Problems in Population Models of Language Change

Jordan Kodner
University of Pennsylvania

FWAV 4, June 29, 2017
University of York
Outline

● Frameworks for Population-Level Change
● Description of our Framework
● Population Size and Assumptions about the Grammar
● Generating S-Curves in Realistic Networks - the Cot-Caught Merger
● Capturing Complex Paths of Change - NCS in the St. Louis Corridor
Important Points

Population models and learning models interact

- Assumptions must be carefully considered when modelling change
- Attested paths of change are governed by these interactions
  - Neither alone provides the full picture
  - Both should be studied to the extent possible
Existing Frameworks
Three Classes of Framework

1. Concrete Frameworks
2. Network Frameworks
3. Algebraic Frameworks
Three Classes of Framework

1. Concrete Frameworks
   - Individual agents on a grid moving randomly and interacting
Three Classes of Framework

1. Concrete Frameworks
   - Individual agents on a grid moving randomly and interacting
     + Gradient interaction probability for free
     + Diffusion is straightforward
     - Not a lot of control over the network
     - Thousands of degrees of freedom -> should run many many times -> slow
     - Unclear how to include a learning model
Three Classes of Framework

1. Concrete Frameworks

2. Network Frameworks
   - Speakers are nodes in a graph, edges are possibility of interaction
Three Classes of Framework

1. **Concrete Frameworks**

2. **Network Frameworks**
   - Speakers are nodes in a graph, edges are possibility of interaction
   - Much more control over network structure
   - Easy to model concepts from the sociolinguistic lit. (e.g., Milroy & Milroy)

   - Nodes only interact with immediate neighbors -> slow and less realistic?
   - Practically implemented as random interactions between neighbors -> same problem as #1
Three Classes of Framework

1. Concrete Frameworks
2. Network Frameworks
3. Algebraic Frameworks
   - Expected outcome of interactions in a perfectly mixed population is calculated analytically
Three Classes of Framework

1. Concrete Frameworks
2. Network Frameworks
3. Algebraic Frameworks

- Expected outcome of interactions in a perfectly mixed population is calculated analytically
  
  + Less reliance on random processes -> faster and more direct
  + Clear how to insert learning models into the framework
  
  - No network structure! Always implemented over perfectly mixed populations
Our Framework
Best of Both Worlds

- An algebraic model operating on network graphs
Best of Both Worlds

- An **algebraic model** operating on **network graphs**
  - No random process in the core algorithm
  - Fast and efficient
Best of Both Worlds

- An **algebraic model** operating on **network graphs**
  - No random process in the core algorithm
  - Fast and efficient
  - Models language change in social structures
Formal Description

Each iteration has two steps

1. **Diffusion** - calculate how variants propagate
2. **Transmission** - calculate how variants are learned
Diffusion

\[ P_{t+1} = B^\top \alpha (I - (1 - \alpha)A)^{-1} H (H^\top H)^{-1} \]

- **A**  \( n \times n \) adjacency matrix
- **\( \alpha \)** jump parameter
- **H**  \( n \times c \) community-membership
- **B**  \( c \times g \) distr. of grammars in comms
- **P**  \( c \times g \) distr. of grammars in inputs
Diffusion

\[ P_{t+1} = B^\top \alpha (I - (1 - \alpha)A)^{-1} H H^\top (H H^\top)^{-1} \]

- \( A \)  \( n \times n \) adjacency matrix
- \( \alpha \) jump parameter
- \( H \)  \( n \times c \) community-membership
- \( B \)  \( c \times g \) distr. of grammars in comms
- \( P \)  \( c \times g \) distr. of grammars in inputs

The network graph

Who speaks what in what proportion
Who hears what in what proportion
Transmission

- Dependent on the learning model
- Our implementation is modular, so many learning models can be slotted in
  - e.g., trigger-based learner (Gibson & Wexler 1994)
  - Variational learner (Yang 2000)
Transmission

- Dependent on the learning model
- Our implementation is modular, so many learning models can be slotted in
  - e.g., trigger-based learner (Gibson & Wexler 1994)
  - Variational learner (Yang 2000)
- Let $L$ be the distribution of grammars internalized by a learner who heard $P$
  - $L$ is a matrix consisting of $g$ vectors $l_1, l_2, \ldots, l_g$
- Define $g$ transition matrices $T_1, T_2, \ldots, T_g$, one for each potential target grammar

$$l_i = \text{dominant eigenvector of } \sum_{j=1}^{g} P_{t+1; j,i} T_j$$
Transmission and Grammatical Advantage

- If $L = P$, learners internalize variants at the rate they hear them
  - This yields neutral change
- Otherwise, learners choose variants in a way that biases some over others
  - Some variants have an advantage over others
  - This yields S-curve change in perfectly mixed populations
Population Size and Grammars
Background

- Simulations typically run with a few hundred agents
  - Kauhanen 2016, Stanford & Kenny 2013, Blythe & Croft 2012, etc.
- Is this true of actual speech communities?
Background

● **Simulations typically run with *a few hundred agents***
  ○ Kauhanen 2016, Stanford & Kenny 2013, Blythe & Croft 2012, etc.

● **Is this true of actual speech communities?**
  ○ *Maybe sometimes!*

*Anuta island*

Just one-sixth of a square mile in area.
Background

- **Simulations typically run with a few hundred agents**
  - Kauhanen 2016, Stanford & Kenny 2013, Blythe & Croft 2012, etc.

- **Is this true of actual speech communities?**
  - Maybe sometimes!
  - But not typically true of the communities under study

- **Martha’s Vineyard (Labov 1963)**
  - ~5,500 in winter → ~42,000 in summer c. 1960
  - Summer population largely from New England (cf Massachusetts 5.1mil in 1960)
Background

- Simulations typically run with **a few hundred agents**
  - Kauhanen 2016, Stanford & Kenny 2013, Blythe & Croft 2012, etc.
- Is this true of actual speech communities?
  - Maybe sometimes!
  - But not typically true of the communities under study
- Martha’s Vineyard (Labov 1963)
  - ~5,500 in winter → ~42,000 in summer c. 1960
  - Summer population largely from New England (cf Massachusetts *5.1mil in 1960*)
- Do-Support (Ellegård 1953)
  - Rise of do-support constructions in English **1400-1700**
  - Involved **millions** of individuals
When is this a Problem?

- If learners internalize a distribution of grammars (i.e. competing grammars) and the population is (approximately) uniformly mixed, it is *not a problem*
  - Change closely approximates the path followed in infinite populations
  - So small-population models are a useful convenience
When is this a Problem?

- If learners internalize a distribution of grammars (i.e. competing grammars) and the population is (approximately) uniformly mixed, it is not a problem
  - Change closely approximates the path followed in infinite populations
  - So small-population models are a useful convenience

- But, if either of the above does not hold, it is a problem (maybe)
  - It becomes impossible to untangle population and learning effects
Demonstration: Neutral Change

- Assume two connected communities
  - C1 begins with 100% variant 1
  - C2 begins with 100% variant 2
Demonstration: Neutral Change

- **Assume** two connected communities
  - C1 begins with 100% variant 1
  - C2 begins with 100% variant 2
- **Neutral change**
Demonstration: Neutral Change

● Assume **two connected communities**
  ○ C1 begins with 100% variant 1
  ○ C2 begins with 100% variant 2

● **Neutral change**

● Over time, each community should approach 50/50 mix
Demonstration: Neutral Change

- **Assume** two connected communities
  - C1 begins with 100% variant 1
  - C2 begins with 100% variant 2
- **Neutral change**
- **Over time, each community should approach 50/50 mix**
- **Assume** speakers internalize a single grammar
  - Chosen probabilistically
  - Weighted by rate in their input
Demonstration: Neutral Change

- **Assume two connected communities**
  - C1 begins with 100% variant 1
  - C2 begins with 100% variant 2
- **Neutral change**
- **Over time, each community should approach 50/50 mix**
- **Assume speakers internalize a single grammar**
  - Chosen probabilistically
  - weighted by rate in their input
  - cf Kauhanen 2016

![Image](Rise of Variant 2 in C1
$n = 200$)

Red curve predicted
Blue curves first 10 trials
$n = 200$

$\begin{align*}
\text{Most trials fix at 0\% or 100\%}
\end{align*}$

$n = 2,000$

$\begin{align*}
\text{Most trials hover near 50\%}
\end{align*}$

$n = 20,000$
Demonstration: Advantage

- Repeating the previous test but with an advantage
  - Single community beginning at 1% innovative grammar
  - Learners choose a single grammar probabilistically, weighted toward innovative
  - Logistic curve predicted

\[ n = 200 \]  \[ n = 2,000 \]  \[ n = 20,000 \]
Demonstration: Advantage

- At small $n$, S-curve change cannot arise

$n = 200$

$n = 2,000$

$n = 20,000$

Looks a lot like neutral change did!
Demonstration: Advantage

- At small $n$, S-curve change cannot arise
- At large $n$, S-curves become smooth

Looks a lot like neutral change did!
Takeaways

Population models and learning models interact
Takeaways

Population models and learning models interact

- “Innocuous” assumptions may dominate behavior
  - Here, choice of population size and single-grammar assumptions
  - Conclusions drawable for $n=200$ do not scale to $n=20,000$ or visa-versa
Takeaways

Population models and learning models interact

- “Innocuous” assumptions may dominate behavior
  - Here, choice of population size and single-grammar assumptions
  - Conclusions drawable for $n=200$ do not scale to $n=20,000$ or visa-versa

- Slightly different assumptions yield drastically different conclusions
  - Is neutral change well-behaved?
  - Do we expect to see S-curve change?
Complex Networks and S-Curves:
The Cot-Caught Merger in New England
Single-Grammar Learners

- The previous section pointed out a problem with single-grammar learners
- But it is not an indictment
Single-Grammar Learners

- The previous section pointed out a problem with single-grammar learners
- But it is not an indictment
- Some changes are neatly modeled as single-grammar processes
  - E.g., the spread of mergers, e.g., cot-caught on the RI/MA border (Johnson 2007, Yang 2009)
The Cot-Caught Merger

- /ɒ/ “cot” merges with /ɔ/ “caught”
- Usually unconditioned
The Cot-Caught Merger

- /d/ “cot” merges with /ɔ/ “caught”
- Usually unconditioned
- Present in some dialects of North American English
  - Eastern New England
  - Western PA
  - Lower Midwest
  - The West
  - Canada (even NL!)

Do you pronounce “cot” and “caught” the same?

Merged
Unmerged

Joshua Katz, Dept of Statistics, NC State University
The Cot-Caught Merger

- /ɒ/ “cot” merges with /ɔ/ “caught”
- Usually unconditioned
- Present in some dialects of North American English
  - Eastern New England
  - Western PA
  - Lower Midwest
  - The West
  - Canada (even NL!)
- It is spreading into Rhode Island (Johnson 2007)
Modeling Merger Acquisition

- **Claim:** Mergers tend to spread because the merged grammar has a processing advantage
Modeling Merger Acquisition

- **Claim:** Mergers tend to spread because the merged grammar has a processing advantage
- **Asymmetric**
  - If a listener is unmerged, merged speakers create misunderstandings
  - If a listener is merged, unmerged speakers do not create misunderstandings

\[
\pi_0 = p_0 M_+ + q_0 M_-
\]

Yang 2009
Modeling Merger Acquisition

- **Claim:** Mergers tend to spread because the merged grammar has a processing advantage
- **Asymmetric**
  - If a listener is unmerged, merged speakers create misunderstandings
  - If a listener is merged, unmerged speakers do not create misunderstandings
- **Calculated for cot-caught, if at least ~17% of input is merged, the learner acquires the merged grammar**

Yang 2009
The Problem

- Except under incredibly specific network settings, a near-uniform population fixes at 0% or 100% in a couple iterations
  - In our model, alpha must be within a 0.005 window to avoid this
  - alpha is never so finicky otherwise
- Not what has happened empirically
The Solution

- A more realistic network!
The Solution

- A more realistic network!
- Large populations are not homogeneous
  - Tend to consist of many tight clusters loosely connected together
  - Echoes of Milroy & Milroy’s “strong and weak connections”
The Solution

- A more realistic network!
- Large populations are not homogeneous
  - Tend to consist of many tight clusters loosely connected together
  - Echos of Milroy & Milroy’s “strong and weak connections”
  - Homophily
  - Physical geography
  - etc.
The Solution

- A more realistic network!
- Large populations are not homogeneous
  - Tend to consist of many tight clusters loosely connected together
  - Echos of Milroy & Milroy’s “strong and weak connections”
  - Homophily
  - Physical geography
  - etc.
- So we consider a loosely connected network of centralized clusters
The Solution

- A network of 39 loosely connected centralized clusters - all unmerged
- Plus one merged cluster
The Solution

- A network of 39 loosely connected centralized clusters - all unmerged
- Plus one merged cluster
- Clusters merges rapidly in succession
The Solution

- A network of 39 loosely connected centralized clusters - all unmerged
- Plus one merged cluster
- Clusters merges rapidly in succession
- But the community average is an S-curve
Properties of Change

The averaged S-curve slope:
- depends on the grammatical advantage and the network
Properties of Change

The averaged S-curve slope

- depends on the grammatical advantage and the network
- is improved by evolving the network
Properties of Change

The averaged S-curve slope
- depends on the grammatical advantage \textit{and} the network
- is improved by evolving the network
- is preserved when introduced with a time offset
Properties of Change

The averaged S-curve slope

- depends on the grammatical advantage and the network
- is improved by evolving the network
- is preserved when introduced with a time offset
  - Is compatible with the Constant Rate Effect
Takeaways

Population models and learning models interact!
Takeaways

Population models and learning models interact!

- They conspire to yield empirically attested rates of change
Takeaways

Population models and learning models interact!

- They conspire to yield empirically attested rates of change
- S-curve change is possible outside competing-grammars scenarios
Takeaways

Population models and learning models interact!

- They conspire to yield empirically attested rates of change
- S-curve change is possible outside competing-grammars scenarios
- But competing- and single-grammars behave differently on small scales
Takeaways

Population models and learning models interact!

- They conspire to yield empirically attested rates of change
- S-curve change is possible outside competing-grammars scenarios
- But competing- and single-grammars behave differently on small scales
- Population effects preserve CRE across simultaneous changes with the same advantage
Complex Paths of Change: NCS in the St. Louis Corridor
Not all Change is Ideal

- An empirical fact
- Some change does not reach completion
- So it is obviously not S-shaped
The St. Louis Corridor

- Dialect region within US Midlands between Chicago and St. Louis
- But has features from the Inland North
  - Northern Cities Shift (NCS)
  - Has advanced and retreated

ANAE 2006
The St. Louis Corridor

- NCS entered the Corridor via Route 66 during the Great Depression

Friedman 2014
The St. Louis Corridor

- NCS entered the Corridor via Route 66 during the Great Depression
- Path of change is different On-Route and Off-Route
  - NCS peaks first On-Route
  - NCS peaks higher On-Route

Friedman 2014
The St. Louis Corridor

- NCS entered the Corridor via Route 66 during the Great Depression
- Path of change is different On-Route and Off-Route
  - NCS peaks first On-Route
  - NCS peaks higher On-Route

Friedman 2014
The St. Louis Corridor

- NCS entered the Corridor via Route 66 during the Great Depression
- Path of change is different On-Route and Off-Route
  - NCS peaks first On-Route
  - NCS peaks higher On-Route
- Typical of two-compartment systems

Wikipedia

“On-Route”
“Off-Route”
Modelling the Corridor: Network Structure

Community Types:
- Midlands (1; “background”)
- Chicago (1)
- On-Route (19)
- Off-Route (19)
Modelling the Corridor: Network Structure

Community Types:
- Midlands (1; “background”)
- Chicago (1)
- **On-Route** (19)
- **Off-Route** (19)

Connections:
- Midlands to all **On-Route** and **Off-Route**
- Chicago to all **On-Route**
- **On-Route** to two adjacent **On-Route**
- **On-Route** to one adjacent **Off-Route**
- **Off-Route** to one adjacent **Off-Route**
Modelling the Corridor: History

- Vary a single parameter: Direction of movement to On-Route communities
Modelling the Corridor: History

- Vary a single parameter: **Direction of movement to On-Route communities**
- Tests Great Depression hypothesis
Modelling the Corridor: History

- Vary a single parameter: Direction of movement to On-Route communities
- Tests Great Depression hypothesis
- It would be too “easy” if we could vary multiple parameters
  - Movement Off-Route
  - Strength of connections between On-Route and Off-Route
  - Strength of connections between On/Off-Route and Chicago/Midlands
  - Advantage of NCS
  - Etc.
Modelling the Corridor: History

- Vary a single parameter: Direction of movement to On-Route communities
- Tests Great Depression hypothesis
- It would be too “easy” if we could vary multiple parameters
  - Movement Off-Route
  - Strength of connections between On-Route and Off-Route
  - Strength of connections between On/Off-Route and Chicago/Midlands
  - Advantage of NCS
  - Etc.
- And the results would be less meaningful
Modelling the Corridor: History

- Vary a single parameter: Direction of movement to On-Route communities
- Tests Great Depression hypothesis

Stage 1 - 5 iterations
   No movement (speaker interaction only)

Stage 2 - 20 iterations
   2% movement from Chicago to On-Route “Great Depression”

Stage 3 - 75 iterations
   2% movement from Midlands to On-Route “Post-Depression”
Modelling the Corridor: The Variable

- Treating the NCS as a single binary variable subject to competing grammars
Modelling the Corridor: The Variable

- Treating the NCS as a single binary variable subject to competing grammars
- **Community Variable Distributions:**
  - Chicago fixed at 100% NCS+
  - Midlands fixed at 100% NCS-
  - On/Off-Route begins 100% NCS- but is allowed to vary
Modelling the Corridor: The Variable

- Treating the NCS as a single binary variable subject to competing grammars
- Community Variable Distributions:
  - Chicago fixed at 100% NCS+
  - Midlands fixed at 100% NCS-
  - On/Off-Route begins 100% NCS- but is allowed to vary
- Tested as neutral, slightly advantaged, and heavily advantaged change
Results: Neutral Change

- A classic two-compartment pattern arises
Results: Neutral Change

- A classic two-compartment pattern arises
- NCS peaks higher and earlier **On-Route** than **Off-Route**
Results: Neutral Change

- A classic two-compartment pattern arises
- NCS peaks higher and earlier on-Route than off-Route
- NCS continues to increase off-Route even after on-Route population movements are reversed
Results: Advantaged Change

- Advantaged change resists being “tamped down” Off-Route
  - NCS recedes given a slight advantage
  - NCS advances given a heavy advantage
Results: Advantaged Change

- Advantaged change resists being “tamped down” **Off-Route**
  - NCS recedes given a slight advantage
  - NCS advances given a heavy advantage
- Exists some threshold above which indirect action **On-Route** is insufficient
Results: Advantaged Change

- Advantaged change resists being “tamped down” Off-Route
  - NCS recedes given a slight advantage
  - NCS advances given a heavy advantage
- Exists some threshold above which indirect action On-Route is insufficient
- Can be solved with additional model parameters
  - Rate of movement Off-Route
  - The advantage itself
  - etc.

Slight Advantage
\[ a=0.80, \quad b=0.82 \]

Heavy Advantage
\[ a=0.80, \quad b=0.85 \]
Final Takeaways

Population models and learning models interact!
Final Takeaways

Population models and learning models interact!

- Assumptions must be carefully considered when modelling change
  - Under what assumptions are results generalizable?
Final Takeaways

Population models and learning models interact!

● Assumptions must be carefully considered when modelling change
  ○ Under what assumptions are results generalizable?

● Attested paths of change are governed by these interactions
  ○ Sometimes explicitly e.g., the St. Louis Corridor
  ○ Sometimes implicitly e.g., New England cot-caught
End

Code Available here:

github.com/jkodner05/NetworksAndLangChange
Extra slides: Diffusion
Diffusion

\[ P_{t+1} = B^\top \alpha (I - (1 - \alpha)A)^{-1} H(H^\top H)^{-1} \]

- **A**: \( n \times n \) adjacency matrix
- **\( \alpha \)**: jump parameter
- **H**: \( n \times c \) community-membership
- **B**: \( c \times g \) distr. of grammars in comms
- **P**: \( c \times g \) distr. of grammars in inputs

- Indicates directed weighted edges between speakers in network
- Column stochastic
- Easy to make undirected or unweighted
Diffusion

\[ P_{t+1} = B^\top \alpha (I - (1 - \alpha)A)^{-1} H(H^\top H)^{-1} \]

- \( A \) \( n \times n \) adjacency matrix
- \( \alpha \) jump parameter
- \( H \) \( n \times c \) community-membership
- \( B \) \( c \times g \) distr. of grammars in comms
- \( P \) \( c \times g \) distr. of grammars in inputs

- Decides “fluidity” of interactions
- Jump distances follow a geometric distribution
  - Speakers are most likely to interact with adjacent speakers
  - But occasionally talk to others far away
- Also implemented with Poisson distribution
Diffusion

\[ P_{t+1} = B^\top \alpha (I - (1 - \alpha)A)^{-1} H(H^\top H)^{-1} \]

- **A** \( n \times n \) adjacency matrix
- **\( \alpha \)** jump parameter
- **H** \( n \times c \) community-membership
- **B** \( c \times g \) distr. of grammars in comms
- **P** \( c \times g \) distr. of grammars in inputs
- **Indicator matrix**
- **Defines “community” membership**
- **More on this later...**
**Diffusion**

\[ P_{t+1} = B^\top \alpha (I - (1 - \alpha)A)^{-1} H(H^\top H)^{-1} \]

- **A** \( n \times n \) adjacency matrix
- **\( \alpha \)** jump parameter
- **H** \( n \times c \) community-membership
- **B** \( c \times g \) distr. of grammars in comms
- **P** \( c \times g \) distr. of grammars in inputs

- Distribution of grammars
- According to which community members produce utterances
Diffusion

$$P_{t+1} = B^\top \alpha (I - (1 - \alpha)A)^{-1} H(H^\top H)^{-1}$$

- A \( n \times n \) adjacency matrix
- \( \alpha \) jump parameter
- H \( n \times c \) community-membership
- B \( c \times g \) distr. of grammars in comms
- P \( c \times g \) distr. of grammars in inputs
- Distribution of grammars
- Heard by learners of each community
Tracking Individuals

- The model can track the average behavior of “communities” rather than individuals.
- If \( c = n \), then \( H \) is \( n \times n \), and the full descriptive detail of the model is available.
  - \( H \) becomes the identity matrix, and the formula for \( P \) can be rewritten as:
    \[
    P_{t+1} = B^\top \alpha (I - (1 - \alpha)A)^{-1}
    \]
Tracking Communities

- If fine-grain detail is unnecessary, tracking community averages provides substantial computational speedup when $c << n$
- If each community is internally uniform, $n \times n \textbf{A}$ admits a $c \times c$ equitable-partition $\textbf{A}^\pi$
- Yielding a more efficient but equivalent update formula for $\textbf{P}$

\[
\textbf{A}^\pi = (\textbf{H}^\top \textbf{H})^{-1} \textbf{H}^\top \textbf{A} \textbf{H}
\]
\[
\textbf{P}_{t+1} = \alpha \textbf{B}^\top \textbf{H} (\textbf{I} - (1 - \alpha) \textbf{A}^\pi)^{-1} (\textbf{H}^\top \textbf{H})^{-1}
\]

Anecdotally, I can run $n = 20,000$ nets on my laptop with $\textbf{A}^\pi$ about as fast as $n = 2,000$ nets with $\textbf{A}$
Extra Slides: Transmission
Transmission Example

- Let there be two languages $L_1$ and $L_2$, the extensions of $g_1$ and $g_2$, produced with probabilities $P_1$ and $P_2$.
- $a = P_1[L_1 \cup L_2]$ \hspace{1cm} $1 - a = P_1[L_1 \setminus L_2]$  
- $b = P_2[L_1 \cup L_2]$ \hspace{1cm} $1 - b = P_2[L_2 \setminus L_1]$
Transmission Example

- Let there be two languages $L_1$ and $L_2$, the extensions of $g_1$ and $g_2$, produced with probabilities $P_1$ and $P_2$.

- $a = P_1[L_1 \cup L_2]$ \hspace{1cm} $1 - a = P_1[L_1 \setminus L_2]$

- $b = P_2[L_1 \cup L_2]$ \hspace{1cm} $1 - b = P_2[L_2 \setminus L_1]$

- Let $T_1$ and $T_2$ be transition matrices assuming $g_1$ and $g_2$ are the target grammars respectively

- $T_1 = [1 \ 0 ; 1-a \ a]$ \hspace{1cm} $T_2 = [b \ 1-b ; 0 \ 1]$
Transmission Example

$T_1 = 1 \ 0$

$1-a \ a$

$T_2 = b \ 1-b$

$0 \ 1$

- If the target grammar is $g_1$, then in the limit...
Transmission Example

$T_1 = \begin{bmatrix} 1 & 0 \\ 1-a & a \end{bmatrix}$

$T_2 = \begin{bmatrix} b & 1-b \\ 0 & 1 \end{bmatrix}$

- If the target grammar is $g_1$, then in the limit...
  - Learners who initially hypothesize $g_1$ will always remain in $g_1$
Transmission Example

\[ T_1 = \begin{bmatrix} 1 & 0 \\ 1-a & a \end{bmatrix} \]

\[ T_2 = \begin{bmatrix} b & 1-b \\ 0 & 1 \end{bmatrix} \]

- If the target grammar is \( g_1 \), then in the limit...
  - Learners who initially hypothesize \( g_1 \) will always remain in \( g_1 \)
  - Learners who initially hypothesize \( g_2 \) will remain at \( g_2 \) with probability \( a \)
Transmission Example

$T_1 = \begin{bmatrix} 1 & 0 \\ 1-a & a \end{bmatrix}$

$T_2 = \begin{bmatrix} b & 1-b \\ 0 & 1 \end{bmatrix}$

- If the target grammar is $g_1$, then in the limit...
  - Learners who initially hypothesize $g_1$ will always remain in $g_1$
  - Learners who initially hypothesize $g_2$ will remain at $g_2$ with probability $a$
  - Or switch to $g_1$ with probability $1-a$