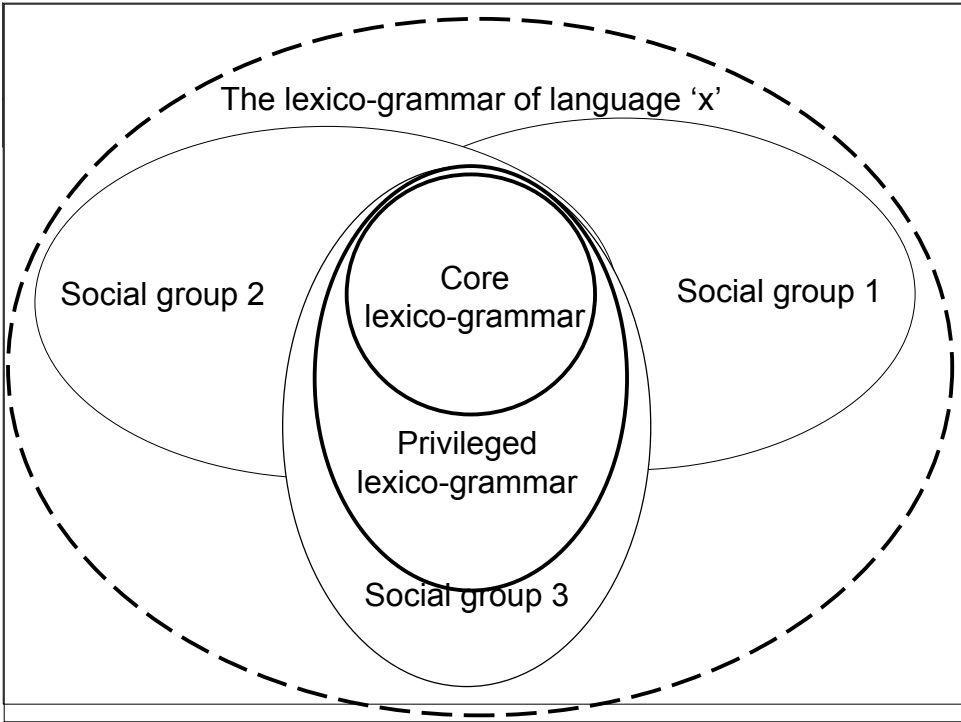
Outline

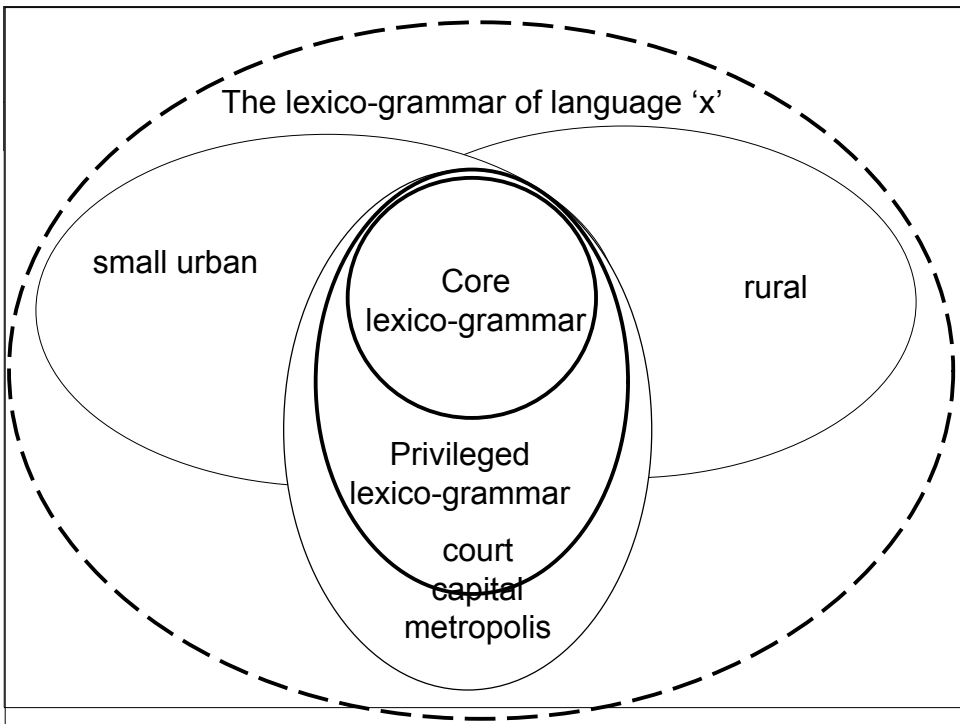
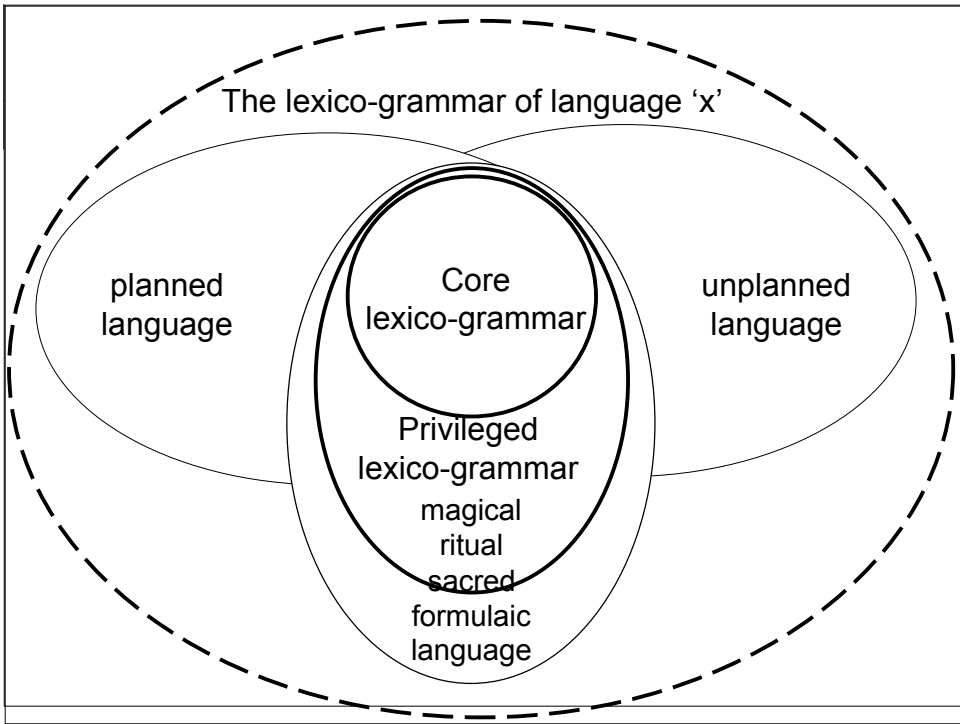
- Introduction
- The case for SL corpus linguistics
- Corpus-based SL research
- Conclusion

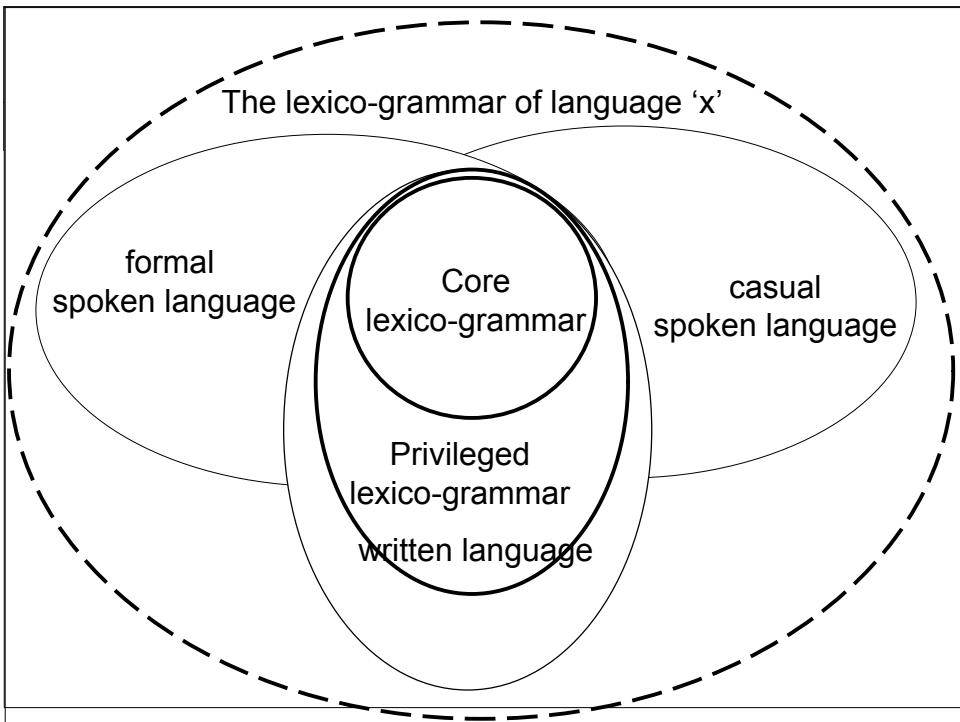
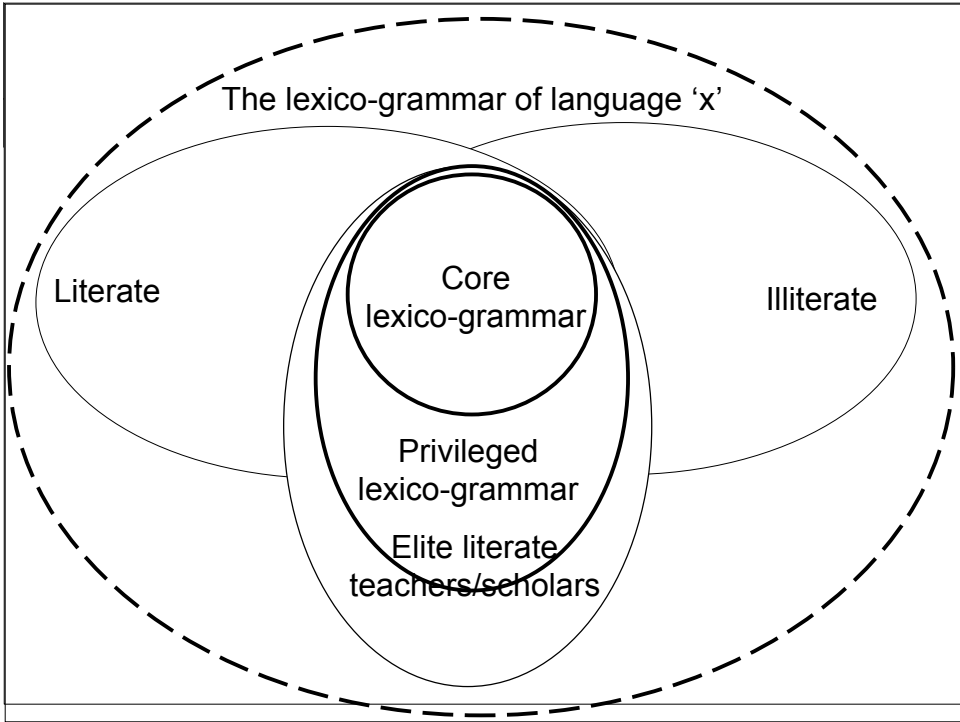
Introduction

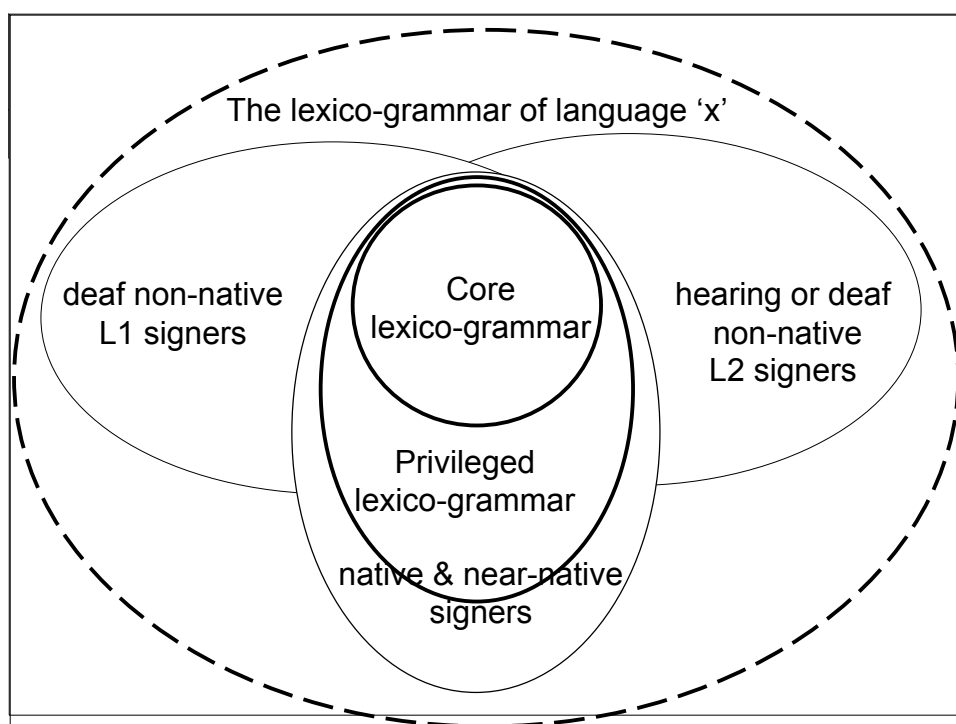
Linguistic atheism











Trevor Johnston Introduction The case for SL corpus linguistics Corpus-based SL research Conclusion	annotating a sign language corpus
The case for SL corpus linguistics	
<ul style="list-style-type: none"> • What do we want to do? • Why do we want to do it? • How do we do it? 	


Trevor Johnston Introduction The case for SL corpus linguistics Corpus-based SL research Conclusion	annotating a sign language corpus What do we want to do? Why do we want to do it? How do we do it?
What do we want to do?	
<ul style="list-style-type: none"> • empirically ground SL description • validate previous research • generate new observations • document linguistic community • create teaching/learning resources 	

Trevor Johnston Introduction The case for SL corpus linguistics Corpus-based SL research Conclusion	annotating a sign language corpus What do we want to do? Why do we want to do it? How do we do it?
Why do we want to do it?	
<ul style="list-style-type: none"> • no easily or commonly used written form • lack of language documentation <ul style="list-style-type: none"> – cf. preservation • language endangerment <ul style="list-style-type: none"> – cf. maintenance, revitalization • limits to intuitions and introspection • unique usage/acquisition environments • difficult for learners to gain exposure 	

Trevor Johnston Introduction The case for SL corpus linguistics Corpus-based SL research Conclusion	annotating a sign language corpus What do we want to do? Why do we want to do it? How do we do it?
How do we do it?	
<ul style="list-style-type: none"> • create language archives <ul style="list-style-type: none"> – i.e., documentary linguistics • adopt a corpus-based approach • value-add to language archives using <ul style="list-style-type: none"> – multi-media annotation software – annotation, not necessarily transcription – systematic linguistic tagging – controlled gloss-based annotations (ID-glosses) • open access for researchers and community <ul style="list-style-type: none"> – learners and teachers 	

Trevor Johnston Introduction The case for SL corpus linguistics Corpus-based SL research Conclusion	annotating a sign language corpus What do we want to do? Why do we want to do it? How do we do it?
Annotation, not necessarily transcription	
<ul style="list-style-type: none"> • Notation = Symbol system • Transcription = Writing system • Annotation = Appended notes • Tagging = Appended codes 	

Trevor Johnston Introduction The case for SL corpus linguistics Corpus-based SL research Conclusion	annotating a sign language corpus What do we want to do? Why do we want to do it? How do we do it?
Notation & transcription	
<ul style="list-style-type: none"> • Notation: representation of language units (e.g., phonemes, morphemes, words or signs) using a dedicated graphic symbol system <ul style="list-style-type: none"> – enables the reader reconstruct the uttered unit, depending on the degree of detail in the system 	

Trevor Johnston Introduction The case for SL corpus linguistics Corpus-based SL research Conclusion	annotating a sign language corpus What do we want to do? Why do we want to do it? How do we do it?
Notation using HamNoSys	
<p>GREEN</p>  <p>↓⁵ ^ 0 n (← →) +</p>	

Trevor Johnston Introduction The case for SL corpus linguistics Corpus-based SL research Conclusion	annotating a sign language corpus What do we want to do? Why do we want to do it? How do we do it?
Notation & transcription	
<ul style="list-style-type: none"> • Notation: representation of language units (e.g., phonemes, morphemes, words or signs) using a dedicated graphic symbol system <ul style="list-style-type: none"> – enables the reader reconstruct the uttered unit, depending on the degree of detail in the system • Transcription overlaps with notation, but <ul style="list-style-type: none"> – usually refers to representation of extended utterances (texts) rather than just isolated words/signs – consciously tries to capture much more of the act of articulation than any writing system ever does 	

Trevor Johnston Introduction The case for SL corpus linguistics Corpus-based SL research Conclusion	annotating a sign language corpus What do we want to do? Why do we want to do it? How do we do it?				
SL transcription					
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 2px;">PRO.1</td> <td style="padding: 2px;">finish</td> <td style="padding: 2px;">give</td> <td style="padding: 2px;">week-PL.2-fut.TEMP.past</td> </tr> </table>		PRO.1	finish	give	week-PL.2-fut.TEMP.past
PRO.1	finish	give	week-PL.2-fut.TEMP.past		
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 2px;">I gave it back to you two weeks ago</td> </tr> </table>		I gave it back to you two weeks ago			
I gave it back to you two weeks ago					
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 2px;"><i>I gave it from me to you two weeks ago</i></td> </tr> </table>		<i>I gave it from me to you two weeks ago</i>			
<i>I gave it from me to you two weeks ago</i>					
<p>The above is an example of interlinear text with</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr><td style="padding: 2px;">1. transcription</td></tr> <tr><td style="padding: 2px;">2. glossing</td></tr> <tr><td style="padding: 2px;">3. free translation</td></tr> <tr><td style="padding: 2px;">4. literal translation</td></tr> </table>		1. transcription	2. glossing	3. free translation	4. literal translation
1. transcription					
2. glossing					
3. free translation					
4. literal translation					

Trevor Johnston Introduction The case for SL corpus linguistics Corpus-based SL research Conclusion	annotating a sign language corpus What do we want to do? Why do we want to do it? How do we do it?
Is notation/transcription necessary?	
<ul style="list-style-type: none"> • YES, notation is required <ul style="list-style-type: none"> – for detailed phonological analysis – for sorting lexical entries by form (pronunciation) • NO, transcription is not necessary <ul style="list-style-type: none"> – a (written) text is not essential prerequisite for multi-media corpus linguistics <ul style="list-style-type: none"> • sign form can be seen in time-aligned video – one simply needs to identify relevant linguistic units (words/signs) and one can then undertake morphosyntactic, phrase, clause, utterance or discourse level analysis of constructions or structures <ul style="list-style-type: none"> • i.e., the sign or extended utterance does not have to be represented (transcribed) before it can be analysed 	

Trevor Johnston Introduction The case for SL corpus linguistics Corpus-based SL research Conclusion	annotating a sign language corpus What do we want to do? Why do we want to do it? How do we do it?
Is this transcription or annotation?	
<p>PRO1sg FINISH 1-GIVE-2 TWO-WEEKS-AGO I gave it (back) to you two weeks ago</p> <p>It is neither:</p> <p>≠ transcription because apart from the attempt to specify the beginning and end points of GIVE (as “1” and “2”) nothing indicates the form of the utterance</p> <p>≠ annotation because there are no utterance units (no recording or no transcription) to which the annotations are attached or appended</p>	

Trevor Johnston Introduction The case for SL corpus linguistics Corpus-based SL research Conclusion	annotating a sign language corpus What do we want to do? Why do we want to do it? How do we do it?
<h2>Is this transcription?</h2>	
<p>The screenshot shows a video player window titled 'Elan - STCA1c2b.eaf'. On the left, a man is signing. On the right, a list of glosses is displayed, including 'RABBIT', 'TURTLE', and 'STORY'. A large, bold, black text overlay reads 'NO, IT IS ANNOTATION' across the center of the video player. Below the video, a timeline and a list of glosses are visible, with 'RABBIT', 'TURTLE', and 'STORY' highlighted.</p>	

Trevor Johnston Introduction The case for SL corpus linguistics Corpus-based SL research Conclusion	annotating a sign language corpus What do we want to do? Why do we want to do it? How do we do it?
<h2>Annotation</h2>	
<ul style="list-style-type: none"> • linguistic ‘commentaries’ appended to identified units in a language • add phonological, lexical, morphological, syntactic, semantic, pragmatic and discourse information about linguistic forms • invaluable aid in helping linguists discern patterns in language at many different levels, with or without the aid of computers 	

Trevor Johnston Introduction The case for SL corpus linguistics Corpus-based SL research Conclusion	annotating a sign language corpus What do we want to do? Why do we want to do it? How do we do it?
Tagging	
<ul style="list-style-type: none"> • no clear cut distinction between an annotation and a tag <ul style="list-style-type: none"> – both are linguistically relevant information appended to a unit of language • however, what is now commonly called ‘tagging’ refers particularly to the kind of automatic annotations appended to written texts after they have been digitized and then processed using computers <ul style="list-style-type: none"> – e.g., part of speech tagging 	

Trevor Johnston Introduction The case for SL corpus linguistics Corpus-based SL research Conclusion	annotating a sign language corpus What do we want to do? Why do we want to do it? How do we do it?
Tags: horizontal v. ‘vertical’	
<p><i>Horizontal, e.g.,</i></p> <p>Joanna_NP stubbed_VBD out_RP her_PP\$ cigarette_NN with_IN unnecessary_JJ fierceness_NN ._. – tags, e.g. _NP for singular proper noun appended to the written text</p> <p><i>Vertical, e.g.,</i></p> <p>ELAN annotations/tags are tiered or ‘vertical’ rather than sequential.</p>	

Trevor Johnston Introduction The case for SL corpus linguistics Corpus-based SL research Conclusion	annotating a sign language corpus What do we want to do? Why do we want to do it? How do we do it?
---	---

Tiers & annotation/tags

RH ID gloss

- RH CA co-occ
- RH mod
- RH rec R
- RH-gram cls
- RH-h/s
- RH-orient
- RH-mov
- RH-ar/ment
- RH meaning
- RH mouthing
- R mouth-gc
- RH brow
- R aspect-form

R aspect-meaning

LH ID gloss

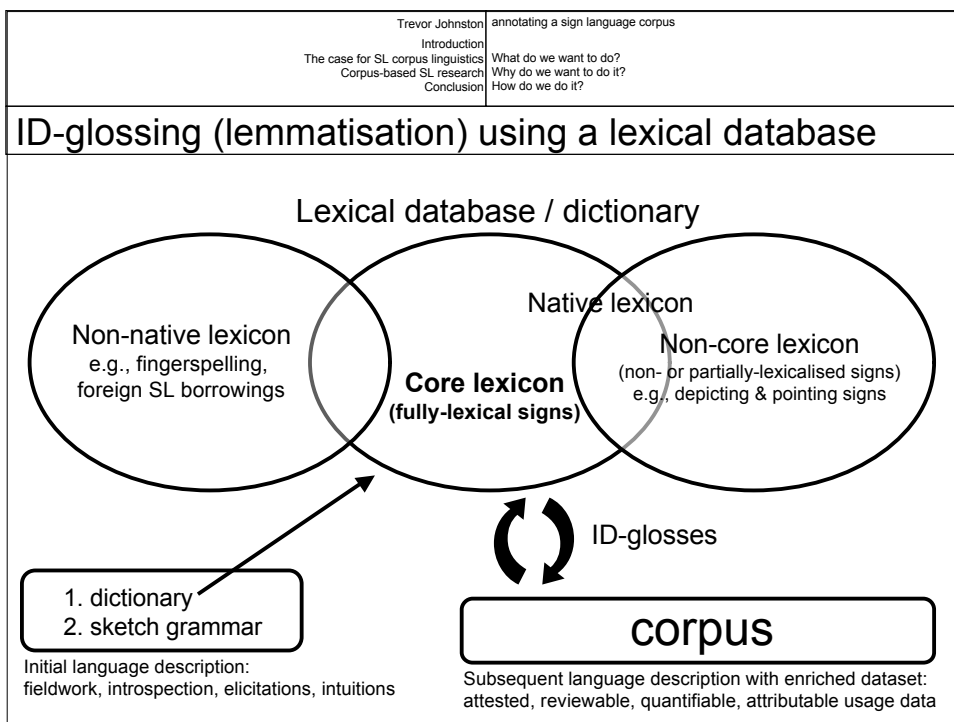
- LH CA co-occ
- LH mod
- Ref rec L
- LH-gram cls
- LH-loc
- LH-h/s
- LH-mov
- LH-orient
- LH-ar/ment
- LH meaning
- LH mouthing
- L mouth-gc
- L aspect-form
- L aspect-meaning
- LH brow

Notes

- CA/roleshift
- Body
- Gaze
- Head
- lit t/lation
- free t/lation
- transcription
- Annotator
- clause

- RH ID gloss = unique identifying gloss
- RH-gram cls = grammatical class
 - NP = plain noun
 - VP = plain verbs
 - VDir = indicating directional verb
 - VILoc = indicating locatable verb
 - ADJ = adjective
- RH mod = spatial modification
 - m = modified
 - n = not modified
 - na = not applicable

Trevor Johnston Introduction The case for SL corpus linguistics Corpus-based SL research Conclusion	annotating a sign language corpus What do we want to do? Why do we want to do it? How do we do it?
ID-glossing as lemmatization	
<ul style="list-style-type: none"> • Lemmatization <ul style="list-style-type: none"> – ‘book’, ‘books’ are forms of the lemma BOOK – ‘walk’, ‘walks’, ‘walked’, ‘walking’ forms of lemma WALK • ID-glossing (“lexical annotation”) is essentially lemmatization <ul style="list-style-type: none"> – for SLs, the citation form is analogous to the lemma – note: explicit lexical annotation conventions are needed for use with partly- or non-lexicalized signs (e.g., points, depicting signs, etc.) • Other tiers contain formational and grammatical information about the signs <ul style="list-style-type: none"> – grammatical class – grammatical/semantic/thematic roles – modification – phonetic/phonological transcriptions (or simply tags) <p style="text-align: center;">So no information is lost</p>	



Trevor Johnston Introduction The case for SL corpus linguistics Corpus-based SL research Conclusion	annotating a sign language corpus
Corpus-based SL research	
<p style="text-align: center;">Some example searches based on annotations</p> <ul style="list-style-type: none"> • Single sign searches <ul style="list-style-type: none"> – Types/tokens – Frequency statistics • Multiple sign searches <ul style="list-style-type: none"> – Concordance patterns and/or constructional schemas – Contextual constraints 	

Trevor Johnston Introduction The case for SL corpus linguistics Corpus-based SL research Conclusion	annotating a sign language corpus Single sign searches: types/tokens Single sign searches: frequency statistics Multiple sign searches: concordance patterns Multiple sign searches: contextually constrained
Single sign searches: types/tokens	
<ul style="list-style-type: none"> • A search for any ID-gloss is a search based on a type <ul style="list-style-type: none"> – the hits are the tokens which may be viewed in context (concordance) 	

Substring Search Single Layer Search Multiple Layer Search

Domain: 186 eaf files Define New Domain

Query History: < >

Mode: Annotation ↓ case sensitive ↓ exact match ↓

Find Tier Name: RH ID gloss ↓

Frequency view would be meaningless since all hits (matches) = LOOK (i.e., 100%)

#hits : 589
 #annotations with a hit : 589
 #annotations investigated : 126177

Ready

>

hit 1 - 11 of 589

*BOY PT:PRO3sg CHAIR2 LOOK GREY BIG DS(1):upright-animal-walk
 GREY BIG DS(1):upright-animal-walk LOOK WOLF YELL2-2h WOLF
 DO G:well JUST-SE LOOK SHEEP G:well-that's-it GRAZE
 GRAZE DS(bO):sheep-looking DS(2bent):ear-of-sheep LOOK WHAT G:well-what? DS(5):sheep
 WHAT G:well-what? DS(5):sheep LOOK DS(6):hold-stick WHAT DS(5):people
 NOT FS:BUT SHEEP LOOK G:well PEACE G:that's-it
 TWO FARM THAT'S-ALL LOOK G:well LOOK
 THAT'S-ALL LOOK G:well LOOK SAY REAL
 DS(A):turtle-plodding RABBIT DS(2bent):rabbit-move LOOK WHERE PT:LOC3sg SOON
 FS:HILL HERD GRAZE LOOK SUNSET AFTERNOON SUNSET
 LAUGH CRY LAUGH LOOK ANGRY MAKE PT:PRO3pl*

Trevor Johnston Introduction The case for SL corpus linguistics Corpus-based SL research Conclusion	annotating a sign language corpus Single sign searches: types/tokens Single sign searches: frequency statistics Multiple sign searches: concordance patterns Multiple sign searches: contextually constrained
Single sign searches: types/tokens	
<ul style="list-style-type: none"> • A search for any ID-gloss is a search based on a type <ul style="list-style-type: none"> – the hits are the tokens which may be viewed in context (concordance) • Searches may be constrained by features of the token tagged on other tiers <ul style="list-style-type: none"> – e.g., <ol style="list-style-type: none"> 1. RH ID-gloss = “x” 2. RH mod = n or m (m n) 	

Substring Search Single Layer Search Multiple Layer Search

Domain: 186 eaf files Define New Domain

Query History: < >

Mode: case sensitive regular expression Clear

Minimal Duration Maximal Duration Begin After End Before

^LOOK\$ Tier Name: RH ID gloss
 Fully aligned
 mln Tier Name: RH mod
 All Tiers

Find

**Frequency view is used here as it is very meaningful (there is more than one match):
LOOK modified = 88%, LOOK unmodified = 6%**

#hits : 94
#annotations with a hit : 94
#annotations investigated : 126177

Ready

frequency 1 - 2 of 2

Annotation	Percentage	Count
#1 LOOK #2 ml #3	93.62%	88
#1 LOOK #2 ln #3	6.38%	6

Trevor Johnston	annotating a sign language corpus
Introduction	Single sign searches: types/tokens
The case for SL corpus linguistics	Single sign searches: frequency statistics
Corpus-based SL research	Multiple sign searches: concordance patterns
Conclusion	Multiple sign searches: contextually constrained

Single signs searches: frequency

- Hits from a search which finds more than one type or sorts different tokens can be viewed (meaningfully) as a frequency list

Trevor Johnston		annotating a sign language corpus
Introduction		Single sign searches: types/tokens
The case for SL corpus linguistics		Single sign searches: frequency statistics
Corpus-based SL research		Multiple sign searches: concordance patterns
Conclusion		Multiple sign searches: contextually constrained

Auslan corpus frequency (50 text subset of 202)

rank	gloss	numberBOTH
1	PT:PRO1	406
2	PT:PRO3	233
3	PT:DEM/LOC	221
4	G:well	218
5	LOOK	121
6	TURTLE	116
7	GOOD	115
8	BOY	107
9	WOLF	100
10	RABBIT	90
11	PEOPLE	86
12	SAY	80
13	REAL	70
14	HAVE	67
15	PT:PRO2	67
16	ONE	64
17	SHEEP	61
18	AGAIN	58
19	SAME	56
20	WHAT	56
21	LAUGH	53
22	DAY	51
23	PT:POSS1	51
24	NOTHING	49
25	THINK	48
26	WHY	48
27	ALL	45
28	CALL	44
29	COME	44
30	NOT	44
31	SLEEP	44
32	WANT	43
33	ARRIVE	42
34	WITH	40
35	SLOW	37
36	KNOW	35
37	SEE	35
38	ALWAYS	33
39	WILL	33
40	GO	32
41	SPRINT	32
42	FS:VILLAGE	31
43	CONTINUE	29
44	NIGHT	28
45	DS:(B)turtle-moving	27
46	MORNING	27
47	STORY	25
48	BAD	24
49	CAN-NOT	23
50	GRAZE	23
51	PAST	23
52	STOP	23

- 5,000 type hit limit!
- 50 texts (c. 10,000 tokens)
- 202 texts (c. 43,000 tokens)
- Signs ranked 1, 2, 3, 4, 15, 23, 30, 39, 51 are grammatical (function) signs
- summed they are comparable to most frequent as % of all tokens in SpL corpora where grammatical or function words occupy the top frequencies

Trevor Johnston		annotating a sign language corpus
Introduction		Single sign searches: types/tokens
The case for SL corpus linguistics		Single sign searches: frequency statistics
Corpus-based SL research		Multiple sign searches: concordance patterns
Conclusion		Multiple sign searches: contextually constrained

^[\^QDS\E|\QG:\E|\QFS\E|\QPT\E|\Q?\E]

Domain: 186 eaf files

Query History: < >

Mode: Annotation | case sensitive | regular expression

Find: ^[\^QDS\E|\QG:\E|\QFS\E|\QPT\E|\Q?\E] Tier Name: RH ID gloss

#hits : 27112

#annotations with a hit : 27112

#annotations investigated : 126177

Ready

hit 1 - 16 of 27112

```

ONE BOY DS(1):person-walk HAVE
ONE BOY DS(1):person-walk HAVE FS-VILLAGE
ONE BOY DS(1):person-walk HAVE FS-VILLAGE DS(5):houses-located-in-a-group i
HAVE FS-VILLAGE DS(5):houses-located-in-a-group BOY PT-LOC ALWAYS MORNING
DS(5):houses-located-in-a-group BOY PT-LOC ALWAYS MORNING ALWAYS NIGHT2
BOY PT-LOC ALWAYS MORNING ALWAYS NIGHT2 DS(c):move-group
PT-LOC ALWAYS MORNING ALWAYS NIGHT2 DS(c):move-group SHEEP2
ALWAYS MORNING ALWAYS NIGHT2 DS(c):move-group SHEEP2 DS(C):move-group
ALWAYS NIGHT2 DS(c):move-group SHEEP2 DS(C):move-group FS-PASTURE DINNER
SHEEP2 DS(C):move-group FS-PASTURE DINNER FS-GRASS DINNER ALWAYS
FS-PASTURE DINNER FS-GRASS DINNER ALWAYS NIGHT2 DS(5):move-group BOY
DINNER FS-GRASS DINNER ALWAYS NIGHT2 DS(5):move-group BOY
FS-GRASS DINNER ALWAYS NIGHT2 DS(5):move-group BOY PT-LOC
ALWAYS NIGHT2 DS(5):move-group BOY PT-LOC BORED WHY
DS(5):move-group BOY PT-LOC BORED WHY DINNER ALL-DAY
BOY PT-LOC BORED WHY DINNER ALL-DAY MORNING

```

Domain: 186 eaf files

Query History: < >

Mode: Annotation | case sensitive | regular expression

Find: ^[\^QDS\E|\QG:\E|\QFS\E|\QPT\E|\Q?\E] Tier Name: RH ID gloss

#hits : 27112

#annotations with a hit : 27112

#annotations investigated : 126177

Ready

frequency 1 - 15 of 2791

Annotation	Percentage	Count
DEAF	2.27%	615
LOOK	2.17%	589
BOY	1.84%	498
SAME	1.75%	475
HAVE	1.58%	429
THINK	1.20%	326
NOTHING	1.19%	322
GOOD	1.18%	319
WHAT	1.05%	284
WHY	1.04%	282
REAL	1.03%	278
NOT	1.00%	272
PEOPLE	0.98%	266
SIGN	0.96%	260
FROG	0.94%	254

Substring Search Single Layer Search Multiple Layer Search

Domain: 186 eaf files **186 files** Define New Domain

Query History: < >

Mode: Annotation case sensitive regular expression

Find: ^[^\QDS\ENQG:\EINQFS\EINQPT\EINQ?E] Tier Name: RH ID gloss

#hits : 27112 **27,112 annotations**
 #annotations with a hit : 27112
 #annotations investigated : 126177 Ready

frequency 1 - 15 of 2791 **2,791 types**

Annotation	Percentage	Count
DEAF	2.27%	615
LOOK	2.17%	589
BOY	1.84%	498
SAME	1.75%	475
HAVE	1.58%	429
THINK	1.20%	326
NOTHING	1.19%	322
GOOD	1.18%	319
WHAT	1.05%	284
WHY	1.04%	282
REAL	1.03%	278
NOT	1.00%	272
PEOPLE	0.98%	266
SIGN	0.96%	260
FROG	0.94%	254

Trevor Johnston annotating a sign language corpus

Introduction The case for SL corpus linguistics Single sign searches: types/tokens

Corpus-based SL research Multiple sign searches: frequency statistics

Conclusion Multiple sign searches: concordance patterns

Multiple sign searches: contextually constrained

PT:PRO.+?(

Search eaf files

Substring Search Single Layer Search Multiple Layer Search

Domain: 201 eaf files Define New Domain

Query History: < >

Mode: Annotation case insensitive regular expression

Find: PT:PRO.+?(\ Tier Name: RH ID gloss

#hits : 162
 #annotations with a hit : 162
 #annotations investigated : 129693 Ready

frequency 1 - 19 of 19

Annotation	Percentage	Count
PT:PRO1sg(B)	67.28%	109
PT:PRO1sg(5)	8.64%	14
PT:PRO3sg(7)	3.70%	6
PT:PRO1sg(7)	2.47%	4
PT:PRO1pl(2)	2.47%	4
PT:PRO1sg(A)	1.85%	3
PT:PRO1sg(7)-2h	1.85%	3
PT:PRO1sg(6)	1.85%	3
PT:PRO3sg(H)	1.23%	2
PT:PRO3sg(B)	1.23%	2
PT:PRO3sg(5)	1.23%	2
PT:PRO3pl(P)	1.23%	2
PT:PRO1pl(B)	1.23%	2
PT:PRO3sg(A)	0.62%	1
PT:PRO3sg(6)	0.62%	1
PT:PRO2sg(7)	0.62%	1
PT:PRO1sg(B)-2h	0.62%	1
PT:PRO1sg(8)	0.62%	1
PT:PRO1pl(3)	0.62%	1

- Finds all pronouns coded with a variant handshape
 - i.e., ~1 handshape.
 - Of course, variants must be coded FIRST!
- Can also be narrowed further:
 - ^PT:PRO1.*?(\
 - ^PT:PRO1sg.*?(\

Search eaf files

Substring Search | Single Layer Search | Multiple Layer Search

Domain: 201 eaf files Define New Domain

Query History: < >

Mode: Annotation | case insensitive | regular expression

Find Tier Name: RH ID gloss

#hits : 162
#annotations with a hit : 162
#annotations investigated : 129693

Ready

PT:PRO1sg(B)

frequency 1 - 19 of 19

Annotation	Percentage	Count
PT:PRO1sg(B)	67.28%	109
PT:PRO1sg(3)	8.64%	14
PT:PRO3sg(7)	3.70%	6
PT:PRO1sg(7)	2.47%	4
PT:PRO1pl(2)	2.47%	4
PT:PRO1sg(A)	1.85%	3
PT:PRO1sg(7)-2h	1.85%	3
PT:PRO1sg(6)	1.85%	3
PT:PRO3sg(H)	1.23%	2
PT:PRO3sg(B)	1.23%	2
PT:PRO3sg(5)	1.23%	2
PT:PRO3pl(P)	1.23%	2
PT:PRO1pl(B)	1.23%	2
PT:PRO3sg(A)	0.62%	1
PT:PRO3sg(6)	0.62%	1
PT:PRO2sg(7)	0.62%	1
PT:PRO1sg(B)-2h	0.62%	1
PT:PRO1sg(8)	0.62%	1
PT:PRO1pl(3)	0.62%	1

Trevor Johnston | annotating a sign language corpus

Introduction | The case for SL corpus linguistics | Single sign searches: types/tokens

Corpus-based SL research | Single sign searches: frequency statistics

Conclusion | Multiple sign searches: concordance patterns

Multiple sign searches: contextually constrained

Comparing subordinate hand annotation

Substring Search | Single Layer Search | Multiple Layer Search

Domain: 35 eaf files Define New Domain

Query History: < >

Mode: case sensitive | regular expression | Clear

Minimal Duration | Maximal Duration | Begin After | End Before

FINISH Tier Name: RH ID gloss

Fully aligned

^.*? \$ Tier Name: LH ID gloss

All Tiers

Find

#hits : 46
#annotations with a hit : 46
#annotations investigated : 45279

Ready

frequency 1 - 6 of 6

Annotation	Percentage	Count
#1 FINISH-GOOD-2h #2 FINISH-GOOD-2h ...	63.04%	29
#1 FINISH-GOOD #2 #3 ...	17.39%	8
#1 FINISH-FIVE #2 FINISH-FIVE #3 ...	8.70%	4
FINISH-GOODto5-2h #2 FINISH-GOODto5>...	6.32%	3
#1 FS:FINISH #2 FS:FINISH #3 ...	2.17%	1
#1 FINISH-GOODto5 #2 #3 ...	2.17%	1

- Reg. exp. `^.*?$` will find an empty or full field
 - it finds a beginning to a field and an end to a field with or without anything in between
- Useful for identifying alternative 1 and 2 handed forms of signs (cf. weak drop, weak prop)

Substring Search Single Layer Search Multiple Layer Search

Domain: 35 eaf files Define New Domain

Query History: < >

Mode: case sensitive regular expression Clear

Minimal Duration Maximal Duration Begin After End Before

FINISH Fully aligned ^.*?\$

Tier Name: RH ID gloss Tier Name: LH ID gloss All Tiers

Find

FINISH-GOOD-2h = 29 hits
FINISH-GOOD = 8 hits

#hits : 46
 #annotations with a hit : 46
 #annotations investigated : 45279 Ready

frequency 1 - 6 of 6

Annotation	Percentage	Count
#1 FINISH-GOOD-2hl #2 FINISH-GOOD-2hl ...	63.04%	29
#1 FINISH-GOOD #2 #3	17.39%	8
#1 FINISH-FIVE #2 FINISH-FIVE #3	8.70%	4
FINISH-GOOD(to5)-2hl #2 FINISH-GOOD(to5)-...	6.52%	3
#1 IFS:FINISH #2 IFS:FINISH #3	2.17%	1
#1 FINISH-GOOD(to5) #2 #3	2.17%	1

Does citation form as one-handed need to be reassessed?

Trevor Johnston	annotating a sign language corpus
Introduction	
The case for SL corpus linguistics	Single sign searches: types/tokens
Corpus-based SL research	Single sign searches: frequency statistics
Conclusion	Multiple sign searches: concordance patterns
	Multiple sign searches: contextually constrained

Multiple sign searches: concordance patterns

- Searches may also be constrained for signs occurring before or after a specific sign

Trevor Johnston annotating a sign language corpus
 Introduction Single sign searches: types/tokens
 The case for SL corpus linguistics Single sign searches: frequency statistics
 Corpus-based SL research Multiple sign searches: concordance patterns
 Conclusion Multiple sign searches: contextually constrained

Compound or collocation? 1

Substring Search Single Layer Search Multiple Layer Search

Domain: 204 eaf files Define New Domain

Query History: < >

Mode: N-gram within annotation case sensitive regular expression

Find: DEAF+CLUB Tier Name: RH ID gloss

#hits : 8
 #annotations with a hit : 8
 #annotations investigated : 133559 Ready

hit 1 - 8 of 8

SAME LOOK-SOMEONE G:come-here DEAF-CLUB LEADER DS5):many-people-go KNOW
 :many-people-go INTEGRATE KNOW-NOT-2h DEAF-CLUB FUTURE G:well KNOW-NOT-2h
 G:umm PT GO DEAF-CLUB PT G:go-away PT
 REAL DEAF THINK DEAF-CLUB BIG PART-SE BIG
 FS:IF PT WANT DEAF-CLUB ? MAKE ONE
 THOUSAND AVAILABLE FOR DEAF-CLUB PEOPLE NEED APPLY
 HAVE-NOT FS:DO NOT DEAF-CLUB PT:POSS FAULT PT:POSS
 YEAR LATER HAVE-NOT DEAF-CLUB PISS-OFF-2h G:you-see FS:SO

- Search for a sequence of DEAF and CLUB
 - search in single layer
 - looking WITHIN annotations
 - “N-gram within annotation”
 - separated by “any kind of matter”
 - regular expression = .+
 - X-Y is one convention for writing a compound
 - the other is to use a unique gloss (such as TOMATO for RED-BALL)
 - one can view hits individually to confirm compound status
 - e.g., is there really phonological reduced in one of the compounded units?

Trevor Johnston annotating a sign language corpus
 Introduction Single sign searches: types/tokens
 The case for SL corpus linguistics Single sign searches: frequency statistics
 Corpus-based SL research Multiple sign searches: concordance patterns
 Conclusion Multiple sign searches: contextually constrained

Compound or collocation? 1

Substring Search Single Layer Search Multiple Layer Search

Domain: 204 eaf files Define New Domain

Query History: < >

Mode: N-gram over annotations case sensitive substring match

Find: DEAF CLUB Tier Name: RH ID gloss

#hits : 9
 #annotations with a hit : 9
 #annotations investigated : 133559 Ready

frequency 1 - 3 of 3

Annotation	Percentage	Count
DEAF FS:CLUB	55.56%	5
DEAF CLUB1	22.22%	2
DEAF CLUB	22.22%	2

- Search for a sequence of DEAF and CLUB
 - search in single layer
 - looking OVER annotations
 - “N-gram over annotation”
 - i.e. one separate annotation after another
 - a sequence of DEAF CLUB is a potential candidate for a compound
 - there are 2 that fit the criteria
 - one can view hits individually to investigate
 - e.g., if there is phonological reduction in one of the elements in these two annotations (DEAF CLUB) then it may need to be combined as one annotation (DEAF-CLUB)

Trevor Johnston | annotating a sign language corpus
 Introduction | Single sign searches: types/tokens
 The case for SL corpus linguistics | Single sign searches: frequency statistics
 Corpus-based SL research | Multiple sign searches: concordance patterns
 Conclusion | Multiple sign searches: contextually constrained

Compound or collocation? 2 (alternative search)

Substring Search | Single Layer Search | Multiple Layer Search

Domain: 204 eaf files Define New Domain

Query History: < >

Mode: case sensitive | substring match Clear

Minimal Duration | Maximal Duration | Begin After | End Before

DEAF = 0 annotations | CLUB Tier Name: RH ID gloss

All Tiers

All Tiers

Find

#hits : 9
 #annotations with a hit : 9
 #annotations investigated : 133559 Ready

Annotation	Percentage	Count
#1 IDEAFI FS:CLUBI #2 #3 #4 #5 #6 #7 #8 #9 #10	55.56%	5
#1 IDEAFI CLUBI #2 #3 #4 #5 #6 #7 #8 #9 #10	22.22%	2
#1 IDEAFI CLUBI #2 #3 #4 #5 #6 #7 #8 #9 #10	22.22%	2

- Search finds potential candidates for compound status
 - e.g., a sequence of DEAF and CLUB over two contiguous annotations.

Trevor Johnston | annotating a sign language corpus
 Introduction | Single sign searches: types/tokens
 The case for SL corpus linguistics | Single sign searches: frequency statistics
 Corpus-based SL research | Multiple sign searches: concordance patterns
 Conclusion | Multiple sign searches: contextually constrained

Anything between two know annotations

Substring Search | Single Layer Search | Multiple Layer Search

Domain: 204 eaf files Define New Domain

Query History: < >

Mode: N-gram over annotations | case sensitive | regular expression

Find: PT:PRO1 # THINK Tier Name: RH ID gloss

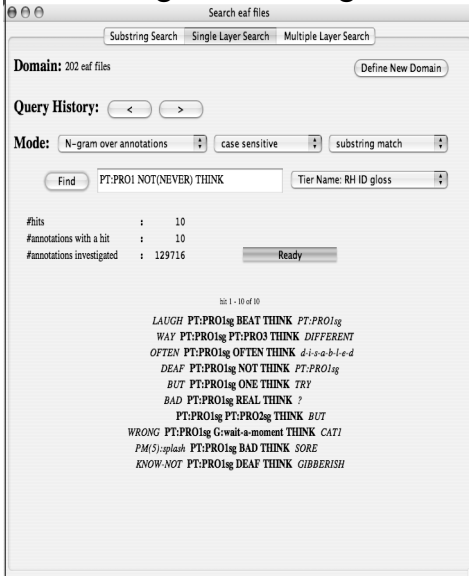
#hits : 14
 #annotations with a hit : 14
 #annotations investigated : 133559 Ready

hit 1 - 14 of 14

```
PT:PRO1p(2) RABBIT LAUGH PT:PRO1sg BEAT THINK PT:PRO1sg BEAT LAUGH
PT:PRO2sg NOT WORTH PT:PRO1sg NOT THINK ARRIVE ALWAYS ARRIVE
2):dive-backwards DS(B):ground DS(5):splash PT:PRO1sg BAD THINK SORE FALL BAD
PT:POSS1sg PARENT KNOW-NOT PT:PRO1sg DEAF THINK GIBBERISH G:don't-bother CONTINUE
PT:PRO3 PEOPLE WAY PT:PRO1sg PT:PRO3 THINK DIFFERENT SOME PEOPLE
G:well PT:PRO1sg OFTEN PT:PRO1sg OFTEN THINK FS:DISABLED SIGN LITTLE
CHILDREN2-1h PT:PRO1sg NEVER PT:PRO1sg NEVER THINK SAY PT:DEM FS:DISABLED
DEAF PEOPLE NOTHING PT:PRO1sg NEVER THINK NEVER GET-ATTENTION CAN
WILL FS:BE DEAF PT:PRO1sg NOT THINK PT:PRO1sg KNOW PT:POSS1sg
SAME PT:PRO1sg BUT PT:PRO1sg ONE THINK TRY THINK WHAT
FUTURE DEAF CHILDREN PT:PRO1sg NEVER THINK PT:PRO1sg THINK PERHAPS
G:so BUT BAD PT:PRO1sg REAL THINK > OVER WANT
PT:PRO1sg PT:PRO2sg THINK BUT SIGN PT:PRO1sg
G:oh-no SORRY WRONG PT:PRO1sg G:walt-a-moment THINK CAT DS(1):the-cat-gets-into-the-opening
```

- Finds all strings between two specified annotations
 - uses the wildcard symbol (#)
- Can do the same thing in multiple layer search grid
 - often more than one way in ELAN

Excluding something between two known annotations



- Finds a string with anything between two stated annotations (substrings) except THE stated annotation
 - instead of # (as in previous search) it uses
 - NOT(the unwanted gloss)
- Thus NOT(NEVER) will find all
 - PRO1 *any gloss* THINK but not
 - PRO1 *NEVER* THINK

Multiple sign searches: contextual constraints

- Searches may also be constrained for signs occurring before or after a specific sign
- These sequential constraints can be combined with simultaneous constraints

PT in same clause as modified sign

Substring Search Single Layer Search Multiple Layer Search

Domain: 1 eaf files Define New Domain

Query History: < >

Mode: case sensitive regular expression Clear

Minimal Duration Maximal Duration Begin After End Before

^PT Tier Name: RH ID gloss

Overlap

^\$ Tier Name: clause

Overlap

m Tier Name: RH mod

Find

#hits : 5
 #annotations with a hit : 5
 #annotations investigated : 2634 Ready

hit 1 - 5 of 5

```
#1 || PT:PRO3sg || #2 || #3 || lml ||
#1 || PT:LOC3sg || #2 || #3 || lml ||
#1 || PT:FBUOY1 || #2 || #3 || lml ||
#1 || PT:PRO2sg || #2 || #3 || lml ||
#1 || PT:PRO1sg || #2 || #3 || lml ||
```

- Finds a pointing sign (^PT) which is part of a clause (^\$) which contains a modified sign (m)
 - see next slide for example ELAN hit

Grid Text Subtitles Metadata Controls

RH ID gloss

PAST · SCHOOL · PT:PRO1sg · TROUBLE · PT:PRO1sg · DS(1):person-moves-to · BOSS · FS:OFFICE · REPEAT · FOR2 · MAN · BOSS · FS:MR-FS:DAY · DS(1):leg-twitch · WARN · G:scratch-balls · PT:PRO1sg · DS(F):eye-twitches · REPEAT · ONE · DAY · PT:PRO1sg · SWEET · PT:LOC · HAVE · FS:SCAFFOLD · DS(2):scaffolding · HIGH · PT:PRO1p(2) · BOY · FS:ALFA · SN:ALFA · PT:PRO1p(2) · DECIDE · LET'S-SEE · DS(2):boy-climbs-up-high · DS(2):boy-stops-suddenly-at-top · WHERE · DS(B):passageway · KNOW · FS:HALL · DOOR · HAVE · DS(B):long-pointed-roof · DS(4):long-poles · DS(S):long-poles · DS(c):long-poles · PT:LOC3sg · DS(B):long-pointed-roof · FOR2 · PAINT · PT:FBUOY · PT:PRO1sg · DS(2):person-stand · PT:LOC · DS(S):hold-onto-bars-and-sway · DS(B):scaffold-swaps · LAUGH · DS(S):hold-onto-bars-and-sway · DS(B):scaffold-swaps · BAD · LOUD · DS(B):scaffold-swaps · BETTER · STOP-2h · LAUGH · STOP-2h ·

00:00:13.051 Selection: 00:00:12.390 - 00:00:13.950 1560

Selection Mode Loop Mode

RH ID gloss [245]	PT:PRO	DS(F):eye-twitches	REPEAT	ONE	DAY
RH-gram cls [240]	Pro	VD	Adv	Det	Nloc
RH mod [137]		m			n
RH meaning [2]					
LH ID gloss [232]		DS(F):eye			DAY
LH-gram cls [225]		VD			Nloc
LH mod [122]					n
CA:roleshift [44]		CA:gavin/self			
clause [105]					

Depicting sign in clause with other signs

Substring Search Single Layer Search Multiple Layer Search

Domain: 1 caf files Define New Domain

Query History: < >

Mode: case insensitive regular expression Clear

Minimal Duration Maximal Duration Begin After End Before

Overlap

Tier Name: RH ID gloss

Overlap

Tier Name: Clause

Overlap

Tier Name: RH-gram ...

Find

#hits : 7
 #annotations with a hit : 7
 #annotations investigated : 1674 Ready

hit 1 - 7 of 7

```
#1 || ONEI || #2 || 101 || #3 || IVDI ||
#1 || IDAYI || #2 || 101 || #3 || IVDI ||
#1 || IWATI || #2 || 101 || #3 || IVDI ||
#1 || IRABBITI || #2 || 101 || #3 || IVDI ||
#1 || IRABBITI || #2 || 134 || #3 || IVDI ||
#1 || IPT:PRO3sg || #2 || 137 || #3 || IVDI ||
#1 || TREEI || #2 || 146 || #3 || IVDI ||
```

- Finds a RH ID gloss that DOES NOT begin with "DS" (i.e., any gloss but a depicting sign gloss)
 - Reg. exp. = `^[^QDSVE]`
- which overlaps a clause tier annotation which is empty or has something in it (e.g. # or α or β)
 - Reg. exp. = `^,+ $`
- which ALSO overlaps a grammatical class tier label which is VD (i.e., which IS a depicting sign, after all).
- Finds all clauses that have a DS and at least one other sign. With numbered clauses, one can export the hits to Excel and sort the hits temporally.
 - Very useful for finding DS-containing clauses that have other signs as well, so one can inspect them. Can make manual coding and analysis much quicker and easier.

File Edit Annotation Tier Type Search View Options Window Help

Grid Text Subtitles Metadata Controls

RH ID gloss

G:come-on · TURKLE · LOOK · PT:PRO2sg · WANT · BEI · RACE · ARRIVE · FIRST · WANT · BET · PT:PRO2sg · RABBIT · YES · WANT · PT:PRO2sg · G:why-not · FINE · GOOD · BYE · SPEED-DUST · DS(2bent):animal-runs · SPEED-DUST · G:there-you-go · RABBIT · TURKLE · G:forget-about-him · DS(B):turtle-moves · RABBIT · DS(2bent):animal-runs · PT:DEM · COINCIDENCE · PT:DEM · DAY · WARM · SUN-SHINES · PT:PRO3sg · DS(2bent):animal-runs-then-slows · LITTLE · TIRED · WORK-OUT · WHY-NOT · SLEEP-2h · WHILE · WHY · FAR · TURKLE · G:forget-about-him · SLEEP-2h · WHILE

00:00:51.590 Selection: 00:00:51.590 - 00:00:53.295 1705

RH ID gloss [193] RAB DS(2bent):animal-runs

Destination tie [193]

RH-gram cls [188] NP VD

LH ID gloss [124] RAB DS(2bent):animal-runs

Clause [79] 34

Trevor Johnston Introduction The case for SL corpus linguistics Corpus-based SL research Conclusion	annotating a sign language corpus Single sign searches: types/tokens Single sign searches: frequency statistics Multiple sign searches: concordance patterns Multiple sign searches: contextually constrained
Clause arguments tier	
A A1 A2 A3 A4 V V1 V2 V3 V4 nonA	a single overt argument of a verb a first overt argument of a verb (when there are more than one) a second overt argument of a verb a third overt argument of a verb a fourth overt argument of a verb a verb a first verb in a serial verb construction a second verb in a serial verb construction a third verb in a serial verb construction a fourth verb in a serial verb construction, and so on. an element of a clause which cannot be construed as an argument. It contributes temporal, location, purposive/reason, verbal auxiliary etc. information to the clause, but is not a 'participant' (argument) or 'process' (verb), as such.

Trevor Johnston Introduction The case for SL corpus linguistics Corpus-based SL research Conclusion	annotating a sign language corpus Single sign searches: types/tokens Single sign searches: frequency statistics Multiple sign searches: concordance patterns Multiple sign searches: contextually constrained
Semantic-macro roles	
ACTR UNDR UNDR1 UNDR2 UNDR3 CARRIER ATTRIB	an Actor-like argument of a verb ('Subject*') an Undergoer, i.e., a non-Actor-like argument of a verb ('Object') a first Undergoer when there is more than one ('Indirect Object') a second Undergoer ('other Object') a third Undergoer ('yet another Object'). argument in verbless clause of which the other argument is the attribute argument in verbless clause which names an attribute of the other argument
<p>* Note: 'Subject' and 'Object' terminology is meant in only the most general possible way. Essentially, at this level of analysis the terminology is misleading. It does not mean the grammatical relations of subject and object.</p>	

Trevor Johnston	annotating a sign language corpus
Introduction	
The case for SL corpus linguistics	Single sign searches: types/tokens
Corpus-based SL research	Single sign searches: frequency statistics
Conclusion	Multiple sign searches: concordance patterns
	Multiple sign searches: contextually constrained

Basic semantic roles

AGENT	agent
BEN	benefactive, recipient
EXP	experiencer
GOAL	goal
INST	instrument
LOC	locative
PATIENT	patient
SOURCE	source

The screenshot shows the Elan software interface with a video of a sign language user on the left. The right side displays a detailed annotation timeline. The timeline includes the following fields:

- RH ID gloss: RABBI | DS(2bent):animal-runs
- Destination tie: NP | VD
- RH-gram cis: RABBI | DS(2bent):animal-runs
- LH ID gloss: RABBI | DS(2bent):animal-runs
- Clause: A | V
- ClauseArgument: ACTR | UNDR
- SemanticRole: AGEN |

A large oval highlights the 'A' (AGENT) role in the ACTR field, and a smaller oval highlights the 'V' (Verb) in the ClauseArgument field.

Search eaf files

Substring Search | Single Layer Search | Multiple Layer Search

Domain: 9 eaf files

Query History: < >

Mode: case sensitive | regular expression

Minimal Duration | Maximal Duration | Begin After | End Before

PT: = 0 annotations

Overlap

V

Overlap

m

Tier Name: RH ID gloss

Tier Name: RH ID gloss

Tier Name: RH mod

Find

#hits : 28
#annotations with a hit : 28
#annotations investigated : 14797

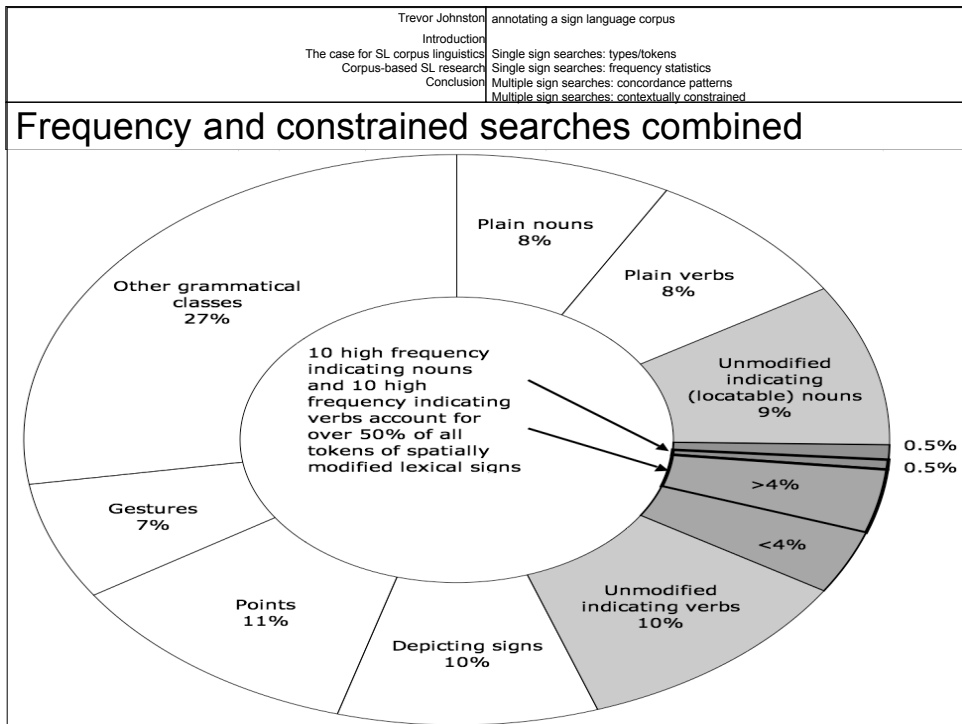
Ready

hit 1 - 16 of 28

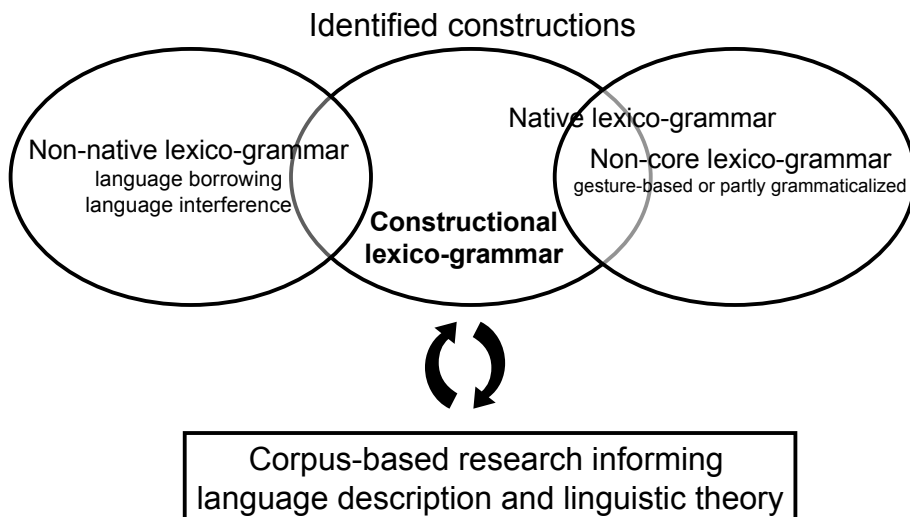
```
#1 IPT:PRO1sg|GO-point-2h| #2 ||IVLocl| #3 ||lml||
#1 IPT:PRO1sg|LOOKI| #2 ||IVDir| #3 ||lml||
#1 IPT:PRO1sg|LOOKI| #2 ||IVDir| #3 ||lml||
#1 IPT:PRO1sg|LOOKI| #2 ||IVLocl| #3 ||lml||
#1 IPT:PRO1sg|LOOKI| #2 ||IVLocl| #3 ||lml||
#1 IPT:PRO3pl|CHATTERBOXI| #2 ||IVLocl| #3 ||lml||
#1 IPT:PRO1sg|HAVEI| #2 ||IVLocl| #3 ||lml||
#1 IPT:PRO1sg|ARGUEI| #2 ||IVLocl| #3 ||lml||
#1 IPT:PRO3sg|ARGUEI| #2 ||IVLocl| #3 ||lml||
#1 IPT:PRO3sg|ATTRACTI| #2 ||IVDir| #3 ||lml||
#1 IPT:PRO3sg|BOTHERI| #2 ||IVDir| #3 ||lml||
#1 IPT:PRO3sg|GET-ATTENTIONI| #2 ||IVDir| #3 ||lml||
#1 IPT:PRO1sg|TEACHI| #2 ||IVDir| #3 ||lml||
#1 IPT:PRO1sg|ISAYI| #2 ||IVDir| #3 ||lml||
#1 IPT:PRO3sg(present referent)|LOOKI| #2 ||IVDir| #3 ||lml||
#1 IPT:PRO1sg(B)|SHAKE-HANDSI| #2 ||IVLocl| #3 ||lml||
```

Three columns

Three rows



Constructions in the lexico-grammatical continuum



Discovering, not searching for patterns (constructions)

- Pattern testing (existing capabilities)
 - Enriching the corpus
 - Testing hypotheses
 - Research observations
- Pattern recognition (desirable capabilities)
 - e.g., CREAGEST team (e.g., Antonio Balvet)
 - need for plug in or software improvement to detect patterns/constructions constrained both ‘vertically’ and ‘horizontally’ by 2, 3 or more values
 - linguistic analysis, new hypotheses etc.

Trevor Johnston Introduction The case for SL corpus linguistics Corpus-based SL research Conclusion	annotating a sign language corpus
Conclusion	
<ul style="list-style-type: none"> • Cross-linguistic & typological SL research • Towards a SL corpus linguistics 	

Trevor Johnston Introduction The case for SL corpus linguistics Corpus-based SL research Conclusion	annotating a sign language corpus Cross-linguistic & typological research Towards a SL corpus linguistics
Cross-linguistic & typological research	
<ul style="list-style-type: none"> • Consistency <ul style="list-style-type: none"> – needed at two levels <ul style="list-style-type: none"> • language-internal & cross-linguistic consistency – documented practice, guidelines or standards? <ul style="list-style-type: none"> • standards desirable, but well-documented internally consistent local practice must not be neglected in the meantime • Comparability <ul style="list-style-type: none"> – descriptive adequacy & typological observations <ul style="list-style-type: none"> • cross-linguistic comparisons are only as strong (valid) as the weakest language-specific description – validation > comparison > re-evaluation <ul style="list-style-type: none"> • testing and validation of language-specific observations should precede cross-linguistic generalization • cross-linguistic comparison nonetheless vital to open new perspectives enabling possible re-evaluation of local descriptions and leading more robust typological generalizations 	

Towards a SL corpus linguistics

- Insist upon corpus-based SL research
 - due to the unique sociolinguistic situation of SL-using communities, corpus-based research is vitally important
- Create true corpora
 - a linguistic corpus is not simply a data-set
 - it is a collection of language which has accurate metadata and is representative, machine readable, accessible and able to be further enriched
- Prioritize annotation above transcription
 - preliminary lexical research necessary to do this effectively
 - use ID-glosses and restricted set of conventions for partly-lexical and non-lexical signs
 - use other tiers to annotate for linguistically salient information
- Use in-built search routines and SQL query language to extract patterns or test generalizations



Acknowledgments

- Hans Rausing Endangered Languages Documentation Program, School of Oriental and African Studies, University of London
 - Project grant #MDP0088 awarded to Trevor Johnston (2004-07)
- Australian Research Council
 - Project grant #LP0346973 awarded to Adam Schembri and Trevor Johnston: *Sociolinguistic Variation in Auslan: Theoretical and applied dimensions* (2003-05)
 - Project grant #DP0665254 awarded to Louise de Beuzeville and Trevor Johnston: *The linguistic use of space in Auslan: semantic roles and grammatical relations in three dimensions* (2006-09)
- Macquarie University
 - Grant #MQSNG9200729255 awarded to Trevor Johnston (2008-09): *The emergence of grammar in human languages: evidence from video-recordings of variation in vocabulary and grammar in a signed language (pilot study).*



Contact information

A/Prof Trevor Johnston
Centre for Language Sciences
Department of Linguistics
Macquarie University
Balaclava Road, North Ryde
Sydney, NSW, Australia 2109

trevor.johnston@mq.edu.au