Modeling Language Change from the Population's Perspective

Jordan Kodner
Stony Brook University

SYNC 2023
December 2, 2023

Outline for Today

- Populations and Change
- Computational Modeling
- A Case Study:
  Northern Cities Features
  in the St. Louis Corridor

Populations and Change

The Individual and the Population

A long-standing paradox

- **Language in the colloquial sense, like “English” is an essentialist notion**
  e.g., there is no “English” as such. As linguists, you know that! 👍

The Individual and the Population

A long-standing paradox

- Language in the colloquial sense, like “English” is an essentialist notion
- The “real” language is the language of the individual

Converging perspectives from the Neogrammarians[1]
to American Structuralism[2] to Generative Grammar[3]
→ There are (at least) as many languages as there are individuals

[1] Paul 1880, [2] Bloomfield 1933, [3] Chomsky 1955

The Individual and the Population

A long-standing paradox

- Language in the colloquial sense, like “English” is an essentialist notion
- The “real” language is the language of the individual
- But when we say language change, we mean still change to “English”

The Individual and the Population

A long-standing paradox

- Language in the colloquial sense, like “English” is an essentialist notion
- The “real” language is the language of the individual
- But when we say language change, we mean still change to “English”
- But “English” has no independent existence, so it is individuals who change

The Individual and the Population

A long-standing paradox

- Language in the colloquial sense, like “English” is an essentialist notion
- The “real” language is the language of the individual
- But when we say language change, we mean still change to “English”
- But “English” has no independent existence, so it is individuals who change
- But if individuals change, then they don’t speak “English” anymore

The Individual and the Population

A long-standing paradox

- Language in the colloquial sense, like “English” is an essentialist notion
- The “real” language is the language of the individual
- But when we say language change, we mean still change to “English”
- But “English” has no independent existence, so it is individuals who change
- But if individuals change, then they don’t speak “English” anymore
- So, “English” doesn’t change. It cannot change…but it does, actually 😑

The Individual and the Population

A long-standing paradox

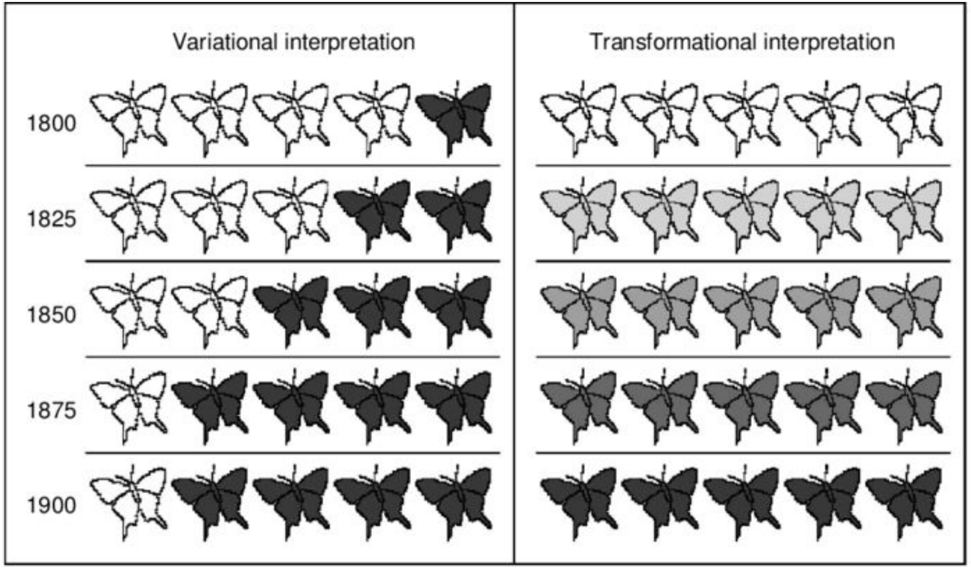
- Language in the colloquial sense, like “English” is an essentialist notion
- The “real” language is the language of the individual
- But when we say language change, we mean still change to “English”
- But “English” has no independent existence, so it is individuals who change
- But if individuals change, then they don’t speak “English” anymore
- So, “English” doesn’t change. It cannot change…but it does, actually 😑

The paradox is the consequence of essentialist thinking

Solving the Paradox: Variationism

A fundamental principle of population-level change

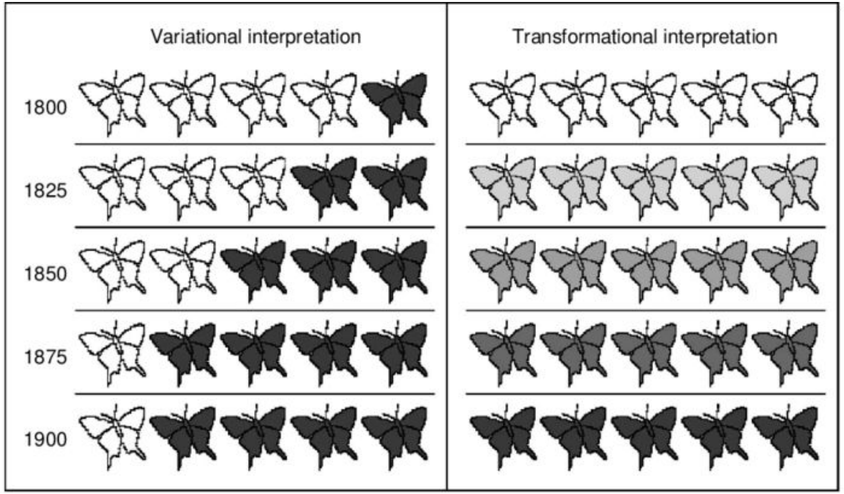
- **Contra an essentialist view of language/species**
- **Variational vs transformational change**



Solving the Paradox: Variationism

A fundamental principle of population-level change

- **Contra an essentialist view of language/species**
- **Variational vs transformational change**
- **Fundamental insight of**
  - **Darwinian evolution[1]**
  - **Variationist sociolinguistics[2]**
  - **Diachronic generative linguistics[3]**
  - **Diachronic usage-based linguistics[4]**
- **Weirdly, much less fundamental to**
  - **Cultural evolution of language[5]**



[1] Mayr 1982, [2] Weinreich, Labov, & Herzog 1968, *et seq*, [3] Kroch 1989, *et seq*, [3] Croft 2000, [5] Kirby 1999, *et seq*

Innovation vs Propagation

Two different sides of change that should not be conflated

Innovation - An Individual Phenomenon

- Where/how/with whom does an innovative variant originate?
- Language acquisition, individual creativity…
- The moment of innovation rarely appears in the historical record

Innovation vs Propagation

Two different sides of change that should not be conflated

Innovation - An Individual Phenomenon

- Where/how/with whom does an innovative variant originate?
- Language acquisition, individual creativity...
- The moment of innovation rarely appears in the historical record

Propagation - A Population Phenomenon

- How/why/through whom does an innovative variant spread?
- Both through the population and through an individual's linguistic system
- This may appear in the historical record, especially later stages

Connecting the Individual and Population: Actuation

Actuation = Innovation + uptake into the speech community[1]

(The hand-off from an individual-level process to a population-level one)

[1] definition paraphrased from Labov, Yaeger & Steiner 1972

Connecting the Individual and Population: Actuation

Actuation = Innovation + uptake into the speech community[1]

(The hand-off from an individual-level process to a population-level one)

The Actuation Problem[2]

- We can never know the exact circumstances at the moment that any particular innovation or actuation occurred
- Sociolinguists often (rightly?) have a negative outlook on actuation research
- The attested "innovators" of a change are probably actually early adopters[3]

[1] definition paraphrased from Labov, Yaeger & Steiner 1972, [2] Weinreich, Labov, & Herzog 1968, [3] Milroy & Milroy 1985

Connecting the Individual and Population: Actuation

Actuation = Innovation + uptake into the speech community[1]

(The hand-off from an individual-level process to a population-level one)

The Actuation Problem[2]

- **We can never know the exact circumstances at the moment that any particular innovation or actuation occurred**
- **Sociolinguists often (rightly?) have a negative outlook on actuation research**
- **The attested "innovators" of a change are probably actually early adopters[3]**

We can actually approach solving actuation…

[1] definition paraphrased from Labov, Yaeger & Steiner 1972, [2] Weinreich, Labov, & Herzog 1968, [3] Milroy & Milroy 1985

Connecting the Individual and Population: Actuation

Actuation = Innovation + uptake into the speech community[1]

(The hand-off from an individual-level process to a population-level one)

The Actuation Problem[2]

- We can never know the exact circumstances at the moment that any particular innovation or actuation occurred
- Sociolinguists often (rightly?) have a negative outlook on actuation research
- The attested "innovators" of a change are probably actually early adopters[3]

We can actually approach solving actuation…asymptotically.

We can get close, but we can never get there

[1] definition paraphrased from Labov, Yaeger & Steiner 1972, [2] Weinreich, Labov, & Herzog 1968, [3] Milroy & Milroy 1985

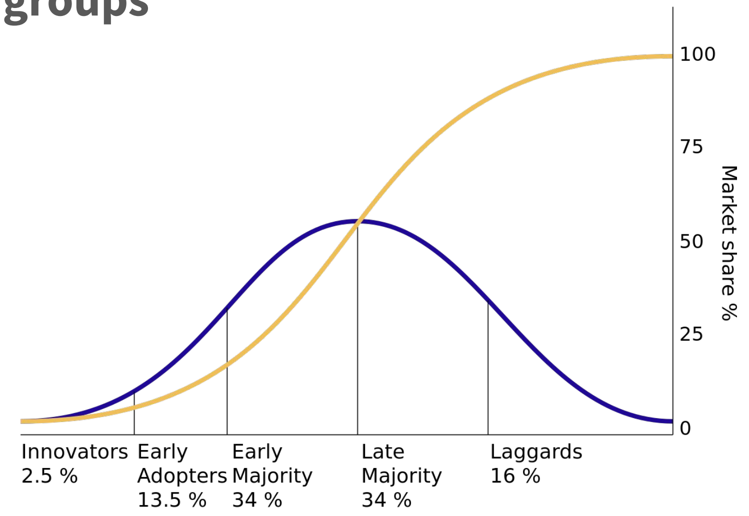
Diffusion of Innovation Theory[1]

Applies to linguistics and other phenomena

- **Innovations do not spread uniformly through groups**
- **Some individuals readily adopt innovations while some tend to resist**

Innovators → Early Adopters → Early Majority → Late Majority → Laggards

- **The population distribution is often normal**
- **The cumulative adoption curve is often a logistic S-curve (also called sigmoid)**



[1] Rogers 1962, *et seq*

Labovian Transmission and Diffusion[1]

Transmission ≈ Language Acquisition ≈ Vertical Transmission

"[The Neogrammarian] unbroken sequence of native-language acquisition by children"

- From parents and older age cohorts to children
- Generally faithful replication/incrementation of linguistic input
- Argued to be the primary source of linguistic diversity

[1] Labov 2007

Labovian Transmission and Diffusion[1]

Transmission ≈ Language Acquisition ≈ Vertical Transmission

"[The Neogrammarian] unbroken sequence of native-language acquisition by children"

- From parents and older age cohorts to children
- Generally faithful replication/incrementation of linguistic input
- Argued to be the primary source of linguistic diversity

- Insight shared by the Neogrammarians,[2] generative linguists,[3] and most child language acquisition researchers
- Tied to the concept of a Critical Period 👶

[1] Labov 2007, [2] Paul 1880, [2] e.g., Lightfoot 1979

Labovian Transmission and Diffusion[1]

Transmission ≈ Language Acquisition ≈ Vertical Transmission

"[The Neogrammarian] unbroken sequence of native-language acquisition by children"

- From parents and older age cohorts to children
- Generally faithful replication/incrementation of linguistic input
- Argued to be the primary source of linguistic diversity

Diffusion ≈ Adult Learning ≈ Horizontal Transmission

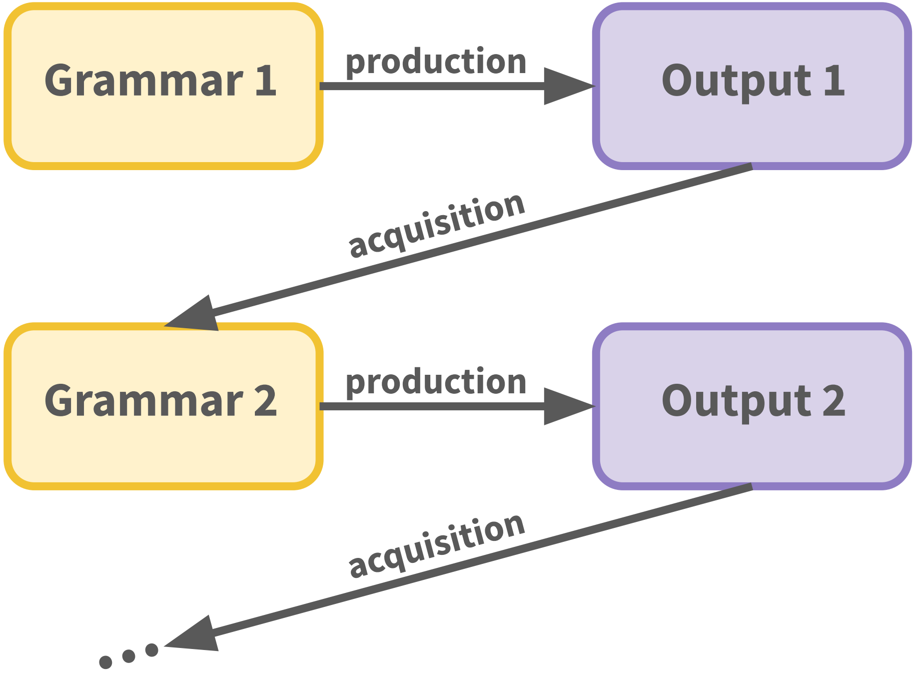
"the result of contact between the speech communities… transfer across branches of the family tree"

- From community to community, among mature speakers within communities
- Subject to social network density effects[2]
- Often manifests degradation of complex structural patterns

[1] Labov 2007, [2] Bloomfield 1933

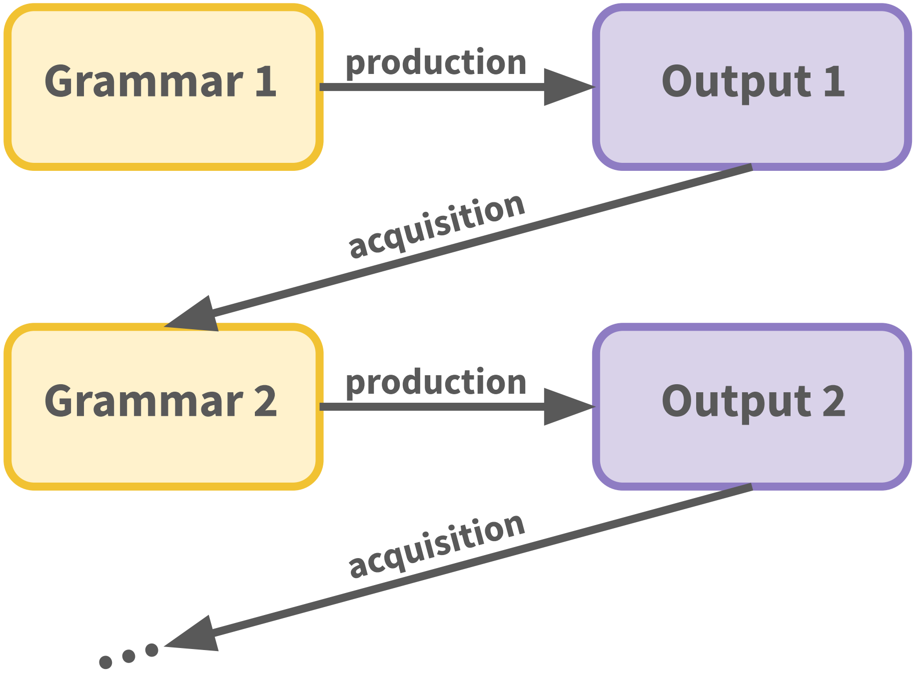
Connecting the Individual and Population: Z-Model

- **Andersen 1973** originally conceived of this as a cycle of error-prone abductive and inductive learning



Connecting the Individual and Population: Z-Model

- Andersen 1973 originally conceived of this as a cycle of error-prone abductive and inductive learning
- Can be interpreted as alternating I-language and E-language
- Presents a role for competence and performance, or representation, learning, and social/diachronic factors
- Primarily captures Labovian transmission

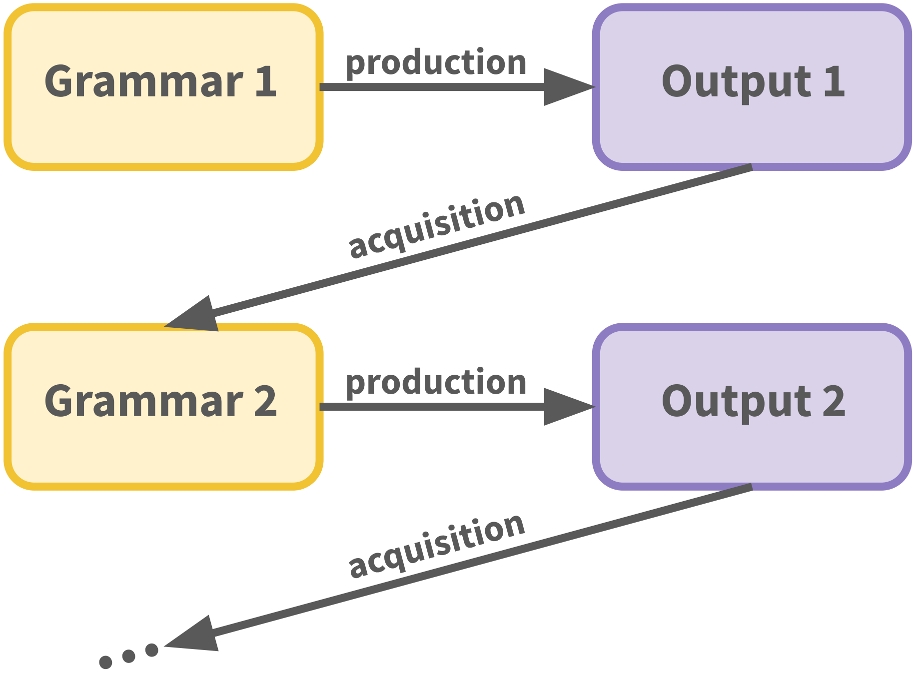


Insufficiency of the Z-Model of Transmission

- Andersen 1973 originally conceived of this as a cycle of error-prone abductive and inductive learning

Insufficiency of the Z-Model

- This is a linear chain
- There is no population!
- It cannot distinguish transformational vs variational change



Generalizing the Z-Model to Populations

Language change is a two-part cycle

1. **Population / Propagation**: How grammars are distributed in the community?
2. **Individual / Learning**: How individuals respond to the community languages

L: That which is transmitted
(≈language≈variety≈*lect≈E-language)

G: That which generates/describes /influences/is learned from L
(≈grammar≈variant≈I-language)

Generalizing the Z-Model to Populations

Language change is a two-part cycle

1. Population / Propagation: How grammars are distributed in the community?
2. Individual / Learning: How individuals respond to the community languages

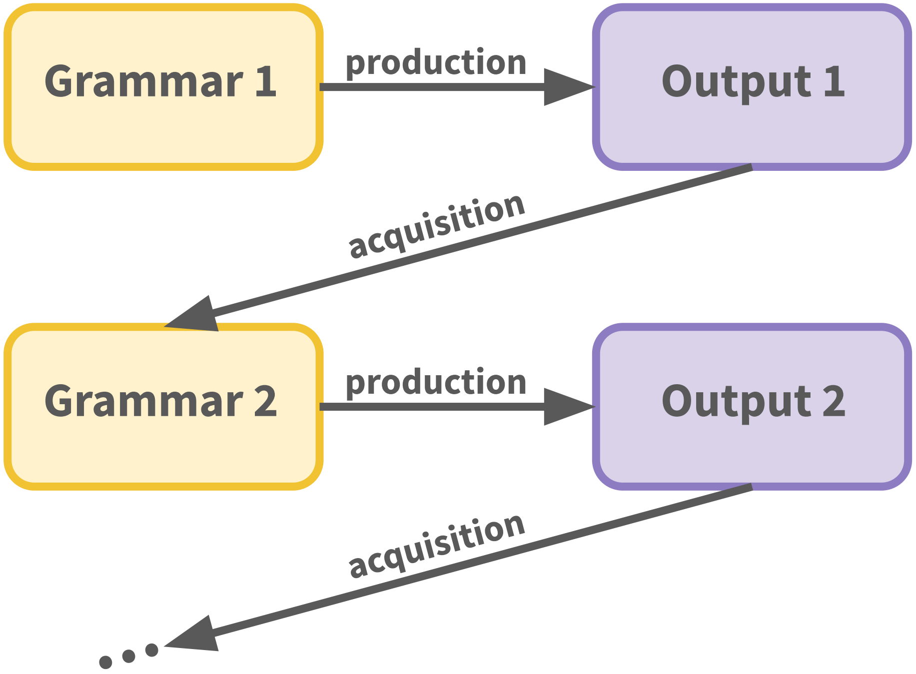
L: That which is transmitted
(≈language≈variety≈*lect≈E-language)

G: That which generates/describes/influences/is learned from L
(≈grammar≈variant≈I-language)

This unfolds on the population-level

→ L_{n-2} → G_{n-2} → L_{n-1} → G_{n-1} → L_n → G_n → L_{n+1} → G_{n+1} → L_{n+2} → G_{n+2} → L_{n+3} →

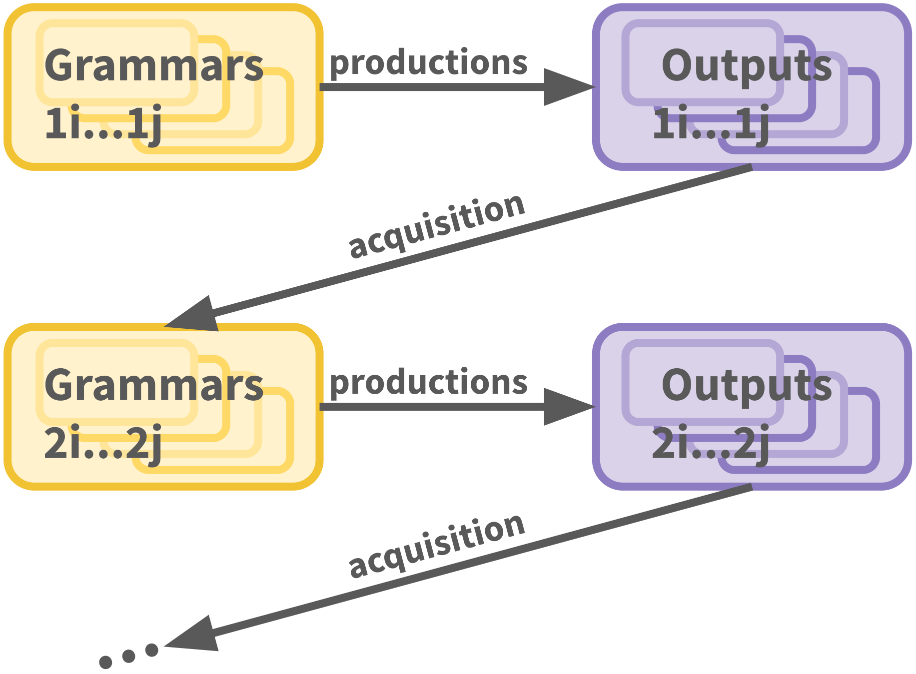
A Population-Level Z-Model of Transmission



A Population-Level Z-Model of Transmission

Individual production

- Variation across social settings
- Variation over lifetimes



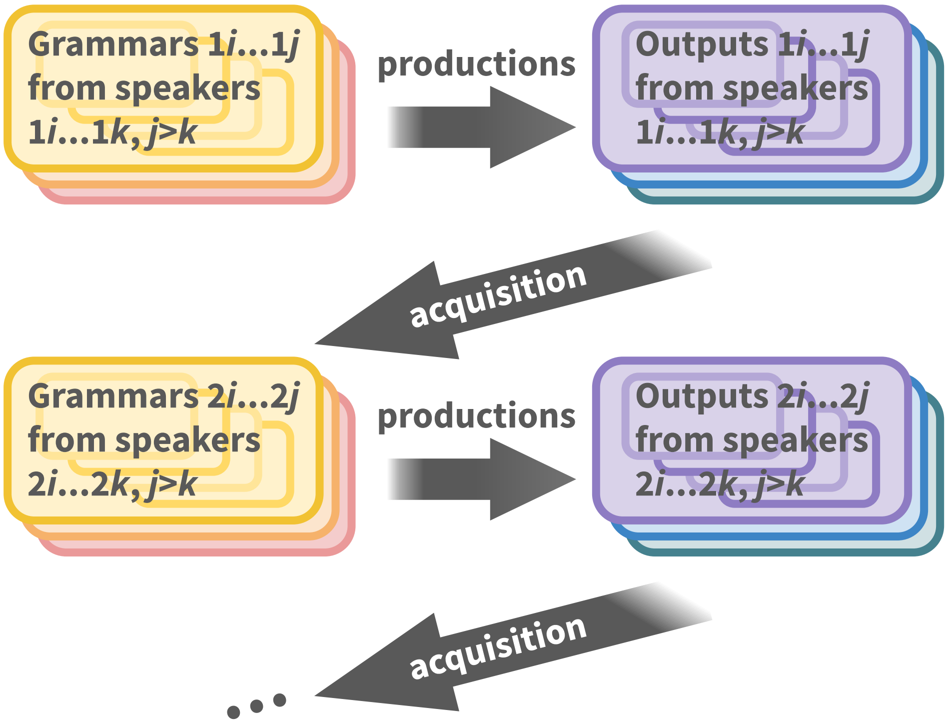
A Population-Level Z-Model of Transmission

Individual production

- Variation across social settings
- Variation over lifetimes

Community Embedding

- Variation across people
- Everyone receives many inputs



A Population-Level Z-Model of Transmission

Individual production

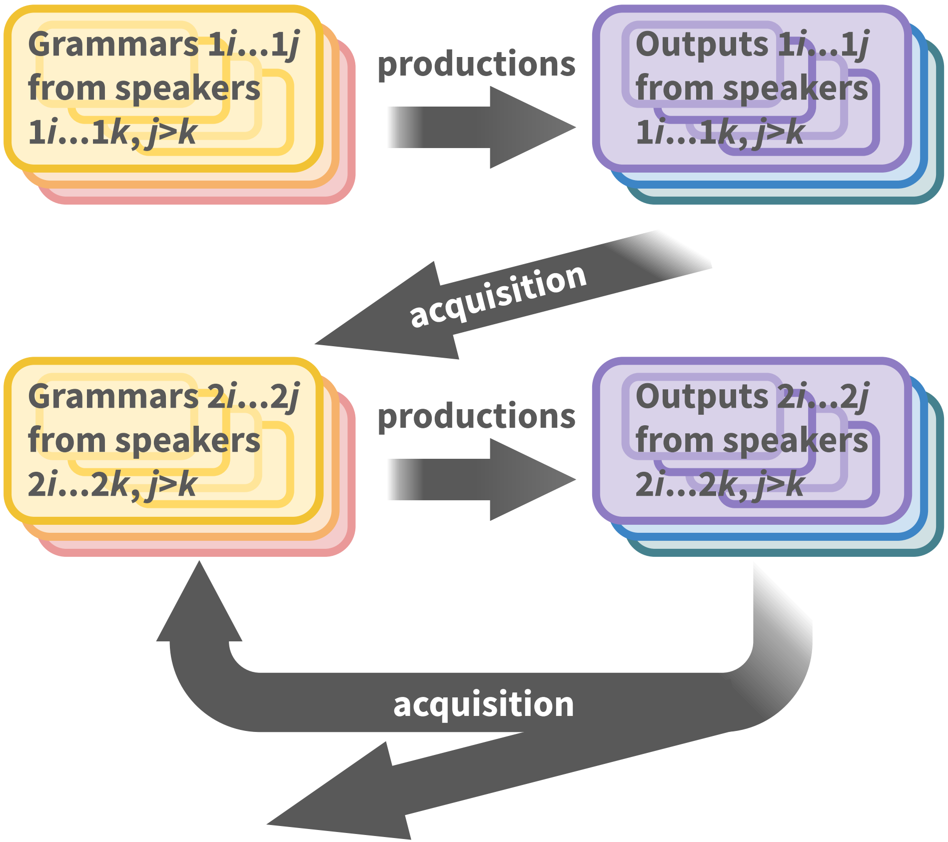
- Variation across social settings
- Variation over lifetimes

Community Embedding

- Variation across people
- Everyone receives many inputs

Gradual Maturation

- Transmission isn't just generational
- Acquisition takes time
- Immature learners influence others



Computational Modeling

Computational Modeling

A complementary methodology in historical/sociolinguistics

- Useful in conjunction with corpus, experimental, and field methods
- Different methodologies have different strengths and weaknesses

Computational Modeling

A complementary methodology in historical/sociolinguistics

- Useful in conjunction with corpus, experimental, and field methods
- Different methodologies have different strengths and weaknesses

Strengths and weaknesses of computational modeling

+ White-box - we know the underlying algorithm and variables
+ Encapsulation - we can observe simulated processes to from start to finish, when the real-life processes are too slow or huge to observe

Computational Modeling

A complementary methodology in historical/sociolinguistics

- Useful in conjunction with corpus, experimental, and field methods
- Different methodologies have different strengths and weaknesses

Strengths and weaknesses of computational modeling

+ White-box - we know the underlying algorithm and variables
+ Encapsulation - we can observe simulated processes to from start to finish, when the real-life processes are too slow or huge to observe
− Artificiality - we can probably simulate anything we want, but does the thing we're simulating actually correspond to something reality?

High-Level Classification of Frameworks

Three Approaches (and One Non-Approach)

1. Swarm Frameworks
2. Network Frameworks
3. Algebraic Frameworks
4. Iterated Learning ← DISQUALIFIED

High-Level Classification of Frameworks

Three Approaches (and One Non-Approach)

1. **Swarm Frameworks**

   e.g., Klein (1966), Schulze et al. (2008), Stanford & Kenny (2013), Hartmann (2023)

- Individual agents moving randomly on a grid and interacting
- Often what is meant when people say Agent-Based Modeling (ABM)

High-Level Classification of Frameworks

Three Approaches (and One Non-Approach)

1. Swarm Frameworks

   e.g., Klein (1966), Schulze et al. (2008), Stanford & Kenny (2013), Hartmann (2023)

- Individual agents moving randomly on a grid and interacting
- Often what is meant when people say Agent-Based Modeling (ABM)

\+ Amenable to concrete interpretation

\+ Diffusion is easy → you get Bloomfield's (1933) Principle of Density for free

High-Level Classification of Frameworks

Three Approaches (and One Non-Approach)

1. Swarm Frameworks

   e.g., Klein (1966), Schulze et al. (2008), Stanford & Kenny (2013), Hartmann (2023)

- Individual agents moving randomly on a grid and interacting
- Often what is meant when people say Agent-Based Modeling (ABM)

\+ Amenable to concrete interpretation

\+ Diffusion is easy → you get Bloomfield's (1933) Principle of Density for free

− Lacks fine-grained control over the network

− Thousands of degrees of freedom

→ Should be run many many times → Slow and expensive!

→ Prone to simulational overfitting

High-Level Classification of Frameworks

Three Approaches (and One Non-Approach)

1. Swarm Frameworks
2. Network Frameworks

   e.g., Minett & Wang (2008), Fagyal et al. (2010), Blythe & Croft (2012), Kauhanen (2016)

- Speakers are nodes in a graph, edges indicate social ties/interaction
- (Un)directed, (un)weighted vs edges, static vs dynamic graphs, etc

High-Level Classification of Frameworks

Three Approaches (and One Non-Approach)

1. Swarm Frameworks
2. Network Frameworks

   e.g., Minett & Wang (2008), Fagyal et al. (2010), Blythe & Croft (2012), Kauhanen (2016)

- Speakers are nodes in a graph, edges indicate social ties/interaction
- (Un)directed, (un)weighted vs edges, static vs dynamic graphs, etc

\+ Fine-grained control over network structure

→ easy to model community structures, test proposals from sociolinguistics

High-Level Classification of Frameworks

Three Approaches (and One Non-Approach)

1. Swarm Frameworks
2. Network Frameworks

   e.g., Minett & Wang (2008), Fagyal et al. (2010), Blythe & Croft (2012), Kauhanen (2016)

- Speakers are nodes in a graph, edges indicate social ties/interaction
- (Un)directed, (un)weighted vs edges, static vs dynamic graphs, etc

\+ Fine-grained control over network structure

→ easy to model community structures, test proposals from sociolinguistics

− Fully connected graphs have $O(|V|^2)$ edges → unwieldy to design

− In practice in linguistics, implemented with random interactions

→ Same problems with random sampling as swarm frameworks

High-Level Classification of Frameworks

Three Approaches (and One Non-Approach)

1. Swarm Frameworks
2. Network Frameworks
3. Algebraic Frameworks
   e.g., Niyogi & Berwick (1997), Abrams & Stroganz (2003), Baxter et al. (2006), Minett & Wang (2008)

- Expected outcomes are calculated analytically for each iteration and end state

High-Level Classification of Frameworks

Three Approaches (and One Non-Approach)

1. Swarm Frameworks
2. Network Frameworks
3. Algebraic Frameworks

   e.g., Niyogi & Berwick (1997), Abrams & Stroganz (2003), Baxter et al. (2006), Minett & Wang (2008)

- Expected outcomes are calculated analytically for each iteration and end state

\+ Closed-form solutions rather than simulation

→ Fast and direct to compute (possibly some large matrix inversions though)

→ Gets at expected outcomes directly rather than approximating them

High-Level Classification of Frameworks

Three Approaches (and One Non-Approach)

1. Swarm Frameworks
2. Network Frameworks
3. Algebraic Frameworks

   e.g., Niyogi & Berwick (1997), Abrams & Stroganz (2003), Baxter et al. (2006), Minett & Wang (2008)

- Expected outcomes are calculated analytically for each iteration and end state

\+ Closed-form solutions rather than simulation

→ Fast and direct to compute (possibly some large matrix inversions though)

→ Gets at expected outcomes directly rather than approximating them

− In practice, “perfectly-mixed” populations with no network structure

→ Cannot model sociolinguistic community structures

High-Level Classification of Frameworks

Three Approaches (and One Non-Approach)

1. Swarm Frameworks
2. Network Frameworks
3. Algebraic Frameworks
4. Iterated Learning

   e.g., Hare & Elman (1995), Kirby (2000), Smith & Wonnacott (2010), Ito & Feldman (2022)

- The classic Z-Model of transmission from single ‘parent’ to ‘child’ agents

[1] Niyogi & Berwick (2009)

High-Level Classification of Frameworks

Three Approaches (and One Non-Approach)

1. Swarm Frameworks
2. Network Frameworks
3. Algebraic Frameworks
4. Iterated Learning

   e.g., Hare & Elman (1995), Kirby (2000), Smith & Wonnacott (2010), Ito & Feldman (2022)

- The classic Z-Model of transmission from single ‘parent’ to ‘child’ agents

\+ Easy to implement in a laboratory setting

[1] Niyogi & Berwick (2009)

High-Level Classification of Frameworks

Three Approaches (and One Non-Approach)

1. Swarm Frameworks
2. Network Frameworks
3. Algebraic Frameworks
4. Iterated Learning

   e.g., Hare & Elman (1995), Kirby (2000), Smith & Wonnacott (2010), Ito & Feldman (2022)

- The classic Z-Model of transmission from single ‘parent’ to ‘child’ agents

+ Easy to implement in a laboratory setting

− Not a population-level model!

  → If learner = community, then variational change = transformational change

  → Does not admit phase changes and bifurcations that population models can[1]

[1] Niyogi & Berwick (2009)

High-Level Classification of Frameworks

Three Approaches (and One Non-Approach)

1. Swarm Frameworks
2. Network Frameworks
3. Algebraic Frameworks
4. Iterated Learning

e.g., Hare & Elman (1995), Kirby (2000), Smith & Wonnacott (2010), Ito & Feldman (2022)

- The classic Z-Model of transmission from single 'parent' to 'child' agents

\+ Easy to implement in a laboratory setting

− Not a population-level model!

→ If learner = community, then variational change = transformational change

→ Does not admit phase changes and bifurcations that population models can[1]

This is grounds for disqualification. Iterated learning is inappropriate for modeling language change. 🈲

[1] Niyogi & Berwick (2009)

Aiming for the Best of All Worlds

Impose density effects on a network structure and calculate the expected outcome of each iteration analytically

- **Modularity between individual learning algorithm and population update**
- **Captures variational population-level change**

Aiming for the Best of All Worlds

Impose density effects on a network structure and calculate the expected outcome of each iteration analytically

- **Modularity between individual learning algorithm and population update**
- **Captures variational population-level change**

Swarm - Captures the Principle of Density

Aiming for the Best of All Worlds

Impose density effects on a network structure and calculate the expected outcome of each iteration analytically

- Modularity between individual learning algorithm and population update
- Captures variational population-level change

Swarm - Captures the Principle of Density

Network - Incorporates social network structures

Aiming for the Best of All Worlds

Impose density effects on a network structure and calculate the expected outcome of each iteration analytically

- Modularity between individual learning algorithm and population update
- Captures variational population-level change

Swarm - Captures the Principle of Density

Network - Incorporates social network structures

Algebraic

- Direct calculation rather than simulation of individual agents
- Reduces to Niyogi & Berwick (1997) if the network is perfectly-mixed

Components of the Model

- The Grammar
- The Community
- The Individual
- The Learning Mechanism

The Grammar

Following typical definitions from formal language theory

$\mathcal{G}$	A family of grammars
g	An specific grammar $g \in \mathcal{G}$
$L(g)$	Language (set of utterances) generated by grammar g $L(g) \subseteq \Sigma^*$

The Grammar

Following typical definitions from formal language theory

$\mathcal{G}$	A family of grammars
g	An specific grammar $g \in \mathcal{G}$
$L(g)$	Language (set of utterances) generated by grammar g $L(g) \subseteq \Sigma^*$

These "grammars" can be interpreted as anything governing the individual's language.

- Formal grammars
  e.g., the space of possible natural language grammars
- Sociolinguistic variants
- …

The Community

Assuming n individuals in the community,

A $n \times n$ column stochastic matrix ← A for "adjacency"

Element a_{ij} indicates weight of connection from individual j to individual i

The Community

Assuming n individuals in the community,

A $n \times n$ column stochastic matrix ← A for "adjacency"

Element a_{ij} indicates weight of connection from individual j to individual i

> This assumes the network is static, but **A** could be replaced with updatable $\mathbf{A}_t$
>
> The network is
> - Undirected iff every $a_{ij}=a_{ji}$, $i≠j$
> - Unweighted iff every $a_{ij} = a_{kj}$, $i≠j≠k$
> - Perfectly mixed iff every $a_{ij} = 1/n$

The Community

Assuming n individuals in the community,

A $n \times n$ column stochastic matrix ← A for "adjacency"

Element a_{ij} indicates weight of connection from individual j to individual i

$\mathbf{G}_t$ $n \times |\mathcal{G}|$ row stochastic matrix ← G for "grammar"

Row $G_{t,i}$ indicates distribution of grammars expressed by individual i at time t

The Community

Assuming n individuals in the community,

A — $n \times n$ column stochastic matrix ← A for "adjacency"

Element a_{ij} indicates weight of connection from individual j to individual i

$\mathbf{G}_t$ — $n \times |\mathcal{G}|$ row stochastic matrix ← G for "grammar"

Row $G_{t,i}$ indicates distribution of grammars expressed by individual i at time t

$\mathbf{E}_t$ — $|\mathcal{G}| \times n$ column stochastic matrix ← E for "environment"

Column $E_{t,i}$ indicates distribution of grammars exposed to individual i at time t

The Individual

What is the relationship between the individual and the grammar?

D_k Input sequence of length $k \subseteq L(\mathcal{G})$ ← D for "data"

Sampled according to $E_{t,i}$ for individual i

where E_t is a function of G_t and A

The Individual

What is the relationship between the individual and the grammar?

D_k Input sequence of length $k \subseteq L(\mathcal{G})$
Sampled according to $E_{t,i}$ for individual i
where E_t is a function of G_t and A

k could be infinite to model **learning in the limit** or finite to model "**thou art mortal**"

The Individual and Learning Mechanism

What is the relationship between the individual and the grammar?

D_k Input sequence of length $k \subseteq L(\mathcal{G})$
Sampled according to $E_{t,i}$ for individual i
where E_t is a function of G_t and A

$\mathcal{A}$ A learning algorithm $\mathcal{A}$: $D \rightarrow \mathcal{G}$ ← $\mathcal{A}$ for "acquisition"
hypothesizes a grammar $\mathcal{A}(D_k) = h \in \mathcal{G}$
$\mathcal{A}$ is a function of $E_{t,i}$ that yields $G_{t+1,i}$

The Individual and Learning Mechanism

What is the relationship between the individual and the grammar?

D_k Input sequence of length $k \subseteq L(\mathcal{G})$
Sampled according to $E_{t,i}$ for individual i
where E_t is a function of G_t and A

$\mathcal{A}$ A learning algorithm $\mathcal{A}$: $D \rightarrow \mathcal{G}$
hypothesizes a grammar $\mathcal{A}(D_k) = h \in \mathcal{G}$
$\mathcal{A}$ is a function of $E_{t,i}$ that yields $G_{t+1,i}$

$\mathcal{A}$ could model transmission (=acquisition) or diffusion

If $\mathcal{A}(D_k) = h \in \mathcal{G}$, then **G** is an indicator matrix.

Or, learners could acquire distributions over grammars $\mathcal{A}(D_k) = P(\mathcal{G}_H)$, where $\mathcal{G}_H \subseteq \mathcal{G}$

Intuition behind the Propagation Algorithm

Diverging from standard approaches, nodes are "locations," not individuals, and edges encode the probability of individuals "traveling" from node to node

```
For T iterations,
    For the individual at each node,
        Begin traveling;
        While traveling,
            Randomly select outgoing edge by weight;
            Follow it OR stop;
            Increase chance of stopping next time;
        End
        Interact with individual at current node;
    End
End
```

Intuition behind the Propagation Algorithm

Diverging from standard approaches, nodes are "locations," not individuals, and edges encode the probability of individuals "traveling" from node to node

For T iterations,
    For the individual at each node,
        Begin *traveling*;
        While *traveling*,
            Randomly select outgoing edge by weight;
            Follow it OR stop;
            Increase chance of stopping next time;
        End
        Interact with individual at current node;
    End
End

Nodes are not individuals. Individuals "stand on" nodes.

Intuition behind the Propagation Algorithm

Diverging from standard approaches, nodes are "locations," not individuals, and edges encode the probability of individuals "traveling" from node to node

```
For T iterations,
    For the individual at each node,
        Begin traveling;
        While traveling,
            Randomly select outgoing edge by weight;
            Follow it OR stop;
            Increase chance of stopping next time;
        End
        Interact with individual at current node;
    End
End
```

Individuals "travel" along edges and find someone to interact with

Intuition behind the Propagation Algorithm

Diverging from standard approaches, nodes are "locations," not individuals, and edges encode the probability of individuals "traveling" from node to node

```
For T iterations,
    For the individual at each node,
        Begin traveling;
        While traveling,
            Randomly select outgoing edge by weight;
            Follow it OR stop;
            Increase chance of stopping next time;
        End
        Interact with individual at current node;
    End
End
```

Individuals connected by shorter or higher weighted paths are more likely to interact.

For today, travel decay is implemented with a geometric distribution

Intuition behind the Propagation Algorithm

Diverging from standard approaches, nodes are "locations," not individuals, and edges encode the probability of individuals "traveling" from node to node

For T iterations,

For the indiv dual at each node,

Begi ... *ling*;

While *tr* ...

Rando ... le ... going edge by weight;

Follow it ...

Increas ... topping next time;

End

Intera ... individual at ... node;

End

End

This is an algebraic model!

We can calculate the expected outcome of each iteration directly 😎 **No need to simulate!**

The Propagation Function

The linguistic environment of each learner depends on every community member's grammars and their interaction probabilities

$$\mathbf{E}_t = \mathbf{G}_t^{\top} \alpha (\mathbf{I} - (1 - \alpha)\, \mathbf{A})^{-1}$$

The Propagation Function

The **linguistic environment** of each learner depends on every **community member's grammars** and their **interaction probabilities**

$$E_t = G_t^{\top} \alpha (I - (1 - \alpha) A)^{-1}$$

- The probability that an individual at node *i* travels an additional step declines according to a geometric distribution
- $\alpha \in [0,1]$ "mobility parameter" from that distribution
  A greater α corresponds to more mobility
  → a faster simulation

The Learning Function

Learning outcomes depends on input data from the environment

$$\mathbf{G}_{t+1} = \mathcal{A}(\mathbf{E}_t)$$

$$= \mathcal{A}(\mathbf{G}_t^{\top} \alpha (\mathbf{I} - (1 - \alpha)\, \mathbf{A})^{-1})$$

The Learning Function

Learning outcomes depends on input data from the environment

$$G_{t+1} = \mathcal{A}(E_t)$$

$$= \mathcal{A}(G_t^{T}\alpha(I - (1 - \alpha)A)^{-1})$$

The next state of the system is only dependent on the current state. It is a...

- Dynamical system
- First-order Markov process

The Learning Function

Learning outcomes depends on input data from the environment

$$\mathbf{G}_{t+1} = \mathcal{A}(\mathbf{E}_t)$$

$$= \mathcal{A}(\mathbf{G}_t^{\mathsf{T}} \alpha (\mathbf{I} - (1 - \alpha)\, \mathbf{A})^{-1})$$

Sampling D_k from $\mathbf{E}_t$

- **No social valuation ($D_k \sim \mathbf{E}_t$)** It does not matter who the input comes from (cf Principle of Density)
- **With social valuation ($\mathbf{G}_{t+1} = (\sum_g \mathbf{E}_{t,g} \mathbf{S}_g)^{\mathsf{T}}$)** ← S for "socio"
  Some individuals matters more than input from others

The Learning Function

Learning outcomes depends on input data from the environment

$$G_{t+1} = \mathcal{A}(E_t)$$

$$= \mathcal{A}(G_t^{\mathsf{T}} \alpha (I - (1 - \alpha) A)^{-1})$$

Capturing basic kinds of change

- **Neutral change ($G_{t+1} = E_t^{\mathsf{T}}$)** Grammars are learned at the rate they are evidenced
- **Advantaged change ($G_{t+1} = (\sum_g E_{t,g} T_g)^{\mathsf{T}}$)** Some variant is preferred/learned at a higher rate than it is evidenced ← **T for "transition"**

The Learning Function

Learning outcomes depends on input data from the environment

$$G_{t+1} = \mathcal{A}(E_t)$$

$$= \mathcal{A}(G_t^{T} \alpha (I - (1 - \alpha) A)^{-1})$$

Can use any $\mathcal{A}$ for which learning outcomes are a function of the distribution of input grammars[1]

- Trigger Learning
- Cue-Based Learning
- Variational Learning
- Maximum Likelihood Estimation

[1] Niyogi 2006, Niyogi & Berwick (1996, 1997, 2009), etc.

Capturing basic kinds of change

- **Neutral change ($G_{t+1} = E_t^{T}$)** Grammars are learned at the rate they are evidenced
- **Advantaged change ($G_{t+1} = (\sum_g E_{t,g} T_g)^{T}$)** Some variant is preferred/learned at a higher rate than it is evidenced

A Case Study:
Northern Cities Features in the St. Louis Corridor

The St. Louis Corridor

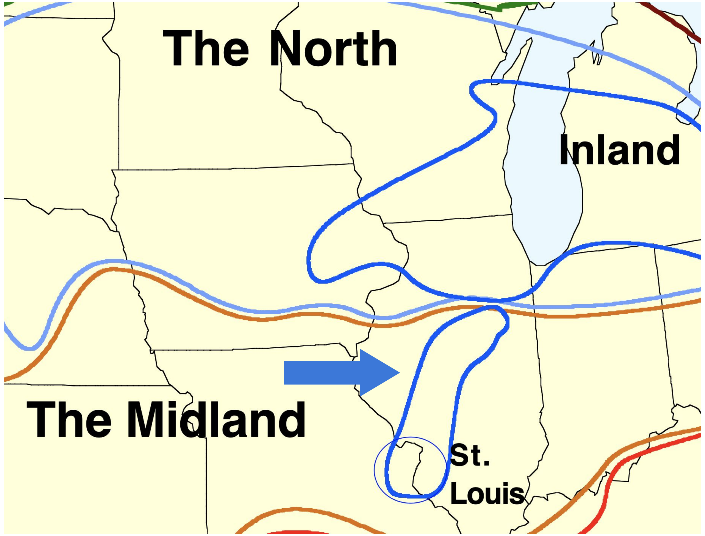
An "Island" of the Inland North within the Midlands

The Inland North contains most of the US Great Lakes Region

- Phonetically back and relatively monophthongal back vowels
- Has the Northern Cities Shift
- No COT-CAUGHT merger...

The Midlands contains the Lower Midwest

- Has the COT-CAUGHT merger
- No Northern Cities Shift
- ...



Section of ANAE map 11.13 (Boberg et al., 2006)

The St. Louis Corridor

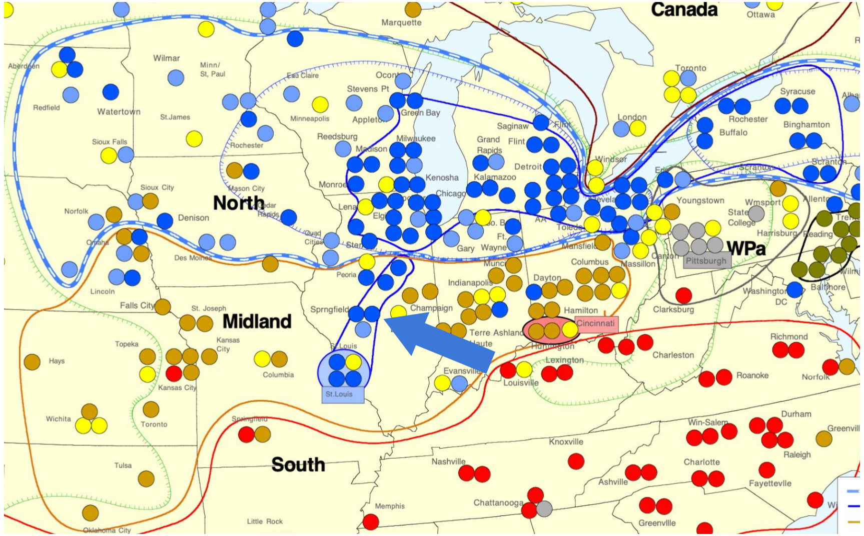
An "Island" of the Inland North within the Midlands

The Inland North contains most of the US Great Lakes Region

- Phonetically back and relatively monophthongal back vowels
- Has the Northern Cities Shift
- No COT-CAUGHT merger...

The Midlands contains the Lower Midwest

- Has the COT-CAUGHT merger
- No Northern Cities Shift
- ...



Section of ANAE map 11.19 (Boberg et al., 2006)

The St. Louis Corridor

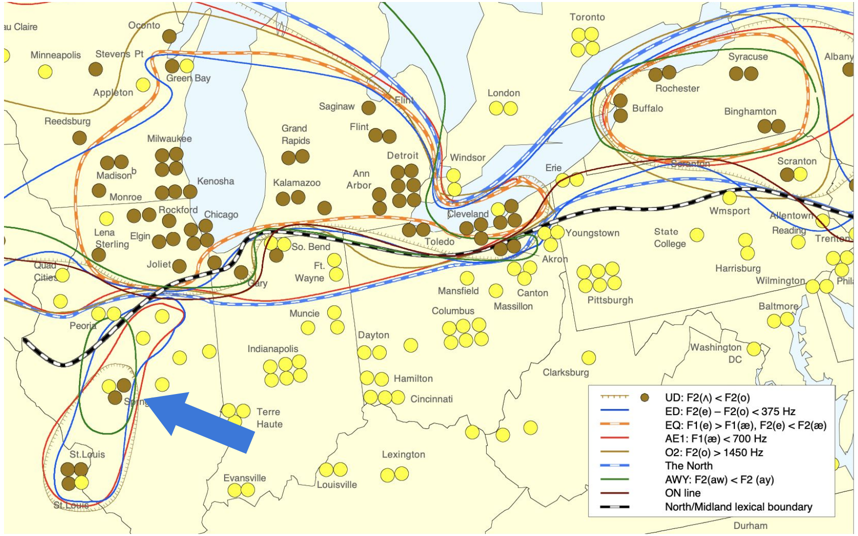
An “Island” of the Inland North within the Midlands

The Inland North contains most of the US Great Lakes Region

- Phonetically back and relatively monophthongal back vowels
- Has the Northern Cities Shift
- No COT-CAUGHT merger…

The Midlands contains the Lower Midwest

- Has the COT-CAUGHT merger
- No Northern Cities Shift
- …

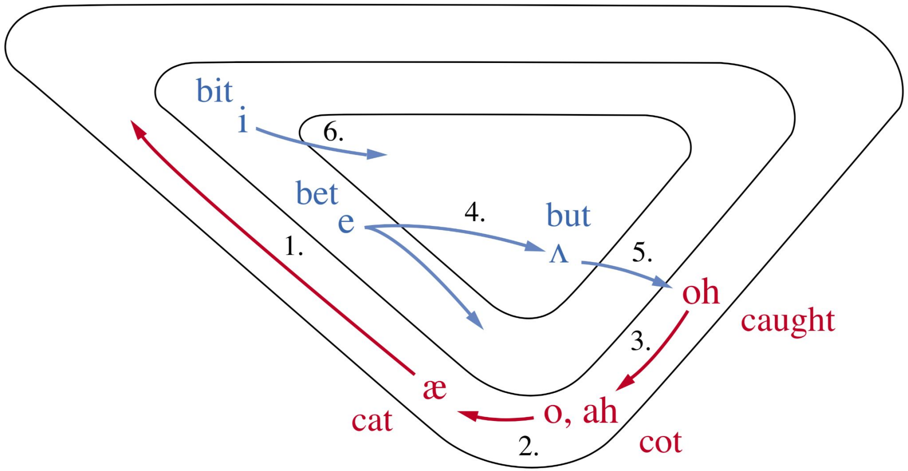


Section of ANAE map 14.11 (Boberg et al., 2006)

The Northern Cities Shift

The most emblematic feature of the Inland North

- A **chain shift** that began with the raising/diphthongization of /æ/ in Upstate NY in the late 19th century
- Fronting of /ɒ/ as /ɔ/ lowered
  → avoided the COT-CAUGHT merger
- Manifests synchronically roughly as an implicational hierarchy in the order that the chain progressed
- Apparently a combination of distinct **pull chains** (1+2+3, 3+5, 4+6) and **push chains** (4+5)



ANAE fig. 14.1 (Boberg et al., 2006)

Linguistic History of the St Louis Corridor[1]

The Corridor's island status is intrinsically linked to Route 66

- **Route 66**, commissioned in 1926, was the first paved road through Illinois

It extended from Chicago to St. Louis on the way to Los Angeles

[1] This section rely heavily on Friedman (2014)

Linguistic History of the St Louis Corridor

The Corridor's island status is intrinsically linked to Route 66

- Route 66, commissioned in 1926, was the first paved road through Illinois
- It was superseded by Interstate 55 in 1977 and decertified in 1985

Route 66 served as the main street of many local towns, but I-55 is controlled-access

→ Motorists used to stop and patronize local businesses

→ Local people used to interact with motorists

→ But I-55 mostly put an end to that interaction

On-Route vs Off-Route Communities

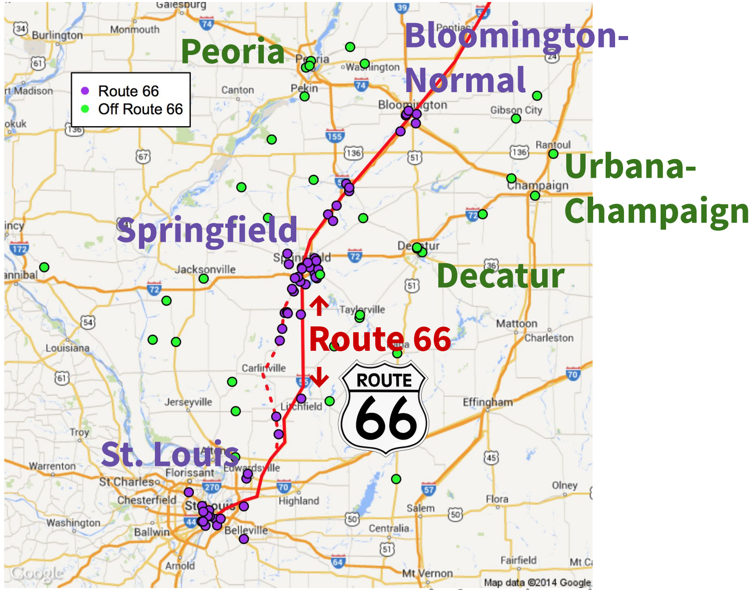
Communities on Route 66 show distinct historical trajectories from communities farther off Route 66

Major On-Route Communities:

- Bloomington-Normal
- Springfield
- St. Louis

Major Off-Route Communities:

- Peoria
- Urbana-Champaign
- Decatur

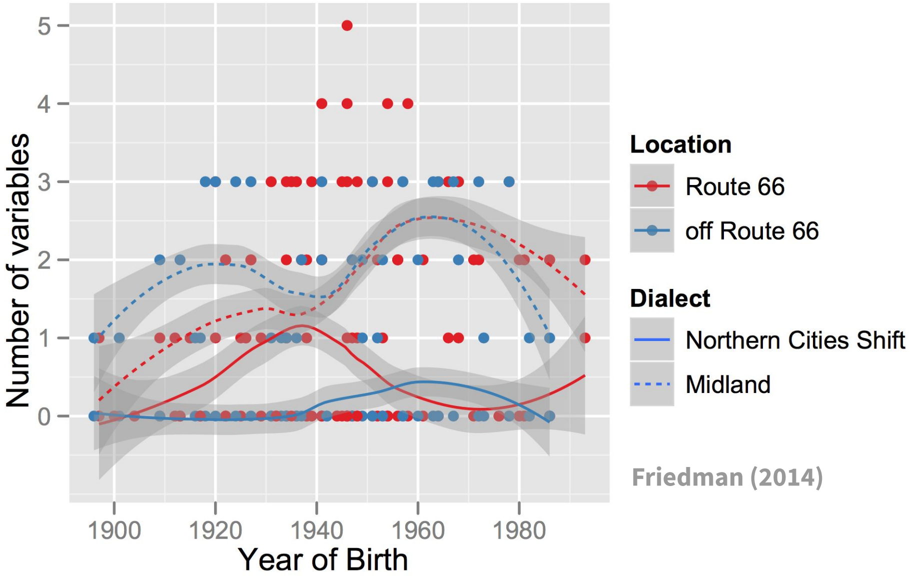


Friedman (2014)

Rise and Retreat of Northern Cities Features

Offset Two-Peak Pattern for Northern Cities Features (NCF)

- Speakers born On-Route around 1940
- Speakers born Off-Route in the 1960s
- Higher On-Route than Off-Route

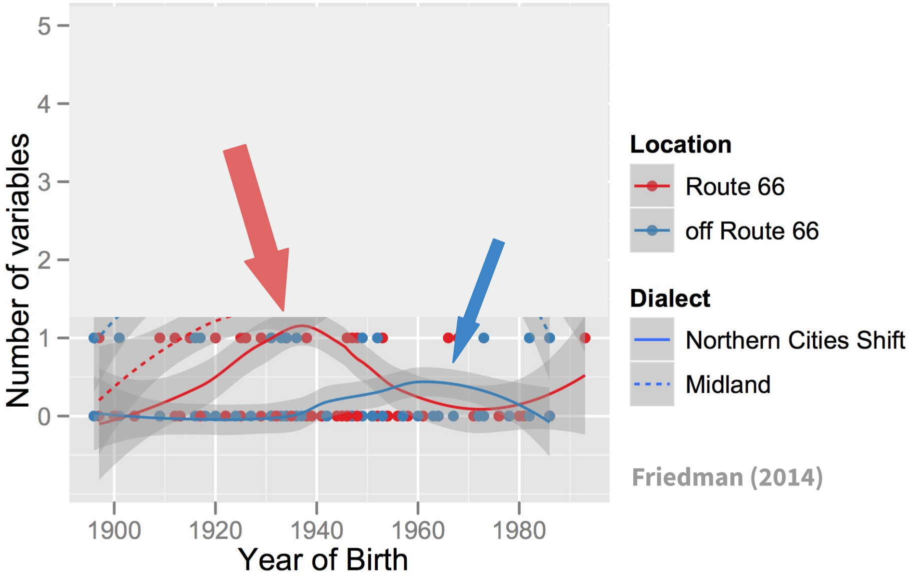


Friedman (2014)

Rise and Retreat of Northern Cities Features

Offset Two-Peak Pattern for Northern Cities Features (NCF)

- Speakers born On-Route around 1940
- Speakers born Off-Route in the 1960s
- Higher On-Route than Off-Route

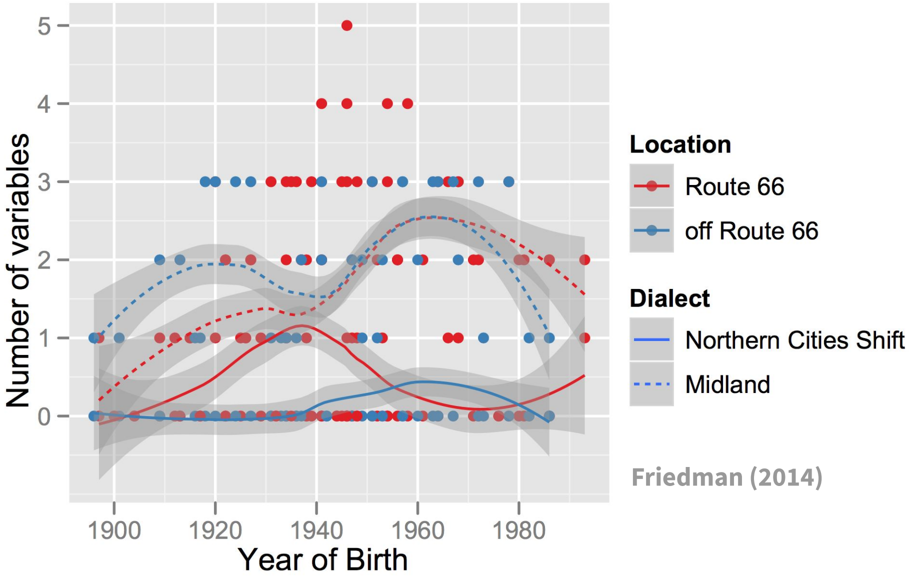


Friedman (2014)

Rise and Retreat of Northern Cities Features

Offset Two-Peak Pattern for Northern Cities Features (NCF)

- Speakers born On-Route around 1940
- Speakers born Off-Route in the 1960s
- Higher On-Route than Off-Route



Friedman (2014)

The Great Depression (1929-1939)

- Commerce on Route 66 declined
- But 1940 Census is the only recorded time in period in which net migration within Illinois was Chicago→Midlands rather than Midlands→Chicago

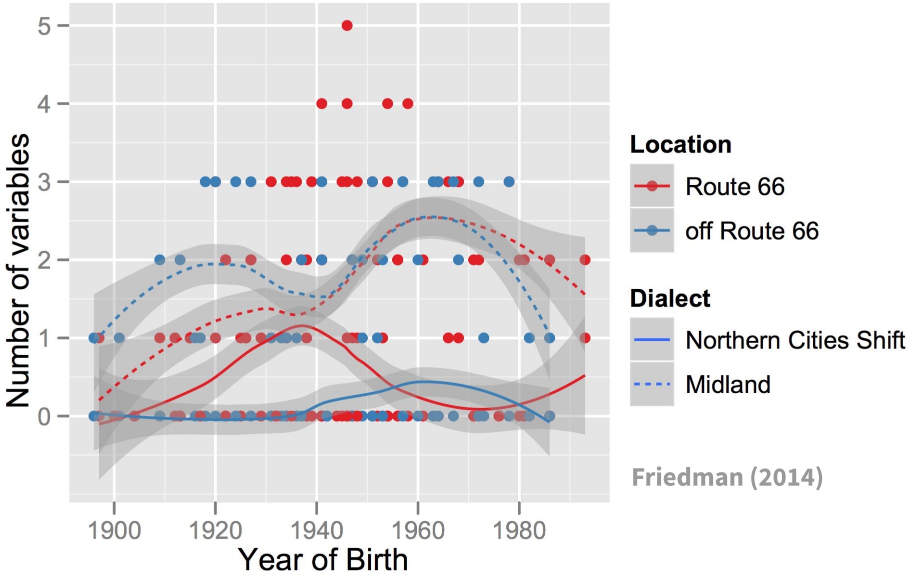
Rise and Retreat of Northern Cities Features

Offset Two-Peak Pattern for Northern Cities Features (NCF)

- Speakers born On-Route around 1940
- Speakers born Off-Route in the 1960s
- Higher On-Route than Off-Route

A Two-Compartment System?

- Direct action to one "compartment" (Chicago→On-Route interaction)
- Has a delayed/moderated effect on the other "compartment" (On-Route→Off-Route)



Friedman (2014)

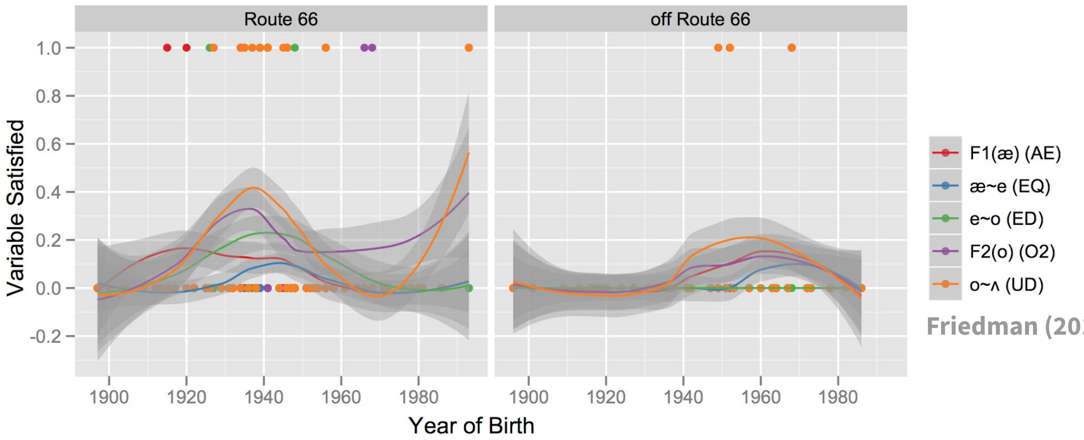
Northern Cities Features, not Northern Cities Shift

Northern Cities Features **did not follow a chain shift** pattern in the Corridor

- The Northern Cities Shift was piecemeal and inconsistent in the Corridor
- They follow a similar diachronic trend but show no sign of an actual shift
- I am calling them Northern Cities Features (NCF) following Friedman (2014)

A collection of independent features

- They were brought in wholesale from Chicago
- They can be analyzed independently



Friedman (2014)

Skipping Ahead to the Conclusion

The Northern Cities Features in the St. Louis Corridor are accounted for by:

Skipping Ahead to the Conclusion

The Northern Cities Features in the St. Louis Corridor are accounted for by:

1. Migration from Chicago to On-Route Communities during the Depression
2. Diffusion among On-Route speakers
3. On-Route speakers migrated to smaller Off-Route communities
4. They transmitted their Northern Cities Features to the next generation
5. The NCF never gained dominance in the Corridor, so they faded at each step as the surrounding Midlands reasserted itself

Simulating the St. Louis Corridor

We can test different migration+diffusion+transmission scenarios

- We can't put Illinois in a lab and run it in ultra fast-forward
- Simulation complements existing sociolinguistic fieldwork
- Which scenarios reproduce the offset two-peak structure of the Corridor NCF?

Simulating the St. Louis Corridor

We can test different migration+diffusion+transmission scenarios

- We can't put Illinois in a lab and run it in ultra fast-forward
- Simulation complements existing sociolinguistic fieldwork
- Which scenarios reproduce the offset two-peak structure of the Corridor NCF?

  Some may fail to reproduce it at all

  Some may reproduce it but only under unreasonable assumptions

  Some may reproduce it under plausible assumptions

  → Successful simulations directed us towards what to study next

Artificiality: The Weakness of Comp Modeling

We can probably simulate anything we want, but does the thing we're simulating actually correspond to something in reality?

Artificiality: The Weakness of Comp Modeling

We can probably simulate anything we want, but does the thing we're simulating actually correspond to something in reality?

A Self-Critical Approach to Computational Modeling

0. **Constantly evaluate your (implicit) assumptions**
   Really, something we should always be doing
   But, something we're all inherently bad at

Artificiality: The Weakness of Comp Modeling

We can probably simulate anything we want, but does the thing we're simulating actually correspond to something in reality?

A Self-Critical Approach to Computational Modeling

0. Constantly evaluate your (implicit) assumptions
1. When possible, motivate model parameters with real-word evidence

   Empirical data is our tether back to reality!

Artificiality: The Weakness of Comp Modeling

We can probably simulate anything we want, but does the thing we're simulating actually correspond to something in reality?

A Self-Critical Approach to Computational Modeling

0. Constantly evaluate your (implicit) assumptions
1. When possible, motivate model parameters with real-word evidence
2. When not possible, perform a parameter sweep

   Are our results actually due to some arbitrary model-internal decision?

   If not, that's great! Be upfront about it

   If so, still be upfront about it. It's another condition on our conclusions

Artificiality: The Weakness of Comp Modeling

We can probably simulate anything we want, but does the thing we're simulating actually correspond to something in reality?

A Self-Critical Approach to Computational Modeling

0. Constantly evaluate your (implicit) assumptions
1. When possible, motivate model parameters with real-word evidence
2. When not possible, perform a parameter sweep
3. Start with simpler models and work up from there

   Analogous problem to overfitting in statistics

   Increased complexity trades a better fit for loss of explanatory power

Artificiality: The Weakness of Comp Modeling

We can probably simulate anything we want, but does the thing we're simulating actually correspond to something in reality?

A Self-Critical Approach to Computational Modeling

0. Constantly evaluate your (implicit) assumptions
1. When possible, motivate model parameters with real-word evidence
2. When not possible, perform a parameter sweep
3. Start with simpler models and work up from there
4. All conclusions are just inferences from the white box model

   Not all outputs have meaningful real-world correlates. That's ok!

   Outputs that do have real meaning only hold as long as the assumptions of the model do too

Representing the Linguistic Variable(s)

As a single continuous variable

- The St. Louis Corridor Northern Cities Features are not part of a chain shift

  → We don't have to / should not represent the chain shift at all

  → Conveniently, we can model a single stand-in variable

Representing the Linguistic Variable(s)

As a single continuous variable

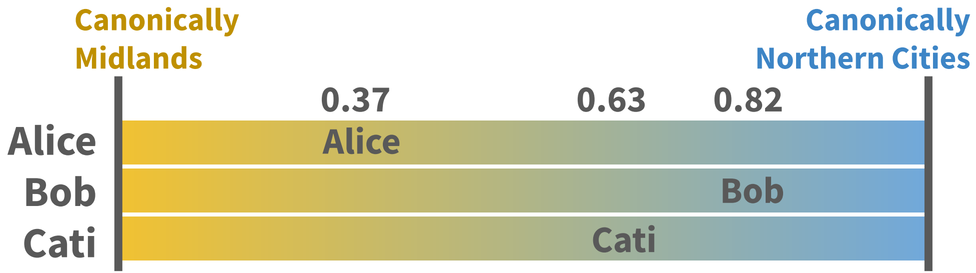
- The St. Louis Corridor Northern Cities Features are not part of a chain shift
  → We don't have to / should not represent the chain shift at all
  → Conveniently, we can model a single stand-in variable
- The variable is continuous because NCF are continuous phonetic phenomena
  → Can be represented with two grammars, g_{NCF} and g_{Mid}, $G_{t,i}$ is non-categorical
  → Proportion g_{NCF} in $G_{t,i}$ is i's innovativeness (raisedness, frontedness…)

Representing the Linguistic Variable(s)

As a single continuous variable

- **The St. Louis Corridor Northern Cities Features are not part of a chain shift**
  - → We don’t have to / should not represent the chain shift at all
  - → Conveniently, we can model a single stand-in variable
- **The variable is continuous because NCF are continuous phonetic phenomena**
  - → Can be represented with two grammars, g_{NCF} and g_{Mid}, $G_{t,i}$ is non-categorical
  - → Proportion g_{NCF} in $G_{t,i}$ is i’s innovativeness (raisedness, frontedness…)

G_t	g_{NCF}	g_{Mid}
Alice	0.37	0.63
Bob	0.82	0.18
Cati	0.63	0.37

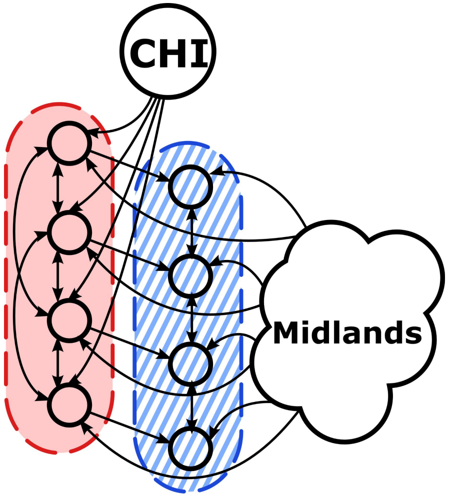


Representing the Physical St. Louis Corridor

Two St. Louis Corridor models at different levels of concreteness

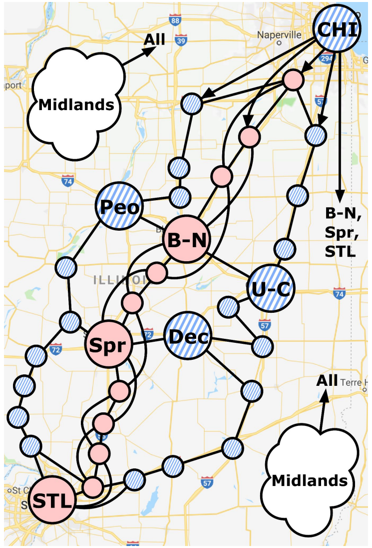
The Schematic Model

Just captures the essence



The Geographic Model

Captures more concrete detail

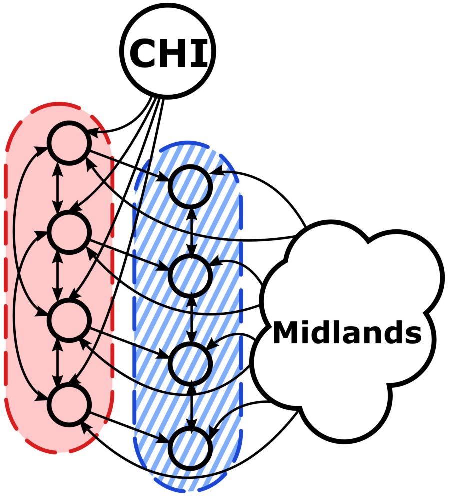


Representing the Physical St. Louis Corridor

Two St. Louis Corridor models at different levels of concreteness

The Schematic Model

Just captures the essence



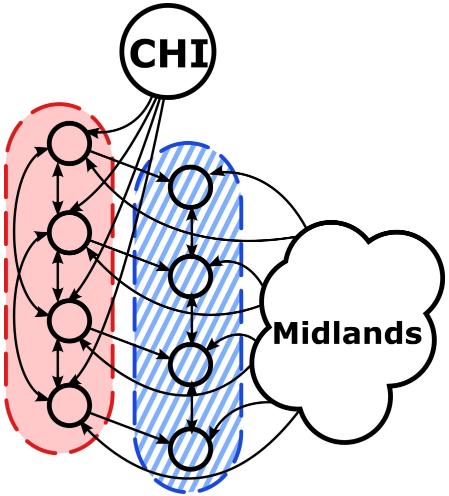
- Two compartments: On-Route and Off-Route
- Both have linear stepping-stone topologies
- Chicago and Midlands have no incoming edges
- On-Route communities have incoming edges from Chicago and edges to their neighbors' neighbors
- Both On-Route and Off-Route communities have incoming edges from the Midlands
- Communities have a partially connected centralized structure

Representing the Physical St. Louis Corridor

Two St. Louis Corridor models at different levels of concreteness

The Schematic Model

Just captures the essence



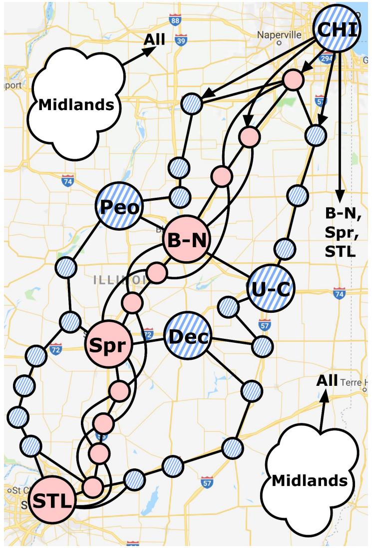
	# Comms	Comms Size
Chicago	1	18 (= ∞)
Midlands	1	18 (= ∞)
On-Route	19	18
Off-Route	19	18
Total	40	720

Representing the Physical St. Louis Corridor

Two St. Louis Corridor models at different levels of concreteness

The Geographic Model

Captures more concrete detail



- Based on the actual geography of the Corridor
- Includes towns with populations >1000 in 1940 at 1:200 scale
- Communities are classified On-Route or Off-Route based on their locations
- Still has a stepping-stone-like topology
- Both adjacent communities and major On-Route communities have incoming edges from Chicago
- All communities have incoming edges from the Midlands

Representing the Physical St. Louis Corridor

Two St. Louis Corridor models at different levels of concreteness

Chicago and Midlands are static and effectively infinite, implemented with size = 1

On-Route	Size
Joliet	210
Dwight	10
Pontiac	35
Chenoa	5
Bloomington-Normal	230
Atlanta	5
Lexington	75
Springfield	380
Farmersville	5
Litchfield	35
Mount Olive	15
Collinsville	50
St. Louis	4080
Total	5135

Off-Route West	Size
Ottawa	80
Minonk	10
El Paso	10
Peoria	520
Havana	20
Pleasant Plains	5
Jacksonville	100
White Hall	15
Carrollton	10
Jerseyville	25
Total	800

Off-Route East	Size
Kankakee	110
Clifton/Onarga	5
Paxton	15
Urbana-Champaign	190
Tuscola	15
Argenta	5
Decatur	295
Mattoon	80
Effingham	30
Vandalia	25
Greenville	15
Total	785

Internal Structure of Communities

Follows a schematization of linguistic community structures

- A loosely connected network of densely connected centralized clusters[1]
- Somewhat fractal: Clusters themselves are centrally organized
  → Central vs peripheral individuals and central vs peripheral clusters
- Implements notion of strong vs weak ties[2]
  Strong ties = higher edge weights, tend to be intra-cluster
  Weak ties = lower edge weights, tend to be inter-cluster
  Innovations crucially spread through weak ties
- This structure promotes an S-curve Diffusion of Innovation pattern
  Innovators → Early Adopters → Early Majority → Late Majority → Laggards

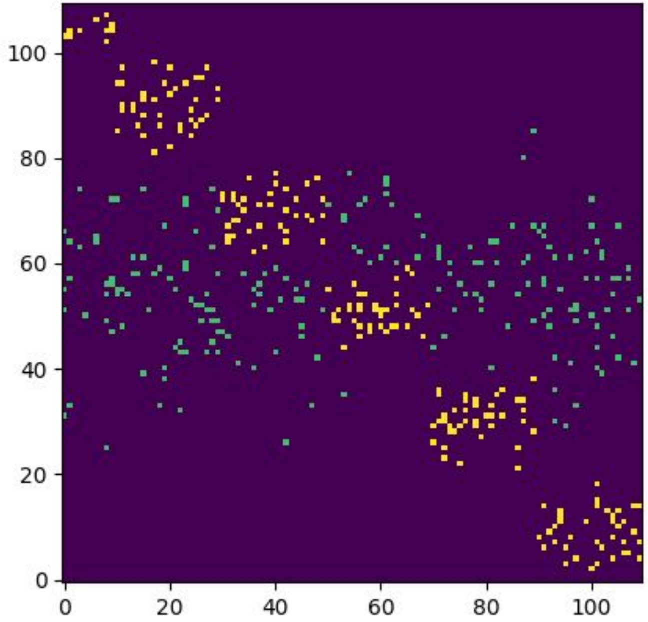
[1] Labov 1974, Milroy & Milroy 1978, Milroy & Milroy 1985, *et seq*, [2] Milroy & Milroy 1985, *et seq*

Internal Structure of Communities

As implemented,

- Communities are divided into clusters of up to size 20
- Centralization of strong ties is achieved by assigning edges following a Gaussian distribution within each cluster
- The same approach is applied to achieve centralized weak ties between clusters

Portion of Adjacency Matrix A for Kankakee, IL (Pop. 110)



Intra-Cluster Edges (Strong Ties)
Inter-Cluster Edges (Weak Ties)

Simulating Four Scenarios

1. Propagation by Diffusion Only
2. Migration Chicago→On-Route and only Diffusion On-Route→Off-Route
3. Migration and Diffusion;
   Manipulating Migration with Advantaged Change
4. Migration Chicago→On-Route and Migration On-Route→Off-Route;
   Manipulating Migration

Simulating Four Scenarios

1. Propagation by Diffusion Only
2. Migration Chicago→On-Route and only Diffusion On-Route→Off-Route
3. Migration and Diffusion;
   Manipulating Migration with Advantaged Change
4. Migration Chicago→On-Route and Migration On-Route→Off-Route;
   Manipulating Migration

All results reported here are for the Geographical Model. Schematic vs. Geographic representation yielded no meaningful differences in predictions 👍

Simulating Four Scenarios

1. **Propagation by Diffusion Only**

Simulation 1: Propagation by Diffusion Only

Hypothesis: NCF were brought into the corridor from Chicago by motorists passing through and interacting with locals[1]

Three-Phase Simulation

Phase 1 (Pre-Depression)	High traffic flow (i.e., edge weights) from Chicago for 5 iterations (≈years)
Phase 2 (Great Depression)	Low traffic flow (1/100x Phase 1) from Chicago for 15 iterations
Phase 3 (Post-Depression)	High traffic flow (= Phase 1) from Chicago for 45 iterations

[1] Stanford & Kenny 2013, Labov 2007

Simulation 1: Propagation by Diffusion Only

Hypothesis: NCF were brought into the corridor from Chicago by motorists passing through and interacting with locals[1]

Three-Phase Simulation

Phase 1 (Pre-Depression)

High traffic flow

Phase 2 (Great Depression)

Low traffic flow

Phase 3 (Post-Depression)

High traffic flow

Some model-internal parameters:

- 1/100x reduction in traffic flow and α were chosen so that the effect of Phase 2 would be clearly visible and so that 1 iteration ≈ 1 year
- Phase boundaries are sharp for simplicity

Consequence for interpretation:

- Neither the absolute *x* nor *y* scale has a real-world interpretation. Only the relative scale does

[1] Stanford & Kenny 2013, Labov 2007

Simulation 1: Propagation by Diffusion Only

Hypothesis: NCF were brought into the corridor from Chicago by motorists passing through and interacting with locals[1]

Three-Phase Simulation

Phase 1 (Pre-Depression)

High traffic flow

Phase 2 (Great Depression)

Low traffic flow

Phase 3 (Post-Depression)

High traffic flow

Choice of $\mathcal{A}$: $G_{t+1} = E_t^T$ (Neutral Change)

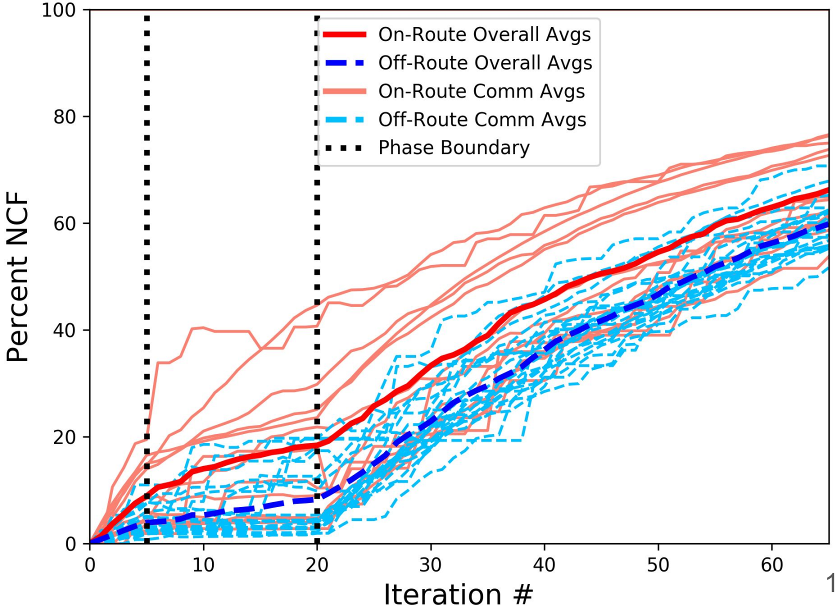
- Individuals probability match their input
- Typical behavior in the face of structured continuous variation[2]
- Common baseline in sociolinguistics
- Common baseline in quantitative population genetics (e.g., "Neutral Theory"[3])

[1] Stanford & Kenny 2013, Labov 2007, [2] Austin et al 2023 for a review, [3] Kimura 1983, *et seq*

Simulation 1: Propagation by Diffusion Only

Hypothesis: NCF were brought into the corridor from Chicago by motorists passing through and interacting with locals[1]

- NCF is more pronounced On-Route than Off-Route
- NCF does not reach 100% (=Chicago) during the simulation

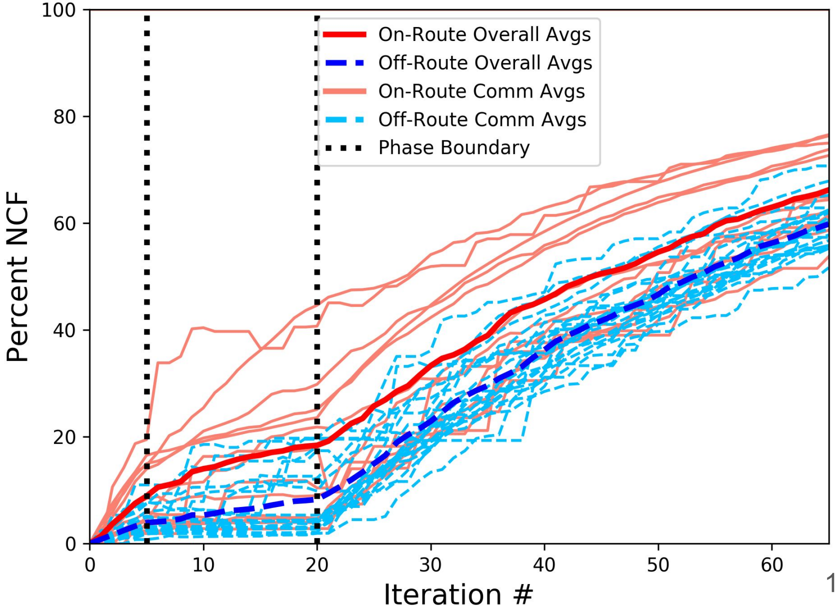


[1] Stanford & Kenny 2013, Labov 2007

Simulation 1: Propagation by Diffusion Only

Hypothesis: NCF were brought into the corridor from Chicago by motorists passing through and interacting with locals[1]

- + NCF is more pronounced On-Route than Off-Route
- + NCF does not reach 100% (=Chicago) during the simulation
- – No offset two-peak pattern
- – Actually, no way for a retreat to occur! Chicago is just way too big



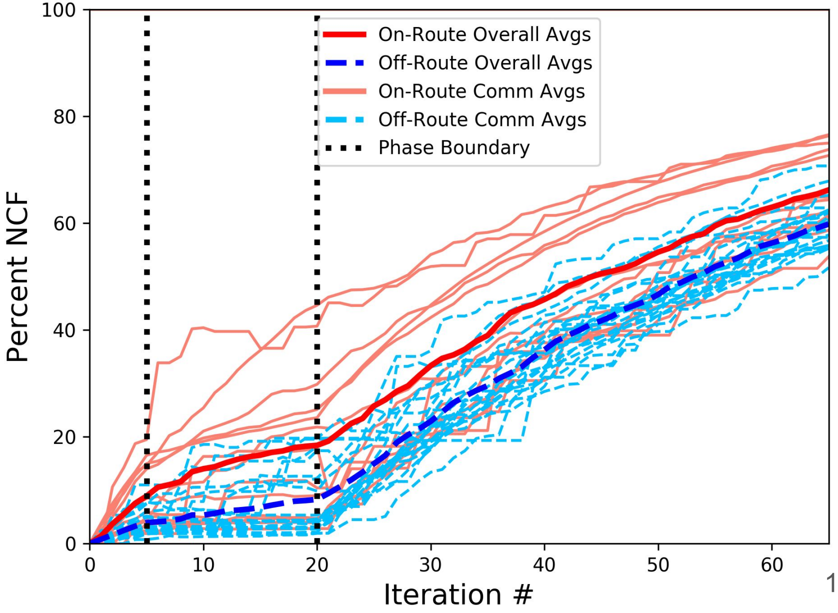
[1] Stanford & Kenny 2013, Labov 2007

Simulation 1: Propagation by Diffusion Only

~~**Hypothesis:** NCF were brought into the corridor from Chicago by motorists passing through and interacting with locals[1]~~

Conclusion

Diffusion alone cannot account for the St. Louis Corridor's linguistic history



[1] Stanford & Kenny 2013, Labov 2007

Simulating Four Scenarios

1. ~~Propagation by Diffusion Only~~

Simulating Four Scenarios

1. ~~Propagation by Diffusion Only~~
2. Migration Chicago→On-Route and only Diffusion On-Route→Off-Route

Sim2: On-Route Migration, Off-Route Diffusion

Hypothesis: NCF were brought into the Corridor and maintained by migrants from Chicago then diffused out to Off-Route communities. Sustained migration from the Midlands caused its retreat

Three-Phase Simulation

Phase 1 (Pre-Depression)	Migration from the Midlands keeps the NCF rate down despite diffusion from Chicago
Phase 2 (Great Depression)	Migration from Chicago to On-Route communities imports the NCF
Phase 3 (Post-Depression)	Migration from the Midlands causes its retreat

Sim2: On-Route Migration, Off-Route Diffusion

Hypothesis: NCF were brought into the Corridor and maintained by migrants from Chicago then diffused out to Off-Route communities. Sustained migration from the Midlands caused its retreat

Three-Phase Simulation

Phase 1 (Pre-Depression)

Migration from the Midlands to On-Route

Phase 2 (Great Depression)

Migration from Chicago to On-Route

Phase 3 (Post-Depression)

Migration from the Midlands to On-Route

Implementing Migration

Nodes have low probability of being replaced
by a Chicagoan ($G_{t,i}$ = [1.0 0.0]) or
by a Midlander ($G_{t,i}$ = [0.0 1.0])
at each iteration

Migration rate is a model-internal parameter

Sim2: On-Route Migration, Off-Route Diffusion

Hypothesis: NCF were brought into the Corridor and maintained by migrants from Chicago then diffused out to Off-Route communities. Sustained migration from the Midlands caused its retreat

Three-Phase Simulation

Phase 1 (Pre-Depression)

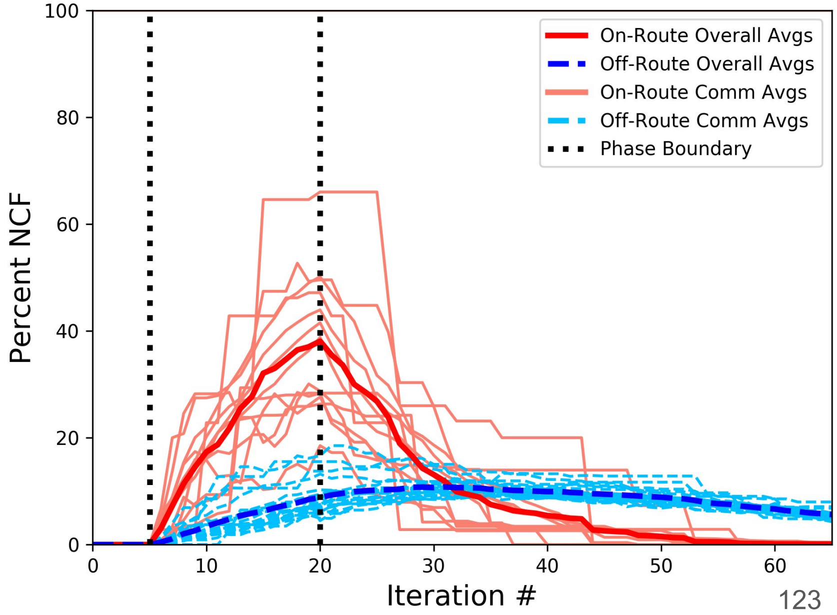
Migration from the Midlands to On-Route

Phase 2 (Great Depression)

Migration from Chicago to On-Route

Phase 3 (Post-Depression)

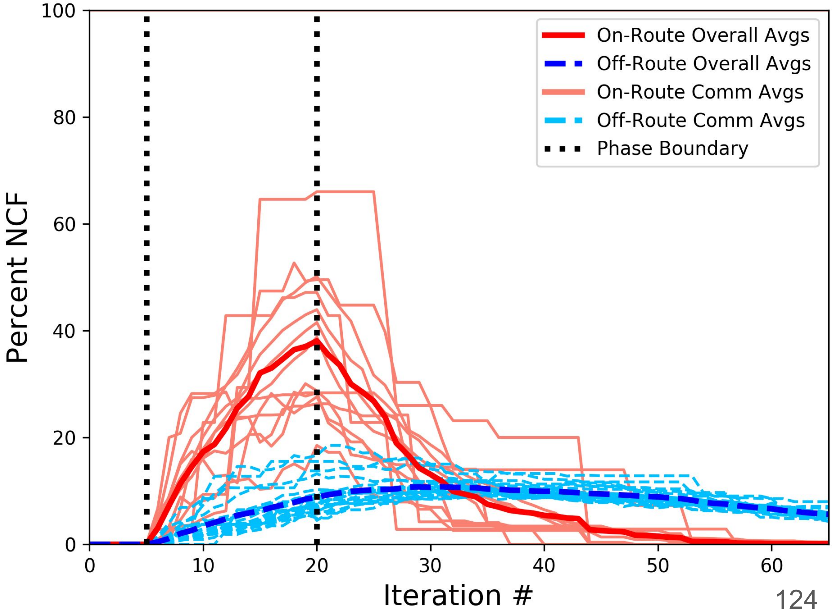
Migration from the Midlands to On-Route



Sim2: On-Route Migration, Off-Route Diffusion

Hypothesis: NCF were brought into the Corridor and maintained by migrants from Chicago then diffused out to Off-Route communities. Sustained migration from the Midlands caused its retreat

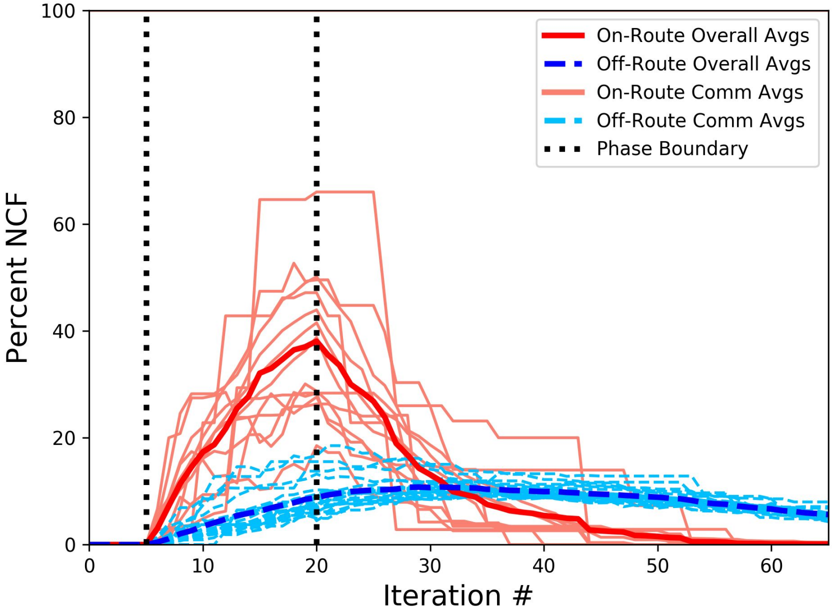
- An offset two-peak pattern!
- Chicago migration only affects On-Route communities directly → Off-Route communities lag as part of a two-compartment system



Sim2: On-Route Migration, Off-Route Diffusion

Hypothesis: NCF were brought into the Corridor and maintained by migrants from Chicago then diffused out to Off-Route communities. Sustained migration from the Midlands caused its retreat

+ An offset two-peak pattern!
+ Chicago migration only affects On-Route communities directly
  → Off-Route communities lag as part of a two-compartment system
− The peaks are too close together. Off-Route should start rising after On-Route peaks
− Off-Route must peak where its curve crosses the On-Route curve
  → There is no way to pull the peaks apart

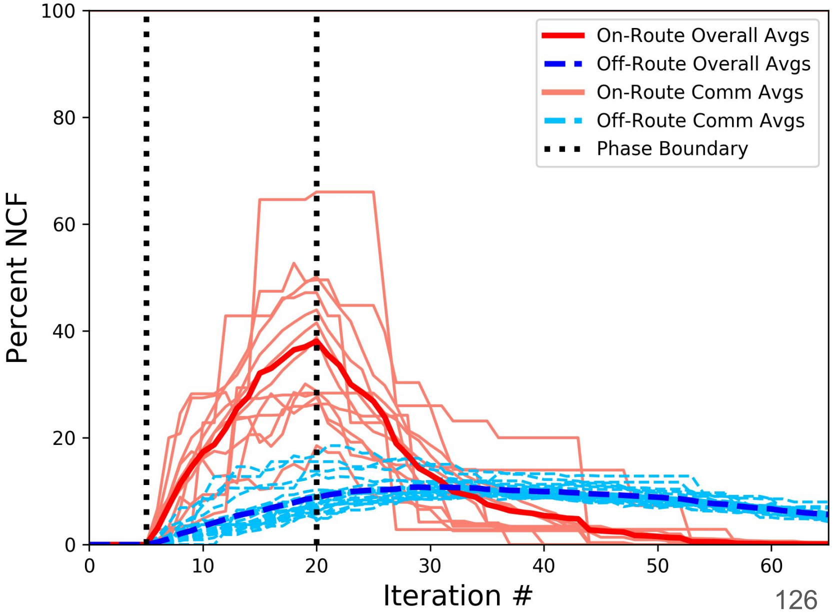


Sim2: On-Route Migration, Off-Route Diffusion

Hypothesis: NCF were brought into the Corridor and maintained by migrants from Chicago then diffused out to Off-Route communities. Sustained migration from the Midlands caused its retreat

Conclusion

This is better, but the two-compartment model is not good enough. Something needs to act on Off-Route communities too.



Simulating Four Scenarios

Need to manipulate Off-Route directly…

1. ~~Propagation by Diffusion Only~~
2. Migration Chicago→On-Route and ~~only Diffusion On-Route→Off-Route~~

Simulating Four Scenarios

Need to manipulate Off-Route directly…

1. ~~Propagation by Diffusion Only~~
2. Migration Chicago→On-Route and ~~only Diffusion On-Route→Off-Route~~
3. Migration and Diffusion;

   Manipulating Migration with Advantaged Change

Sim3: Migration with Advantaged Change

Hypothesis: NCF had an advantage that allowed it to rise Off-Route after it already declined On-Route. It retreated due to migration from the Midlands to both On-Route and Off-Route

Sim3: Migration with Advantaged Change

Hypothesis: NCF had an advantage that allowed it to rise Off-Route after it already declined On-Route. It retreated due to migration from the Midlands to both On-Route and Off-Route

Problem:

- Advantaged NCF will spike towards 100% due to diffusion from Chicago
- NCF can be slowed by curtailing traffic flow from Chicago
- NCF can be forced to retreat with migration from the Midlands

Sim3: Migration with Advantaged Change

Hypothesis: NCF had an advantage that allowed it to rise Off-Route after it already declined On-Route. It retreated due to migration from the Midlands to both On-Route and Off-Route

Problem:

- Advantaged NCF will spike towards 100% due to diffusion from Chicago
- NCF can be slowed by curtailing traffic flow from Chicago
- NCF can be forced to retreat with migration from the Midlands

How severely must traffic flow be restricted?
A plausible or implausible amount?

Sim3: Migration with Advantaged Change

Hypothesis: NCF had an advantage that allowed it to rise Off-Route after it already declined On-Route. It retreated due to migration from the Midlands to both On-Route and Off-Route

Four-Phase Simulation

Phase 1 (Pre-Depression)	Migration from the Midlands keeps the NCF rate down despite diffusion from Chicago
Phase 2 (Great Depression)	Migration from Chicago to On-Route communities imports the NCF. Traffic On-Route→Off-Route curtailed. 1/1000x necessary to delay the Off-Route peak
Phase 3 (Post-Depression)	Migration from the Midlands to On-Route communities
Phase 4 (Late 20th Century)	Migration from the Midlands to Off-Route as well

Sim3: Migration with Advantaged Change

Hypothesis: NCF had an advantage that allowed it to rise Off-Route after it already declined On-Route. It retreated due to migration from the Midlands to both On-Route and Off-Route

Four-Phase Simulation

Phase 1 (Pre-Depression)
Migration from the Midlands

Phase 2 (Great Depression)
Migration from Chicago. 1/1000x traffic

Phase 3 (Post-Depression)
Migration from the Midlands to On-Route

Phase 4 (Late 20th Century)
Migration from the Midlands to Off-Route too

Choice of $\mathcal{A}$: $G_{t+1} = (\sum_g E_{t,g} T_g)^T$ (Advantaged Change)

- **T** is a **transition matrix** specifying the probability of moving from g_{NCF} to g_{Mid} and vice-versa[1]
- Set so individuals pick up g_{NCF} at a slightly higher rate than attested in their environment
- Implements sociolinguistic **incrementation**
- The math is the same whether this advantage is language-internal or language-external

[1] following Niyogi & Berwick 1997

Sim3: Migration with Advantaged Change

Hypothesis: NCF had an advantage that allowed it to rise Off-Route after it already declined On-Route. It retreated due to migration from the Midlands to both On-Route and Off-Route

Four-Phase Simulation

Phase 1 (Pre-Depression)

Migration from the Midlands

Phase 2 (Great Depression)

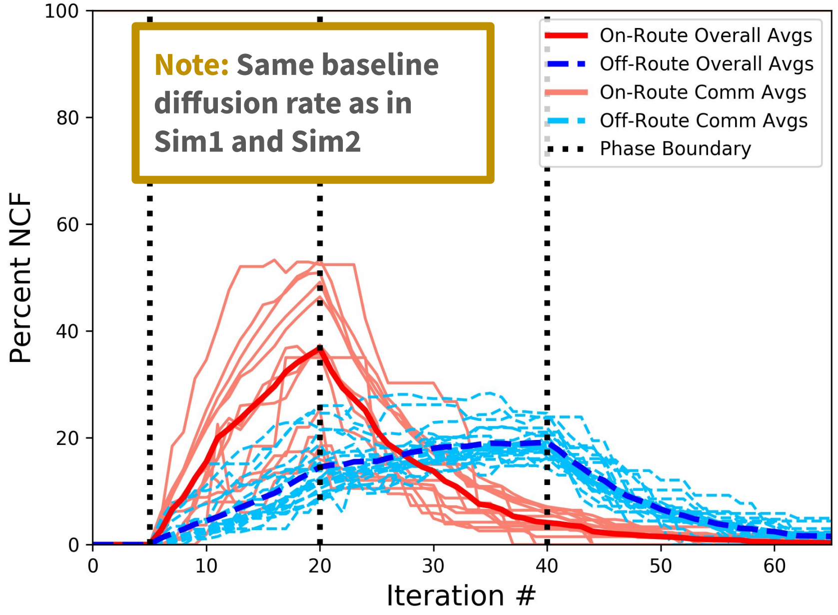
Migration from Chicago. 1/1000x traffic

Phase 3 (Post-Depression)

Migration from the Midlands to On-Route

Phase 4 (Late 20th Century)

Migration from the Midlands to Off-Route too



[1] following Niyogi & Berwick 1997

Sim3: Migration with Advantaged Change

Hypothesis: NCF had an advantage that allowed it to rise Off-Route after it already declined On-Route. It retreated due to migration from the Midlands to both On-Route and Off-Route

Four-Phase Simulation

Phase 1 (Pre-Depression)

Migration from the Midlands

Phase 2 (Great Depression)

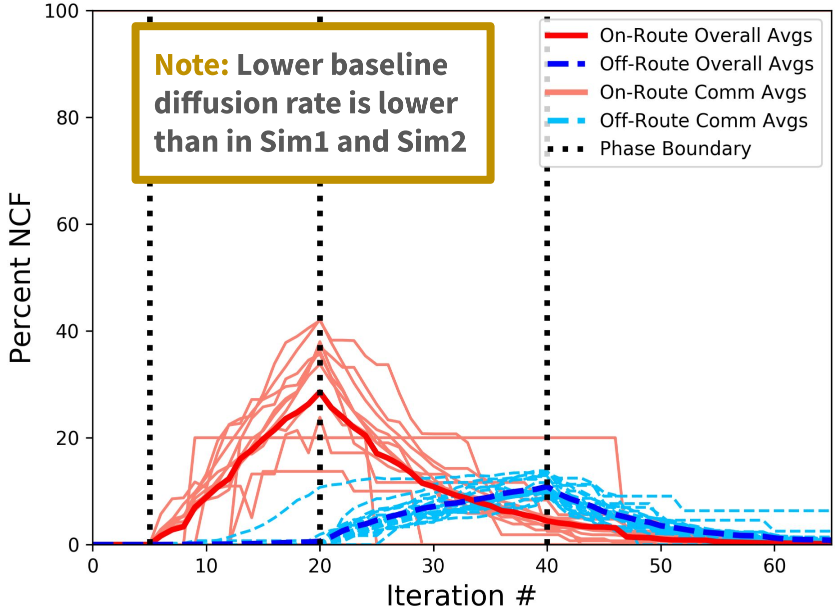
Migration from Chicago. 1/1000x traffic

Phase 3 (Post-Depression)

Migration from the Midlands to On-Route

Phase 4 (Late 20th Century)

Migration from the Midlands to Off-Route too

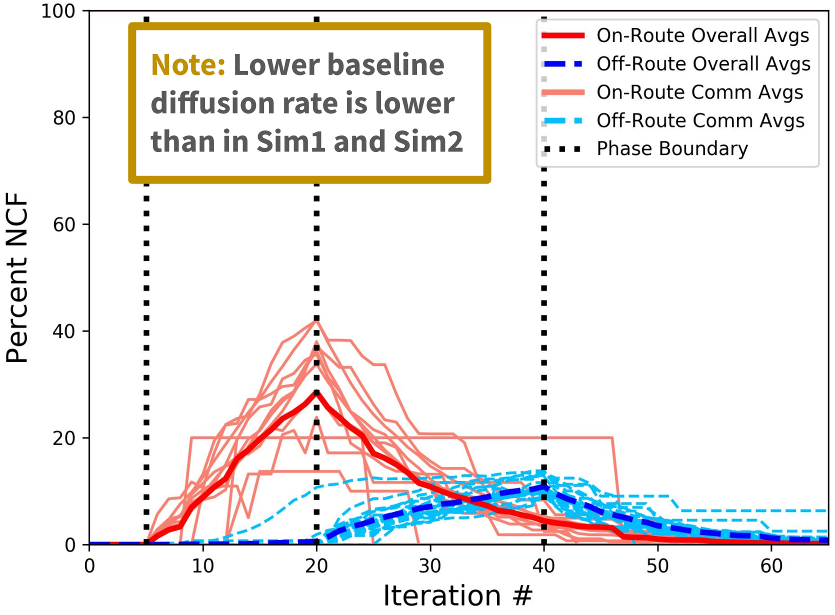


[1] following Niyogi & Berwick 1997

Sim3: Migration with Advantaged Change

Hypothesis: NCF had an advantage that allowed it to rise Off-Route after it already declined On-Route. It retreated due to migration from the Midlands to both On-Route and Off-Route

- + It looks good! The offset two-peak pattern
- + Off-Route's peak is appropriately delayed (delay length is a model-internal parameter)
- + NCS incremented in the Inland North in that era, so advantaged change is plausible

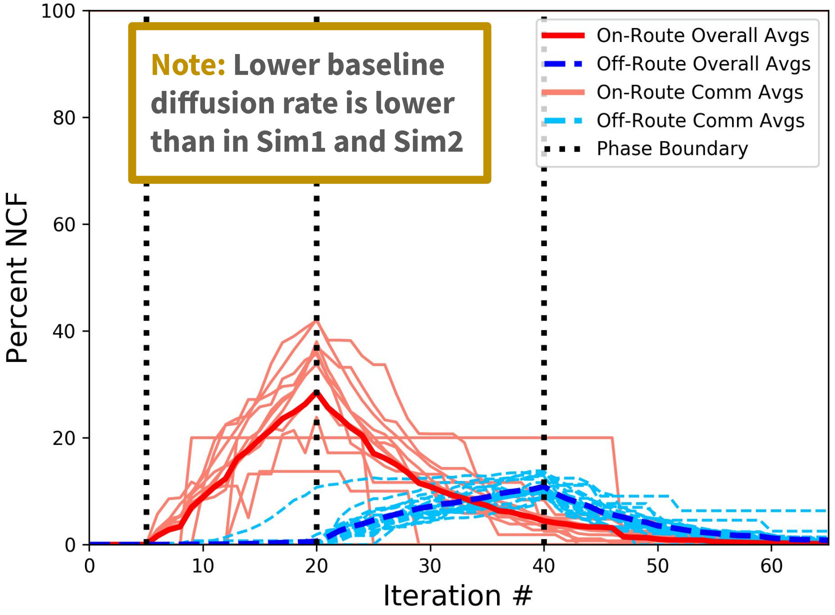


[1] following Niyogi & Berwick 1997

Sim3: Migration with Advantaged Change

Hypothesis: NCF had an advantage that allowed it to rise Off-Route after it already declined On-Route. It retreated due to migration from the Midlands to both On-Route and Off-Route

- \+ It looks good! The offset two-peak pattern
- \+ Off-Route's peak is appropriately delayed (delay length is a model-internal parameter)
- \+ NCS incremented in the Inland North in that era, so advantaged change is plausible
- – Required massive curtailment of traffic flow The exact value is model-internal, but its real-world equivalent is very implausible



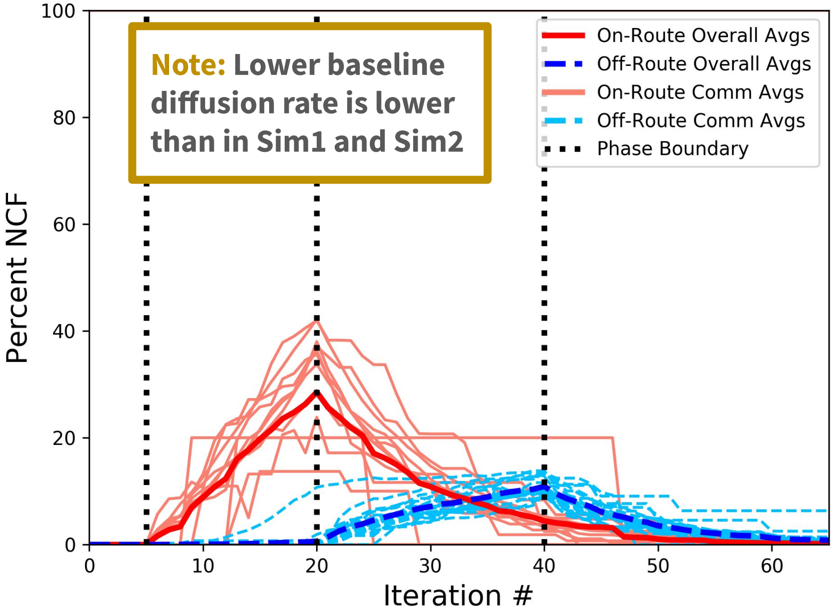
[1] following Niyogi & Berwick 1997

Sim3: Migration with Advantaged Change

Hypothesis: NCF had an advantage that allowed it to rise Off-Route after it already declined On-Route. It retreated due to migration from the Midlands to both On-Route and Off-Route

Conclusion

This advantaged change approach works in principle, but forces us into implausible assumptions in practice



[1] following Niyogi & Berwick 1997

Simulating Four Scenarios

Need to manipulate Off-Route directly…

Requires implausible assumptions about traffic

1. ~~Propagation by Diffusion Only~~
2. Migration Chicago→On-Route and ~~only Diffusion On-Route→Off-Route~~
3. Migration and Diffusion;

   Manipulating Migration with Advantaged Change ??

Simulating Four Scenarios

Need to manipulate Off-Route directly…

Requires implausible assumptions about traffic

1. ~~Propagation by Diffusion Only~~
2. Migration Chicago→On-Route and ~~only Diffusion On-Route→Off-Route~~
3. Migration and Diffusion;
   Manipulating Migration with Advantaged Change ??
4. Migration Chicago→On-Route and Migration On-Route→Off-Route;
   Manipulating Migration

Sim4: On-Route to Off-Route Migration

Hypothesis: NCF arrived On-Route with migration from Chicago and was transmitted to a new generation that then migrated to Off-Route Communities

Sim4: On-Route to Off-Route Migration

Hypothesis: NCF arrived On-Route with migration from Chicago and was transmitted to a new generation that then migrated to Off-Route Communities

Five-Phase Simulation

Phase 1 (Pre-Depression) Migration from the Midlands keeps the NCF rate down despite diffusion from Chicago

Phase 2 (Great Depression) Migration from Chicago to On-Route communities imports the NCF.

Phase 3 (Post-Depression 1) Migration from the Midlands to all communities

Phase 4 (Post-Depression 2) Migration On-Route to Off-Route with end of Phase 2 On-Route NCF rate ← Migration + Transmission

Phase 5 (Late 20th Century) Continuation of Phase 3

Sim4: On-Route to Off-Route Migration

Hypothesis: NCF arrived On-Route with migration from Chicago and was transmitted to a new generation that then migrated to Off-Route Communities

Five-Phase Simulation

Phase 1 (Pre-Depression)

Migration from the Midlands

Phase 2 (Great Depression)

Migration from Chicago

Phase 3 (Post-Depression 1)

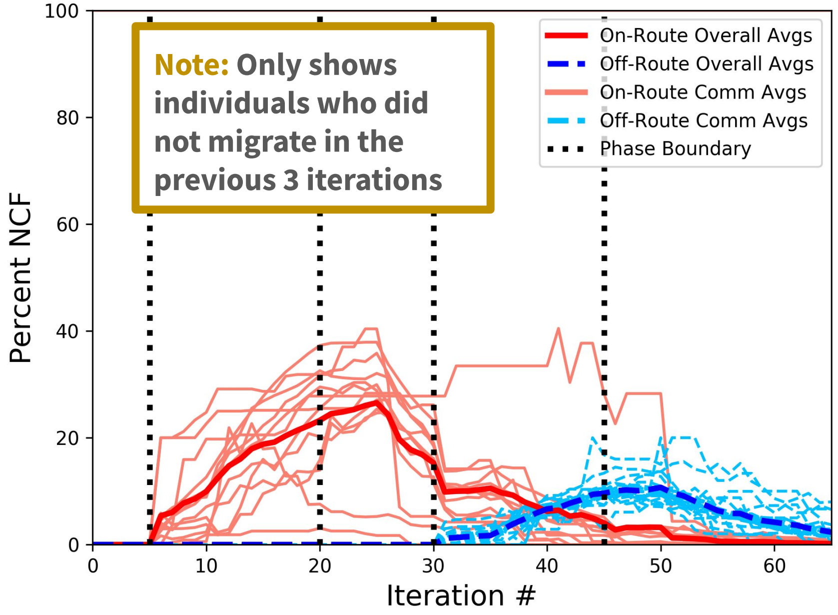
Migration from the Midlands to all

Phase 4 (Post-Depression 2)

Migration from On-Route to Off-Route

Phase 5 (Late 20th Century)

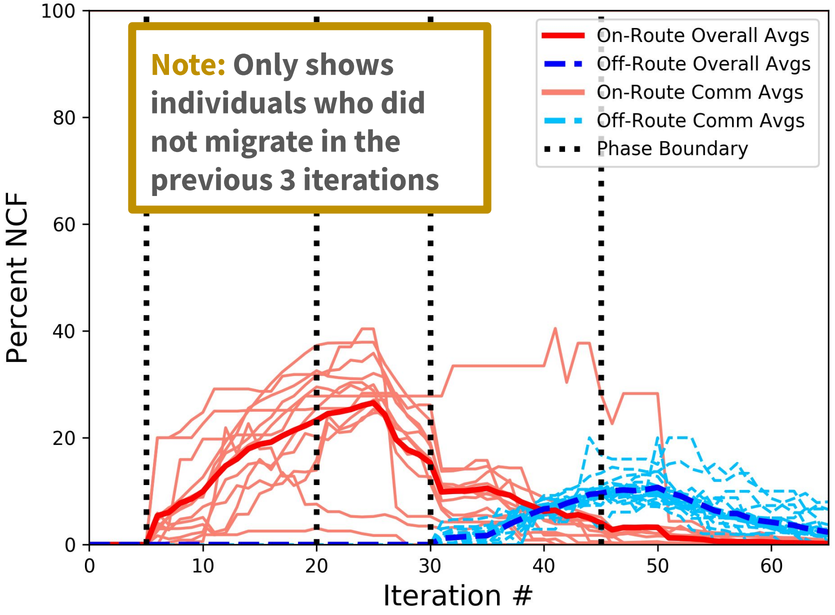
Return to Phase 3



Sim4: On-Route to Off-Route Migration

Hypothesis: NCF arrived On-Route with migration from Chicago and was transmitted to a new generation that then migrated to Off-Route Communities

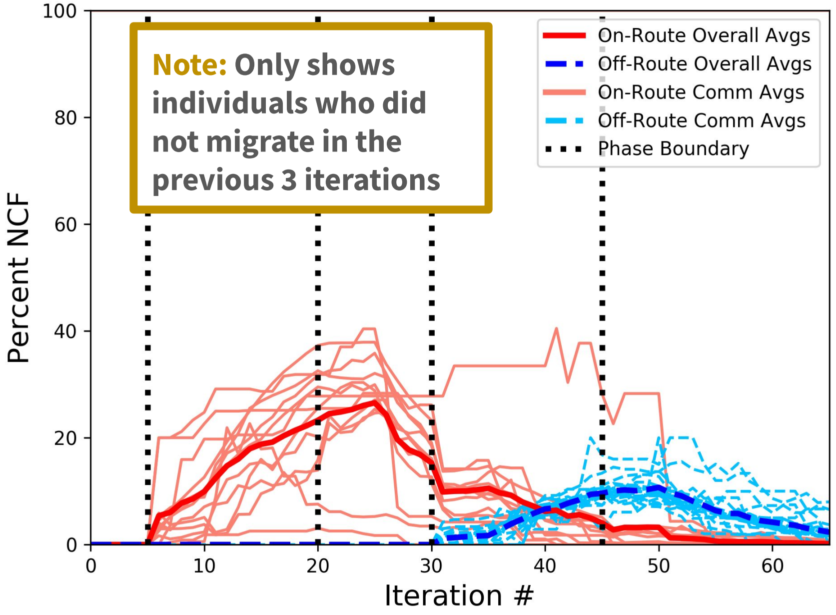
- \+ It looks good! The offset two-peak pattern
- \+ Off-Route's peak is appropriately delayed (delay length is a model-internal parameter)
- \+ Migration is consistent with population trends



Sim4: On-Route to Off-Route Migration

Hypothesis: NCF arrived On-Route with migration from Chicago and was transmitted to a new generation that then migrated to Off-Route Communities

- \+ It looks good! The offset two-peak pattern
- \+ Off-Route's peak is appropriately delayed (delay length is a model-internal parameter)
- \+ Migration is consistent with population trends
- ● No obvious drawbacks :-) but… no reason to assume that this is the only possible successful implementation



Sim4 Redux: Learning instead of Midlands Migration

Hypothesis: NCF arrived On-Route with migration from Chicago and was transmitted to a new generation that then migrated to Off-Route Communities

Five-Phase Simulation

Phase 1 (Pre-Depression)
Migration from the Midlands
Phase 2 (Great Depression)
Migration from Chicago
Phase 3 (Post-Depression 1)
~~Migration from the Midlands to all~~ → No migration
Phase 4 (Post-Depression 2)
Migration from On-Route to Off-Route
Phase 5 (Late 20th Century)
Return to Phase 3

[1] Yang 2009

Sim4 Redux: Learning instead of Midlands Migration

Hypothesis: NCF arrived On-Route with migration from Chicago and was transmitted to a new generation that then migrated to Off-Route Communities

Five-Phase Simulation

Phase 1 (Pre-Depression)
Migration from the Midlands

Phase 2 (Great Depression)
Migration from Chicago

Phase 3 (Post-Depression 1)
~~Migration from the Midlands to all~~ → No migration

Phase 4 (Post-Depression 2)
Migration from On-Route to Off-Route

Phase 5 (Late 20th Century)
Return to Phase 3

[1] Yang 2009

Replace Midlands migration with a threshold learning model[1] suggested by Friedman 2014

- Variants are acquired if they are attested above some threshold in the input ← Empirically estimated from the lexicon
- Related to the Variational Learning and Trigger-Learning algorithms

Test out the predictions of these thresholds:

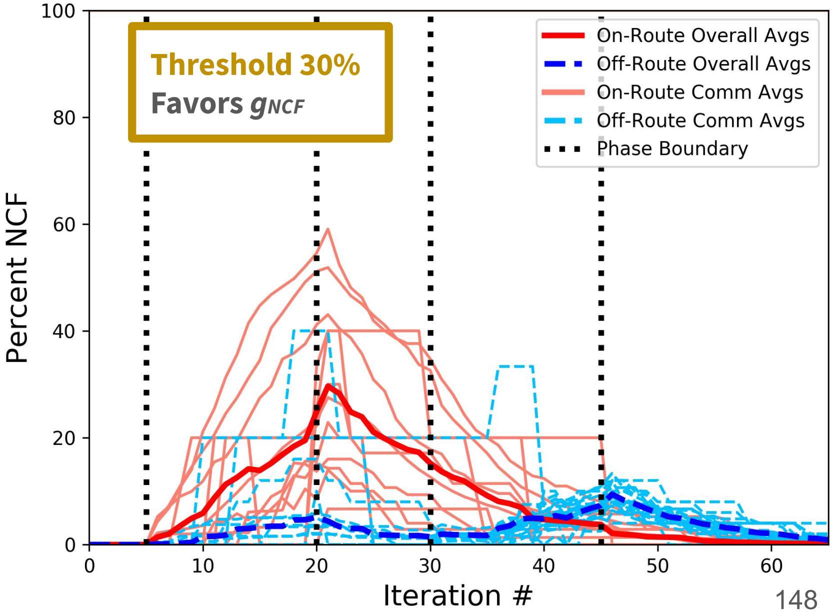
30% to acquire *g*NCF (NCF advantaged)
50% to acquire *g*NCF (neutral)
80% to acquire *g*NCF (NCF disadvantaged)

Sim4 Redux: Learning instead of Midlands Migration

Hypothesis: NCF arrived On-Route with migration from Chicago and was transmitted to a new generation that then migrated to Off-Route Communities

When g_{NCF} is favored

Off-peak rises too early. Reminiscent of the Sim3 advantage problem

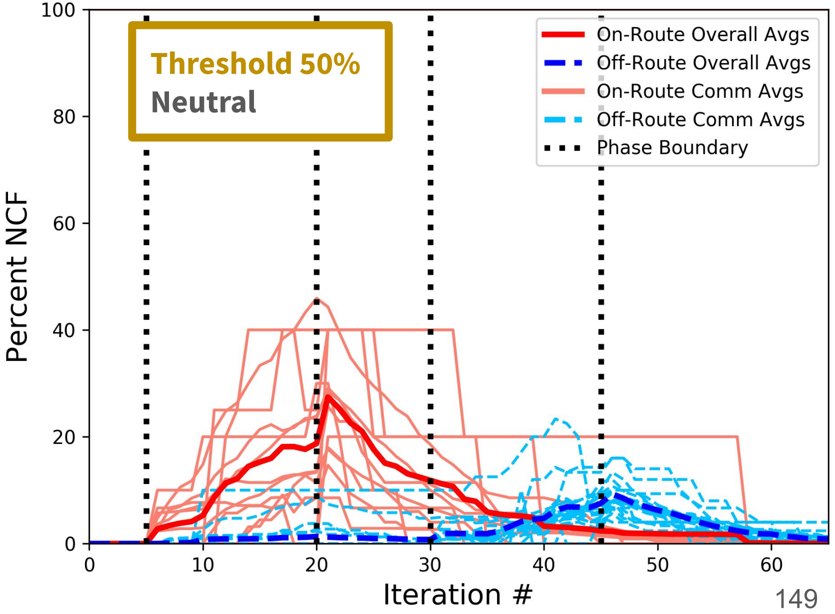


Sim4 Redux: Learning instead of Midlands Migration

Hypothesis: NCF arrived On-Route with migration from Chicago and was transmitted to a new generation that then migrated to Off-Route Communities

When learning is neutral

It works!

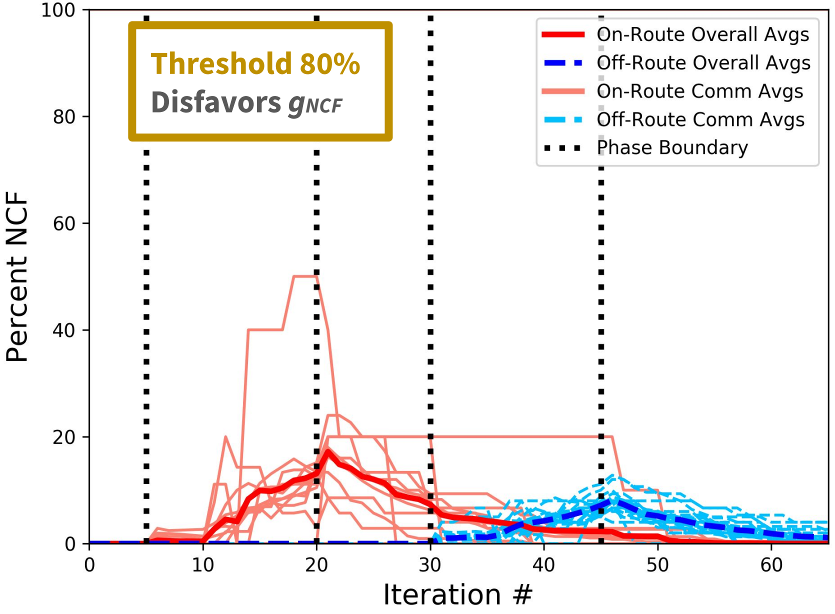


Sim4 Redux: Learning instead of Midlands Migration

Hypothesis: NCF arrived On-Route with migration from Chicago and was transmitted to a new generation that then migrated to Off-Route Communities

When *g*NCF is disfavored

It works!



Sim4 Redux: Learning instead of Midlands Migration

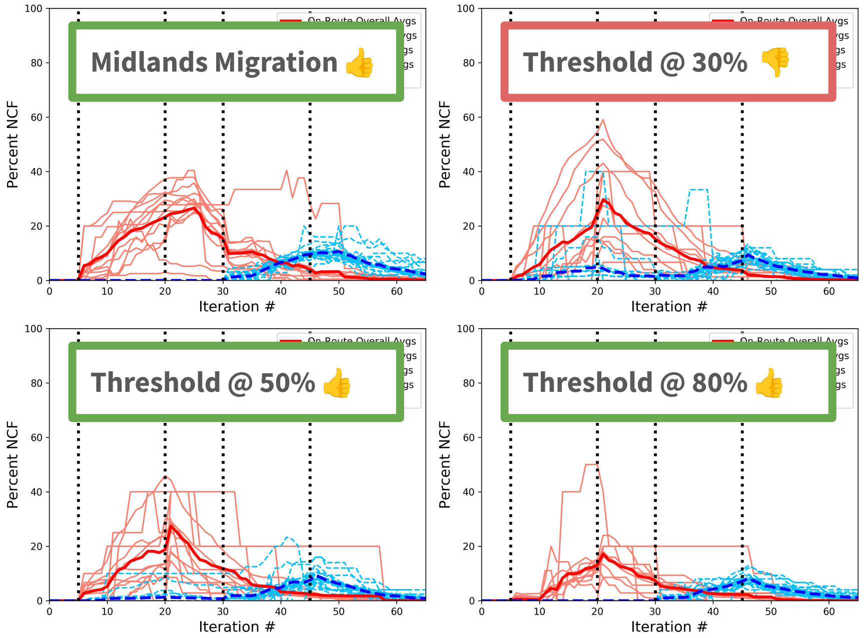
Hypothesis: NCF arrived On-Route with migration from Chicago and was transmitted to a new generation that then migrated to Off-Route Communities

Conclusion

Migration + Transmission works under a range of assumptions, which is great!

But… this means our simulation lacks the power to differentiate between these hypotheses.

Is Midlands migration supported by field data? Is threshold learning supported? If so, which threshold is supported?



Simulating Four Scenarios

Need to manipulate Off-Route directly…

Requires implausible assumptions about traffic

1. ~~Propagation by Diffusion Only~~
2. Migration Chicago→On-Route and ~~only Diffusion On-Route→Off-Route~~
3. Migration and Diffusion;
   Manipulating Migration with Advantaged Change ??
4. Migration Chicago→On-Route and Migration On-Route→Off-Route;
   Manipulating Migration

It works under a variety of assumptions, but do these assumptions correspond to the real world?

Takeaways under the Self-Critical Approach

We can make a range of inferences about the St. Louis Corridor

To the extent that our modeling assumptions correspond to the real world,

We can rule out some models

Simulation 1 - Diffusion alone

Simulation 2 - Migration to On-Route
+ Diffusion Off-Route

Simulation 3 - Migration with
Advantaged Change

Takeaways under the Self-Critical Approach

We can make a range of inferences about the St. Louis Corridor

To the extent that our modeling assumptions correspond to the real world,

We can rule out some models

Simulation 1 - Diffusion alone

Simulation 2 - Migration to On-Route
+ Diffusion Off-Route

Simulation 3 - Migration with
Advantaged Change

Cannot produce the empirical offset two-peak pattern

Takeaways under the Self-Critical Approach

We can make a range of inferences about the St. Louis Corridor

To the extent that our modeling assumptions correspond to the real world,

We can rule out some models

Simulation 1 - Diffusion alone

Simulation 2 - Migration to On-Route + Diffusion Off-Route

→ Cannot produce the empirical offset two-peak pattern

Simulation 3 - Migration with Advantaged Change

→ Can produce the empirical offset two-peak pattern only under implausible assumptions

Takeaways under the Self-Critical Approach

We can make a range of inferences about the St. Louis Corridor

To the extent that our modeling assumptions correspond to the real world,

We can identify research questions for follow-up

Simulation 4 - Migration + Transmission

Works under a few variants, but our simulations cannot distinguish them

Takeaways under the Self-Critical Approach

We can make a range of inferences about the St. Louis Corridor

To the extent that our modeling assumptions correspond to the real world,

Successful simulations complete a research cycle

Sociolinguistic fieldwork provided hypotheses that they can't distinguish →

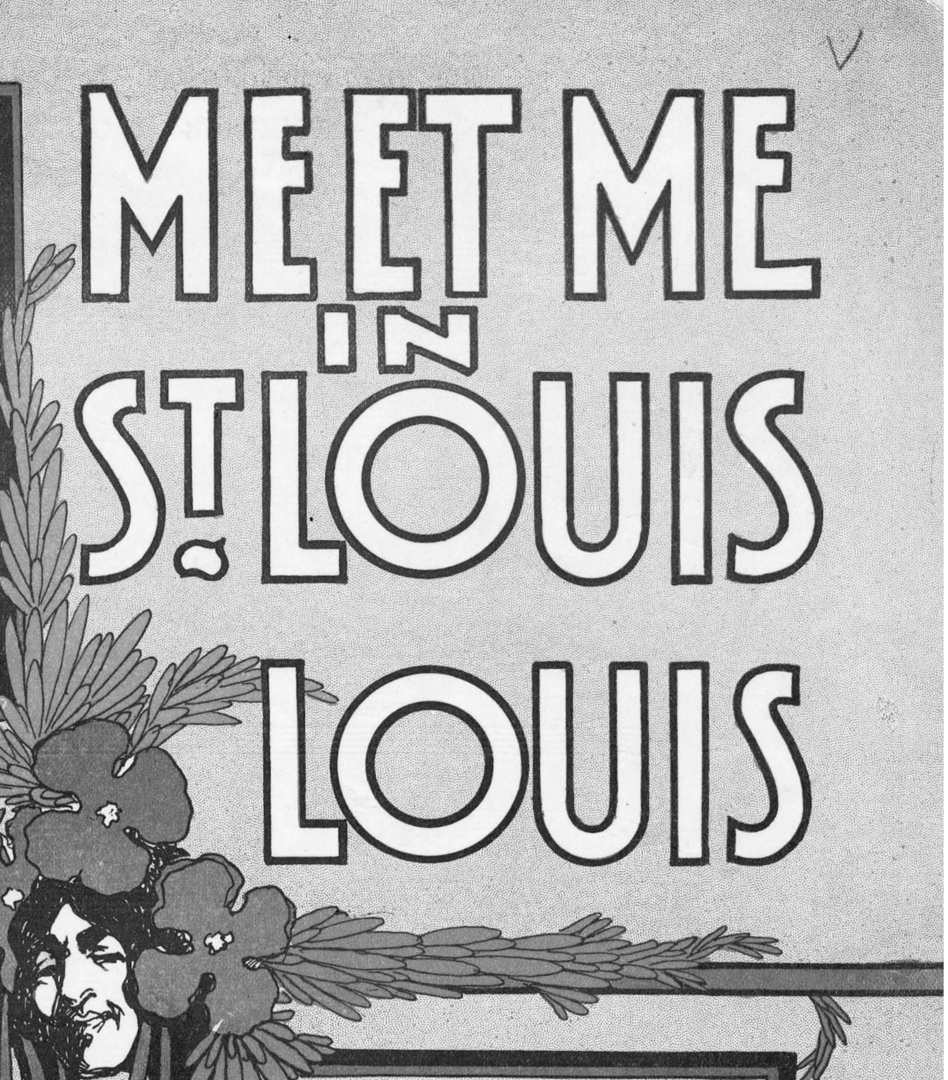
Our simulations were directed by a filtered down those hypotheses

and made some new ones →

Fieldwork is directed by and will filter out our hypotheses

and make some new ones →

…



The End

Thank you!

For more information about my work on population-level modeling of language change (links available at https://jkodner05.github.io/):

- Kodner & Cerezo Falco (2018, *ACL*)
- Kodner (2020, *LVC*)
- Kali & Kodner (2022, *LChange*)
- Kodner (2023, *JHS*)

Code for our framework as well as the St. Louis Corridor adjacency matrices is available at: https://github.com/jkodner05/corridornetwork