Online Experiments for Language Scientists

Lecture 8: Iterated learning

Kenny Smith

kenny.smith@ed.ac.uk

Assessment 2 Q&A

- Due on 8th December
- Read the assignment brief (https://kennysmithed.github.io/oels2022/assessment/AssignmentBrief2022.pdf)
- **Happy to answer questions now, or in next week's lecture**
- We can help with basic coding stuff in extra drop-in labs (see times on github course page)
- No questions after 10am on Monday 5th December (other than in drop-ins)
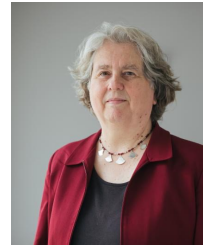
Beckner et al (2017)

Beckner, C., Pierrehumbert, J., & Hay, J. (2017). The emergence of linguistic structure in an online iterated learning task. *Journal of Language Evolution, 2*, 160–176.

An iterated artificial language learning experiment

- Does compositional structure emerge 'for free' from person-to-person transmission?

Clay Beckner
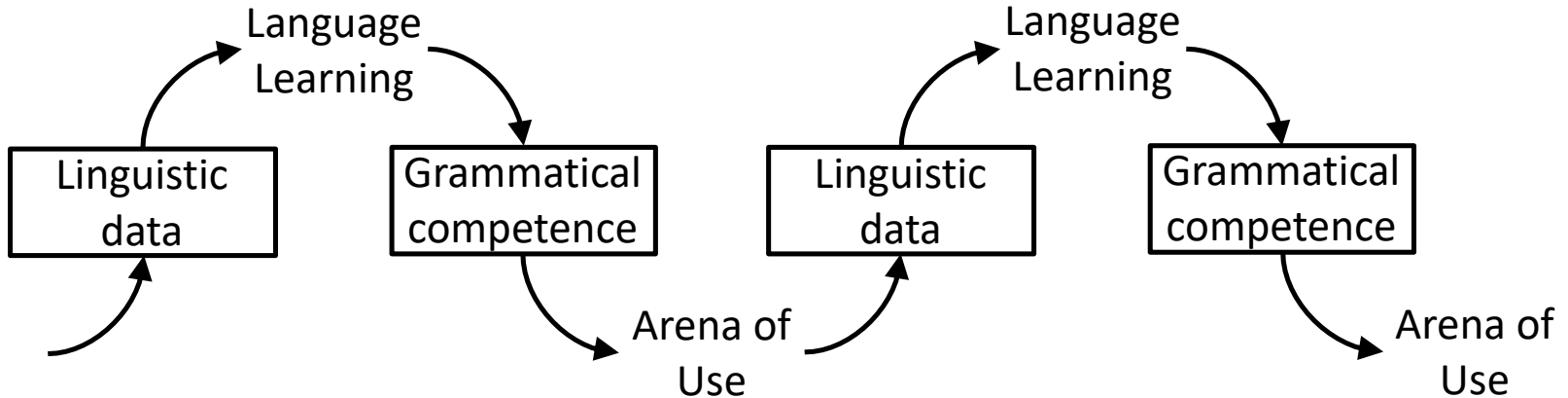*(now at Warwick)*

Janet Pierrehumbert
(Oxford)

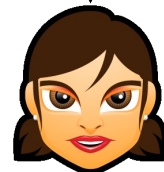**Jen Hay**
(Canterbury, NZ)

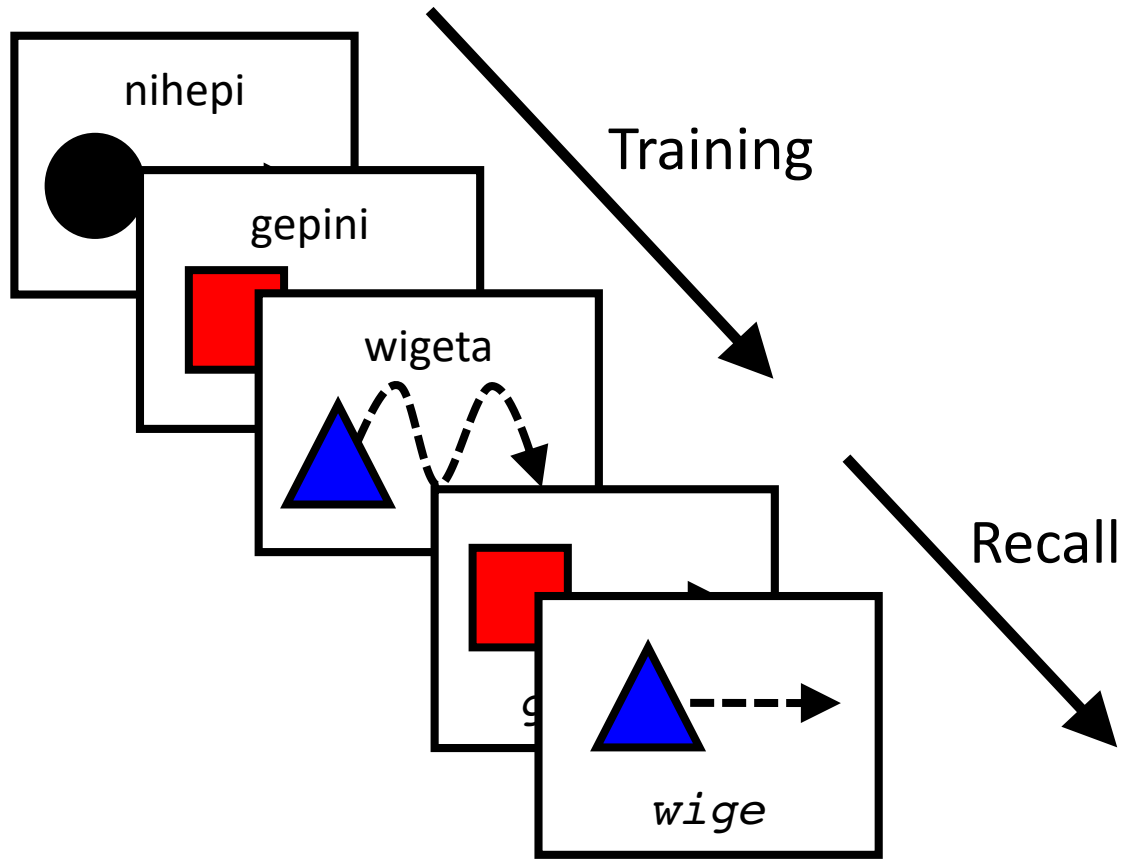Language is transmitted via repeated **learning** and **use**

Language is shaped by these processes

The cycle of learning and use produces structure

Language
Learning

Language
Learning

Linguistic
data

Grammatical
competence

Linguistic
data

Grammatical
competence

Arena of
Use

Arena of
Use

Iterated learning

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *PNAS, 105,* 10681-10686.

Demo using this week's practical code
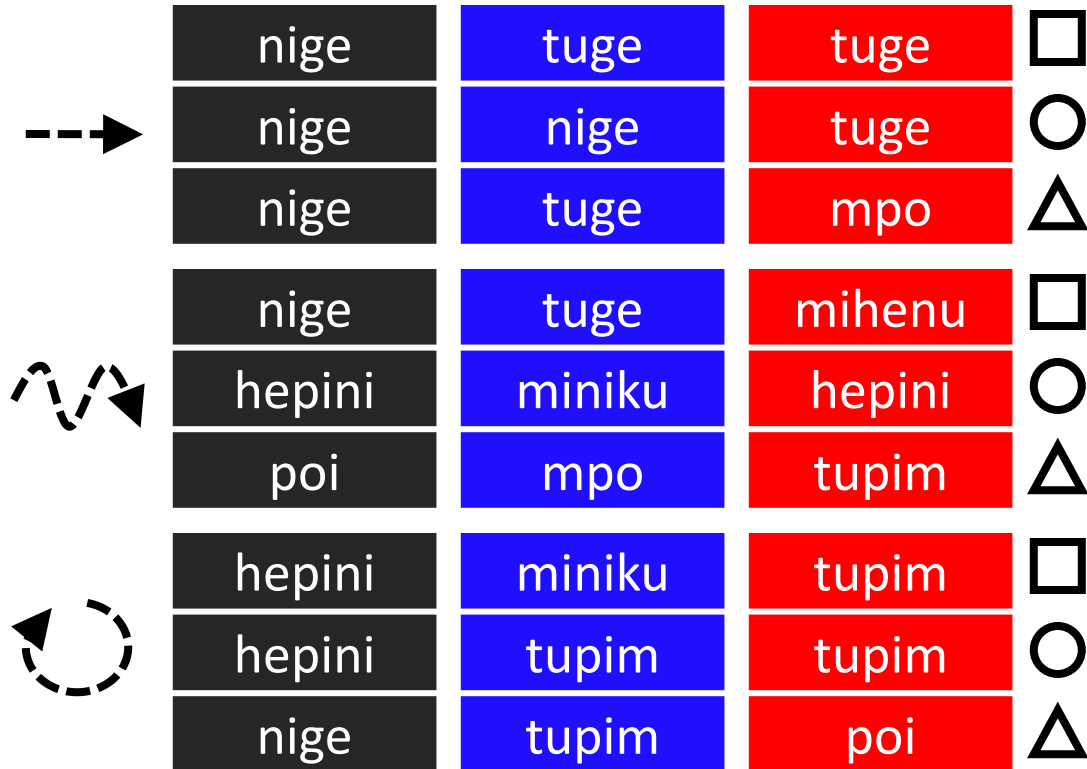
Initial **holistic** language from chain 4

Generation 1 language from chain 4

nige	miniku	poh	□
mip	mpo	miniku	○
tuge	tuge	weg	△

pemini	kupini	pon	□
kimei	miwn	miheniw	○
poi	mhip	kuwpi	△

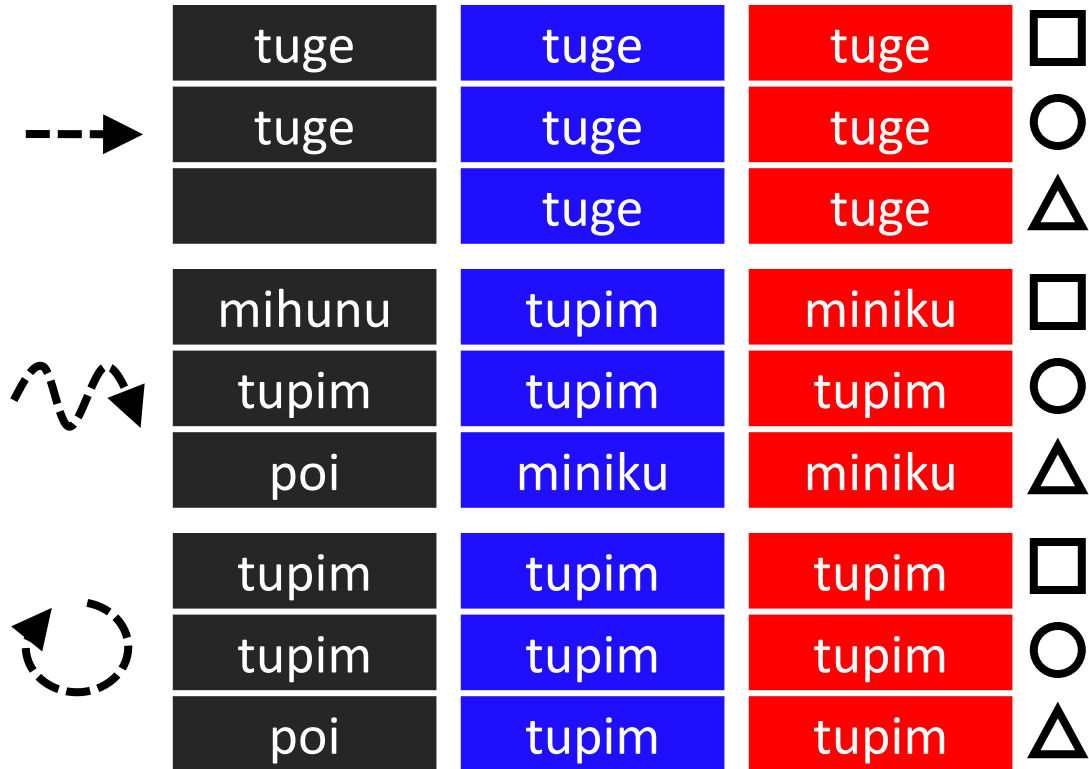|			
hepinimi	himini	hipe	□
kuhepi	wige	mie	○
pobo	tupim	hipe	△

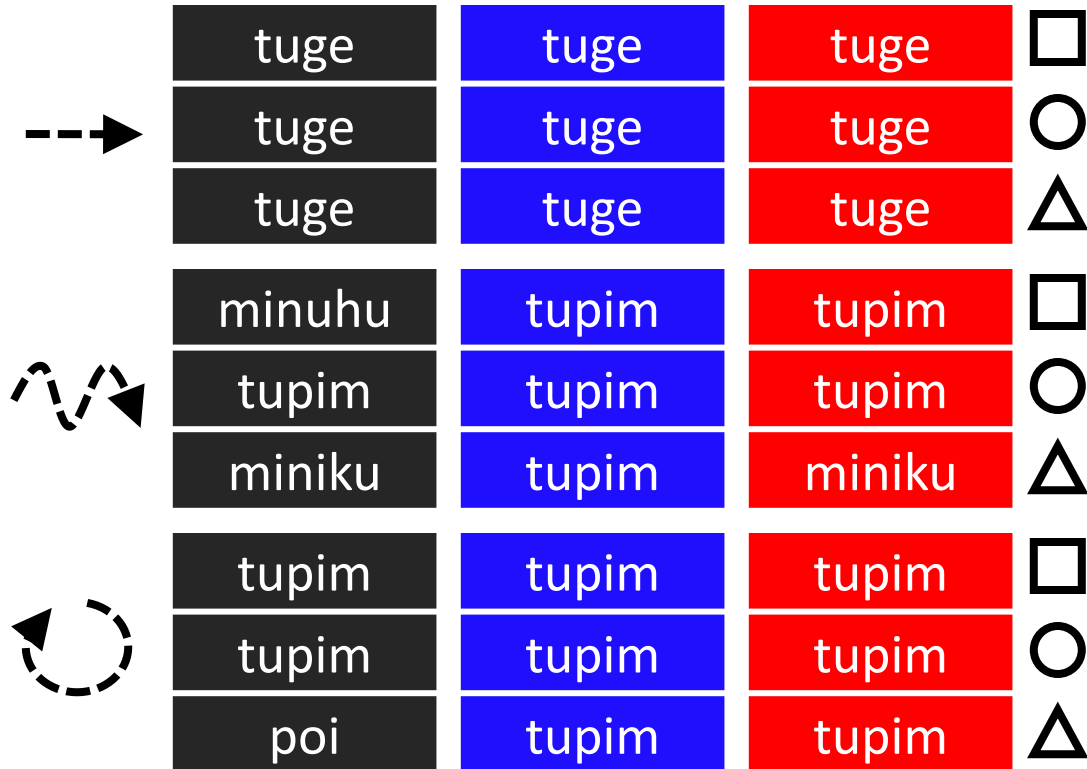# Generation 2 language from chain 4
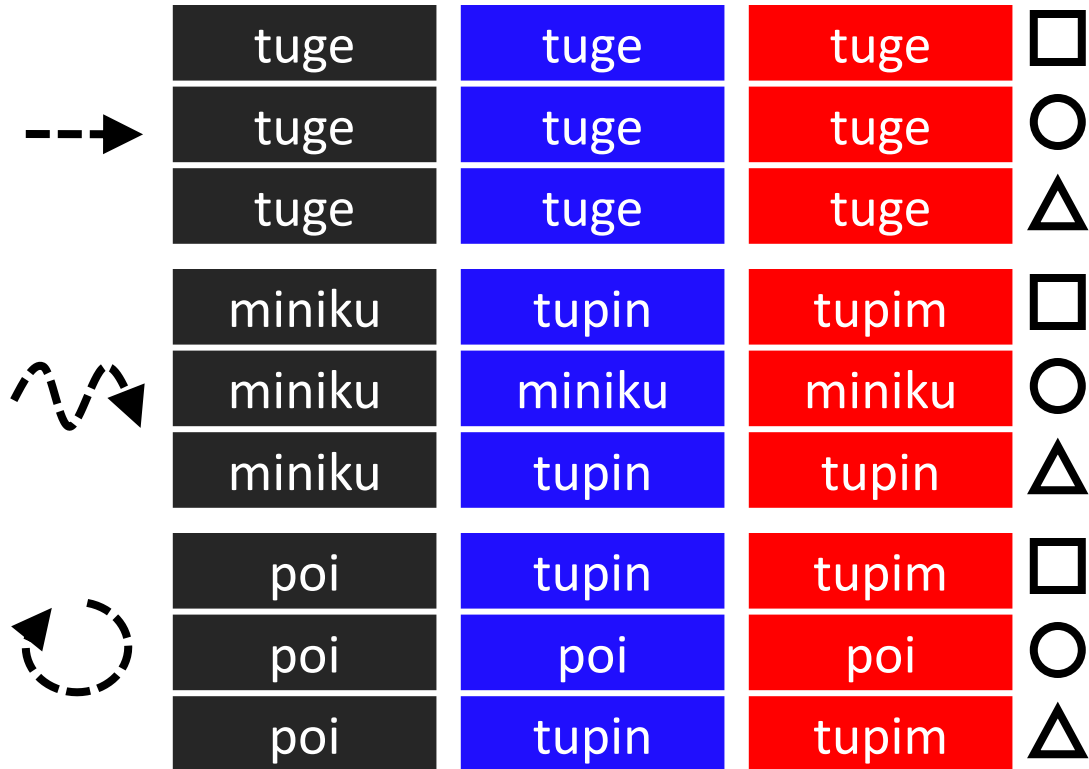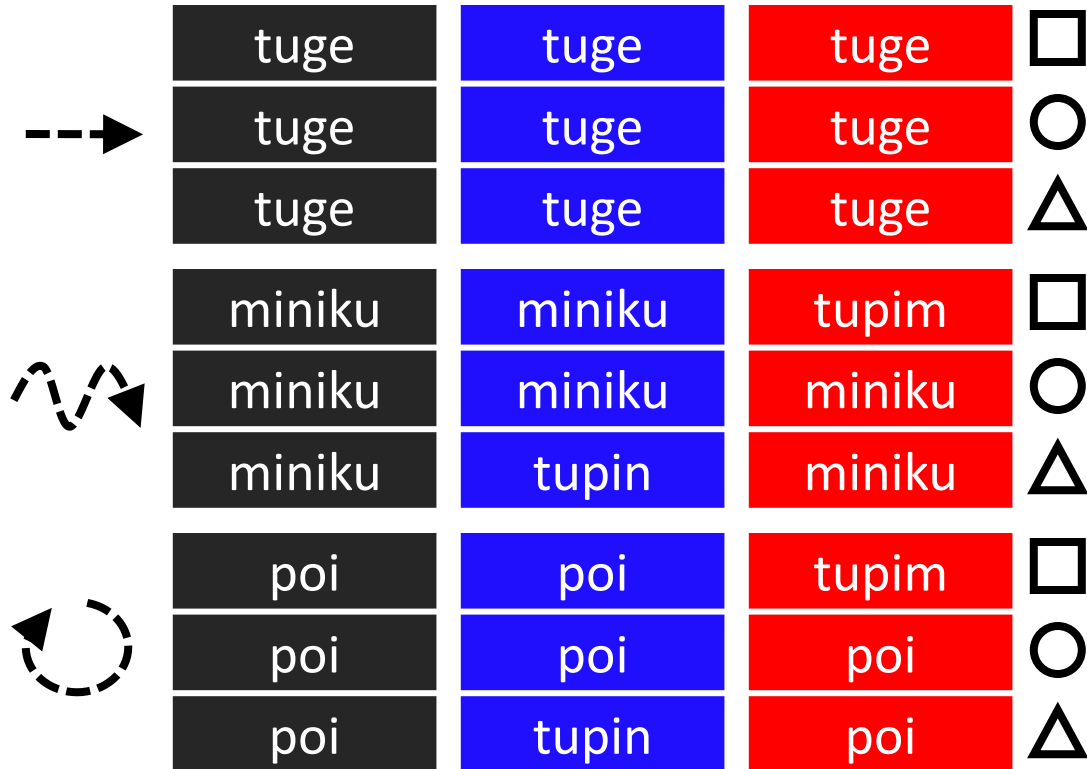
Generation 3 language from chain 4
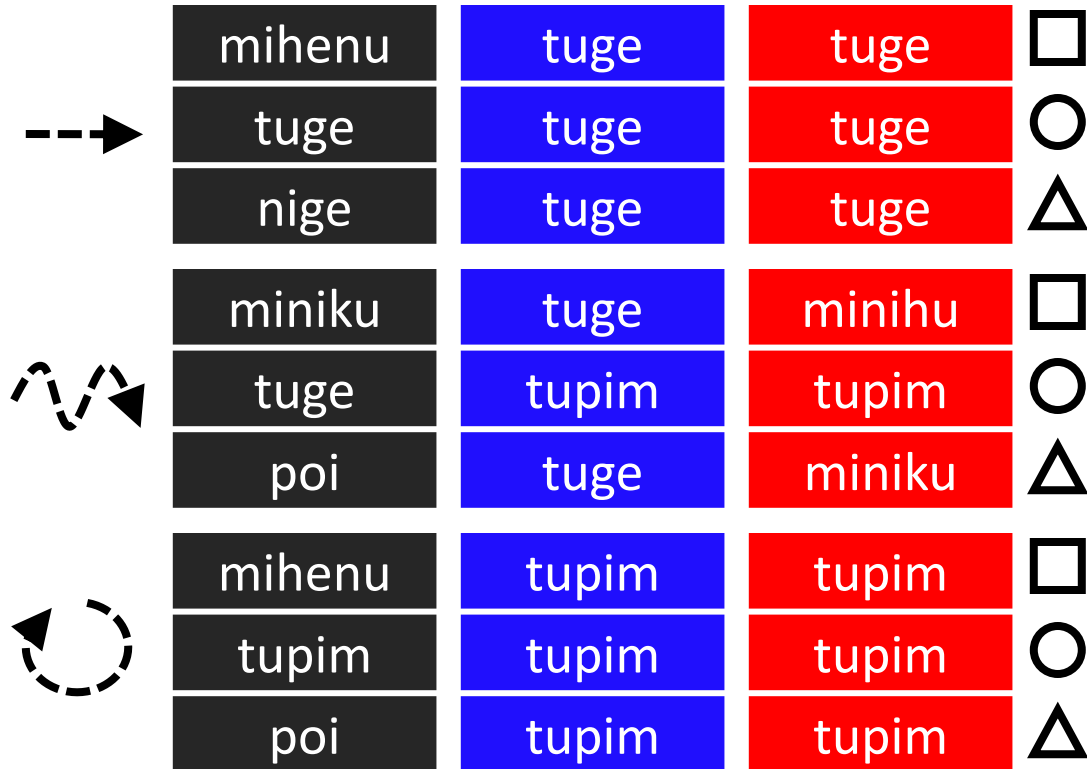
Generation 4 language from chain 4

Generation 5 language from chain 4

Generation 6 language from chain 4

Generation 7 language from chain 4

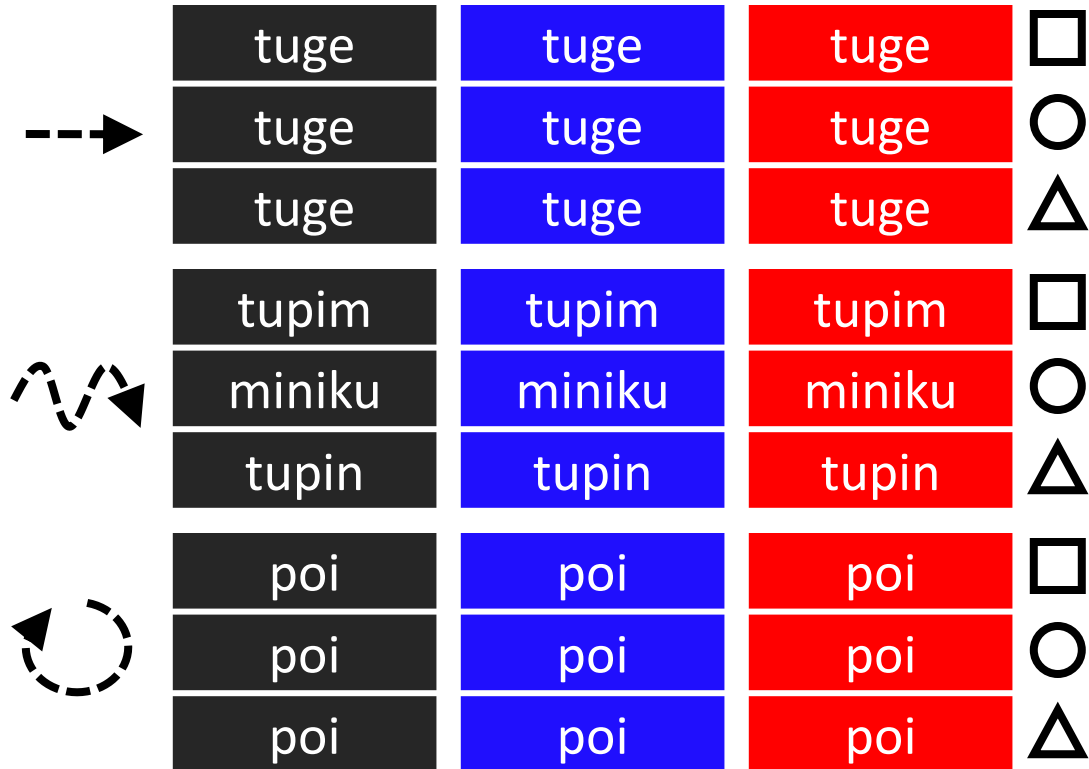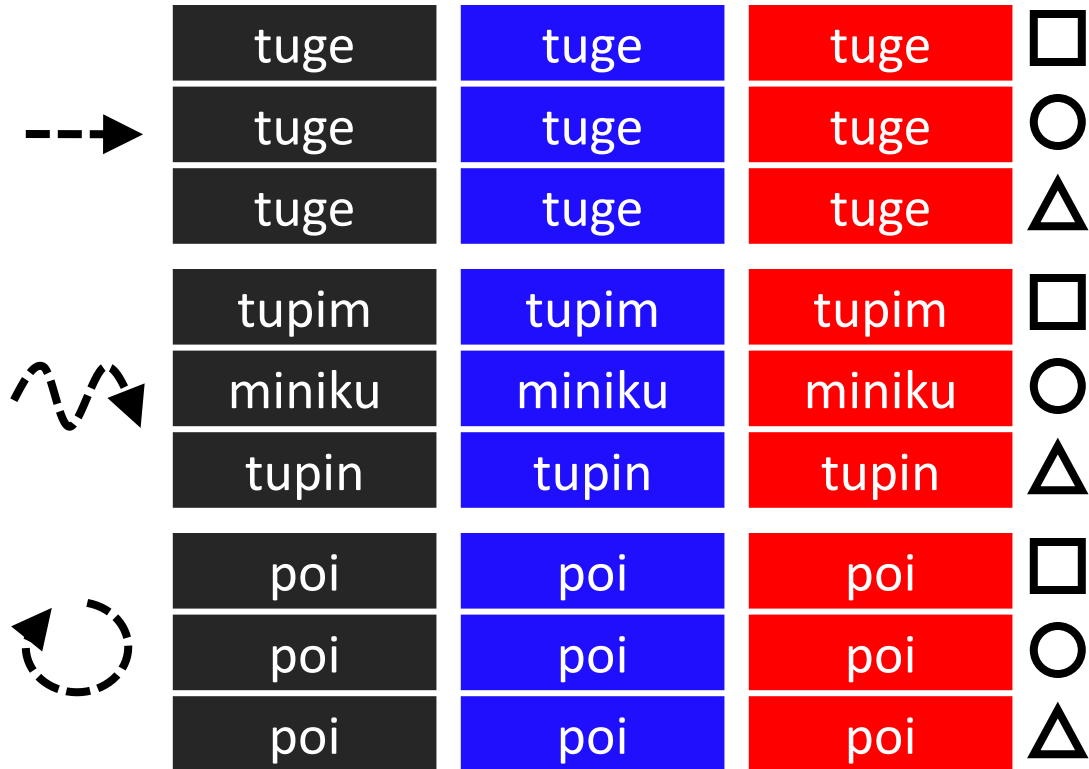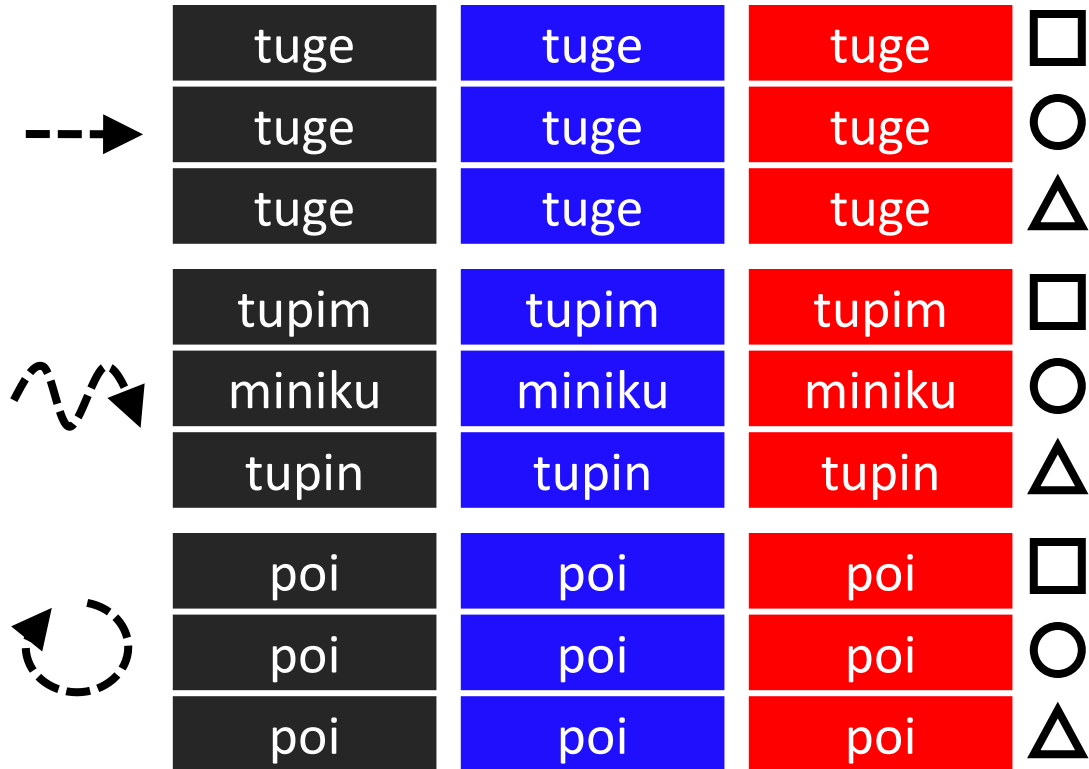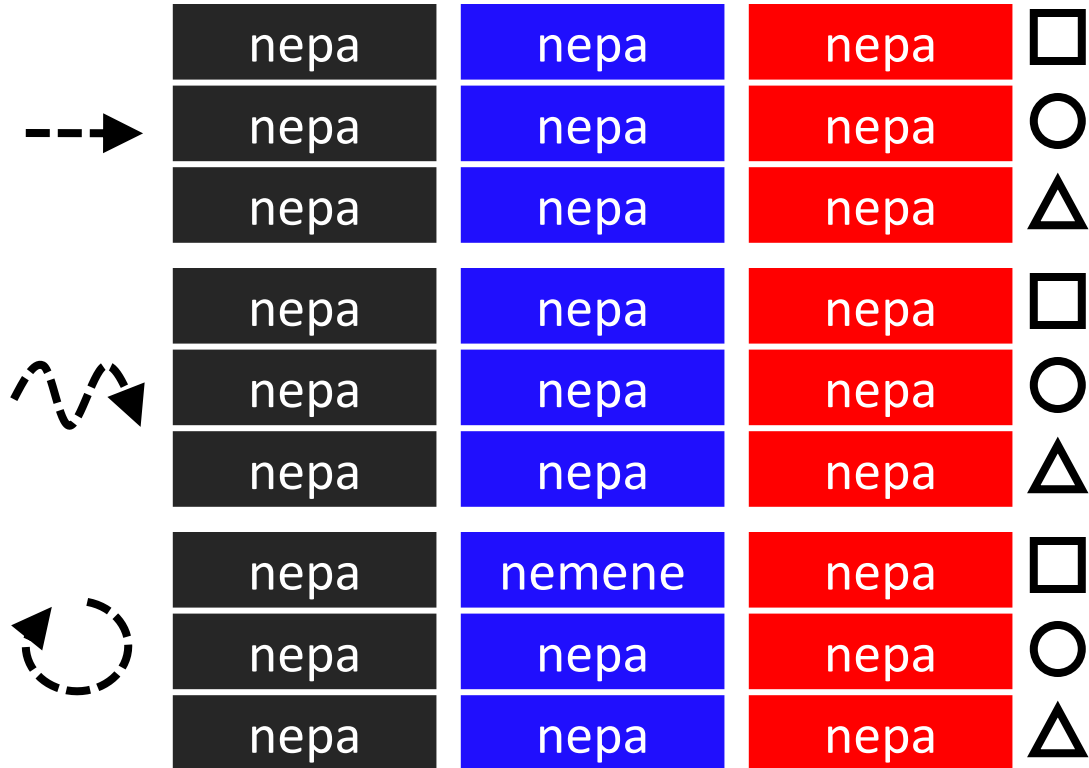|			
tuge	tuge	tuge	□
→ tuge	tuge	tuge	○
tuge	tuge	tuge	△
miniku	miniku	tupim	□
miniku	miniku	miniku	○
miniku	tupin	miniku	△
poi	poi	tupim	□
poi	poi	poi	○
poi	tupin	poi	△

Generation 8 language from chain 4
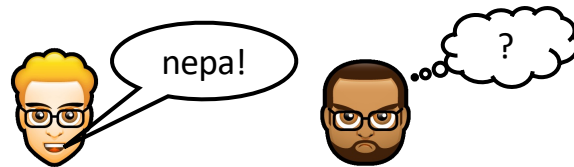
Generation 9 language from chain 4

Generation 10 language from chain 4

Final language from chain 1 (!)

The languages become **degenerate**

Generation 9 language from chain 5 (with homonymy filter)

Beckner et al. (2017)

Reanalysis/gentle roasting of Kirby, Cornish & Smith (2008)

- Our sample size was tiny
- Our statistics were rudimentary
- They find an interesting (?) difference between semantic dimensions

Replication

- Participants recruited from MTurk
- N=240 (2 conditions, 12 chains per condition, 10 participants per chain)
- 22-25 minutes, paid $3

Measuring structure

"the dog chew-ed the bone" – "the dog lick-ed the bone"

Meaning distance = 1 (predicate)

Signal distance = 1 (verb stem)

"the dog chew-ed the bone" - "the dog lick-s the bone"

Meaning distance = 2 (predicate, tense)

Signal distance = 2 (verb stem, suffix)

Pairwise meaning and signal distances will be highly correlated in a compositional system: similar meanings map to similar signals (and dissimilar meanings map to dissimilar signals)

Measuring structure

- For every pair of meaning-signal pairs
  - Measure meaning distance (Hamming distance)
  - Measure signal distance (Levenshtein string-edit distance)
  - Correlate these distances
- Evaluate statistical significance of that correlation
  - Randomise label assignments, recalculate measure, repeat 1000 times to give distribution
  - Calculate z-score of veridical correlation

kunita

kuloja

Meaning Distance = 1
Signal distance = 3
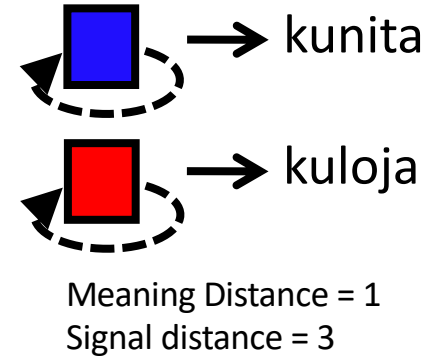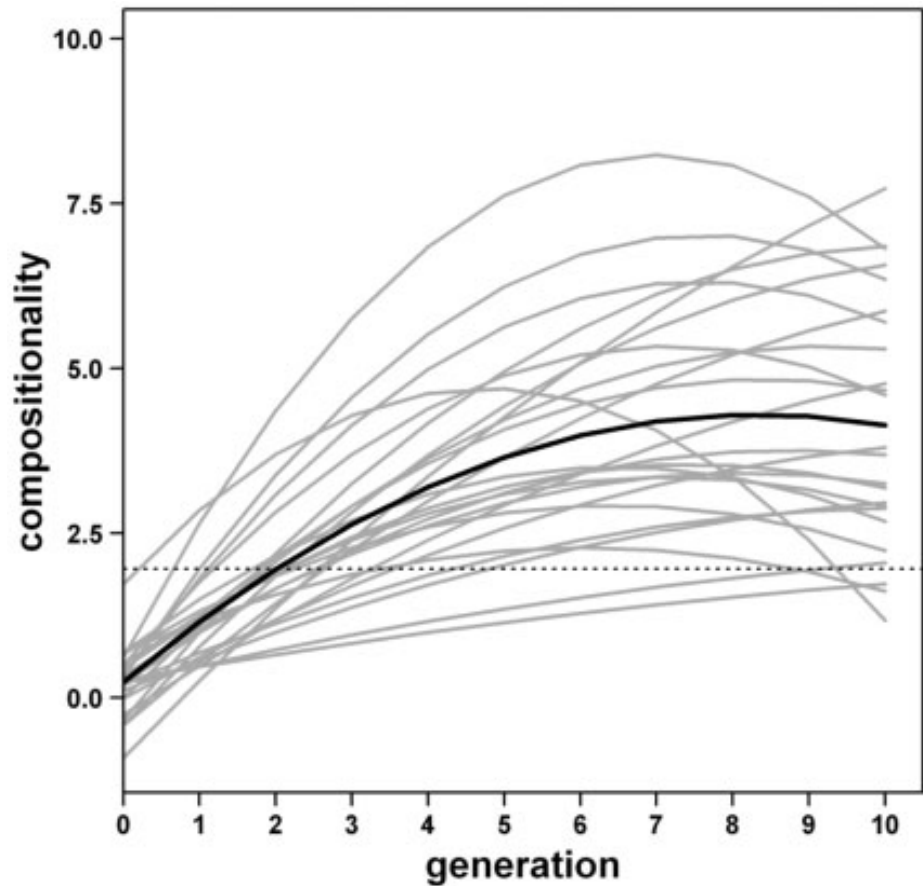
	'red'	'green'	'blue'	
'berry'	shen-to	shen-ta	shen-to	'1'
	shen-tra	shen-tro	shen-tra	'2'
	shen-trio	shen-trio	shen-trio	'3'
'key'	div-tro	div-tro	div-tro	'1'
	dev-tro	dev-tro	dev-etrio	'2'
	dev-stra	div-stra	dev-stra	'3'
'phone'	lolni-tro	lolni-tro	lolni-to	'1'
	lolne-stra	lolni-tro	lolne-stro	'2'
	lolni-tra	lolni-stra	lolni-stra	'3'

Beckner et al.'s conclusions

Iterated learning **does** produce structure

- Our 2008 result replicates with a proper sample size

- The method also works online…

- … but for this kind of challenging task, MTurk data is noisier?

Time for Q&A/discussion on this week's reading

Next up

Wednesday, 9am: lab

• Iterated learning, manipulating CSVs and looping trials

Next week: final lecture and lab 🥲

• Zipf's Law of Abbreviation, real-time dyadic interaction