Modelling contact-induced language change in Angolan Portuguese

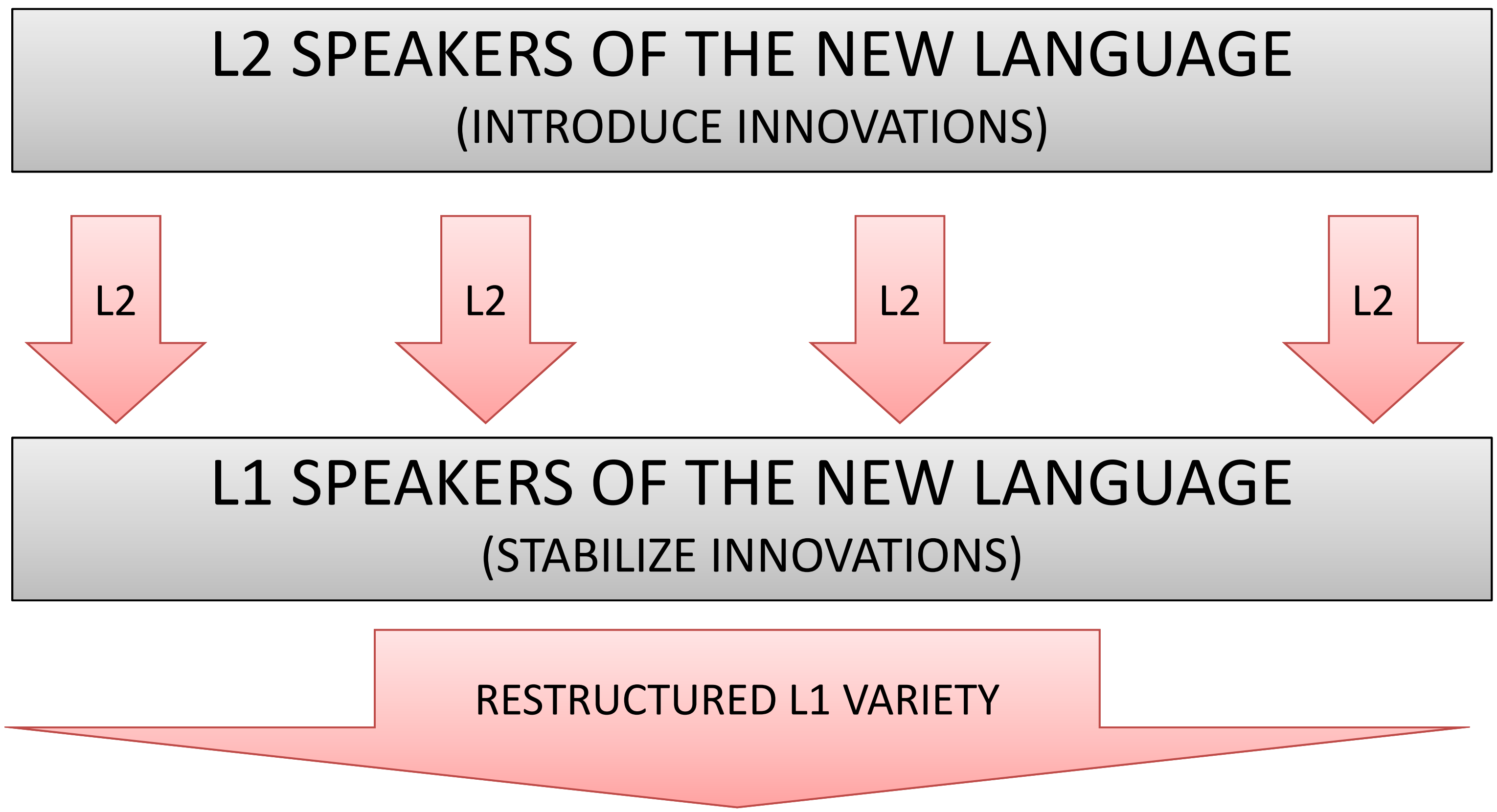
Anna Jon-And (Stockholm University/Dalarna University)

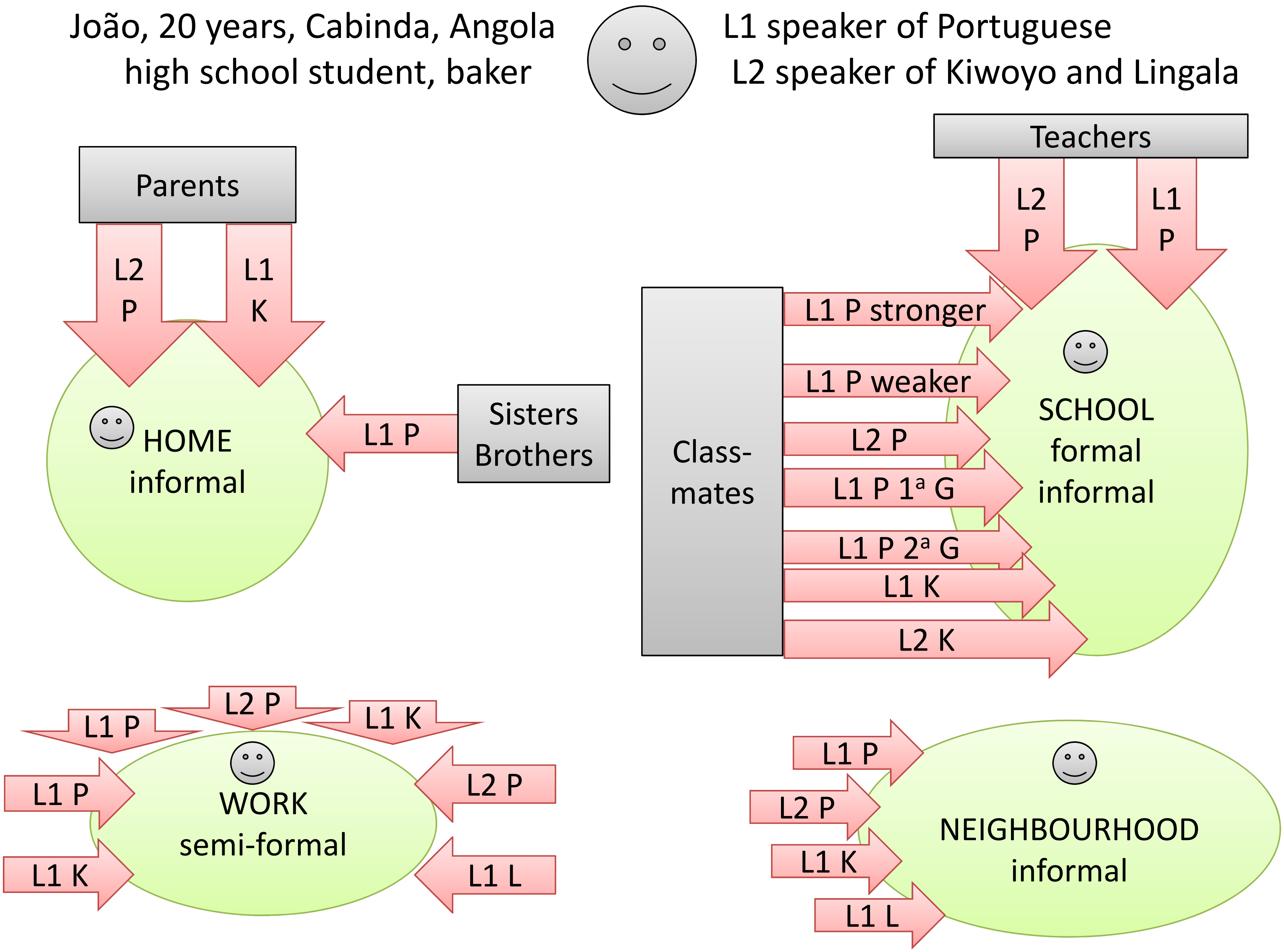
Elliot Aguilar (Graduate Center, City University of New York)

General aim:

To investigate how mathematical modelling and computational simulation (using evolutionary game theory) may help us to understand contact-induced language change.

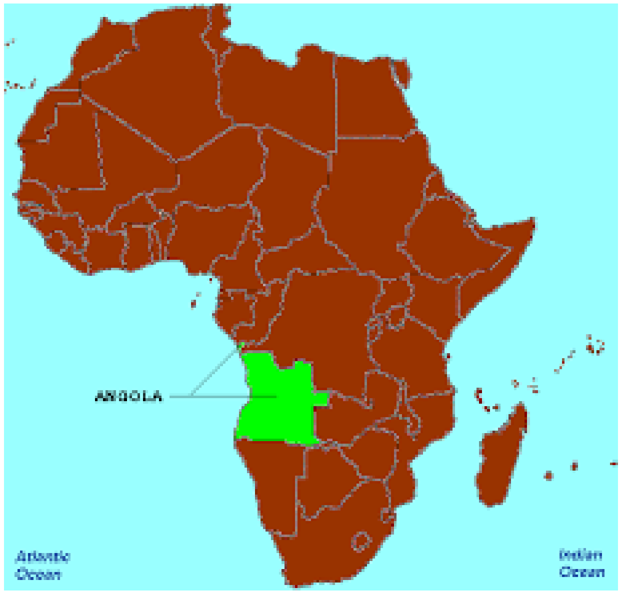
Language shift
Generalization:




João, 20 years, Cabinda, Angola
high school student, baker
L1 speaker of Portuguese
L2 speaker of Kiwoyo and Lingala
Parents
L2 P
L1 K
HOME
informal
Sisters
Brothers
L1 P
Teachers
L2 P
L1 P
Class-
mates
L1 P stronger
L1 P weaker
L2 P
L1 P 1ª G
L1 P 2ª G
L1 K
L2 K
SCHOOL
formal
informal
L1 P
L2 P
L1 K
L1 P
L1 K
L2 P
L1 L
WORK
semi-formal
L1 P
L2 P
L1 K
L1 L
NEIGHBOURHOOD
informal

The field

- Cabinda, Northern Angola
- Language shift: Bantu languages → Portuguese has been going on mainly since independence in 1974
- Today we find a mixture of L1 and L2 speakers of Portuguese among senior high school students of 20-25 years



Cabinda

More specific aims of the project

- Compare statistical analysis of linguistic variation in Cabinda Portuguese with outcomes of computational simulations that reconstruct the growth of the Portuguese speaking population in Angola since 1974, as well as the introduction and spread of contact related linguistic variants
- Formalize and test hypotheses concerning who may introduce, spread and stabilize innovative forms in a population constituted of L1 and L2 speakers due to language shift.

Earlier work on modelling contact-induced language change

- Parkvall, Jansson & Strimling (2014): *Simulating the genesis of Mauritian*
  - Frequencies in input may explain linguistic outcome (stabilization of lexicon, grammar and phonology)
- Baxter, Blythe, Croft & McCane (2009): *Modeling language change: An evaluation of Trudgill's theory of the emergence of New Zealand English*
  - Frequencies in input may *not* explain linguistic outcome (stabilization of phonology)

- Both papers deal with stabilization of variation given by demographical circumstances (migration)

What about innovation and propagation?

- Pierrehumbert, Stonedahl & Daland (2014) *A model of grassroots changes in linguistic systems*
  - Accounts for neutral innovation, propagation and stabilization
  - The spread of a neutral innovation depends on heterogeneous biases and network structure
  - Most innovations do not spread in a non-contact situation

Why does language contact accelerate change?

- Suggestions:
  - The same innovation appears at several places, not just once
  - The population is more biased for innovative forms
- Question for current simulation:
  - What conditions are required for an innovation to spread if only L2 speakers innovate?

The simulation

- Evolutionary game in expanding population
- Portuguese speaking population grows by recruitment of L2 speakers and birth of L1 speakers
- The simulation starts with the conservative variant of the linguistic feature fixed
- Add a probability that some individuals (L2 speakers) introduce innovation
- Individuals interact and update their probabilities of using conservative or innovative variant
- Parameters that vary
  - The influence of interactions between L1 and L2 speakers
  - Recruitment rate of L2 speakers in relation to birth rate of L1 speakers

Model description

An agent *i* chooses from a set of *n* linguistic variants.

$$\mathbf{V} = \{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}\}$$

With probabilities

$$P_i = \{p_{i1}, p_{i2}, \ldots, p_{in}\}$$

After hearing agent *j* utter v_k, agent *i* updates her probability of using v_k by:

$$p'_{ik} = p_{ik} + (1 - p_{ik})l_{ij}$$

And of all over variants m≠k by:

$$p'_{im} = p_{im} - p_{im}l_{ij}$$

Simulation details

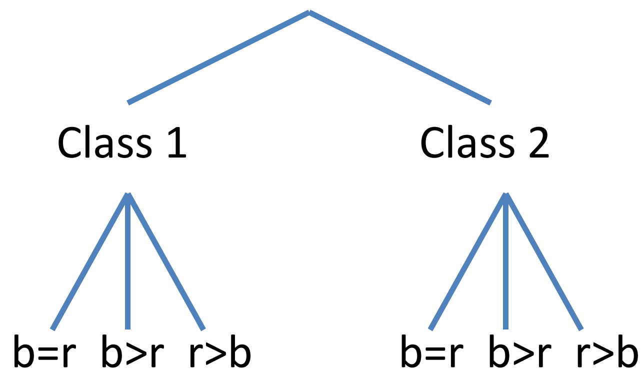
- Agents meet and interact, 365 times/year, for 25 years.
- Every year:
  - Births occur with rate b.
  - Deaths occur with rate d.
  - Recruitments of L2 speakers occur with rate r.

Only L2 speakers can introduce a new variant with probability m.

Run summary

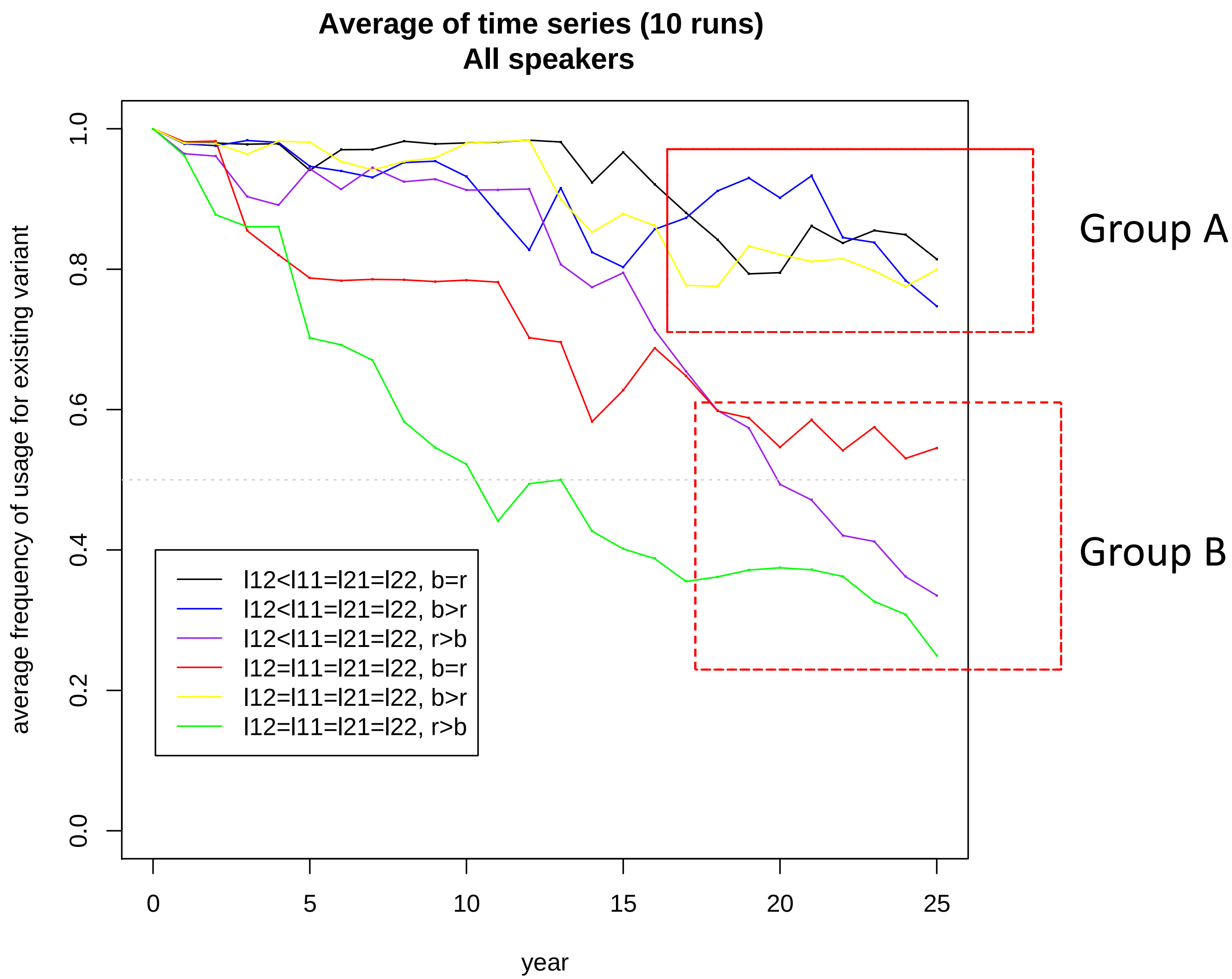
m=.5, d=.015, interactions/year=365, years=25, init. N=100

- We ran two classes of runs with three conditions each:
- Class 1: $l_{12}<l_{11}=l_{21}=l_{22}$
- Class 2: $l_{12}=l_{11}=l_{21}=l_{22}$



- Conditions: b=r, b>r, r>b (b={.046,.1}, r={.046,.1})
- 10 runs each param. setting.

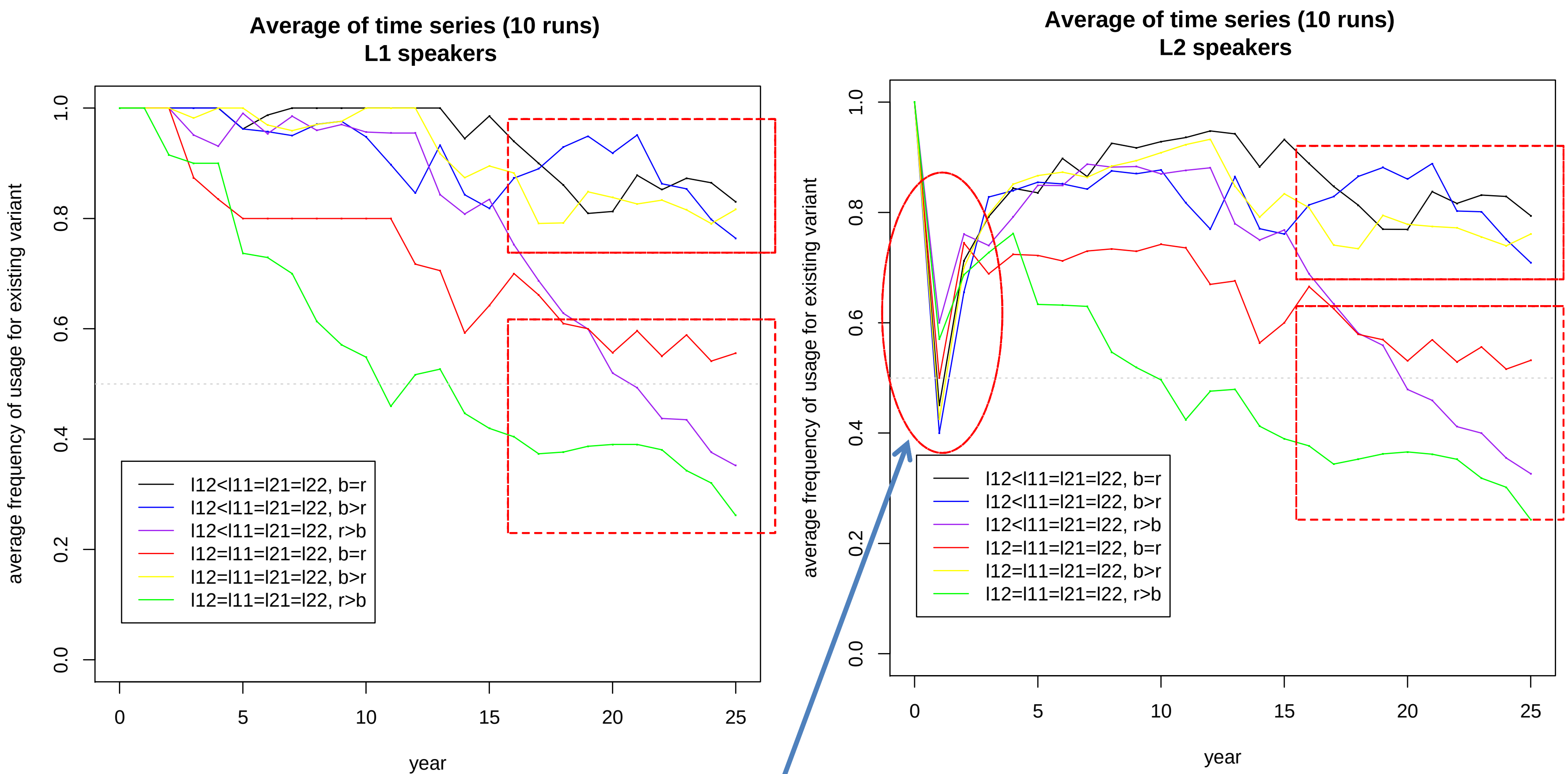
Results


Average of time series (10 runs)
All speakers
average frequency of usage for existing variant
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
year
l12<l11=l21=l22, b=r
l12<l11=l21=l22, b>r
l12<l11=l21=l22, r>b
l12=l11=l21=l22, b=r
l12=l11=l21=l22, b>r
l12=l11=l21=l22, r>b
Group A
Group B

- Group A: Runs where $b>r$ or $b=r$ but $l_{12}<l_{11}=l_{21}=l_{22}$

- Group B: Runs where $r>b$ or $b=r$ but $l_{12}=l_{11}=l_{21}=l_{22}$

No difference between L1 and L2



Caused by small number of L2 speakers early on

- Variable interactions determine probability of replacement.

Frequency of total rereplacement by mutant variant (over 10 runs)		
	$L_{12}=L_{11}=L_{21}=L_{22}$	$L_{12}<L_{11}=L_{21}=L_{22}$
birth rate = recruitment rate	.4	.1
birth rate > recruitment rate	.1	0
recruitment rate > birth rate	.6	.3

- Again, these results mirror the groupings from the time series.

Conclusions

- A variant introduced by L2 speakers will have a greater spread if:
  - recruitment rate of L2 speakers exceeds birth rate for L1 speakers *or*
  - recruitment rate of L2 speakers equals birth rate for L1 speakers *and* L1 speakers are not less influenced by interactions with L2 speakers than with L1 speakers.
- The spread of the innovative variant follows a similar pattern among L1 and L2 speakers

Discussion

- Our runs are likely still in the transient phase (Haven’t reached stationary phase).
- How do we achieve variation within the population?
  - Heterogeneous biases? (+/- sensetive to innovative variants; +/- categorical behaviour)
  - More advanced network structure?