

Cross-Linguistic Acquisition Of Complex Verb Inflection In A Neural Network Model

Felix Engelmann

Sonia Granlund, Joanna Kolak, Ben Ambridge, Julian Pine, Anna Theakston, Elena Lieven

Cross-linguistic work package
University of Manchester



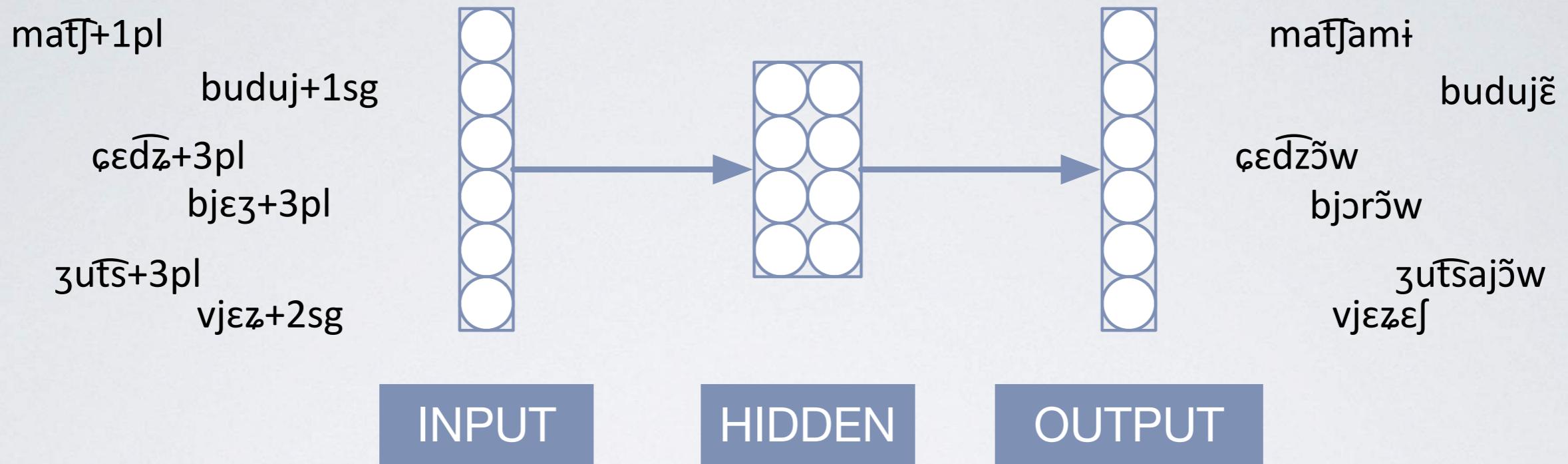
A CONSTRUCTIVIST MODEL

The constructivist theory of learning (from previous talk):

- Gradual construction of grammar on the basis of input
- Language rules constructed in the course of development
- First rote-stored forms, then low-freq ones
- Low overall error rate in production BUT higher error rates in certain parts of the paradigm

A neural network model is a constructivist model!

A NEURAL NETWORK MODEL

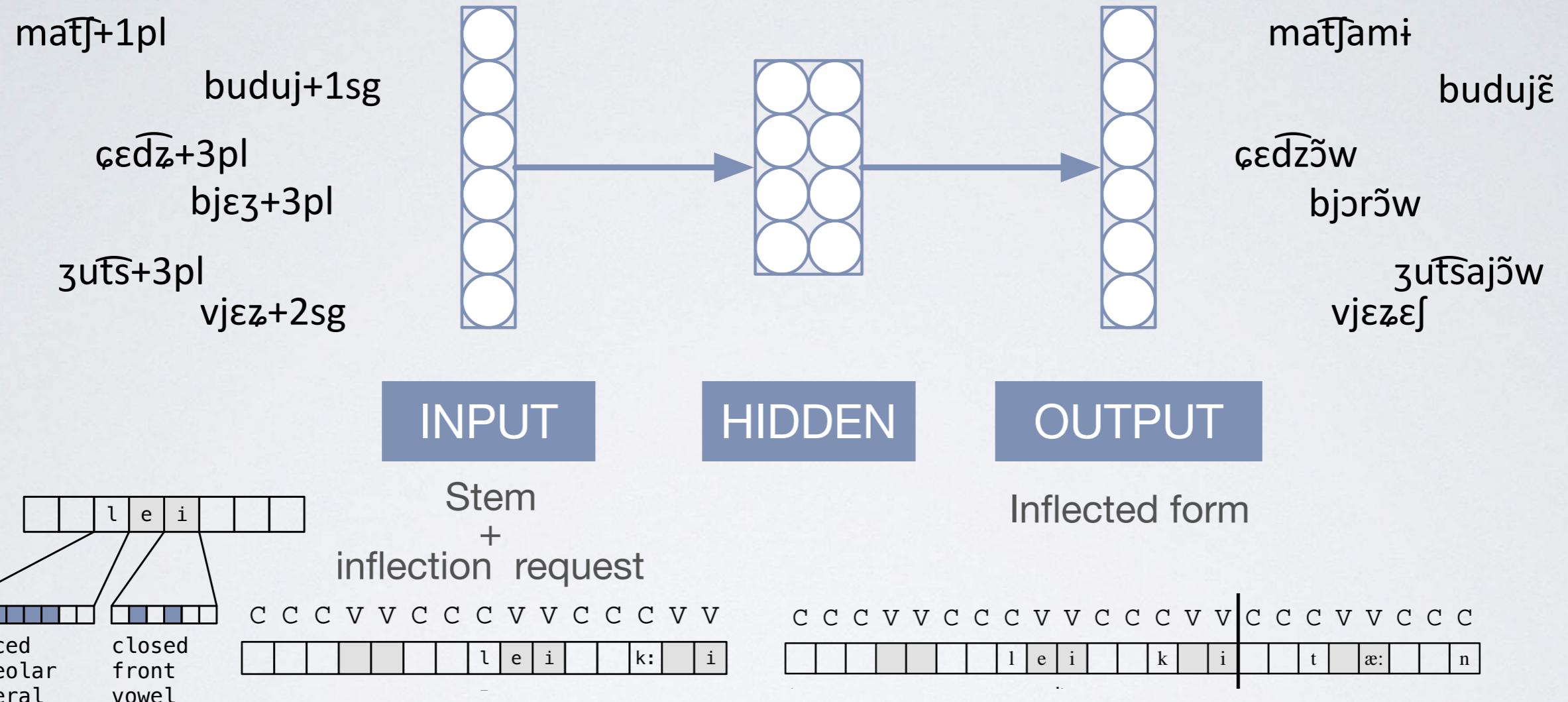


- Purely input-based learning
- Naturally sensitive to frequency of presentation
- No a priori assumptions about rules or mechanisms or similarities
- Rule-like behaviour emerges from abstraction over input regularities (in the hidden layer)

QUESTIONS

- Can a purely exposure-based mechanism learn Finnish and Polish verb inflection?
- Are simulated error patterns similar to children's?
- Are model errors predicted by form frequency and phonological neighbourhood density (PND)?

TRAINING PROCEDURE



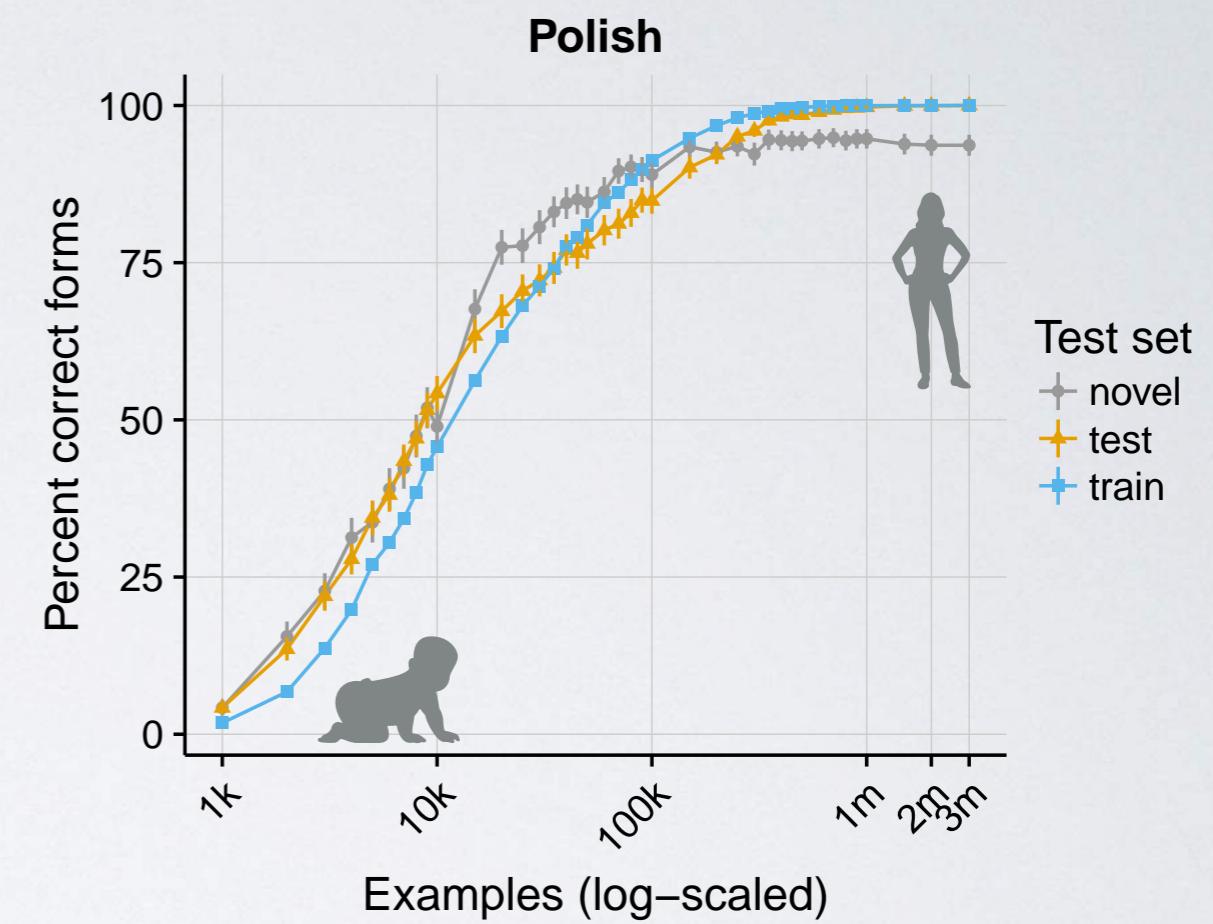
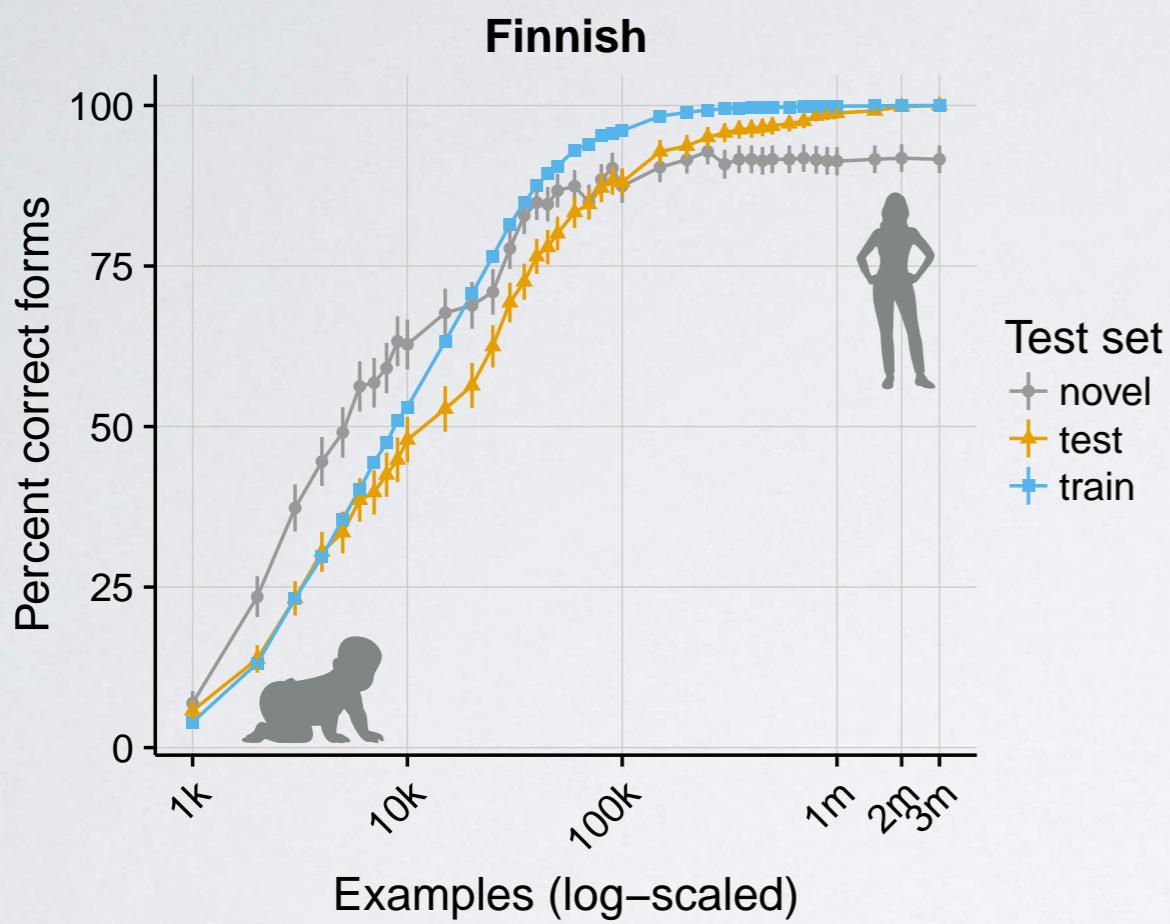
- Verb forms from 800 verbs were presented probabilistically according to CDS frequency
- Forms were presented as phonemes right-justified in syllable template

TRAINING PROCEDURE

	Finnish	Polish
verbs		800
person/number	1sg, 2sg, 3sg, pass, 2pl	1sg, 2sg, 3sg, 1pl, 2pl, 3pl
forms	1784	2431
exp. test forms	160	192
novel forms	165 (33 verbs)	306 (51 verbs)
input stem	4th inf. part. stem	3sg stem
templates	in: CCWCCCWCCW out: CCVCCWCCW CCVVCC	in: CCCWCCCWCCCWCCC out: CCCWCCCWCCWCCC VCC
input units	90	138
output units	136	168
hidden units		200

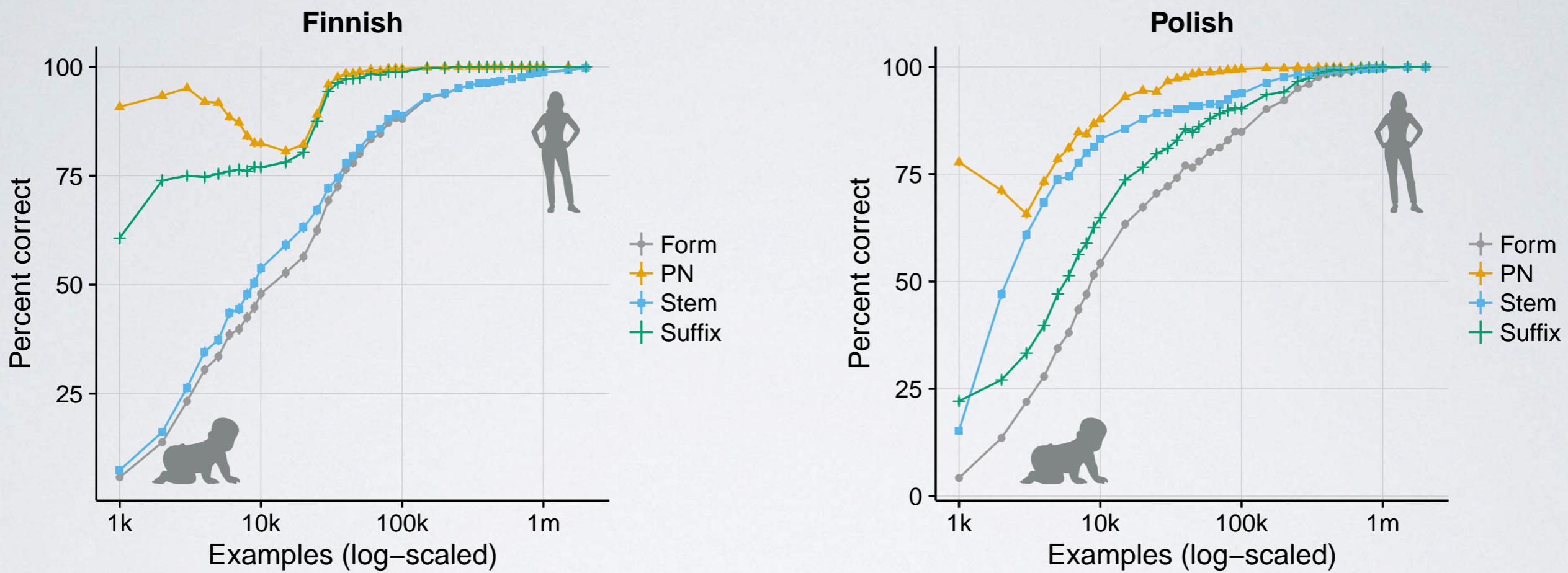
- 10 networks in each language
- For a subset of verbs, one form each was withheld from training (different for each network)

RESULTS OVERVIEW



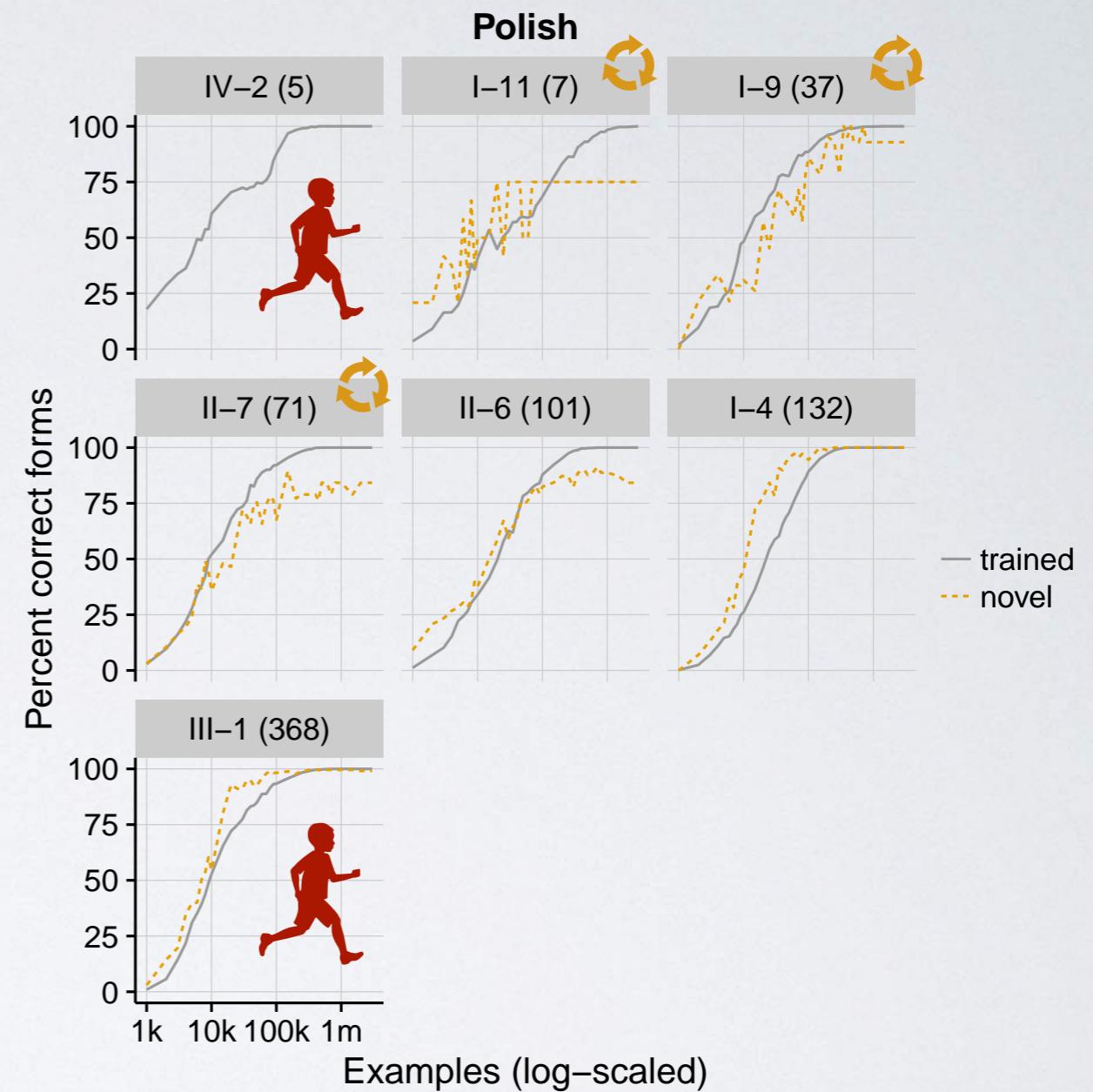
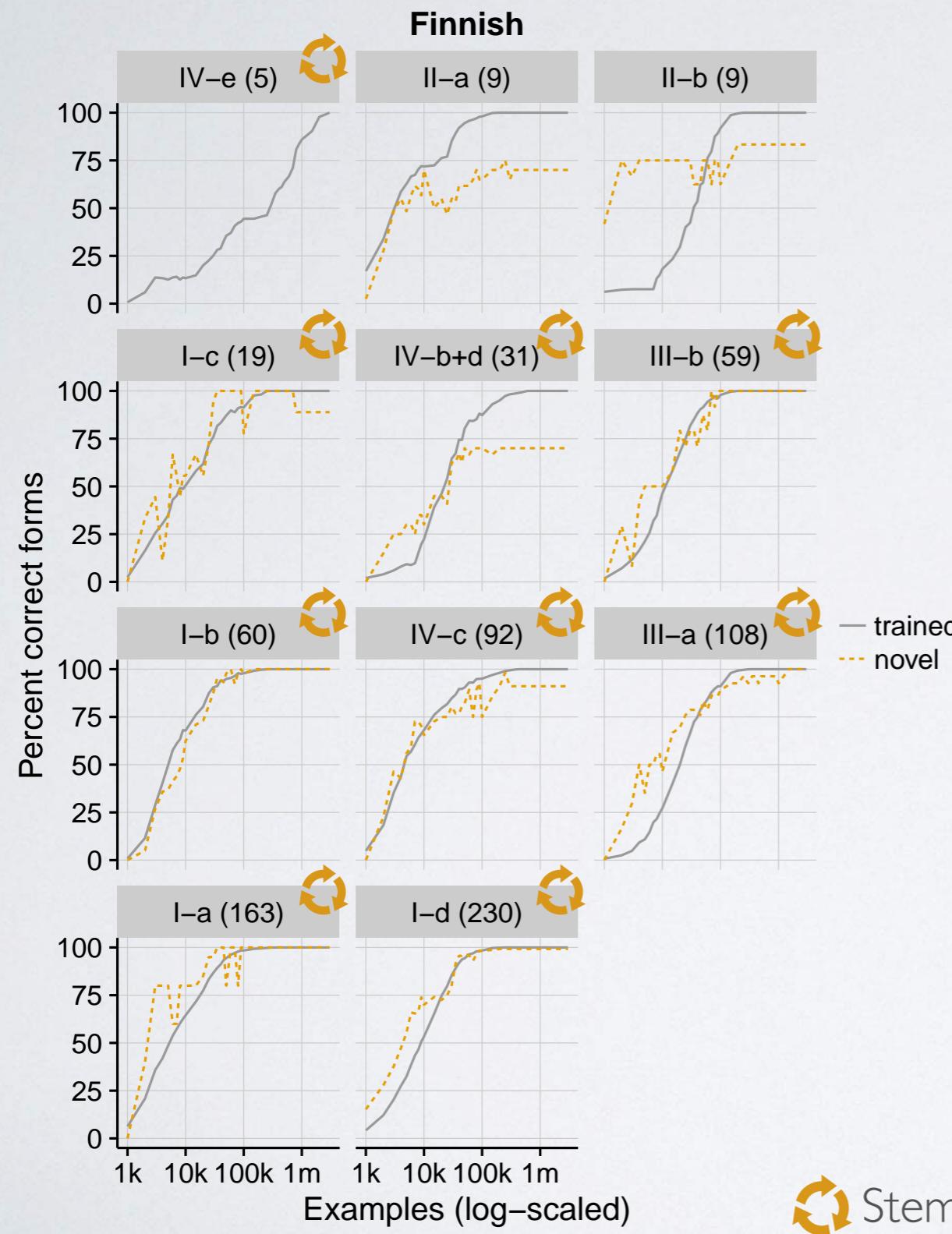
- All forms acquired after seeing 3 million examples
- 93% of unseen forms correctly generalised
- Stems harder in Finnish, suffixes harder in Polish
- Person/number target often correct but wrong suffix (class error)

RESULTS OVERVIEW



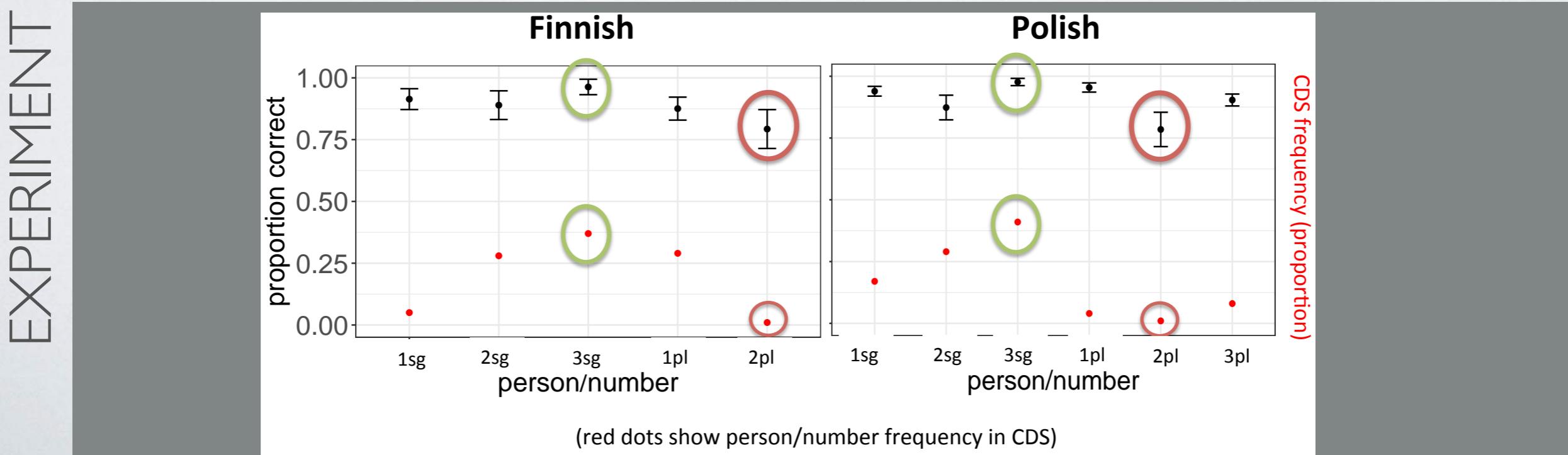
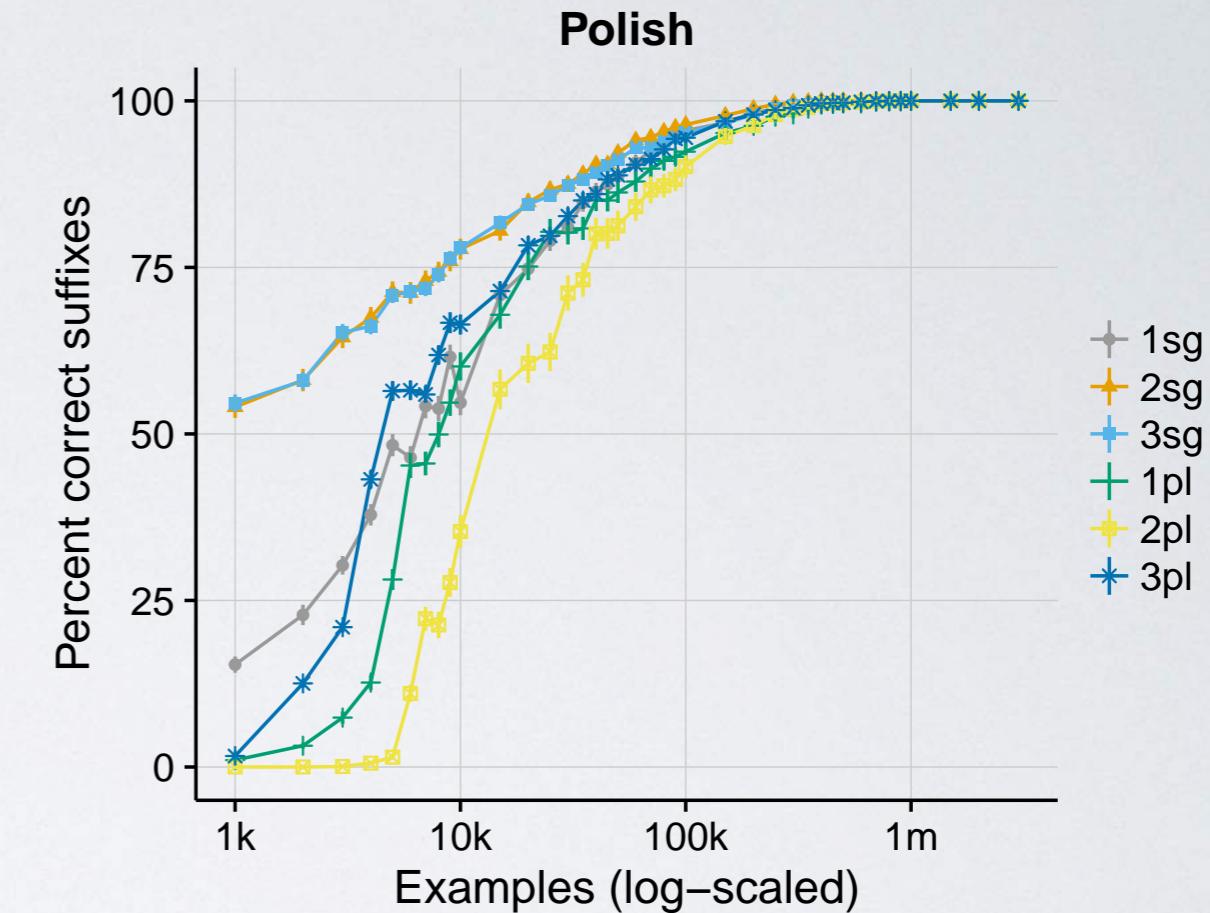
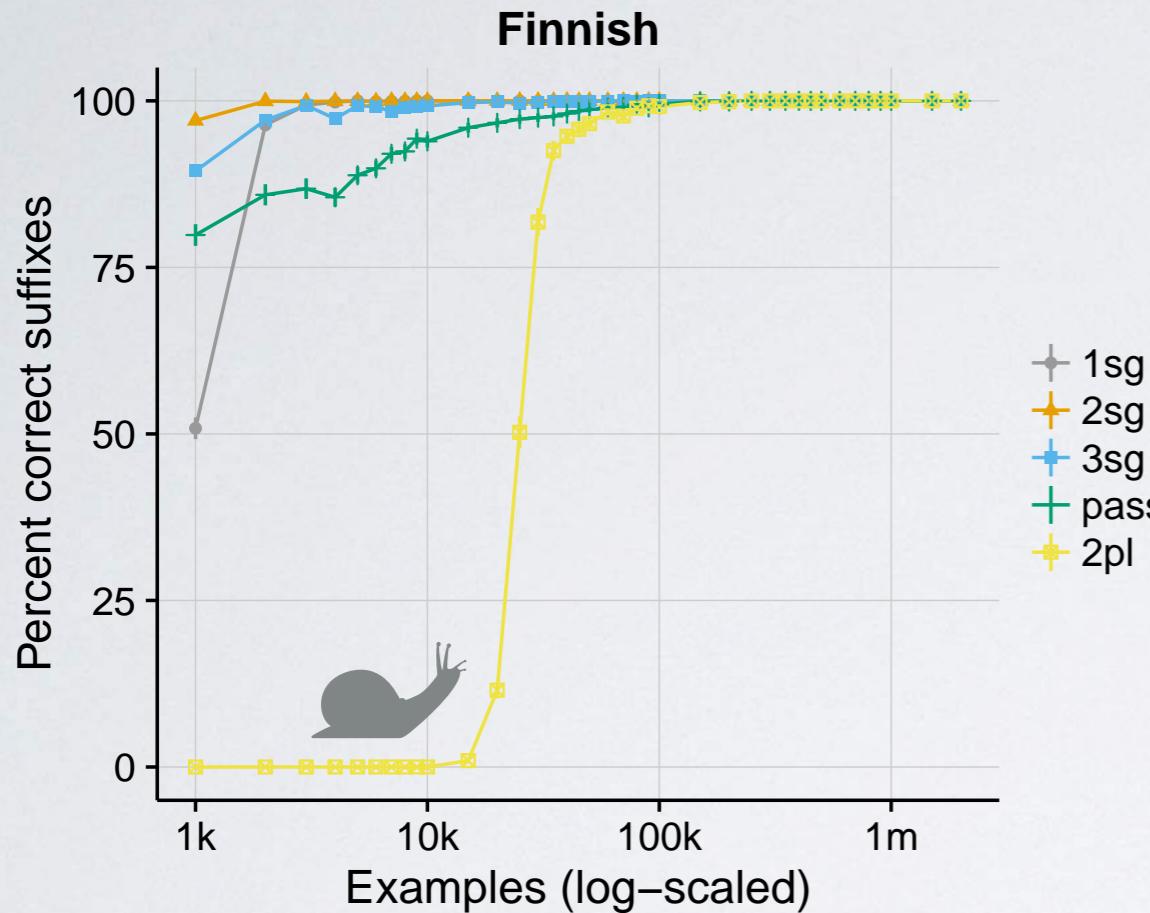
- All forms acquired after seeing 3 million examples
- 93% of unseen forms correctly generalised
- Stems harder in Finnish, suffixes harder in Polish
- Person/number target often correct but wrong suffix (class error)

ACCURACY & GENERALISATION BY CLASS



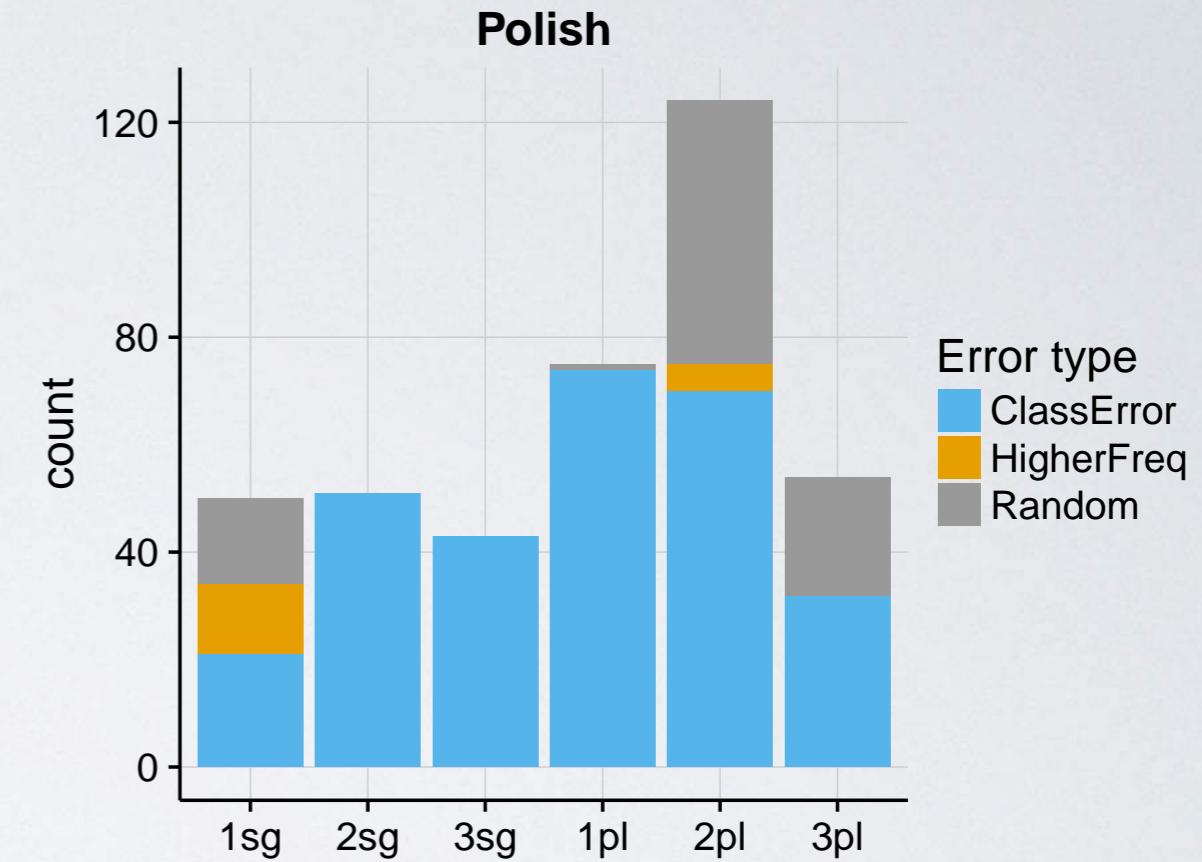
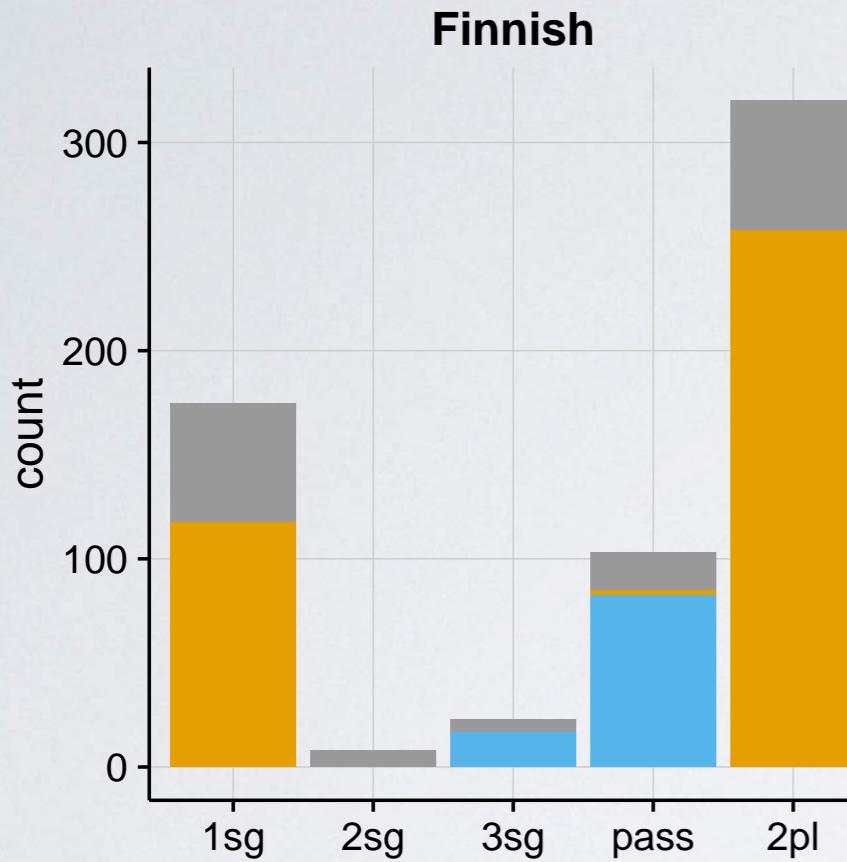
Stem alternations

ACCURACY BY PERSON/NUMBER



SUFFIX ERRORS

at training epoch with 75% correct suffixes



EXPERIMENT

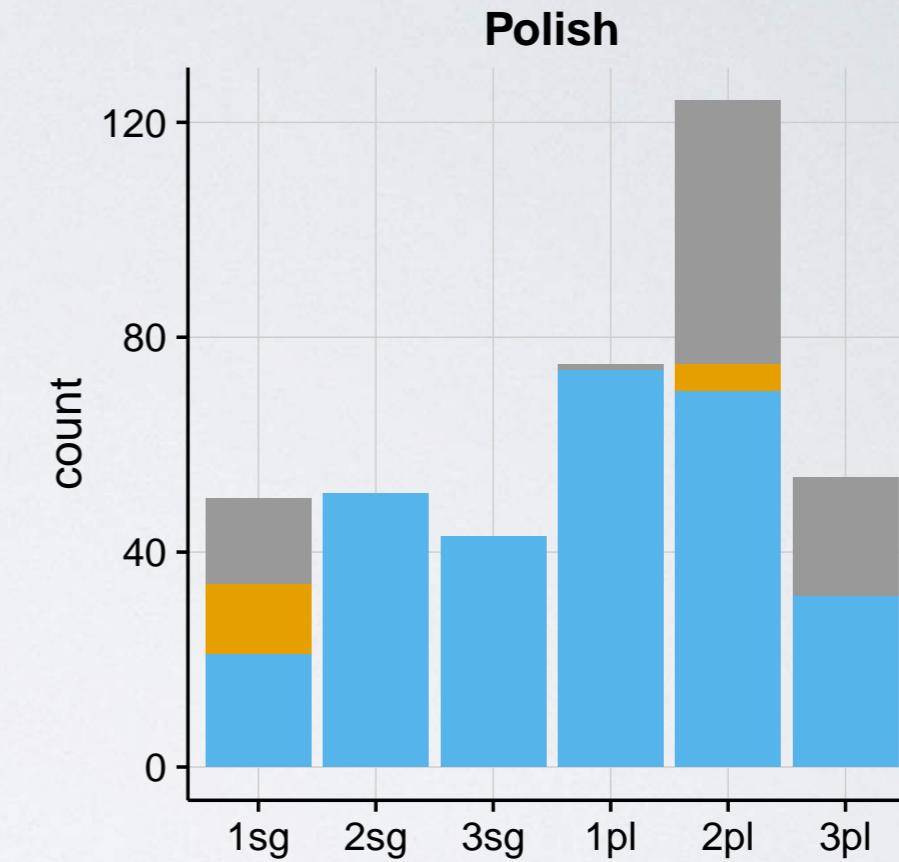
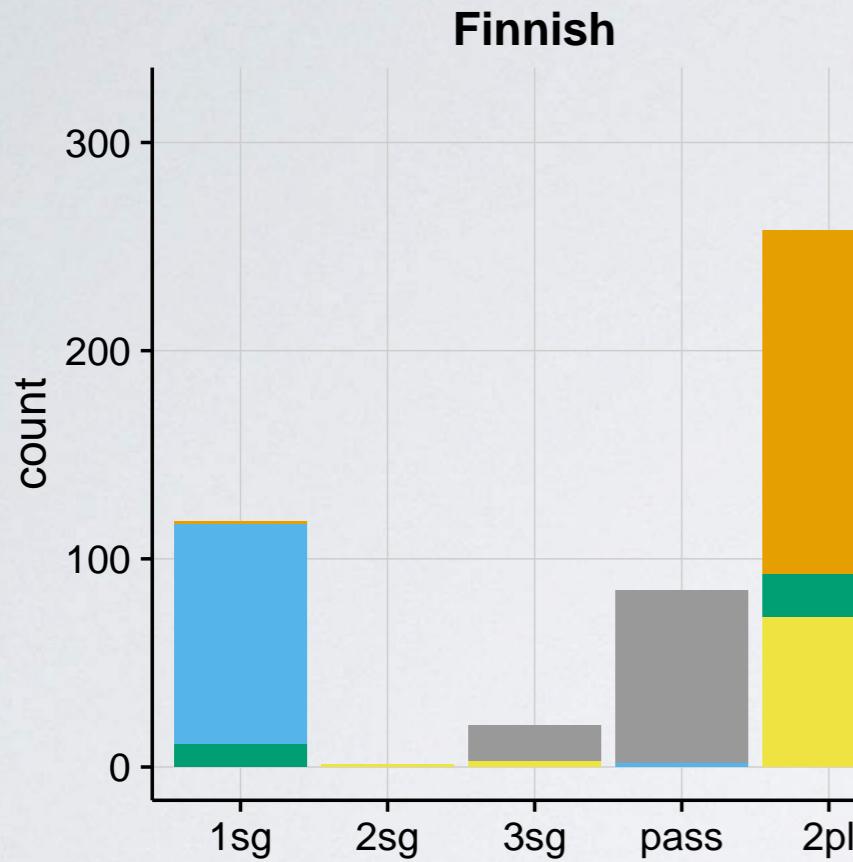
		pn used						
		target pn	1sg	2sg	3sg	1pl	2pl	3pl
target pn		1sg	2sg	3sg	1pl	2pl	3pl	
1sg		3	39	4	0	0	0	
2sg		1	33	8	0	0	0	
3sg		1	1	5	3	0	0	
1pl		1	0	18	2	0	0	
2pl		2	5	40	18	1	0	

+ class error: class 3 instead of
class 1 (77 errors)

		pn used						
		target pn	1sg	2sg	3sg	1pl	2pl	3pl
target pn		1sg	2sg	3sg	1pl	2pl	3pl	
1sg		2	10	5	0	0	0	
2sg		8	26	27	5	1	0	
3sg		2	0	3	0	1	0	
1pl		4	0	2	5	11	0	
2pl		1	2	3	59	43	0	
3pl		2	1	11	2	0	0	

SUFFIX ERRORS

at training epoch with 75% correct suffixes



EXPERIMENT

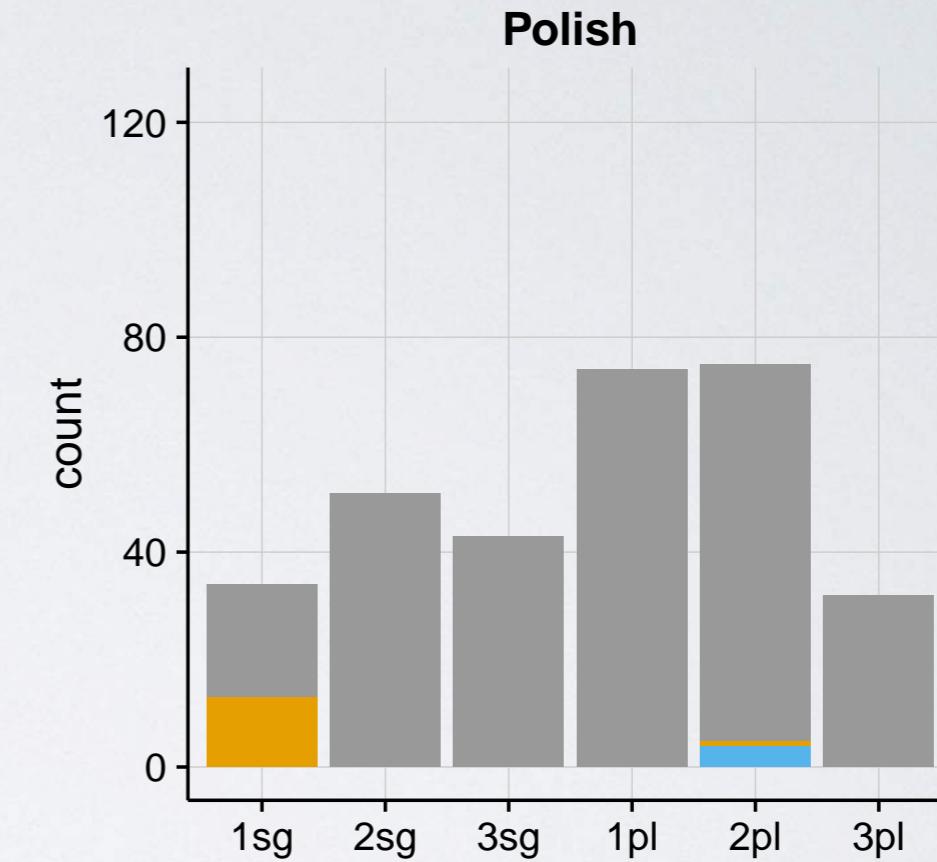
	pn used					
target pn	1sg	2sg	3sg	1pl	2pl	3pl
1sg	3	39	4	0	0	0
2sg	1	33	8	0	0	0
3sg	1	1	5	3	0	0
1pl	1	0	18	2	0	0
2pl	2	5	40	18	1	0

+ class error: class 3 instead of
class 1 (77 errors)

	pn used					
target pn	1sg	2sg	3sg	1pl	2pl	3pl
1sg	2	10	5	0	0	0
2sg	8	26	27	5	1	0
3sg	2	0	3	0	1	0
1pl	4	0	2	5	11	0
2pl	1	2	3	59	43	0
3pl	2	1	1	11	2	0

SUFFIX ERRORS

at training epoch with 75% correct suffixes



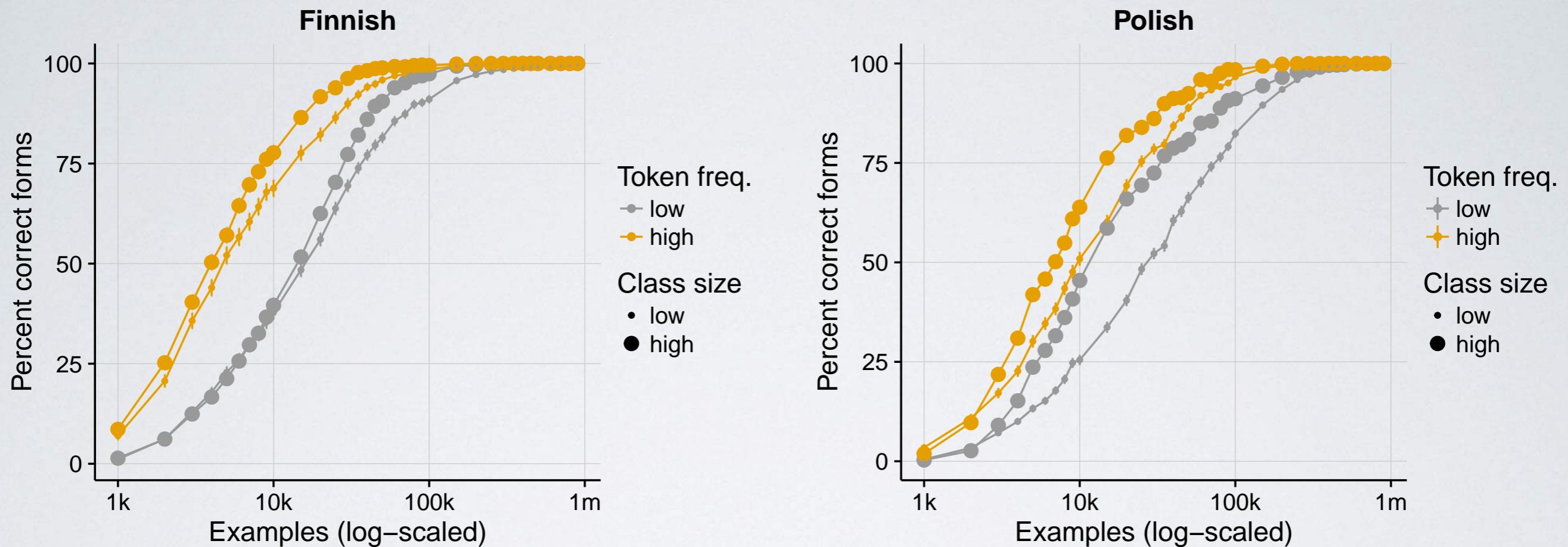
EXPERIMENT

		pn used					
target pn	1sg	2sg	3sg	1pl	2pl	3pl	
1sg	3	39	4	0	0	0	
2sg	1	33	8	0	0	0	
3sg	1	1	5	3	0	0	
1pl	1	0	18	2	0	0	
2pl	2	5	40	18	1	0	

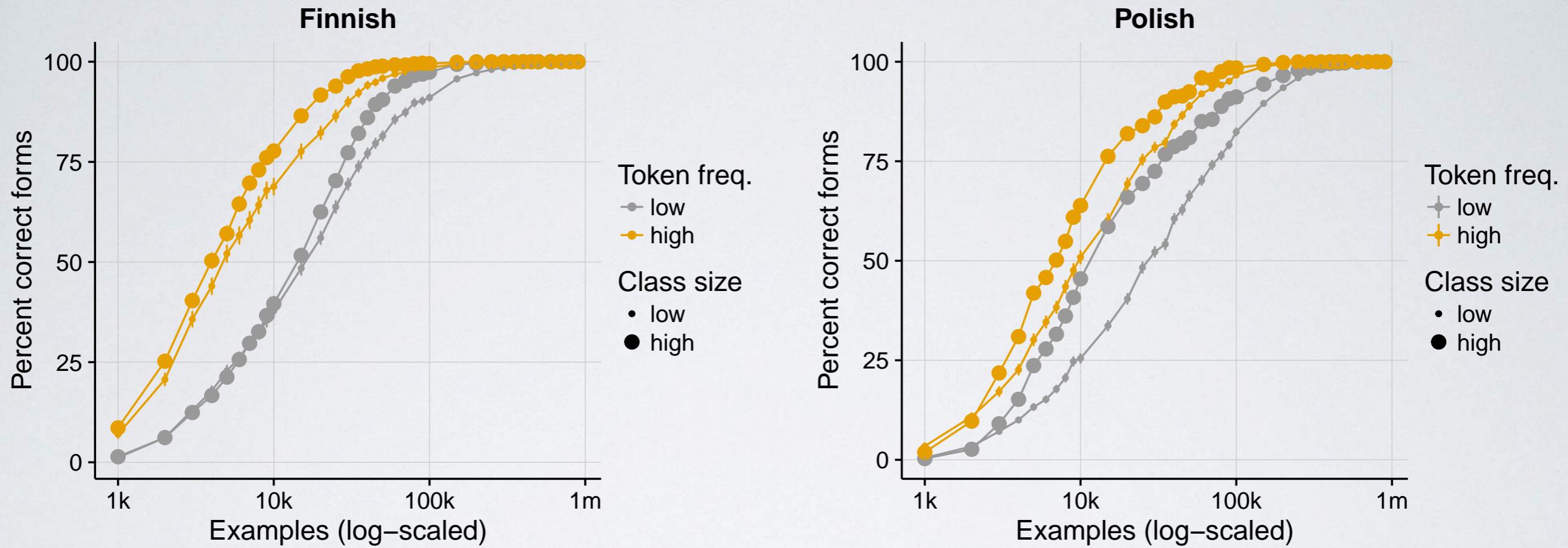
+ class error: class 3 instead of
class 1 (77 errors)

		pn used					
target pn	1sg	2sg	3sg	1pl	2pl	3pl	
1sg	2	10	5	0	0	0	
2sg	8	26	27	5	1	0	
3sg	2	0	3	0	1	0	
1pl	4	0	2	5	11	0	
2pl	1	2	3	59	43	0	
3pl	2	1	1	11	2	0	

FREQUENCY AND PND



FREQUENCY AND PND

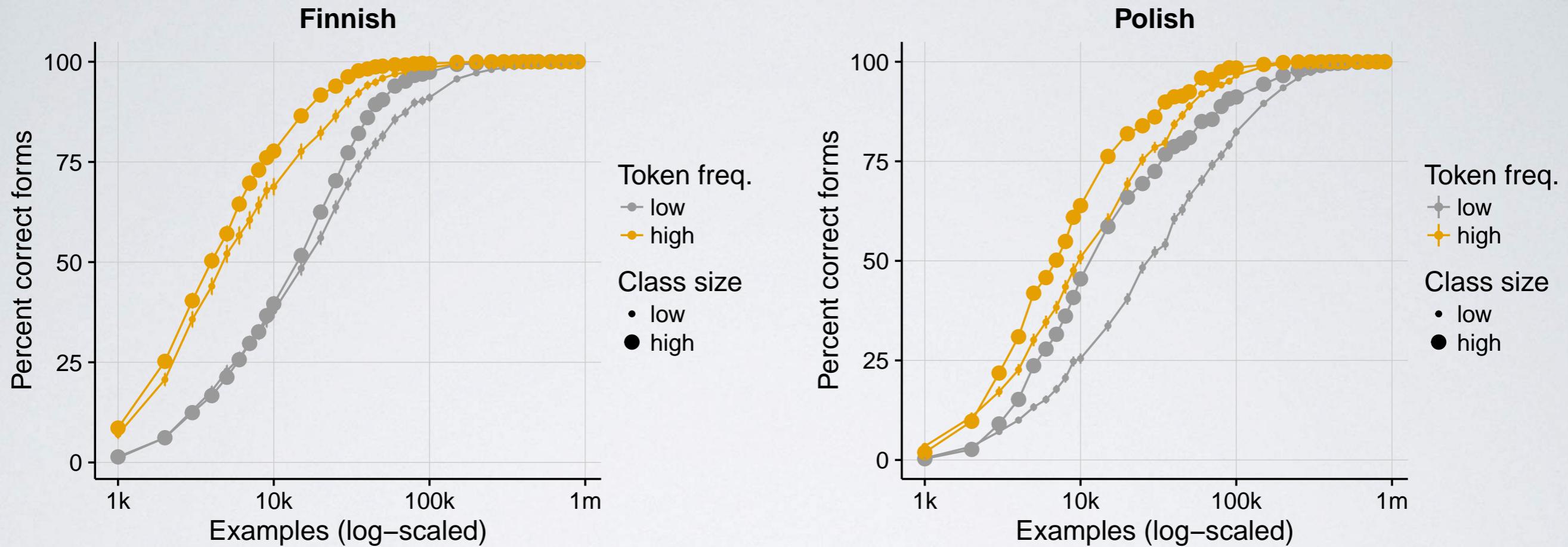


LMM on activation error in **training set**, beginning when suffix accuracy at 75%

	Est.	SE	t value	p value	sign
(Intercept)	-0.35	0.02	-22.90		
PND	-0.08	0.02	-5.08	<0.001	***
TokenFreq	-0.21	0.00	-132.00	<0.001	***
Epoch	-0.72	0.00	-235.36	<0.001	***
PND:TokenFreq	0.06	0.00	42.91	<0.001	***
TokenFreq:Epoch	0.35	0.00	113.97	<0.001	***
PND:Epoch	0.10	0.00	32.32	<0.001	***

	Est.	SE	t value	p value	sign
(Intercept)	-0.23	0.02	-12.89		
PND	-0.14	0.02	-8.13	<0.001	***
TokenFreq	-0.16	0.00	-122.74	<0.001	***
Epoch	-0.92	0.00	-397.25	<0.001	***
PND:TokenFreq	0.01	0.00	8.12	<0.001	***
TokenFreq:Epoch	0.29	0.00	124.99	<0.001	***
PND:Epoch	0.12	0.00	52.39	<0.001	***

FREQUENCY AND PND

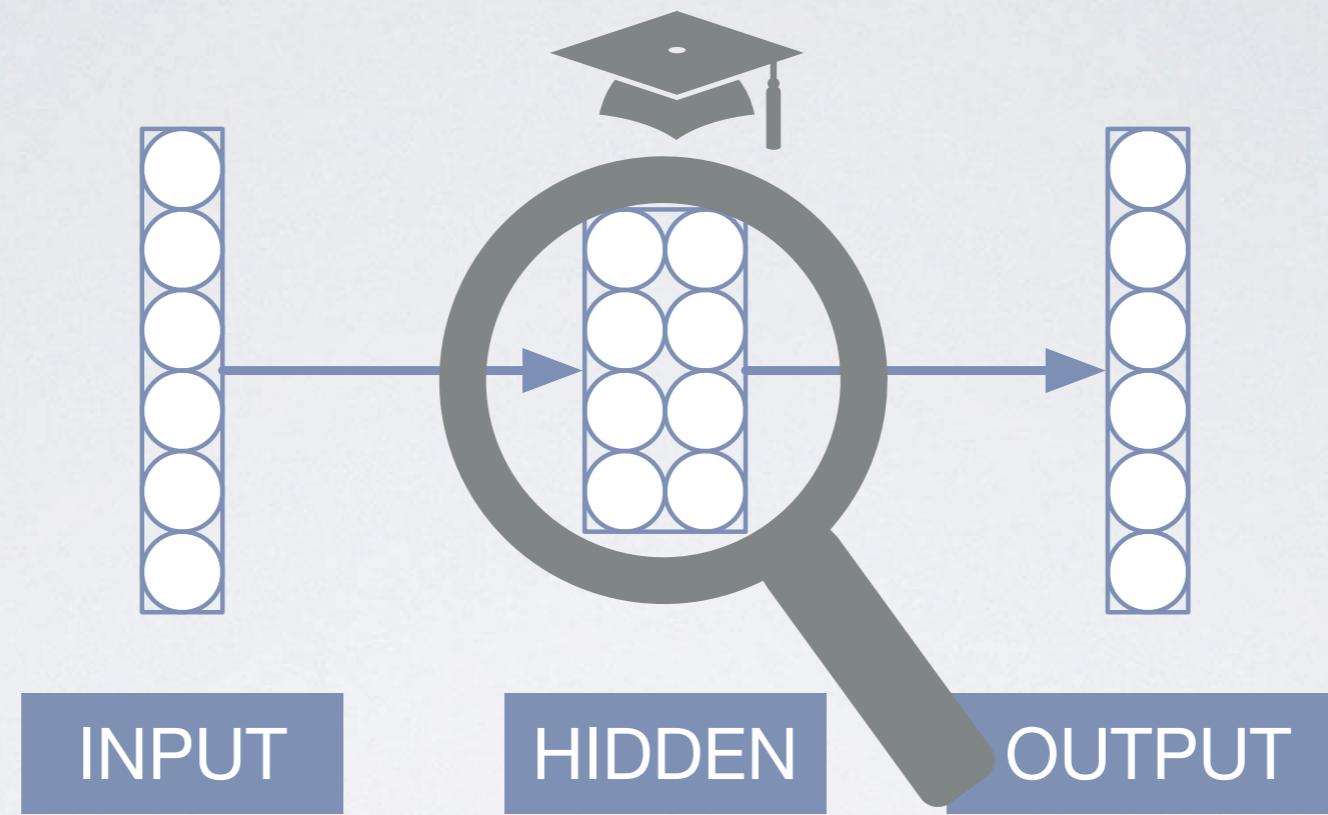


LMM on activation error in **test set**, beginning when suffix accuracy at 75%

	Est.	SE	t value	p value	sign
(Intercept)	-0.24	0.09	-2.61		
PND	-0.19	0.09	-2.01	0.04	*
TokenFreq	-0.25	0.01	-45.23	<0.001	***
Epoch	-0.74	0.01	-70.09	<0.001	***
PND:TokenFreq	0.05	0.00	11.00	<0.001	***
TokenFreq:Epoch	0.36	0.01	33.85	<0.001	***
PND:Epoch	0.15	0.01	14.54	<0.001	***

	Est.	SE	t value	p value	sign
(Intercept)	-0.20	0.09	-2.31		
PND	-0.03	0.09	-0.40	0.68	
TokenFreq	-0.19	0.00	-40.57	<0.001	***
Epoch	-0.70	0.01	-77.40	<0.001	***
PND:TokenFreq	-0.02	0.00	-3.80	<0.001	***
TokenFreq:Epoch	0.33	0.01	35.97	<0.001	***
PND:Epoch	0.10	0.01	10.56	<0.001	***

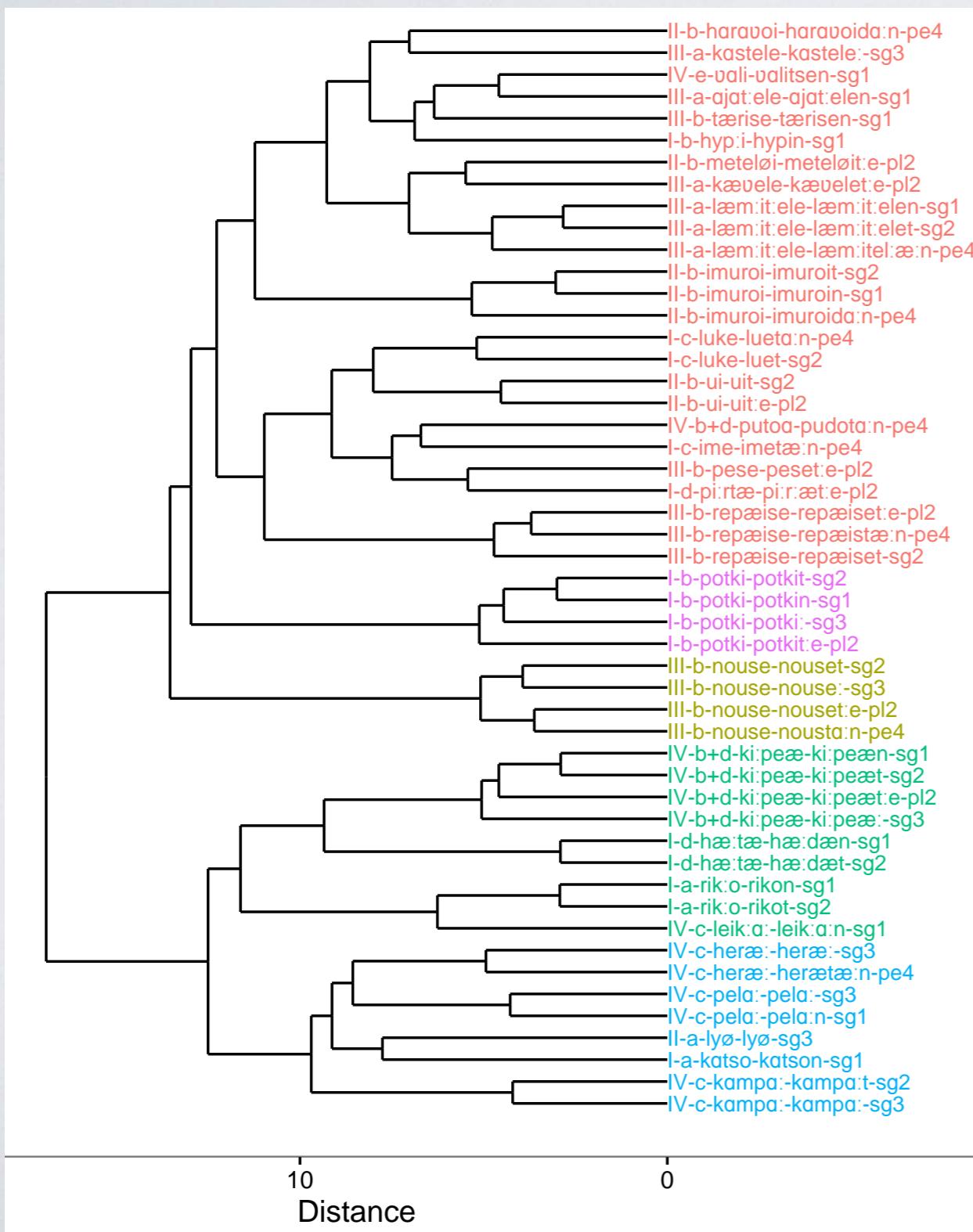
THE NETWORK'S “KNOWLEDGE”



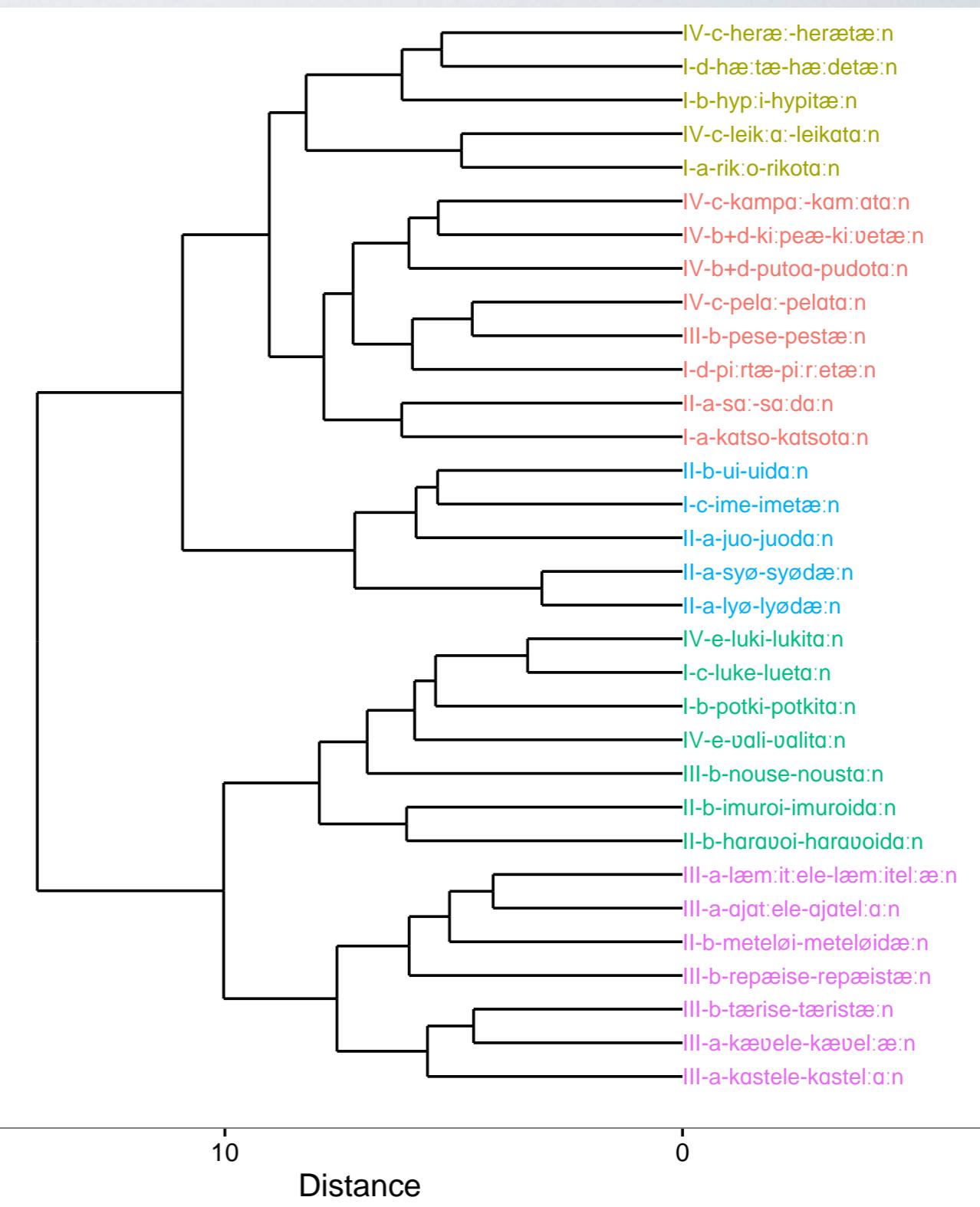
- We cannot see the network's 'rules' but we can see its verb categories in the hidden layer.
- Verbs represented similarly in the hidden layer lead to similar behaviour.
- We calculated the differences between hidden layer activations for all verb forms and clustered them hierarchically.

THE NETWORK'S "KNOWLEDGE"

Finnish

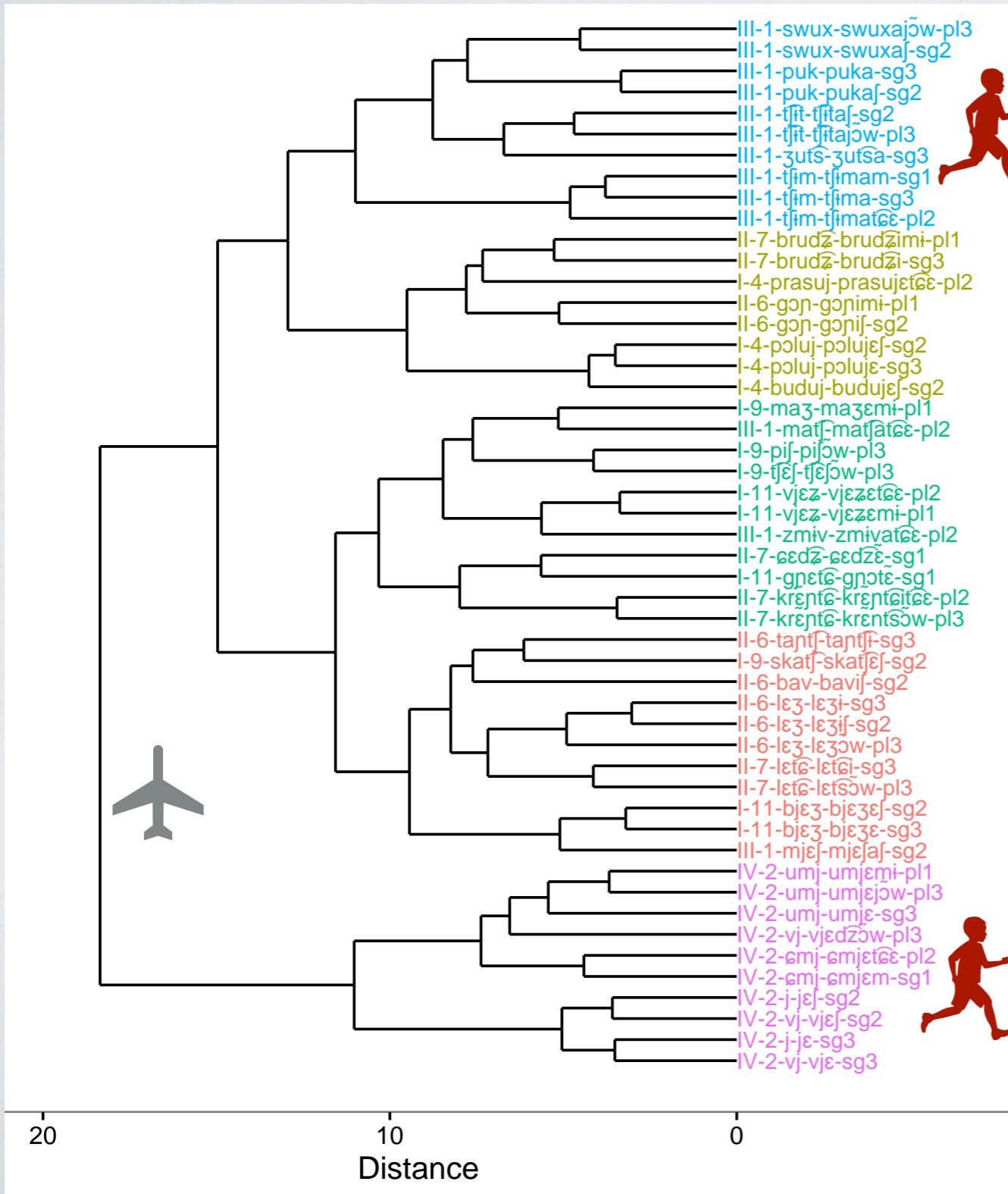


Finnish (pass)



THE NETWORK'S "KNOWLEDGE"

Polish



- The two easiest classes are best represented:
- Class III is the largest class
- Class IV is the smallest but most distinctive

COMPARISON WITH EXPERIMENT

Experiment	Model
Error rates predicted by freq of 'ready-inflected' forms (token freq).	✓
Error rates lower for verbs with more phonological neighbours (PND).	✓
P/n marking errors higher in lower freq contexts.	✓
Most p/n errors involve use of higher freq forms in lower freq p/n contexts.	Most p/n errors are class errors. Some involve high-freq. forms in low-freq. contexts.
No interaction between PND and frequency?	Low-frequency forms benefit more from PND than high-frequency forms (in later development).
	Low-frequency inflections are acquired later.

CONCLUSIONS

- A common exposure-based (constructivist) mechanism can learn complex verb inflection cross-linguistically.
- Error patterns are similar to children's errors.
- Model errors predicted by token frequency and phonological neighbourhood density.

THANK YOU



This research is funded
by the ESRC grant
ES/L008955/1

