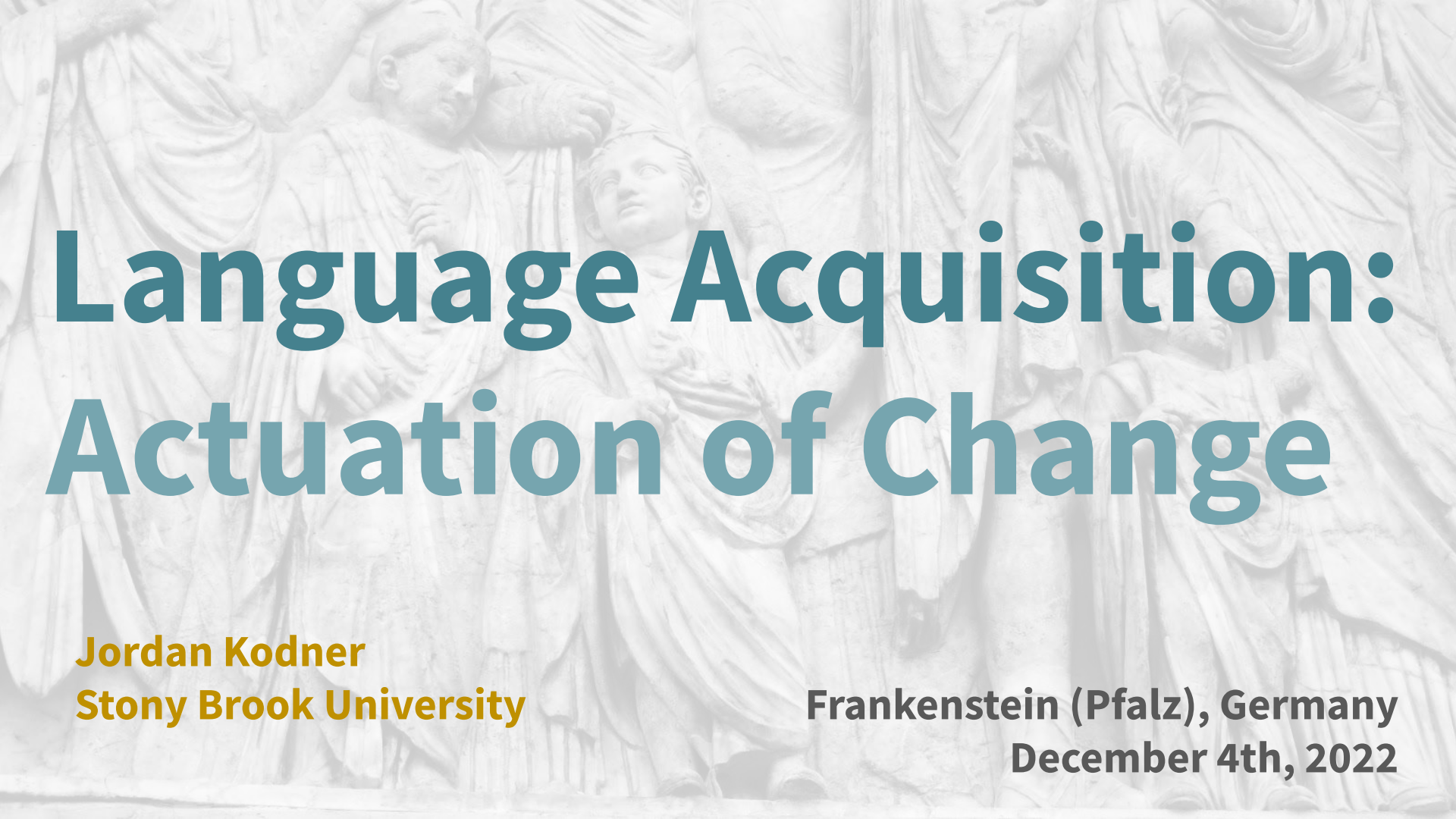


Language Acquisition: Actuation of Change

Jordan Kodner
Stony Brook University

Frankenstein (Pfalz), Germany
December 4th, 2022

Outline

Partial /aɪ/-Raising in North American English

- **Modeling the acquisition of productivity…in the context of variation**
- **Modeling the emergence and disappearance of a new grammar**

Modeling Language Acquisition in the Past

- **Actuation as the primary focus**
- **Historical data and modeling acquisition**

The Rise of the to-Dative in Middle and Early Modern English

- **A multi-generational diachronic case study**
- **Change as over-regularization**

Time Permitting: Some thoughts on Acquisition and Propagation

Outline

Partial /aɪ/-Raising in North American English

- Modeling the acquisition of productivity…in the context of variation
- Modeling the emergence and disappearance of a new grammar

Modeling Language Acquisition in the Past

- Actuation as the primary focus
- Historical data and modeling acquisition

The Rise of the to-Dative in Middle and Early Modern English

- A multi-generational diachronic case study
- Change as over-regularization

Time Permitting: Some thoughts on Acquisition and Propagation

Outline

Partial /aɪ/-Raising in North American English

- Modeling the acquisition of productivity…in the context of variation
- Modeling the emergence and disappearance of a new grammar

Modeling Language Acquisition in the Past

- Actuation as the primary focus
- Historical data and modeling acquisition

The Rise of the to-Dative in Middle and Early Modern English

- A multi-generational diachronic case study
- Change as over-regularization

Time Permitting: Some thoughts on Acquisition and Propagation

Outline

Partial /aɪ/-Raising in North American English

- Modeling the acquisition of productivity…in the context of variation
- Modeling the emergence and disappearance of a new grammar

Modeling Language Acquisition in the Past

- Actuation as the primary focus
- Historical data and modeling acquisition

The Rise of the to-Dative in Middle and Early Modern English

- A multi-generational diachronic case study
- Change as over-regularization

Time Permitting: Some thoughts on Acquisition and Propagation

Emergence of Partial /aɪ/-Raising as a "Contact Phenomenon"

with Caitlin Richter (2020, *PWPL*)

“Canonical” /aɪ/-Raising

“Canadian” Raising of /aɪ/ before (underlyingly) voiceless segments

/taɪd/	“tide”		/tʌɪt/	“tight”
/laɪv/	“live”	vs.	/lʌɪf/	“life”
/raɪz/	“rise”		/rʌɪs/	“rice”

Interacts with /t/-flapping - classic example of phonological opacity

/raɪɾɚ/	“rider”
/rʌɪɾɚ/	“writer”

“Transparent” /aɪ/-Raising

Raising before surface voiceless segments only

Canonical

/raɪd/ “ride” ≠ /rʌɪt/ “write”

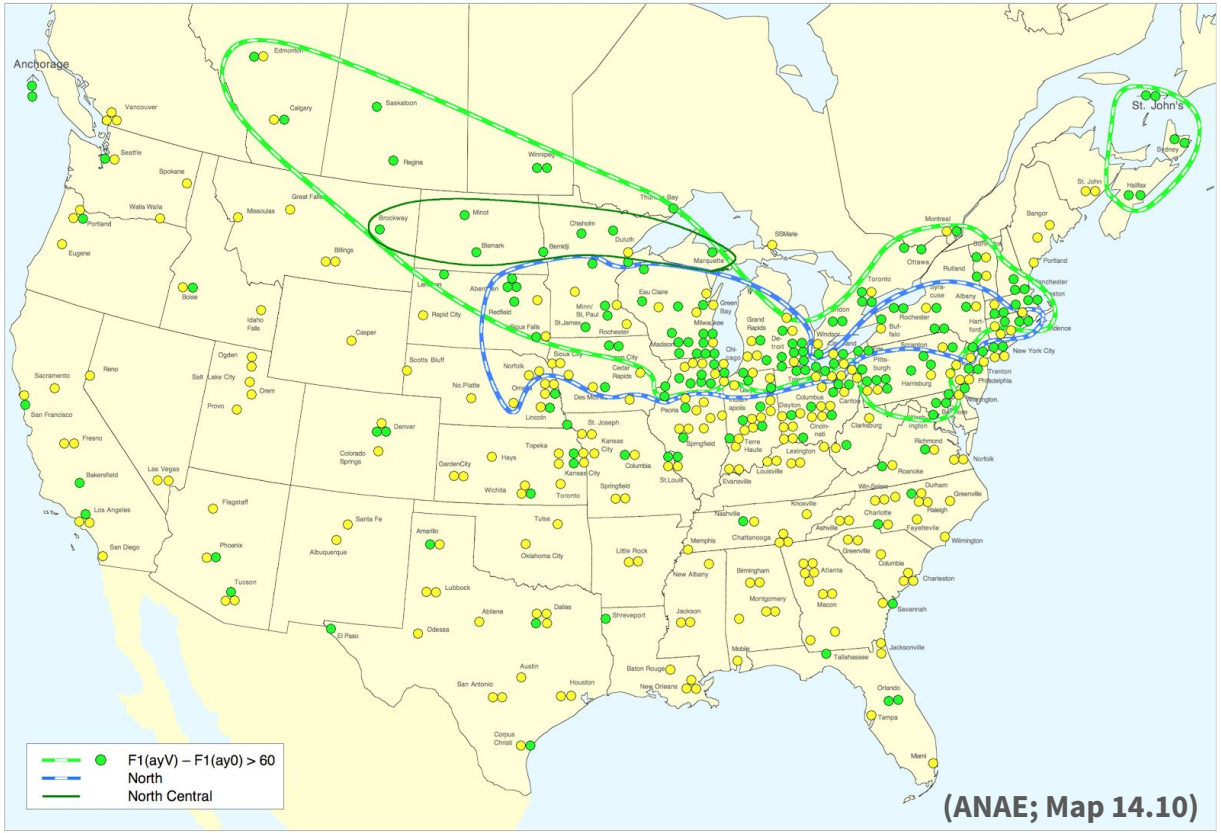
/raɪɾɚ/“rider” ≠ /rʌɪɾɚ/“writer”

Transparent

/raɪd/ “ride” ≠ /rʌɪt/ “write”

/raɪɾɚ/“rider” = /raɪɾɚ/“writer”

“Transparent” /aɪ/-Raising in the Wild

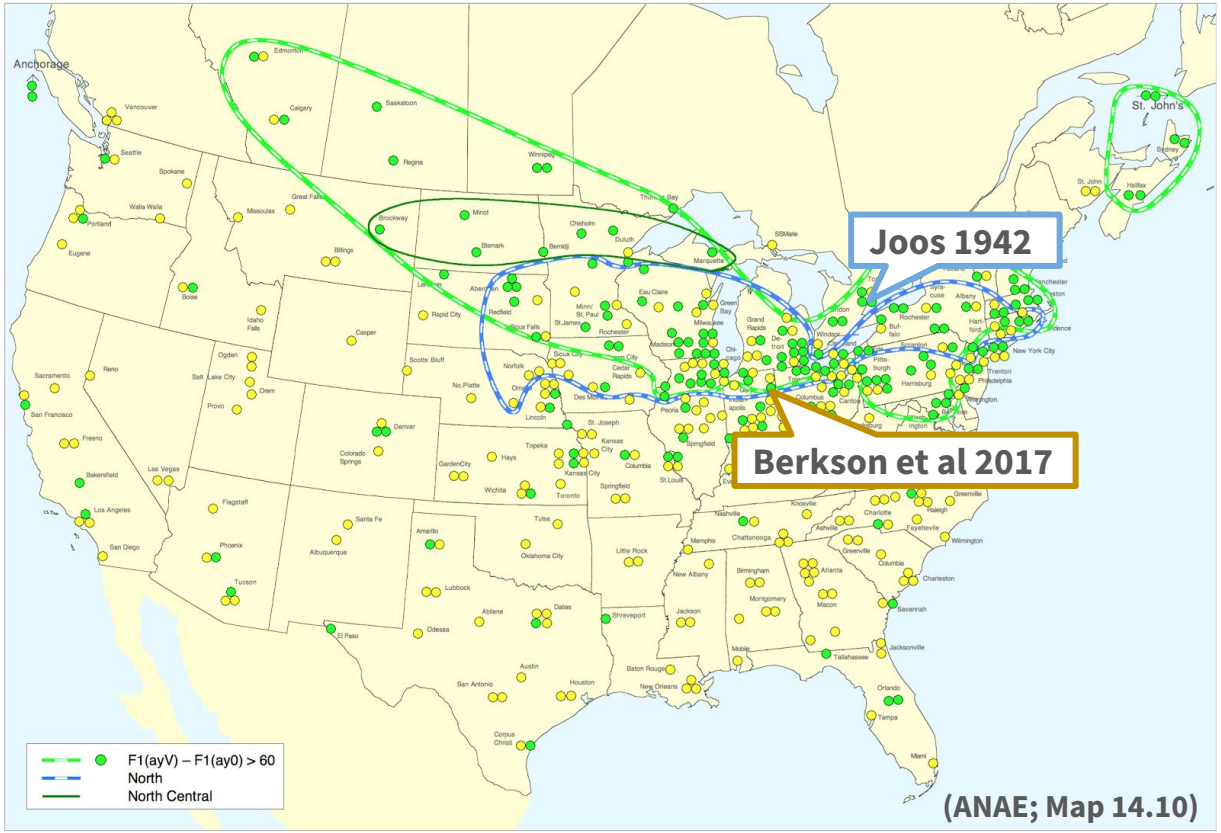


“Transparent” /aɪ/-Raising in the Wild

- Reported just twice
- 75 years apart
- On edges of the raising region

Joos (1942)
in Toronto

Berkson et al (‘17, ‘20, ‘22) in Fort Wayne, IN



“Transparent” /aɪ/-Raising as Incipient /aɪ/-Raising

A phonetically-driven precursor to canonical raising

- **Hypocorrection[1] before surface /t/ spread to flapped /t/**
- **Offglide peripheralization[2]**
- **pre-voiceless shortening[3]**
- **Berkson et al 2017 argue for hypocorrection**

[1] Ohala 1981, [2] Moreton & Thomas 2007, [3] Joos 1942

“Transparent” /aɪ/-Raising as Incipient /aɪ/-Raising

A phonetically-driven precursor to canonical raising

- **Hypocorrection**[1] before surface /t/ spread to flapped /t/
- **Offglide peripheralization**[2]
- **pre-voiceless shortening**[3]
- Berkson et al 2017 argue for hypocorrection

But…

- Fruehwald 2016 finds that /aɪ/-raising was always conditioned by the **underlying** consonant in the Philadelphia Neighborhood Corpus[4]

[1] Ohala 1981, [2] Moreton & Thomas 2007, [3] Joos 1942, [4] Labov & Rosenfelder 2011

An Alternative Account

Transparent raising as a byproduct of child language acquisition in mixed canonical/non-raising input environments

- **If children hardly hear raising, they should not learn it**
- **If they hear it consistently, they should learn raising**
- **And if inconsistently, they may learn transparent raising as a partial system**

What is a “mixed input” environment?

Everybody receives input from multiple grammars

- Navigating variation is a normal part of language acquisition
- Monolingual input is generated from potentially many very similar grammars
- And we know a lot about variation in North American English

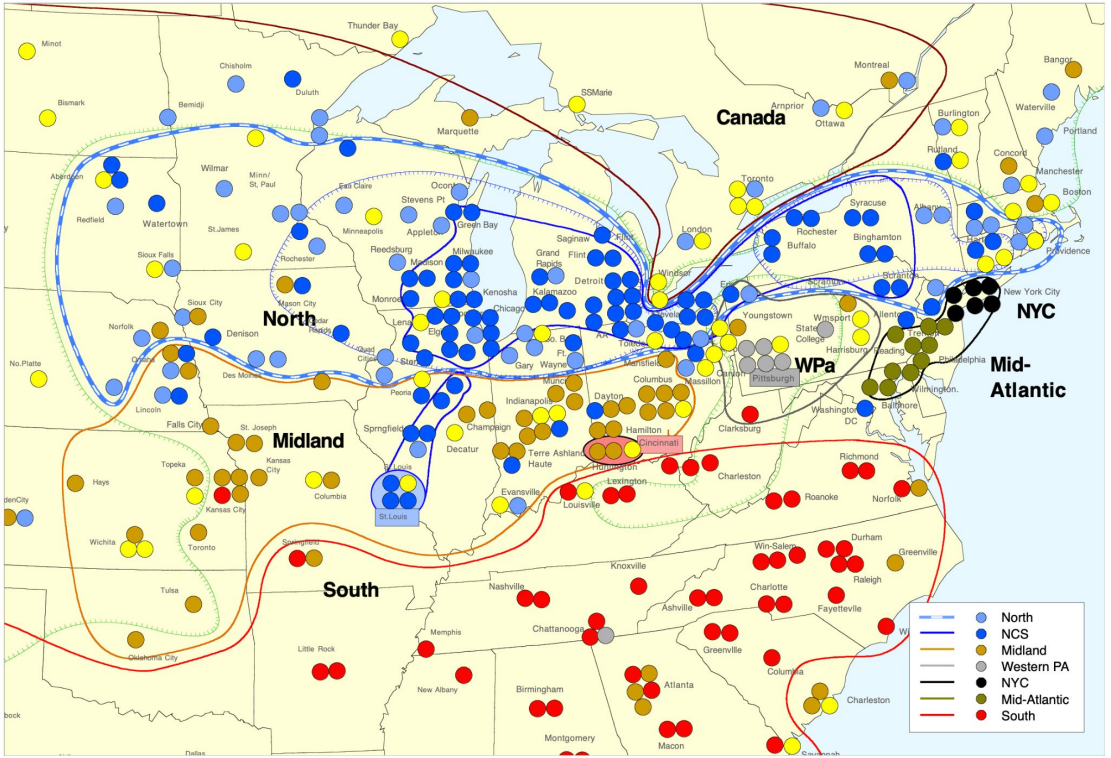
Geographic Diversity in North America

Driven by historical and ongoing patterns of migration

Map 11.15. An overall view of North American dialects

Geographic Diversity in North America

Driven by historical and ongoing patterns of migration

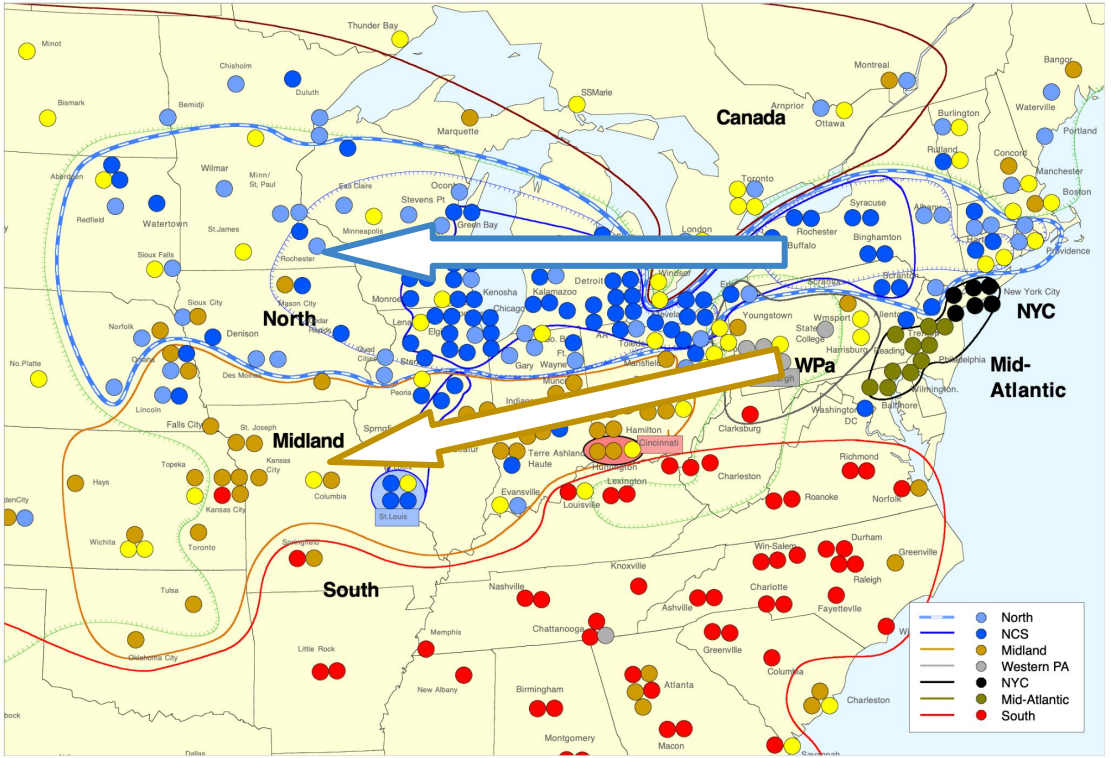


Map 11.9. The Midland and the Mid-Atlantic regions

Geographic Diversity in North America

Driven by historical and ongoing patterns of migration

- **Inland North** and **Midlands** form long E→W bands

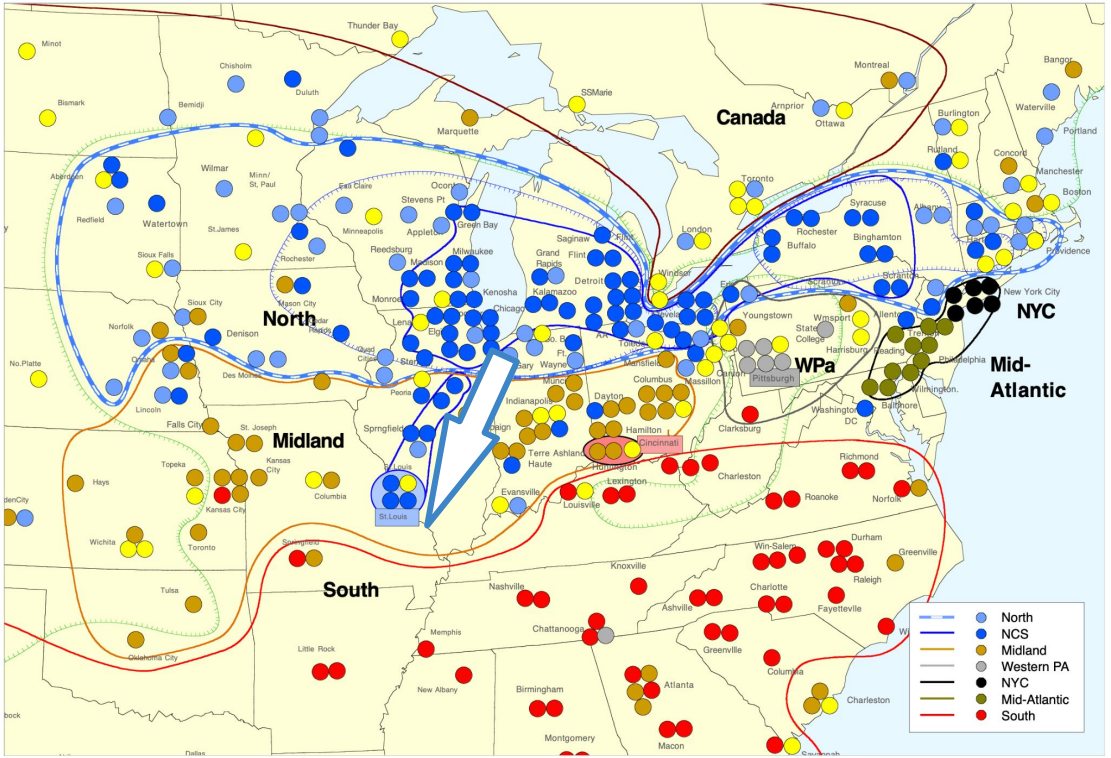


Map 11.9. The Midland and the Mid-Atlantic regions

Geographic Diversity in North America

Driven by historical and ongoing patterns of migration

- Inland North and Midlands form long E→W bands
- St Louis corridor is an island of the InN in the Midlands



Map 11.9. The Midland and the Mid-Atlantic regions

Early Acquisition of Phonology

Children identify inventory of surface segments early in acquisition

- Stable system of contrasts emerges at ~6-12 months[1]

Can learn allophones underlyingly relating some segments

- Certain phenomena, eg aspirated and unaspirated English /p/[2]
- Influence of learned allophones evident in perception ~8 months[3]

[1] Kuhl et al 1992, Werker & Tees 1984, [2] Pierrehumbert 2003, [3] Pegg & Werker 1997

The Tolerance Principle (Yang 2016)

- A concrete model for the acquisition of linguistic generalization
- An evaluation metric over linguistic hypotheses
- Developed in the context of the Past Tense Debate
  But has since been applied across levels of the grammar

The Tolerance Principle (Yang 2016)

Given a hypothesized generalization operating over some class, quantitatively define the number of exceptions below which the generalization is tenable

N = number of types that should obey the generalization

e = number of types that do not obey the generalization

θ = max # of exceptions that can be tolerated

Exceptions are tolerable if

$$e < \theta$$

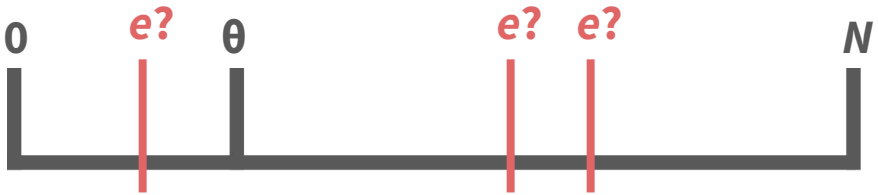
$$\theta = N / \ln N$$

N and *e* Vary over Individual Development

- *N* and *e* are properties of each individual
- *N* is the number of class members a child has learned so far
- *N* and *e* grow as the learner's vocabulary grows
- Can learn generalizations over small *N* not possible over large *N*

Visualization of the Tolerance Principle

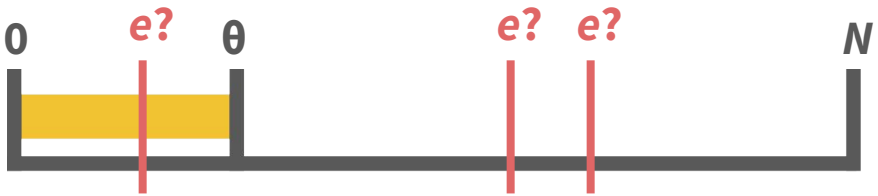
N = types it should apply to
e = types that are exceptions
θ = tolerance threshold



e falls in [0,*N*] and may be less than or greater than θ

Visualization of the Tolerance Principle

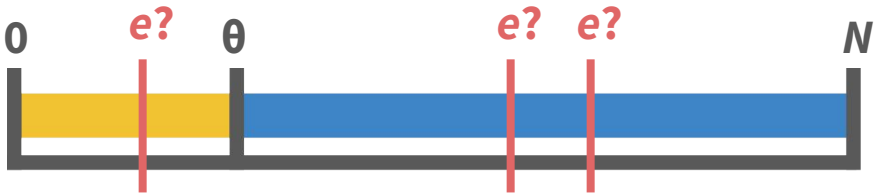
N = types it should apply to
e = types that are exceptions
θ = tolerance threshold



If *e* is below θ,
acquire pattern as rule

Visualization of the Tolerance Principle

N = types it should apply to
e = types that are exceptions
θ = tolerance threshold



If *e* is below θ,
acquire pattern as rule
Otherwise, do not form rule

Visualization of the Tolerance Principle

N = types it should apply to
e = types that are exceptions
θ = tolerance threshold

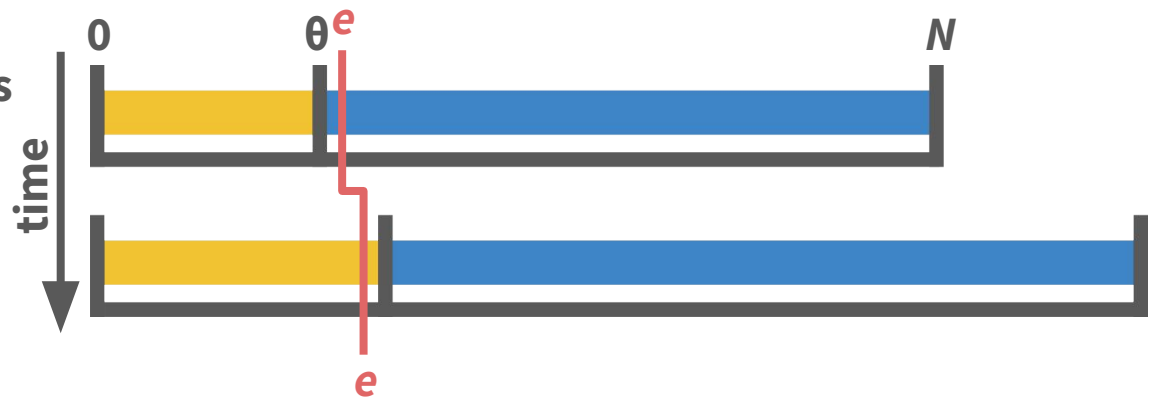
If *e* is below θ,
acquire pattern as rule
Otherwise, do not form rule

- *N* grows over an individual's development, θ grows more slowly

Visualization of the Tolerance Principle

N = types it should apply to
e = types that are exceptions
θ = tolerance threshold

If *e* is below θ,
acquire pattern as rule
Otherwise, do not form rule

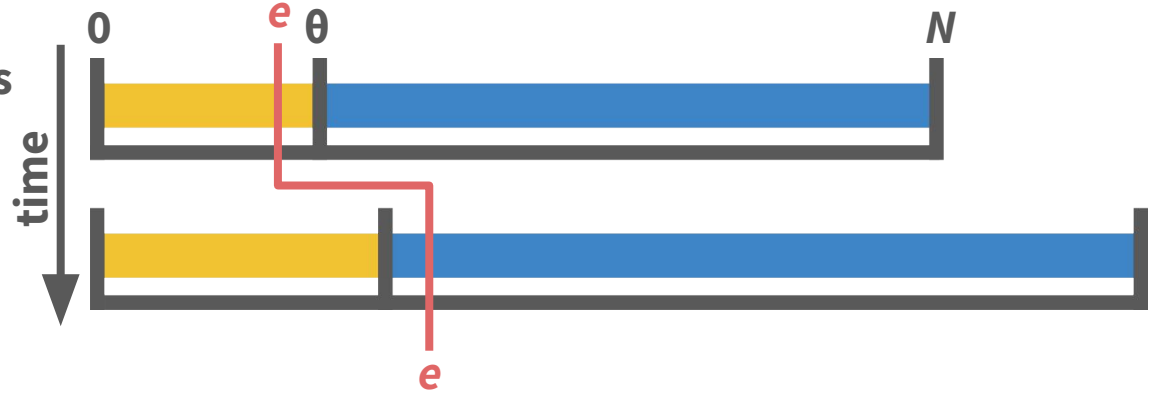


- *N* grows over an individual's development, θ grows more slowly
- If θ grows faster than *e*, a pattern may fall into productivity

Visualization of the Tolerance Principle

N = types it should apply to
e = types that are exceptions
θ = tolerance threshold

If *e* is below θ,
acquire pattern as rule
Otherwise, do not form rule



- *N* grows over an individual's development, θ grows more slowly
- If θ grows faster than *e*, a pattern may fall into productivity
- If *e* grows faster than θ, a pattern may fall out of productivity

Regularizing Mixed Input

- Children often regularize variable input rather than probability matching[1]
  Especially younger children
  And for categorical patterns of the grammar
- Older children do probability match, particularly for continuous variables[2]

For the Tolerance Principle,

- Which variant of a type enters the lexicon is based on token frequency
- Tolerance of generalizations is based on type frequency

In lexical learning with phonological variation, children acquire whichever variant of each lexical type is more frequent in the input[3]

[1,2] Singleton & Newport 2004, Hudson Kam & Newport 2005, Schuler et al 2017, Newport 2019, Austin et al 2022, [3] Sneller et al 2018

Mixed Input Learning is Probabilistic

- Whether a specific lexical item is acquired in one form or another is probabilistic → it depends on the ratio of raising and non-raising input
- Whether a form is learned as raised or non-raised is like flipping a coin
  Binomial distribution weighted by distribution of variants in the population

Mixed Input Learning is Probabilistic

- Whether a specific lexical item is acquired in one form or another is probabilistic → it depends on the ratio of raising and non-raising input
- Whether a form is learned as raised or non-raised is like flipping a coin
  Binomial distribution weighted by distribution of variants in the population

Determining *N* and *e*

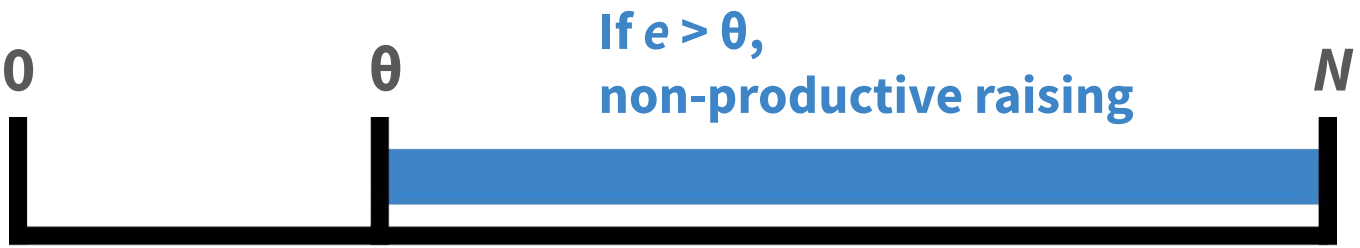
- *N* and *e* vary by child
- *N* the number of raisable words a child has learned so far
- *e* probabilistic. The number of raisable words learned as non-raised

Threshold for Canonical /aɪ/-Raising

- *N* = # of “raisable” words (underlying /aɪt/)
- *e* = # of those *N* not learned as raised
- θ = tolerance threshold

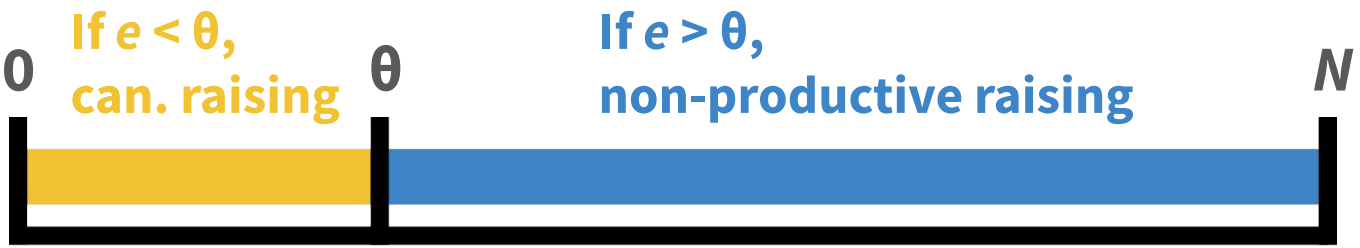
Threshold for Canonical /aɪ/-Raising

- *N* = # of “raisable” words (underlying /aɪt/)
- *e* = # of those *N* not learned as raised
- θ = tolerance threshold



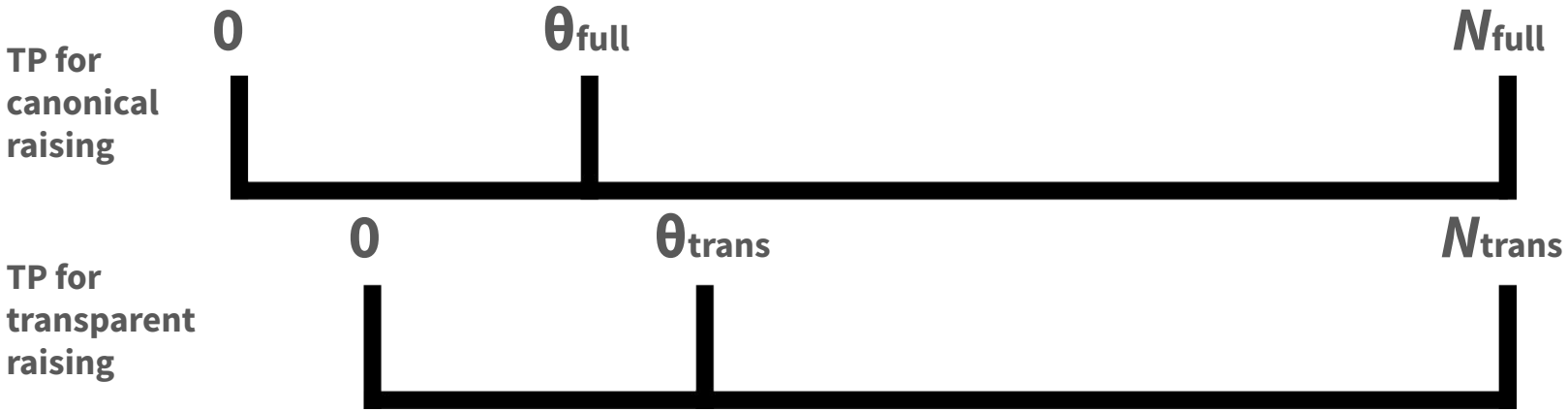
Threshold for Canonical /aɪ/-Raising

- *N* = # of “raisable” words (underlying /aɪt/)
- *e* = # of those *N* not learned as raised
- θ = tolerance threshold



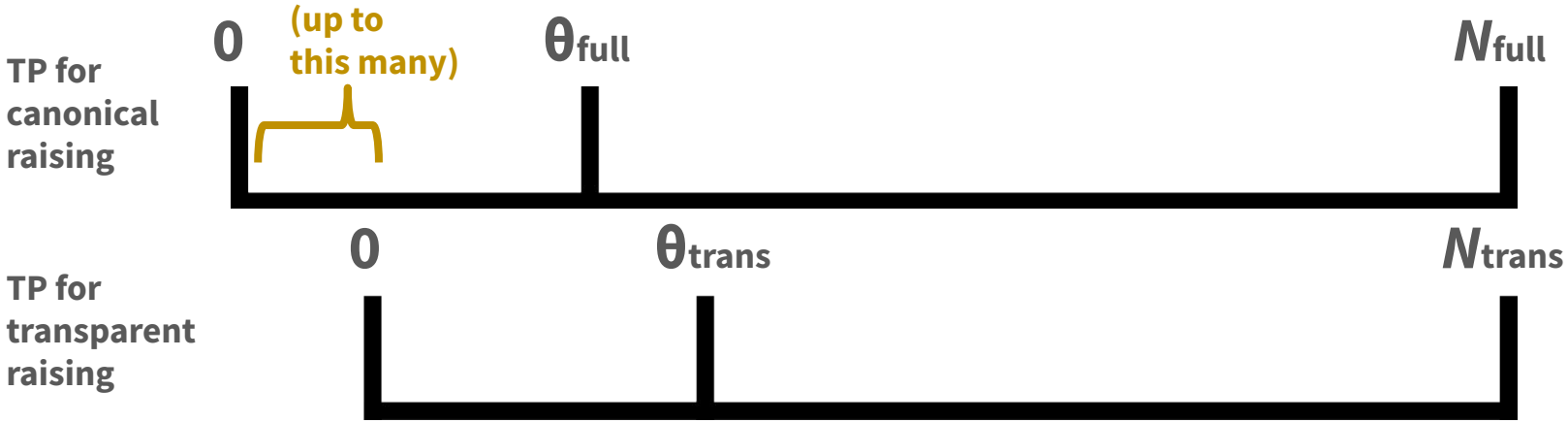
Threshold for Transparent /aɪ/-Raising

- The same calculation but with its own *N*, *e*, θ
- N_{full} = # of surface /t/ /aɪt/ words + # of flapped /aɪt/ words
- N_{trans} = # of surface /t/ /aɪt/ words



Threshold for Transparent /aɪ/-Raising

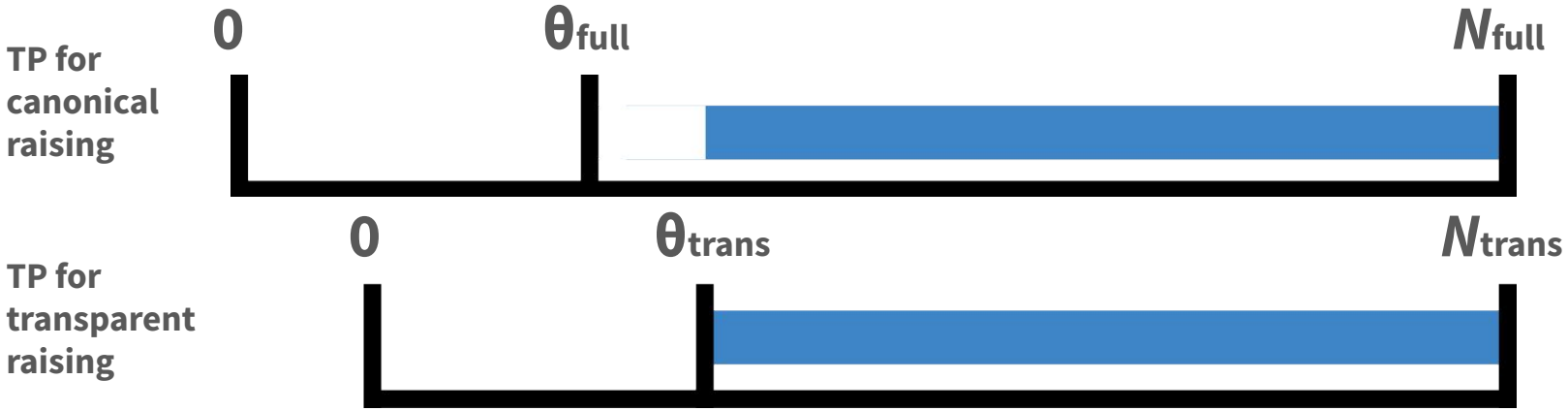
- The same calculation but with its own *N*, *e*, θ
- N_{full} = # of surface /t/ /aɪt/ words + # of flapped /aɪt/ words
- N_{trans} = # of surface /t/ /aɪt/ words
- Canonical raising can have extra exceptions that are irrelevant to transparent



Threshold for Transparent /aɪ/-Raising

- The same calculation but with its own *N*, *e*, θ
- $N_{full} = N_{trans} + N_{flap}$

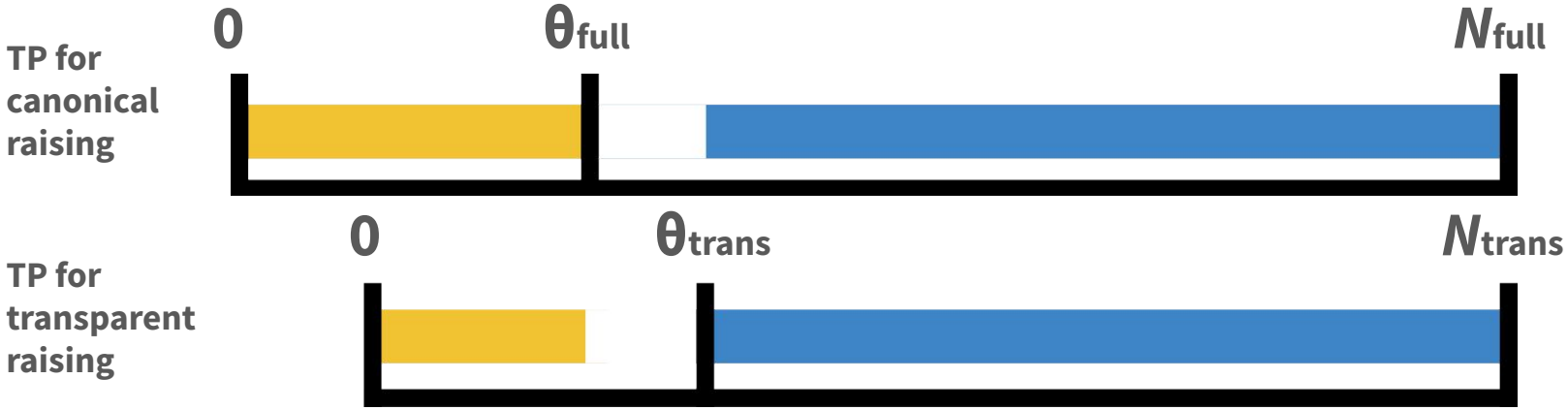
If both thresholds are exceeded, no productive raising



Threshold for Transparent /aɪ/-Raising

- The same calculation but with its own *N*, *e*, θ
- $N_{full} = N_{trans} + N_{flap}$

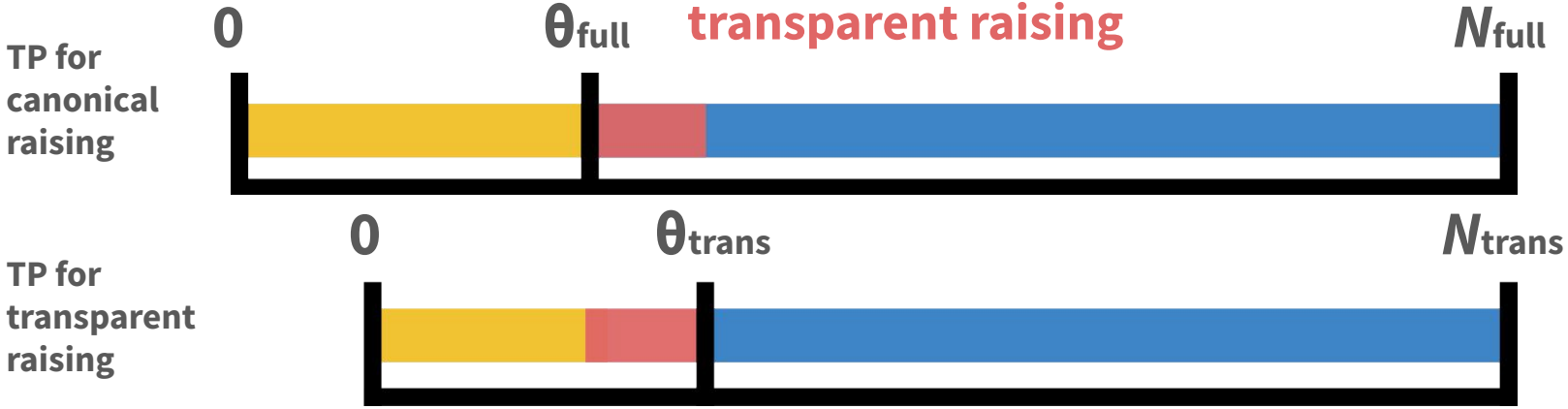
If the broader generalization is upheld, canonical raising



Threshold for Transparent /aɪ/-Raising

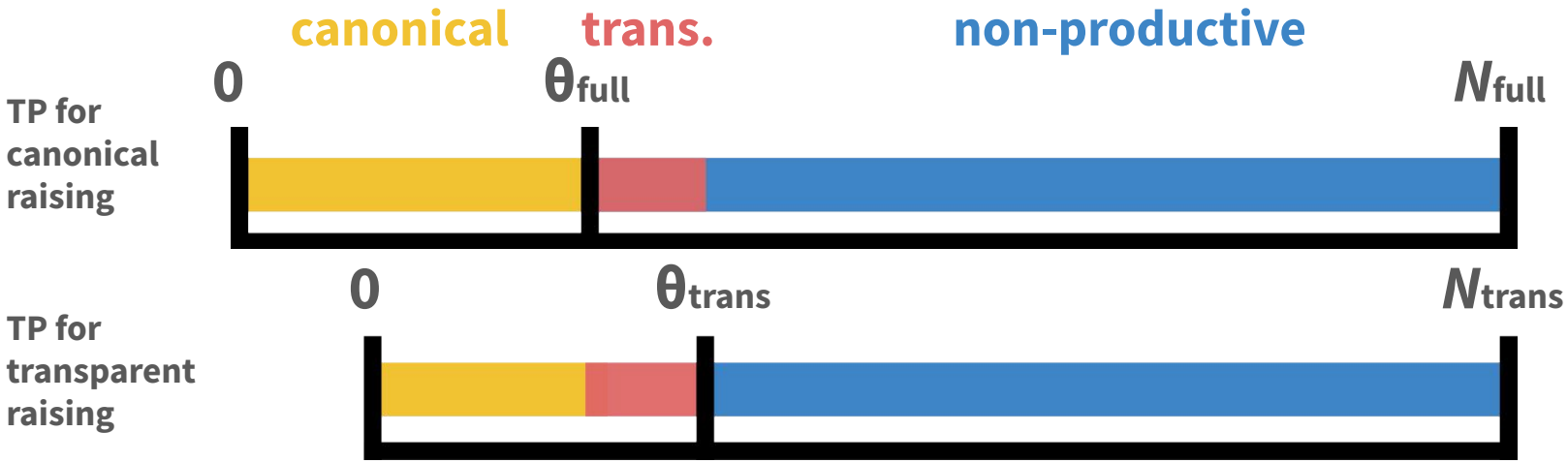
- The same calculation but with its own *N*, *e*, θ
- $N_{full} = N_{trans} + N_{flap}$

If there are too many flapped exceptions
but not too many faithful ones,
transparent raising



Threshold for Transparent /aɪ/-Raising

- The same calculation but with its own *N*, *e*, θ
- It is technically possible for linguistic input to support transparent raising while not supporting canonical raising. How likely is this?



Threshold for Transparent /aɪ/-Raising

- More formally, e_{trans} and e_{full} must fall in the following ranges
- Need “too many” flapped exceptions while still having room to have “not too many” surface faithful ones

$$\theta_{full} - N_{flap} \leq e_{trans} \leq \lfloor \theta_{trans} \rfloor$$

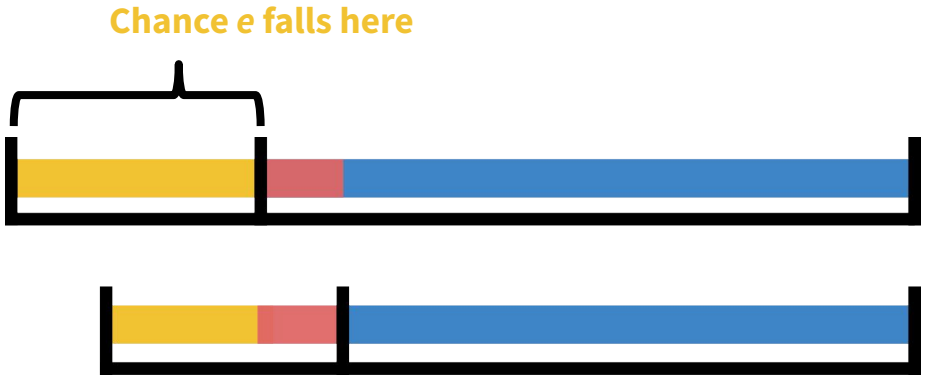
$$\theta_{full} - e_{trans} \leq e_{flap} \leq \lfloor N_{flap} \rfloor$$

Calculating Learner Outcomes

Probability of learning canonical raising

(p_{none} = fraction of non-raisers in community = 1-p_{full})

$$= \sum_{e_{\text{full}}=0}^{\lfloor \theta_{\text{full}} \rfloor} \binom{N_{\text{full}}}{e_{\text{full}}} p_{\text{none}}^{e_{\text{full}}} \, p_{\text{full}}^{N_{\text{full}} - e_{\text{full}}}$$



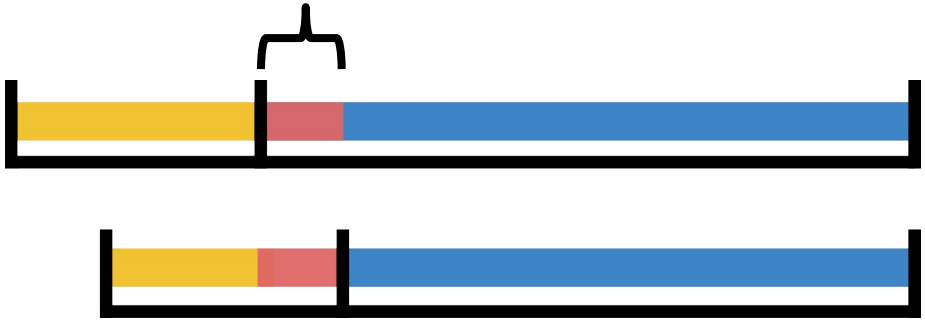
Calculating Learner Outcomes

Probability of learning canonical raising

(p_{none} = fraction of non-raisers in community = 1-p_{full})

$$= \sum_{e_{\text{full}}=0}^{\lfloor \theta_{\text{full}} \rfloor} \binom{N_{\text{full}}}{e_{\text{full}}} p_{\text{none}}^{e_{\text{full}}} \, p_{\text{full}}^{N_{\text{full}} - e_{\text{full}}}$$

Chance *e* falls here
(too many flapped exceptions,
not too many faithful ones)



Probability of learning transparent raising

$$= \sum_{e_{\text{trans}}=\theta_{\text{full}}-N_{\text{flap}}}^{\lfloor \theta_{\text{trans}} \rfloor} \left(\binom{N_{\text{trans}}}{e_{\text{trans}}} p_{\text{none}}^{e_{\text{trans}}} \, p_{\text{full}}^{N_{\text{trans}} - e_{\text{trans}}} \sum_{e_{\text{flap}}=\theta_{\text{full}}-e_{\text{trans}}}^{N_{\text{flap}}} \binom{N_{\text{flap}}}{e_{\text{flap}}} p_{\text{none}}^{e_{\text{flap}}} \, p_{\text{full}}^{N_{\text{flap}} - e_{\text{flap}}} \right)$$

Four Features of First Language Acquisition

1. **All children receive unique input yet exhibit gross developmental uniformity**[1]
2. The type frequency of a pattern is crucial for acquisition of generalizations, as opposed to token frequency or attestation of initial items[2]
3. Token frequencies correlate with relative order of acquisition[3]
4. Early learner vocabularies are small[4]

[1] Labov 1972, [2] Aronoff 1976, MacWhinney 1978, Bybee 1985, Baayen 1993, Elman 1998, Pierrehumbert 2003, Yang 2016, [3] Goodman 2008, [4] Hart & Risley 1995, 2003, Szagun et al. 2006

Four Features of First Language Acquisition

1. All children receive unique input yet exhibit gross developmental uniformity[1]
2. The type frequency of a pattern is crucial for acquisition of generalizations, as opposed to token frequency or attestation of initial items[2]
3. Token frequencies correlate with relative order of acquisition[3]
4. Early learner vocabularies are small[4]

[1] Labov 1972, [2] Aronoff 1976, MacWhinney 1978, Bybee 1985, Baayen 1993, Elman 1998, Pierrehumbert 2003, Yang 2016, [3] Goodman 2008, [4] Hart & Risley 1995, 2003, Szagun et al. 2006

Four Features of First Language Acquisition

1. All children receive unique input yet exhibit gross developmental uniformity[1]
2. The type frequency of a pattern is crucial for acquisition of generalizations, as opposed to token frequency or attestation of initial items[2]
3. Token frequencies correlate with relative order of acquisition[3]
4. Early learner vocabularies are small[4]

[1] Labov 1972, [2] Aronoff 1976, MacWhinney 1978, Bybee 1985, Baayen 1993, Elman 1998, Pierrehumbert 2003, Yang 2016, [3] Goodman 2008, [4] Hart & Risley 1995, 2003, Szagun et al. 2006

Four Features of First Language Acquisition

1. All children receive unique input yet exhibit gross developmental uniformity[1]
2. The type frequency of a pattern is crucial for acquisition of generalizations, as opposed to token frequency or attestation of initial items[2]
3. Token frequencies correlate with relative order of acquisition[3]
4. Early learner vocabularies are small[4]

[1] Labov 1972, [2] Aronoff 1976, MacWhinney 1978, Bybee 1985, Baayen 1993, Elman 1998, Pierrehumbert 2003, Yang 2016, [3] Goodman 2008, [4] Hart & Risley 1995, 2003, Szagun et al. 2006

Four Features of First Language Acquisition

1. All children receive unique input yet exhibit gross developmental uniformity[1]
2. The type frequency of a pattern is crucial for acquisition of generalizations, as opposed to token frequency or attestation of initial items[2]
3. Token frequencies correlate with relative order of acquisition[3]
4. Early learner vocabularies are small[4]

As a result,

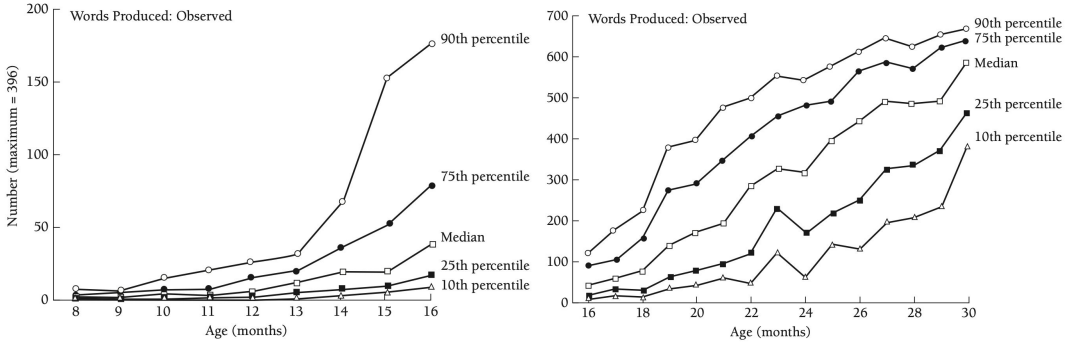
- Applying a frequency cutoff to items in CDS approximates a “typical” child
- Insight taken by type frequency-based models of acquisition[5]

[1] Labov 1972, [2] Aronoff 1976, MacWhinney 1978, Bybee 1985, Baayen 1993, Elman 1998, Pierrehumbert 2003, Yang 2016, [3] Goodman 2008, [4] Hart & Risley 1995, 2003, Szagun et al. 2006, [5] Nagy & Anderson 1984, Yang 2016

Child Lexical Knowledge

- Learners' vocabularies grow over the course of development
- There is significant individual variation, but consistent trends[1]
- **Only on the order of 10^2** for English and German learners by around age 3
- Children have the foundations for language-specific grammars by this point

Language	Estimated |Vocab|
English 2;10-3;0[1]	525-1,116
German 2;6[3]	$\mu = 429, \sigma > 100$



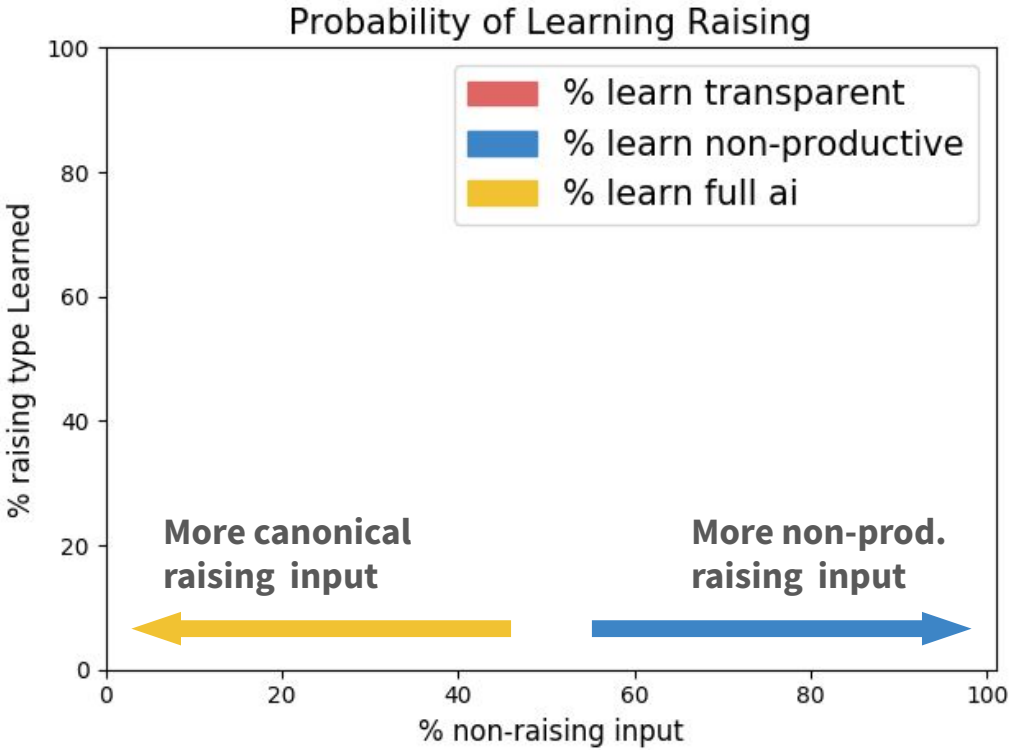
[1] Fenson et al 1994, Hart & Risley 2003, [2] Hart & Risley 2003, [3] Szagun et al 2006, Plots from Fenson et al 1994

Estimating N_{full} and N_{trans}

- From corpora of child-directed speech
- We took multiple estimates from Brown and Brown+Brent (in CHILDES)
- Recall, *N* is calculated over types, not tokens

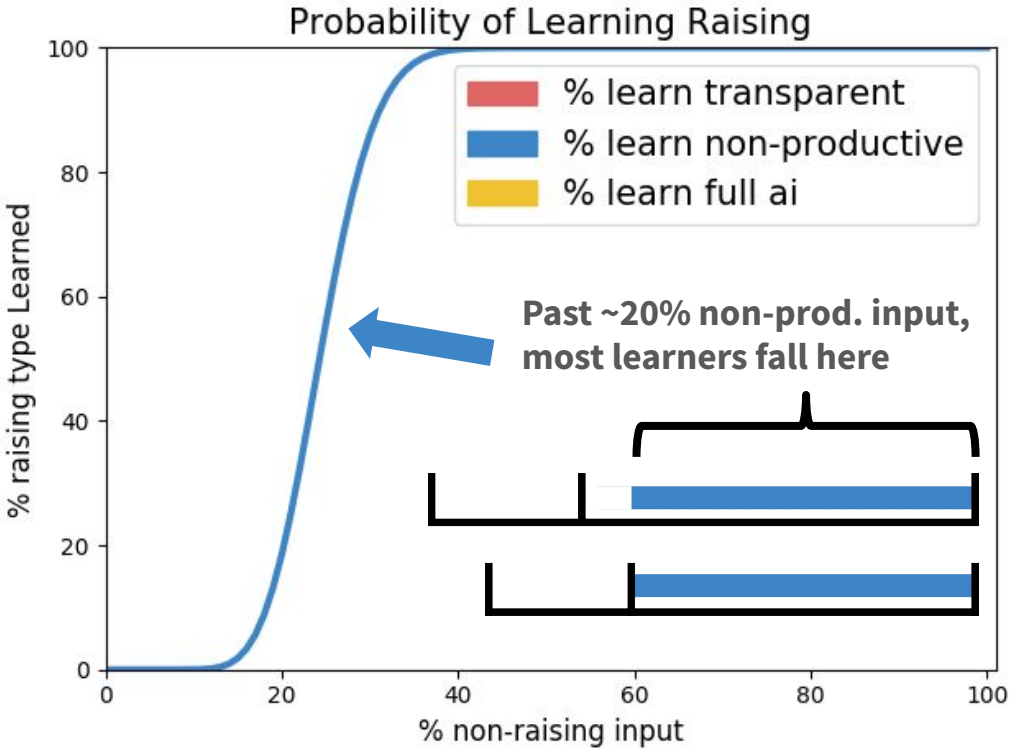
Estimate	N_{full} (# types)	N_{trans} (# types)
Brown (freq ≥ 5)	53	45
B+B (freq ≥ 5)	82	69
Brown (all)	122	103
B+B (all)	182	155

Transparent Raising Emerging *De Novo*



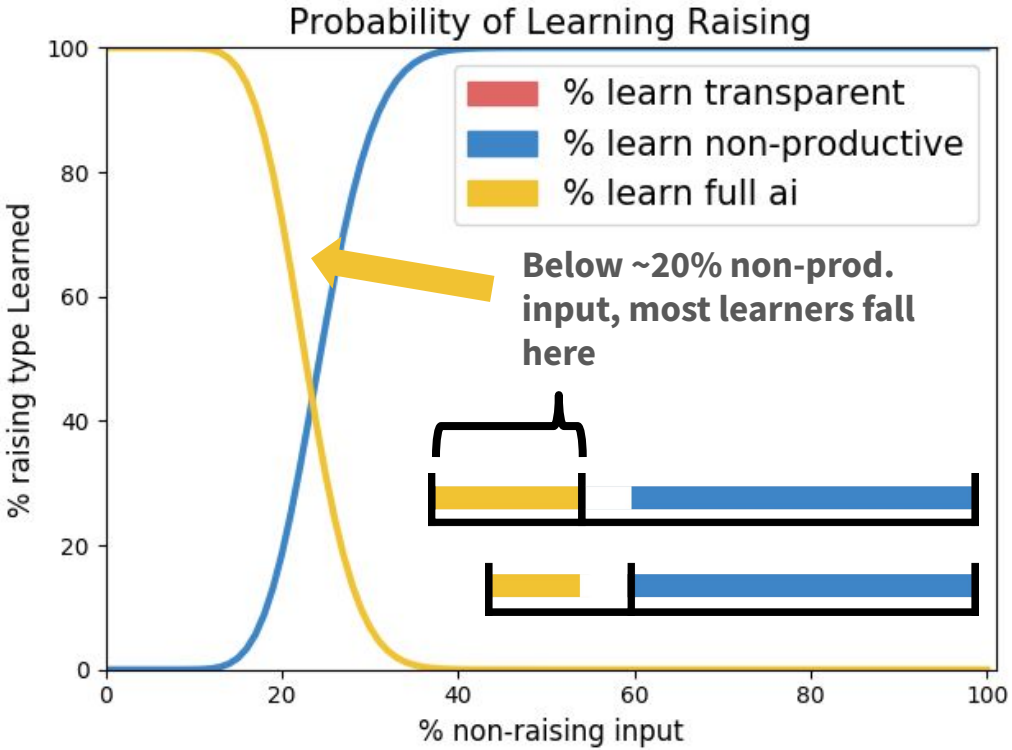
- *x*-axis initial rate of non-raising vs canonical raising in community
- *y*-axis proportion of learners learning each raising type

Transparent Raising Emerging *De Novo*



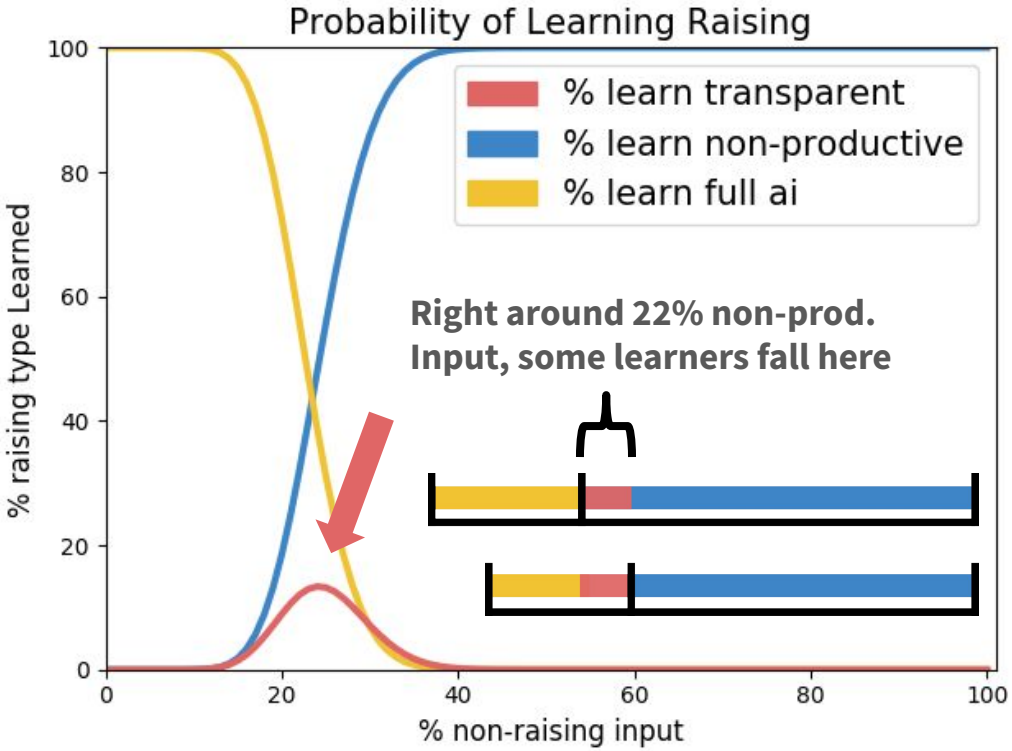
- *x*-axis initial rate of non-raising vs canonical raising in community
- *y*-axis proportion of learners learning each raising type

Transparent Raising Emerging *De Novo*



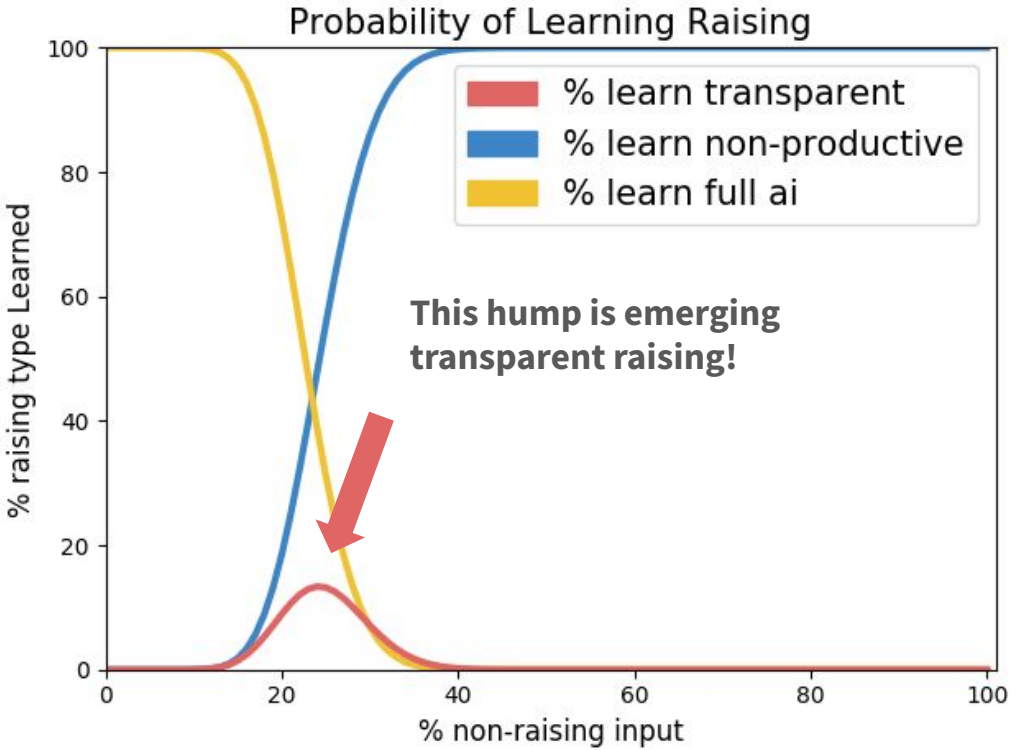
- *x*-axis initial rate of non-raising vs canonical raising in community
- *y*-axis proportion of learners learning each raising type

Transparent Raising Emerging *De Novo*



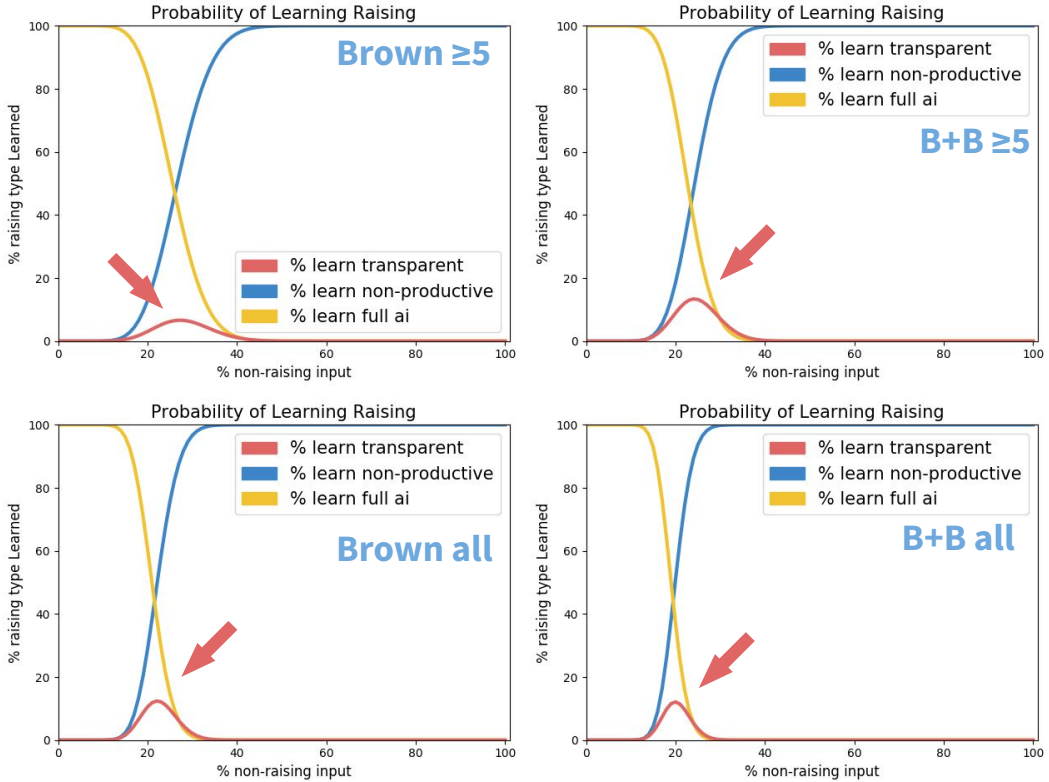
- *x*-axis initial rate of non-raising vs canonical raising in community
- *y*-axis proportion of learners learning each raising type

Transparent Raising Emerging *De Novo*



- *x*-axis initial rate of non-raising vs canonical raising in community
- *y*-axis proportion of learners learning each raising type

Results are Independent of Corpus and Filtering



- **Transparent peaks occur at >20% non-raiser communities**
- **Transparent peaks reach <20% max**

This works because N_{full} tends to be just slightly larger than N_{trans}

The Instability of Transparent /aɪ/-Raising

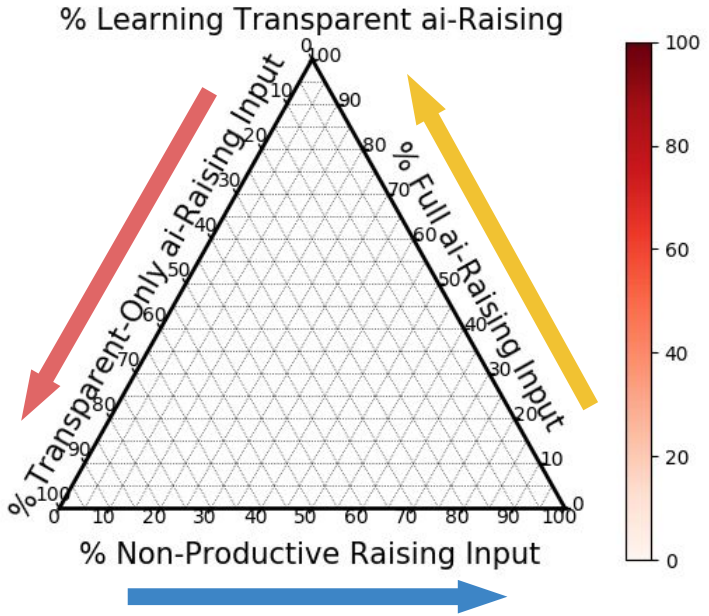
- Berkson et al suggest that transparent raising is rare because it is fleeting
- Our model concurs and provides an explanation for why

Populations of non/trans/canonical raisers are unstable

- They trend toward either non-raising or canonical raising over time
- Transparent raising dies out rapidly

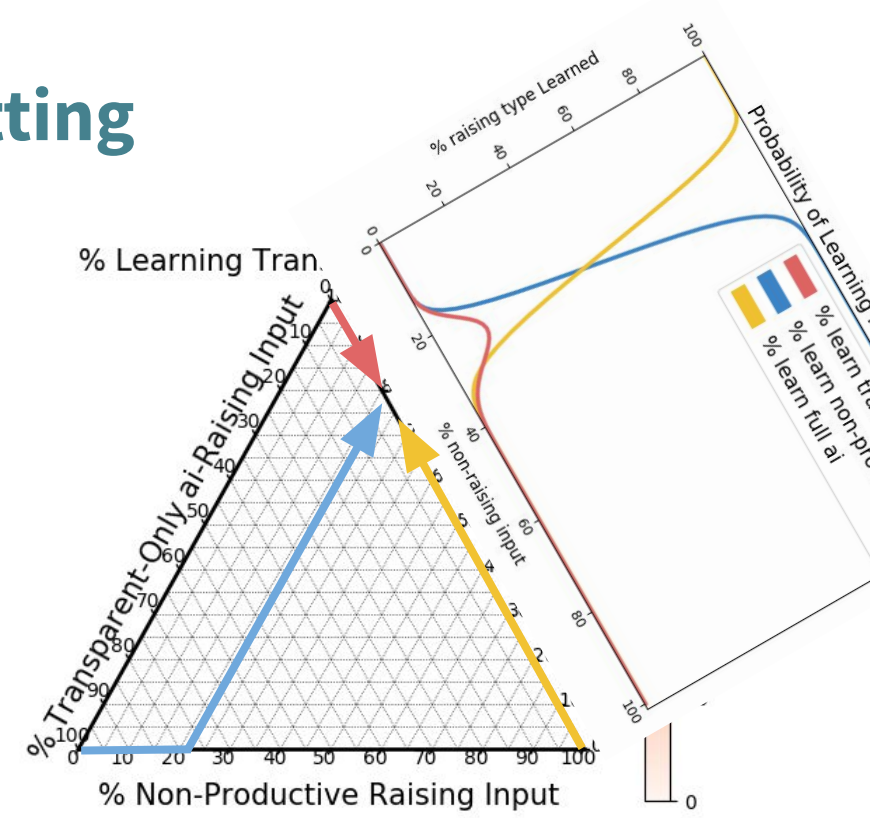
Learning in a 3-way Mixed Setting

- The previous model but allowing for 3-way input mixes
- Run iteratively to show raising evolves in the population over time
- Ternary plot for visualization



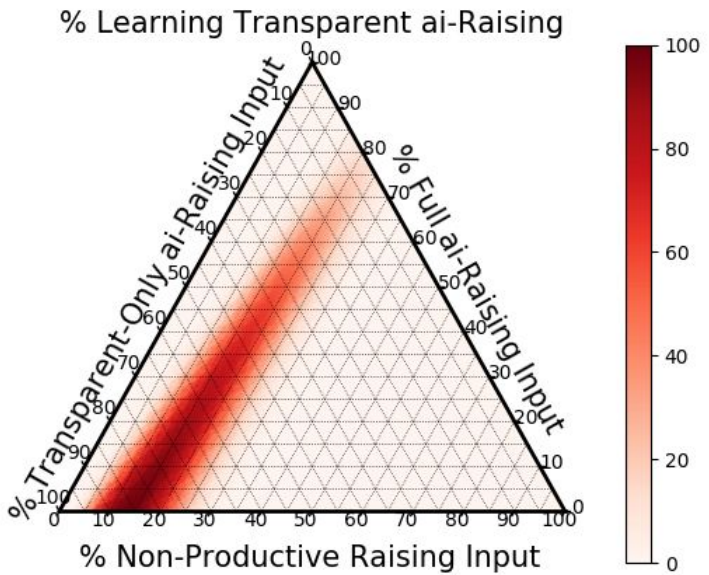
Learning in a 3-way Mixed Setting

- The previous model but allowing for 3-way input mixes
- Run iteratively to show raising evolves in the population over time
- **Ternary plot for visualization**
- Previous plots were on the top right side of the triangle where $p_{trans} = 0$ (no transparent raising input)



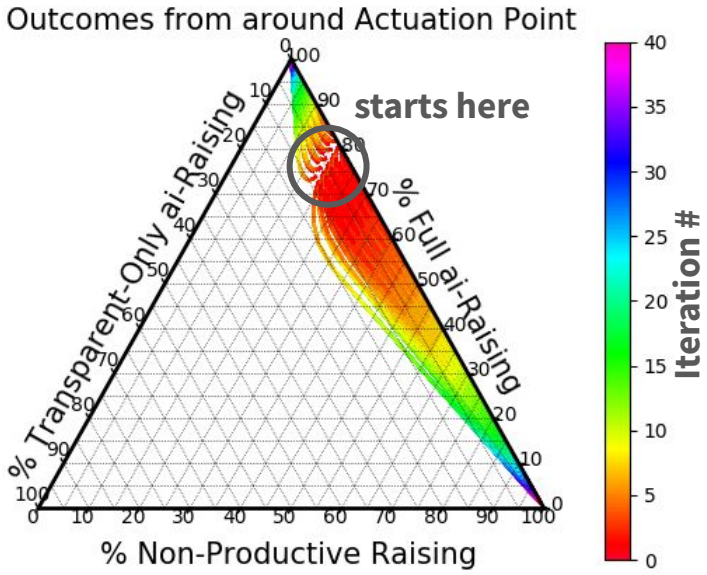
Proportion Learning Transparent /aɪ/-Raising

- **Transparent raising has a narrow band of viability**



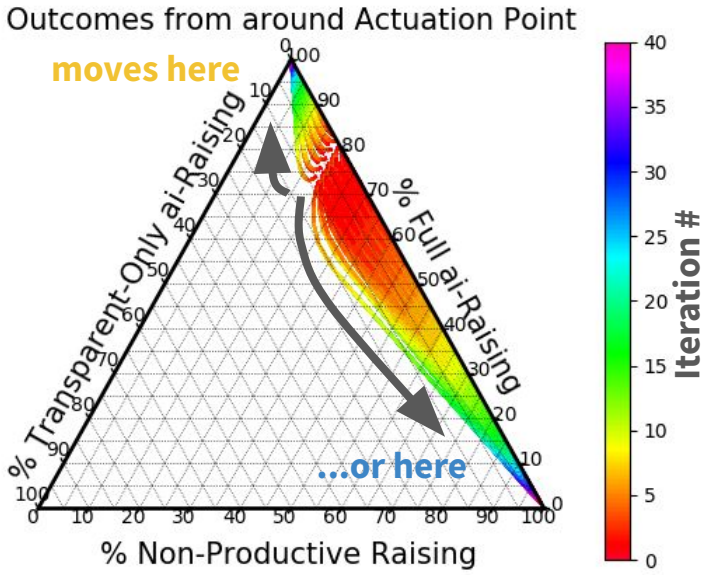
Transparent /aɪ/-Raising over Time

- **Initialize it around 24% non-raising / 76% canonical to give transparent its best shot**
- **Take the output of that calculation and feed it back in to evolve the system**
- **Update 10% of the population each time**



Transparent /aɪ/-Raising over Time

- As it evolves, it rapidly falls off the band of viability depending on the exact starting condition
- Transparent raising dies out and never becomes common



Summary

Transparent /aɪ/-raising as a contact phenomenon

- Language acquisition in a mixed input setting
- Attested at boundaries of the raising region - mixed input is expected
- Crucially the result of receiving multiple inputs - **emergence of a new grammar**
- Transparent raising populations should rapidly transition away

Summary

Transparent /aɪ/-raising as a contact phenomenon

- Language acquisition in a mixed input setting
- Attested at boundaries of the raising region - mixed input is expected
- Crucially the result of receiving multiple inputs - **emergence of a new grammar**
- Transparent raising populations should rapidly transition away

Empirically Verifying the Model

- Our model makes quantitative predictions about the relationship between phonological input and changes in progress
- This renders it falsifiable with empirical investigation
- Suggests where researchers should look to find more transparent raisers

Some Key Points about Studying Acquisition in the Past

Language Change by Language Acquisition

- First language acquisition is one of the primary drivers of language change[1]
- Plays a role in both innovation and propagation

The general idea

- Minor “errors” in acquisition accrue over successive generations
- This eventually yields population-level change, which may be dramatic

Study acquisition as a way of understanding mechanisms of change

[1] Paul 1880, Sweet 1899, Halle 1962, Kiparsky 1965, Andersen 1973, Baron 1977, Lightfoot 1979 *et seq*, Labov 1989, Niyogi 1996 *et seq*, Kroch 2005, Yang 2002 *et seq*, van Gelderen 2011, Cournane 2017, Kodner 2020, *inter multa alia*

Some Principles of Acquisition-Driven Change

- **“Language change” is a collection of phenomena → Not just one explanation**
- **Learning is crucially individual-level. Change is crucially population-level.**
- **Child language acquisition is the primary driver for many types of change**
  Especially, “discrete changes,” “changes to the grammar” *in sensu stricto*
- **Variation is normal as a part of language acquisition, even monolingual**
- **Language acquisition is hard. It is “tractable, not trivial.”**
- **Acquisition is particularly important for innovation**
  → therefore young learners must be important for actuation
- **Learning + sociolinguistics interact → “Sibling-Induced Change”**
- **Incrementation and usage are rightly handled primarily by sociolinguistics**

Some Principles of Acquisition-Driven Change

- **"Language change" is a collection of phenomena → Not just one explanation**
- **Learning is crucially individual-level. Change is crucially population-level.**
- **Child language acquisition is the primary driver for many types of change**
  Especially, "discrete changes," "changes to the grammar" *in sensu stricto*
- **Variation is normal as a part of language acquisition, even monolingual**
- **Language acquisition is hard. It is "tractable, not trivial."**
- **Acquisition is particularly important for innovation**
  → therefore young learners must be important for actuation
- **Learning + sociolinguistics interact → "Sibling-Induced Change"**
- **Incrementation and usage are rightly handled primarily by sociolinguistics**

Some Principles of Acquisition-Driven Change

- **"Language change" is a collection of phenomena → Not just one explanation**
- **Learning is crucially individual-level. Change is crucially population-level.**
- **Child language acquisition is the primary driver for many types of change**
  Especially, "discrete changes," "changes to the grammar" *in sensu stricto*
- **Variation is normal as a part of language acquisition, even monolingual**
- **Language acquisition is hard. It is "tractable, not trivial."**
- **Acquisition is particularly important for innovation**
  → therefore young learners must be important for actuation
- **Learning + sociolinguistics interact → "Sibling-Induced Change"**
- **Incrementation and usage are rightly handled primarily by sociolinguistics**

Some Principles of Acquisition-Driven Change

- **"Language change" is a collection of phenomena → Not just one explanation**
- **Learning is crucially individual-level. Change is crucially population-level.**
- **Child language acquisition is the primary driver for many types of change**
  Especially, "discrete changes," "changes to the grammar" *in sensu stricto*
- **Variation is normal as a part of language acquisition, even monolingual**
- **Language acquisition is hard. It is "tractable, not trivial."**
- **Acquisition is particularly important for innovation**
  → therefore young learners must be important for actuation
- **Learning + sociolinguistics interact → "Sibling-Induced Change"**
- **Incrementation and usage are rightly handled primarily by sociolinguistics**

Some Principles of Acquisition-Driven Change

- **"Language change" is a collection of phenomena → Not just one explanation**
- **Learning is crucially individual-level. Change is crucially population-level.**
- **Child language acquisition is the primary driver for many types of change**
  Especially, "discrete changes," "changes to the grammar" *in sensu stricto*
- **Variation is normal as a part of language acquisition, even monolingual**
- **Language acquisition is hard. It is "tractable, not trivial."**
- **Acquisition is particularly important for innovation**
  → therefore young learners must be important for actuation
- **Learning + sociolinguistics interact → "Sibling-Induced Change"**
- **Incrementation and usage are rightly handled primarily by sociolinguistics**

Some Principles of Acquisition-Driven Change

- **"Language change" is a collection of phenomena → Not just one explanation**
- **Learning is crucially individual-level. Change is crucially population-level.**
- **Child language acquisition is the primary driver for many types of change**
  Especially, "discrete changes," "changes to the grammar" *in sensu stricto*
- **Variation is normal as a part of language acquisition, even monolingual**
- **Language acquisition is hard. It is "tractable, not trivial."**
- **Acquisition is particularly important for innovation**
  → therefore young learners must be important for actuation
- **Learning + sociolinguistics interact → "Sibling-Induced Change"**
- **Incrementation and usage are rightly handled primarily by sociolinguistics**

Some Principles of Acquisition-Driven Change

- **"Language change" is a collection of phenomena → Not just one explanation**
- **Learning is crucially individual-level. Change is crucially population-level.**
- **Child language acquisition is the primary driver for many types of change**
  Especially, "discrete changes," "changes to the grammar" *in sensu stricto*
- **Variation is normal as a part of language acquisition, even monolingual**
- **Language acquisition is hard. It is "tractable, not trivial."**
- **Acquisition is particularly important for innovation**
  → therefore young learners must be important for actuation
- **Learning + sociolinguistics interact → "Sibling-Induced Change"**
- **Incrementation and usage are rightly handled primarily by sociolinguistics**

Criticisms of Acquisition-Driven Change

Comes in three flavors?

Criticisms of Acquisition-Driven Change

Comes in three flavors?

- **Nope. Everything hinges on bilingual contact (cf Meisel 2011)**
  Tantamount to claiming that change requires an external impetus
  Counterpoint: Variation is a spectrum. Bilingualism is special, not *that* special

Criticisms of Acquisition-Driven Change

Comes in three flavors?

- **Nope. Everything hinges on bilingual contact (cf Meisel 2011)**
  Tantamount to claiming that change requires an external impetus
  Counterpoint: Variation is a spectrum. Bilingualism is special, not *that* special
- **Nope. Sociolinguists say young adults drive change (cf Sanford 2014)**
  A focus on what sociolinguists study, which is certain kinds of change
  Counterpoint: Sociolinguistics is important and largely complementary!
  Acq-driven change is not a replacement/threat

Criticisms of Acquisition-Driven Change

Comes in three flavors?

- **Nope. Everything hinges on bilingual contact (cf Meisel 2011)**
  Tantamount to claiming that change requires an external impetus
  Counterpoint: Variation is a spectrum. Bilingualism is special, not *that* special
- **Nope. Sociolinguists say young adults drive change (cf Sanford 2014)**
  A focus on what sociolinguists study, which is certain kinds of change
  Counterpoint: Sociolinguistics is important and largely complementary!
  Acq-driven change is not a replacement/threat
- **Nope. The anti-Chomskyan position? (cf Diessel 2012)**
  Seems to be driven primarily by dogma? I don't know what's happening
  Counterpoint: Too many to list now (good softball question for Q&A 😁)

Actuation[1] and the Paradox of Language Change[2]

If children are so good at acquiring language,
why are they so bad at it?

Helps to have a precise definition of actuation

Actuation = Innovation + uptake into the speech community

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

[1] Definition paraphrased from Labov, Yager & Steiner 1972, [2] term coined by Niyogi & Berwick 1997

Not all Change is Driven by Children

To a 1st-degree approximation, children are responsible for discrete rather than continuous changes

Discrete Changes

- Categorical properties of the grammar
- New or lost structures or constructions
- Virtually fixed over individuals' lifetimes[1]
- The realm of child language acquisition

Continuous Changes

- The stereotypical subjects of variationist sociolinguistics
- Positions in the vowel space, usage frequencies, optionality
- Variable over lifetimes
- Not only child language acquisition

[1] Andersson 1995, Sankoff & Blondeau 2007, Nycz 2013

Discrete and Continuous Changes

Two sides of one coin

- Once a discrete innovation enters the population, it becomes variation[1]
- Variationism concerns [continuous] distribution of discrete choices[2]
- So do competing grammars in historical syntax and morphology[3]
- So the interesting part of discrete changes is closer to **actuation** than **incrementation**[4]

[1] Kroch 2005, [2] Sankoff 1988, [3] Kroch 1994, [4] Weinrich et al 1968 for foundational discussion

Tractable not Trivial Learning

- One cannot acquire language from input alone
- UG renders learning possible in the face of the PoS[1]
- But many language specific patterns must still be acquired from the input[2]

Input is both richer and poorer than typically acknowledged

- Evidenced by the successes and failures of modern NLP[3]
- Zipfian and other long-tailed distributions for all manner of linguistic features

  Most lexical items appear only once even in massive corpora

  Sparsity is consistently worse than our intuitions about sparsity

[1] Chomsky 1959, 1980, [2] eg Baker's Paradox (Baker 1979), [3] eg the successes of distributional semantics vs the failures of coreference

Tractable not Trivial Learning

- One cannot acquire language from input alone
- UG renders learning possible in the face of the PoS[1]
- But many language specific patterns must still be acquired from the input[2]

Abject Poverty

- Cases populations may not converge on a single grammar
  - **Syntax** Interaction of Korean V-raising and negation[3]
  - **Morphology** (non)decomposition of English 'semi-weak' verbs[4]
- Parts of the grammar may go unspecified - paradigmatic gaps[5]

[1] Chomsky 1959, 1980, [2] eg Baker's Paradox (Baker 1979), [3] Han, Litz & Musolino 2007, [4] Guy & Boyd 1990, [5] see Gorman 2019 for summary

Learner Innovation ≠ Learner Error

Innovations need not be due to “errors”

Learner Innovation ≠ Learner Error

Innovations need not be due to “errors”

Errors - “Blame the Child”

- The learner does not act correctly on its input “a buggy algorithm”
- errors presuppose appropriate evidence and an available target

Learner Innovation ≠ Learner Error

Innovations need not be due to “errors”

Errors - “Blame the Child”

- The learner does not act correctly on its input “a buggy algorithm”
- errors presuppose appropriate evidence and an available target

Non-errors - “Blame the Environment”

- The learner acts correctly but is dealt a bad input sample
- Even for a good algorithm, “garbage in, garbage out”
- Change in the face of severely underspecified input or even trivial variation

Transmission is not strictly linear and generational

- Children mature in communities and receive input from multiple speakers
- Community input is formally necessary for attested dynamics of change[1]
- Young children learn sociolinguistic variables[2]
- Children attend to input from older children[3] who are not linguistically mature
- Multiple competing targets may be present in the input

Everybody receives input from multiple grammars

"Monolingual" "Multilingual"

Multi-idiolect multi-dialectal traditional multilingual

[1] Niyogi & Berwick 2009, [2] Labov 1989, Anderson 1990, [3] Manly 1930, Weinreich, Labov & Herzog 1968 p 145, Roberts and Labov 1995, Labov 2001 p449, Nardy, Chevrot & Barbu 2014

Acquisition in the Past

- Children in the past must have acquired language in the same way that modern children do - this is straightforward **uniformitarianism**[1]
- We can reason about acquisition in the past in the same way we do now

[1] Labov 1972 as applied to linguistics, Walkden 2019

Acquisition in the Past

- Children in the past must have acquired language in the same way that modern children do - this is straightforward **uniformitarianism**[1]
- We can reason about acquisition in the past in the same way we do now

But where can we get data about acquisition in the past?

- We can't run experiments on subjects who are no longer alive
  With appropriate caution, we can project experimental results back to the past
- We can't do corpus or modeling work on ancient child-directed speech
  There is none! Overwhelmingly, modern languages don't have CDS either…

[1] Labov 1972 as applied to linguistics, Walkden 2019

Acquisition in the Past

- Children in the past must have acquired language in the same way that modern children do - this is straightforward uniformitarianism[1]
- We can reason about acquisition in the past in the same way we do now

Can non-child-directed speech corpora be substituted for child-directed speech to study the relevant problem?

Yes, for the purposes of lexical acquisition

[1] Labov 1972 as applied to linguistics, Walkden 2019

Taking Estimates from Other Corpora

- Maybe we can estimate child linguistic knowledge from adult and historical corpora when CDS is unavailable
- This is reasonable if CDS and non-CDS are sufficiently similar in respect to relevant linguistic properties

I demonstrate that historical and modern non-CDS are effectively indistinguishable from CDS for the purpose of using them to estimate child linguistic experience

Four Features of First Language Acquisition

1. All children receive unique input yet exhibit gross developmental uniformity[1]
2. The type frequency of a pattern is crucial for acquisition of generalizations, as opposed to token frequency or attestation of specific items[2]
3. Token frequencies correlate with relative order of acquisition[3]
4. Early learner vocabularies are small[4]

As a result,

- Applying a frequency cutoff to lemmas in CDS approximates a “typical” child
- Insight taken by type frequency-based models of acquisition[5]

[1] Labov 1972, [2] Aronoff 1976, MacWhinney 1978, Bybee 1985, Baayen 1993, Elman 1998, Pierrehumbert 2003, Yang 2016, [3] Goodman 2008, [4] Hart & Risley 1995, 2003, Szagun et al. 2006, [5] Nagy & Anderson 1984, Yang 2016

Five Studies

1. Trimming infrequent vocabulary from Mod. English CDS and non-CDS corpora
2. **Morphophonological and syn-sem type freqs across ModE CDS and non-CDS**
3. **Sem. overlap between ModE and Spanish, Latin, and PGmc lexicons**
4. **Morphological sparsity in Modern CDS, adult and historical corpora**
5. **Outcome of learning model applied to Modern English CDS and non-CDS data**

Five Studies

1. Trimming infrequent vocabulary from Mod. English CDS and non-CDS corpora
2. Morphophonological and syn-sem type freqs across ModE CDS and non-CDS
3. Sem. overlap between ModE and Spanish, Latin, and PGmc lexicons
4. Morphological sparsity in Modern CDS, adult and historical corpora
5. Outcome of learning model applied to Modern English CDS and non-CDS data

Study 2 - Data

- **Adult** - Corpus of Contemporary American English (COCA)[1]
- **CDS** - CHILDES[2] Brown, Brent, MacWhinney[3]
- Corpora are POS-tagged and lemmatized. All verb lemmas were extracted.
- Sub-lexicons were made by frequency trimming:
  n = *all*, 1048, 500, 100

Corpus Type	Number	Lexicon size (*n*)
CDS	3	918, 984, 1042
Academic	28	4,917 - 7,786
Fiction	28	5,544 - 8,015
Magazine	28	6,116 - 9,662
News	28	5,080 - 7,365
Spoken	28	4,144 - 5,566

[1] Davies 2009, [2] MacWhinney 2000
[3] Brown 1973, Brent & Siskind 2001, MacWhinney 1991

Study 2 - Type Frequency

- Compares the number of lemmas expressing linguistic properties (ie, type frequencies) across corpora rather than specific lemmas
- Since children develop similar grammars despite surface-unique input

Three Properties

is strong verb	modern reflexes of Germanic Classes I-VII
is DO alternator verb	double object / to-dative alternators[1]
is Latinate verb	polysyllabic Latinate verbs

[1] Levin 1993: §2.1 (119)

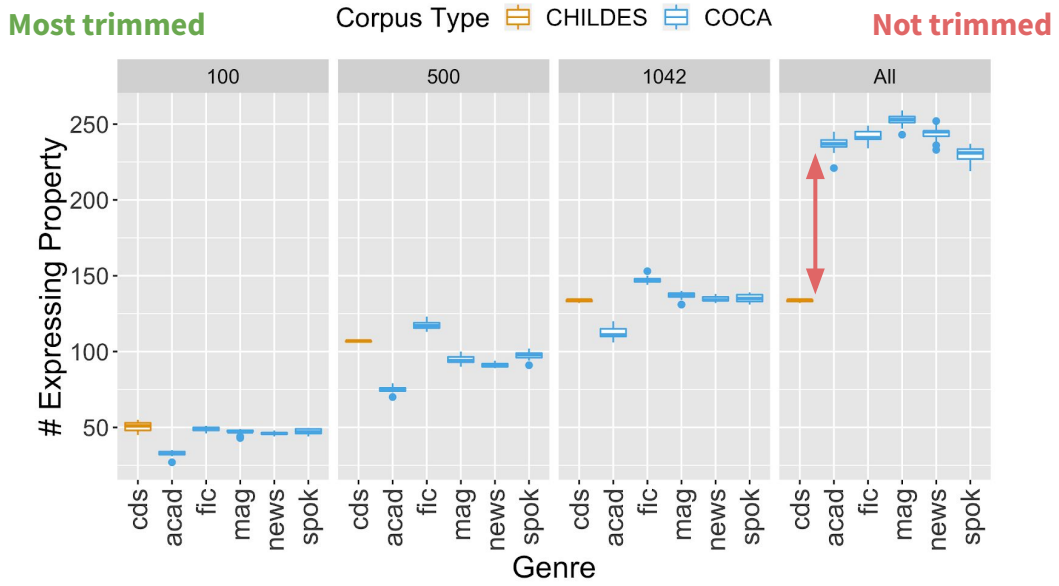
Irregular Verb Type Frequencies

Examples

- *sing ~ sang ~ sung*
- *go ~ went ~ gone*
- *tell ~ told ~ told*

Results

- **In frequency-trimmed conditions, non-CDS falls in line with CDS**
- **Academic remains an outlier**



DO Alternator Verb Type Frequencies

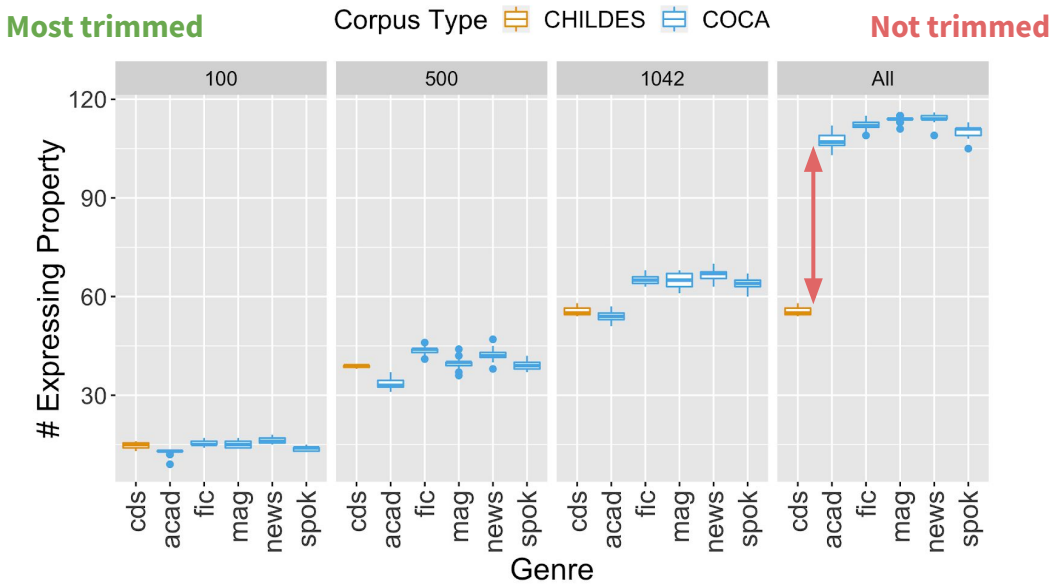
Example

Alice gave Bob the message

Alice gave the message to Bob

Results

- In frequency-trimmed conditions, non-CDS falls in line with CDS
- Academic remains an outlier



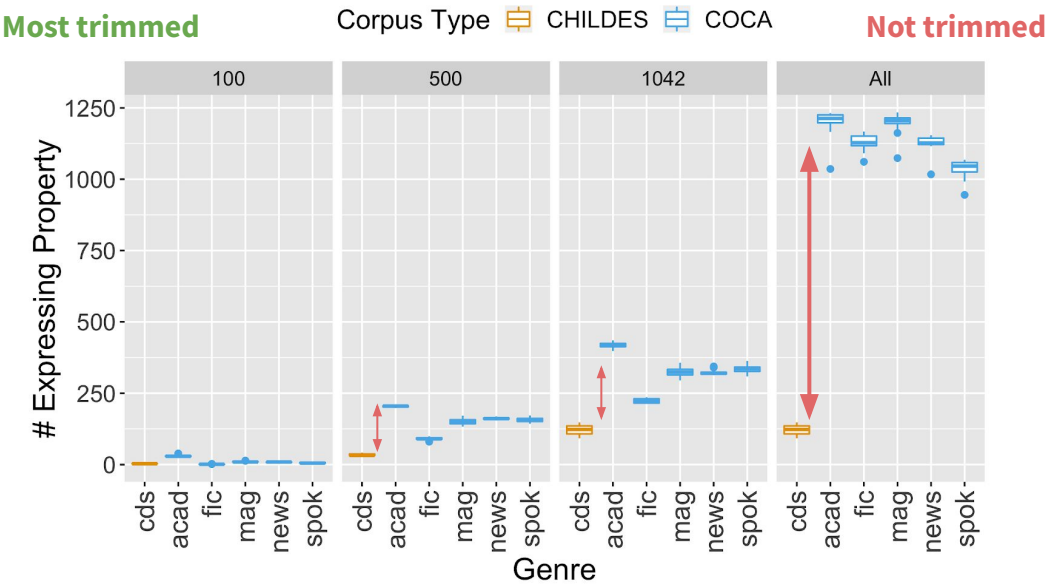
Latinate Verb Type Frequencies

Examples

- Both high- and lower-register
- *encapsulate*, *irradiate*...
- *confuse*, *offer*, *remember*...

Results

- In frequency-trimmed conditions, non-CDS comes much closer to CDS
- Differences consistent with salient genre effects for Latinate vocabulary[1]
- Academic is still the outlier...



[1] Levin et al 1981, Levin & Novak 1991

Study 2 - Results

- **Linguistic type frequencies in CDS and adult genres are quantitatively similar and often statistically the same**
  When infrequent vocabulary is trimmed
  When the relevant feature does not saliently vary by genre
- **Superficial differences in lemma overlap are misleading**
  Suggests that superficial descriptions may not reflect more relevant patterns
  Suggests a reason why children acquire similar language-specific patterns

Non-CDS can be reasonably substituted for CDS when estimating type frequencies of these linguistic properties

Five Studies

1. Trimming infrequent vocabulary from Mod. English CDS and non-CDS corpora
2. Morphophonological and syn-sem type freqs across ModE CDS and non-CDS
3. Sem. overlap between ModE and Spanish, Latin, and PGmc lexicons
4. Morphological sparsity in Modern CDS, adult and historical corpora
5. Outcome of learning model applied to Modern English CDS and non-CDS data

Study 3 - Cross-Language

- **Lexical overlap cannot be computed automatically when vocabularies differ**
- **But lexicons can be compared for translations or items with similar meaning**
- **I perform this comparison manually on languages with which I am familiar**
- **Raw Lexical Overlap. They correlate well when the corpora are similarly sized, but RLO scores are systematically higher.**

For corpus-derived lexicons A and B
where A and B are unordered sets,

$$similarity = |A \cap B| / \min(|A|, |B|)$$

Study 3 - Corpora

- **English CDS** - verb lemmas in CHILDES Brown (and Brent for comparison)
- **Spanish CDS** - verb lemmas in combined CHILDES FernAguado, Hess, OreaPine, Remedi, Romero, SerraSole
- **Proto-Germanic** - securely reconstructable strong verbs from Seebold 1979[1]
- **Classical Latin** - verb lemmas in all Perseus online 3rd BC - 2nd AD (inclusive)
- Frequency cutoffs were employed to bring the others in line with PGmc
- PGmc strong verbs do not form a semantically coherent class

Corpus	Freq Cutoff	Lexicon size (*n*)
English CDS Brown	< 17	260
English CDS Brent	< 17	257
Spanish CDS	< 11	263
Proto-Germanic	-	258
Latin	< 666	260

[1] Credit to Don Ringe for extracting them

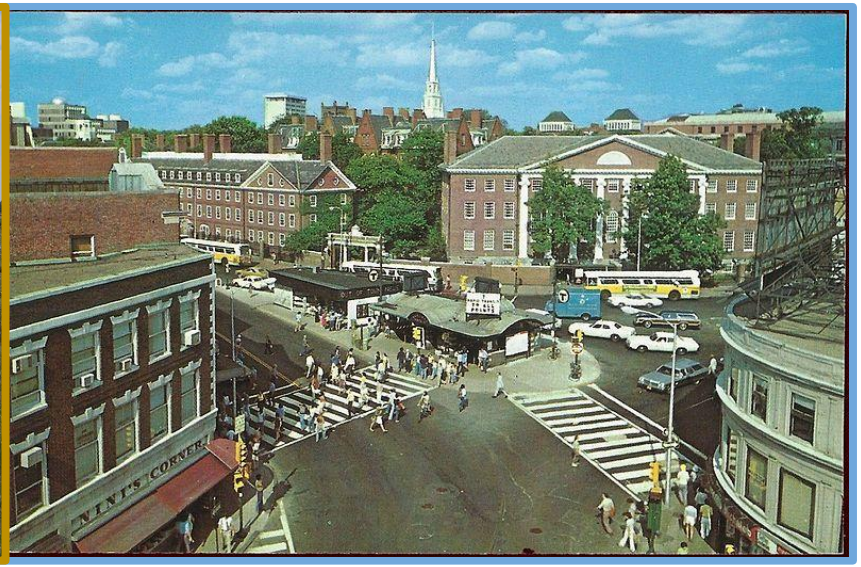
Study 3 - Comparisons

- Baselines: English-English (within-language) English-Spanish (cross-language)
- PGmc comparisons are just a few points lower than English-Spanish

  The kind of terms which are reconstructable are frequent everyday vocabulary which are preserved in daughter languages -
  the same kind likely to be present in CDS
- Latin comparisons are higher than English-Spanish
- Why are the PGmc overlaps not higher?

Comparison	% Overlap
English - EN Brent	81.71%
English - Spanish	73.07%
English - PGmc	66.67%
Spanish - PGmc	71.32%
English - Latin	75.77%
Spanish - Latin	78.62%

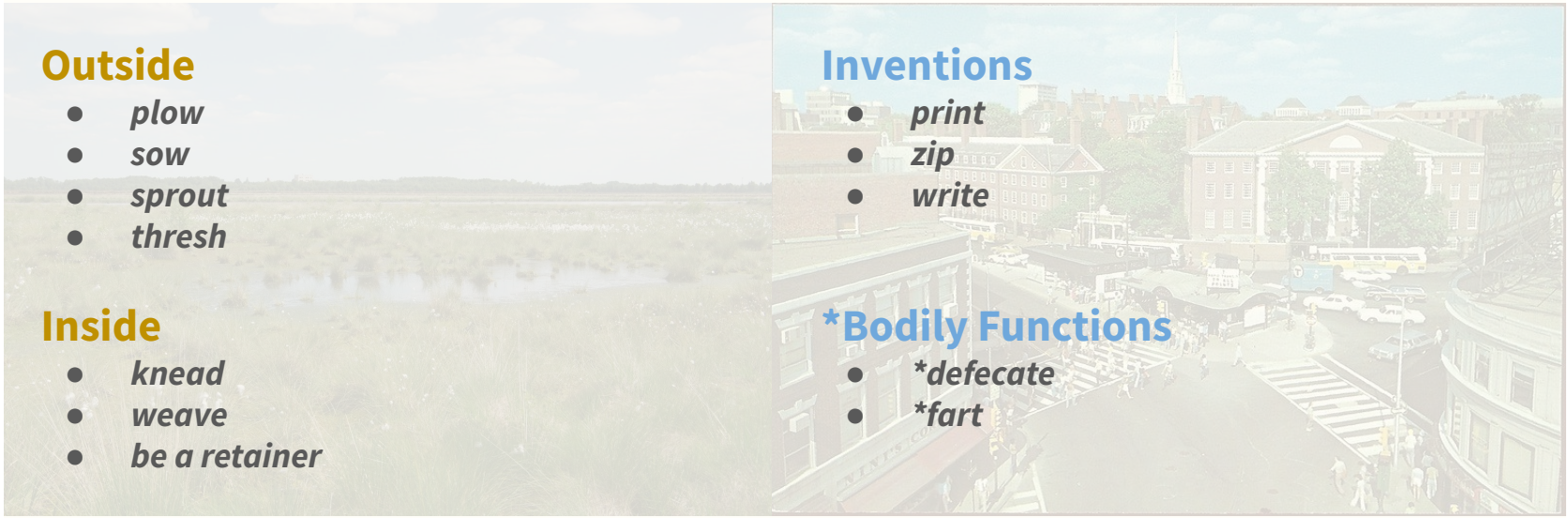
Study 3 - The Proto-Germanic Homeland



***Germanic Urheimat, 1st Millenium BC**
(photo of a Danish peat bog)

Cambridge, MA, c. 1970
(Harvard when Brown was collected)

Study 3 - The Proto-Germanic Homeland



Outside

- *plow*
- *sow*
- *sprout*
- *thresh*

Inside

- *knead*
- *weave*
- *be a retainer*

Inventions

- *print*
- *zip*
- *write*

***Bodily Functions**

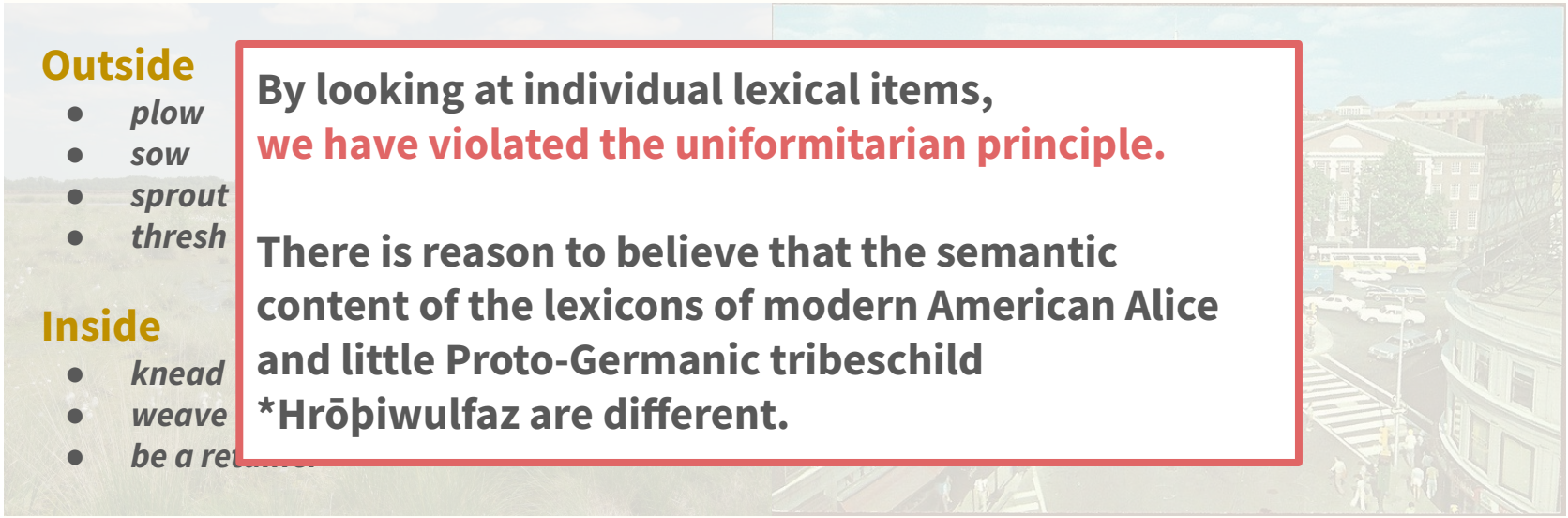
- **defecate*
- **fart*

***Germanic Urheimat, 1st Millenium BC**

Cambridge, MA, c. 1970

Cultural differences between the Iron Age and Atomic Age

Study 3 - The Proto-Germanic Homeland



Outside

- *plow*
- *sow*
- *sprout*
- *thresh*

Inside

- *knead*
- *weave*
- *be a ret…*

By looking at individual lexical items,
we have violated the uniformitarian principle.

There is reason to believe that the semantic content of the lexicons of modern American Alice and little Proto-Germanic tribeschild *Hrōþiwulfaz are different.

*Germanic Urheimat, 1st Millenium BC

Cambridge, MA, c. 1970

Cultural differences between the Iron Age and Atomic Age

Five Studies

1. Trimming infrequent vocabulary from Mod. English CDS and non-CDS corpora
2. Morphophonological and syn-sem type freqs across ModE CDS and non-CDS
3. Sem. overlap between ModE and Spanish, Latin, and PGmc lexicons
4. Morphological sparsity in Modern CDS, adult and historical corpora
5. Outcome of learning model applied to Modern English CDS and non-CDS data

Study 4 - Paradigm Saturation

Paradigm Saturation[1]

Proportion of a word's possible forms actually attested in a corpus

If the size of a language's {noun, verb…} paradigm is *N*
And a {noun, verb} *w* is attested in some corpus in *n* forms, PS(*w*) = *n*/*N*

[1] Chan 2008

Paradigm Saturation Data

- All POS-tagged, lemmatized, morpho feature annotated
- CDS - English (Brown), Spanish and German (CDS Leo[1])
- Modern - UD[2] English, Finnish, German, Spanish, Turkish
- Historical - UD Gothic, Latin
- Order 10^5 verb tokens
- CDS token/type ratios are on the order of 10x higher

Corpus	Lang	# V Tokens	# V Types	Ratio
CDS	English	94,768	916	103.46
CDS	Spanish	96,686	879	110.00
CDS	German	81,351	641	126.91
Modern	English	53,796	3,225	16.67
Modern	Spanish	85,861	5,019	17.11
Modern	German	21,835	2,826	7.73
Modern	Finnish	63,891	3,476	18.38
Modern	Turkish	12,064	968	12.46
Historic	Gothic	12,749	1,172	10.88
Historic	Latin	99,066	2,2833	34.97

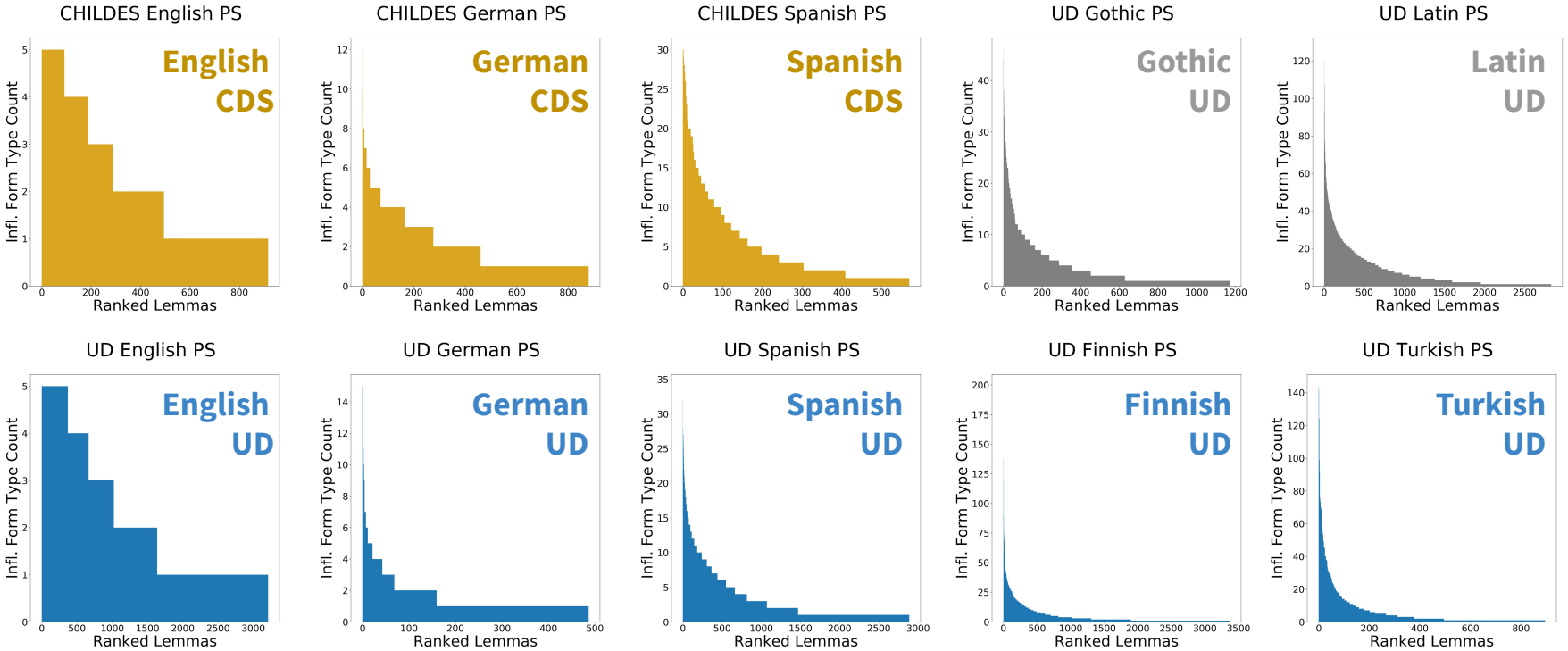
[1] Behrens 2006, [2] Nivre et al 2018

Paradigm Saturations

- **CDS saturations only slightly higher than modern equivs**
- **Despite difference in token/type ratios**
- **Historical corpora similar to modern ones**
- **Saturation appears related to paradigm size if anything**

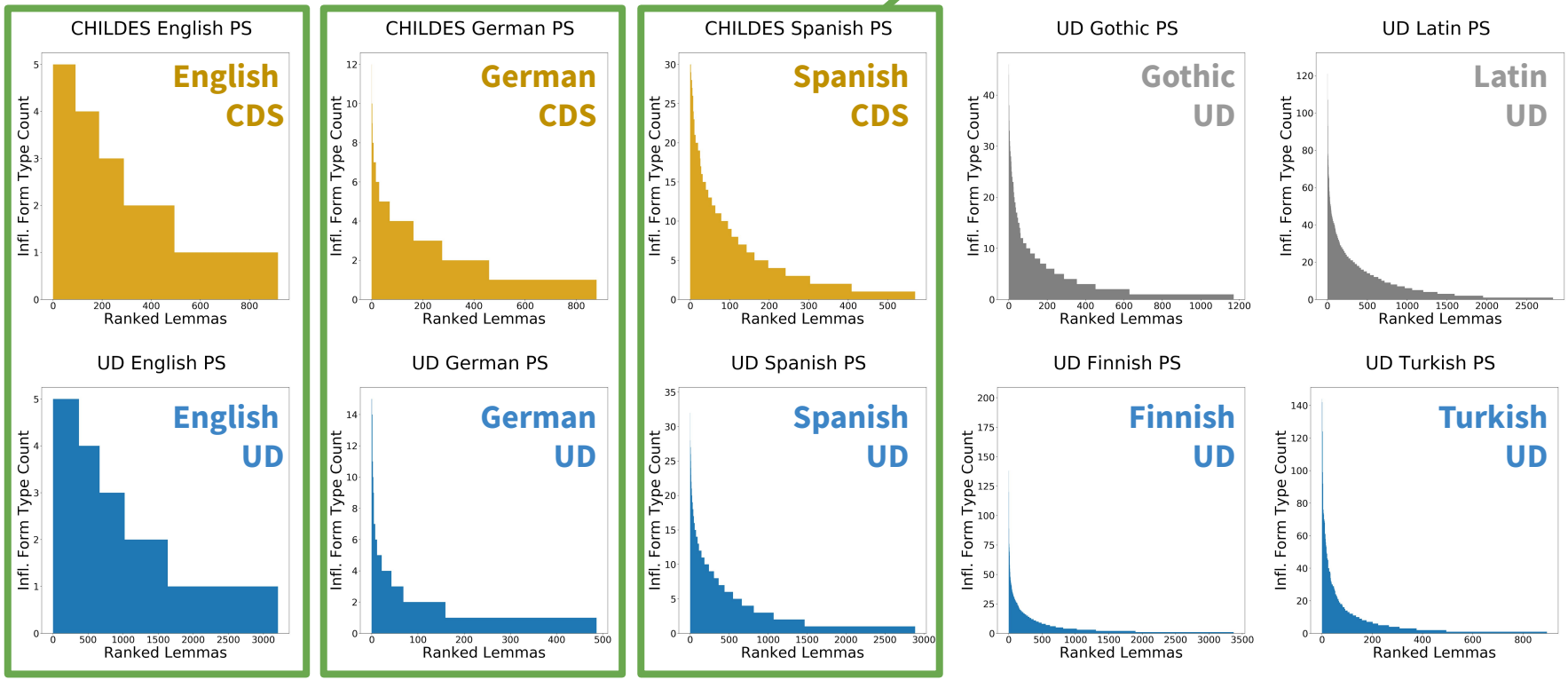
Corpus	Lang	|Paradigm|	Max Sat.	Mean Sat.	Med Sat.
CDS	English	5	100%	43.23%	40.00%
CDS	Spanish	29	44.83%	7.59%	6.90%
CDS	German	67	52.24%	8.31%	4.48%
Modern	English	5	100%	42.80%	40.00%
Modern	Spanish	67	43.28%	4.91%	1.49%
Modern	German	29	51.72%	5.83%	3.45%
Modern	Finnish	150	27.33%	2.46%	1.33%
Modern	Turkish	120	99.17%	4.83%	1.67%
Historic	Gothic	52	53.85%	6.31%	3.85%
Historic	Latin	113	81.42%	5.90%	2.65%

Zipfian Distributions



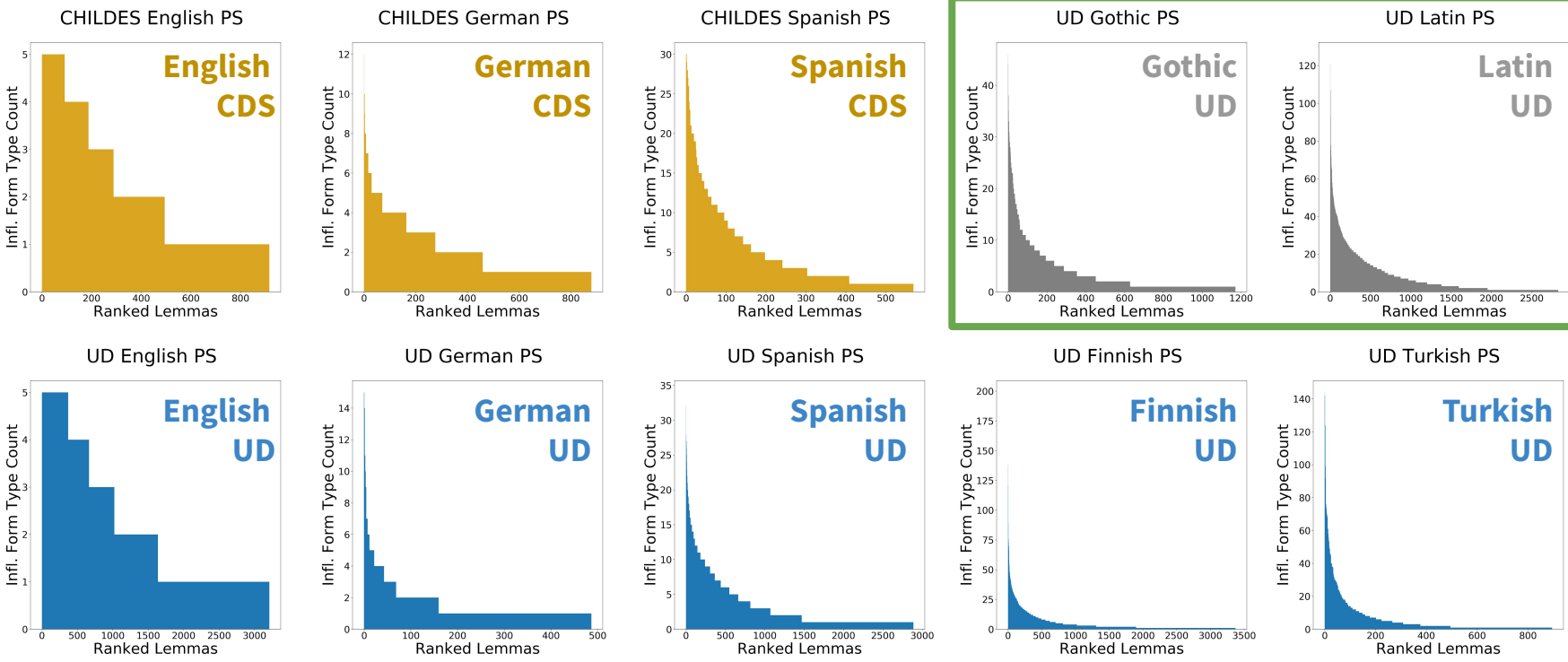
Zipfian Distributions

CDS and UD distributions correspond by language

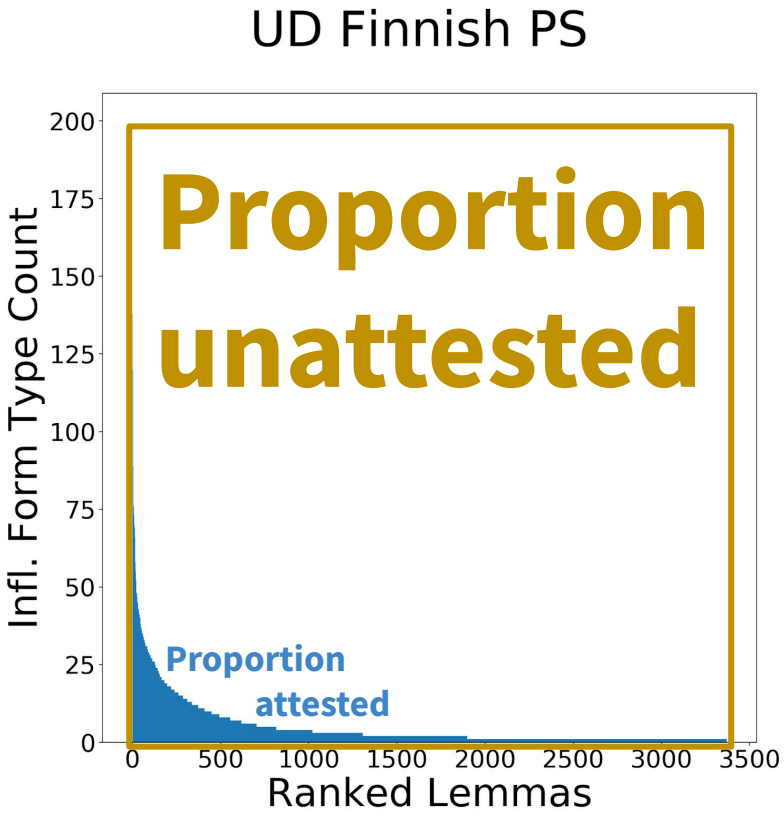


Zipfian Distributions

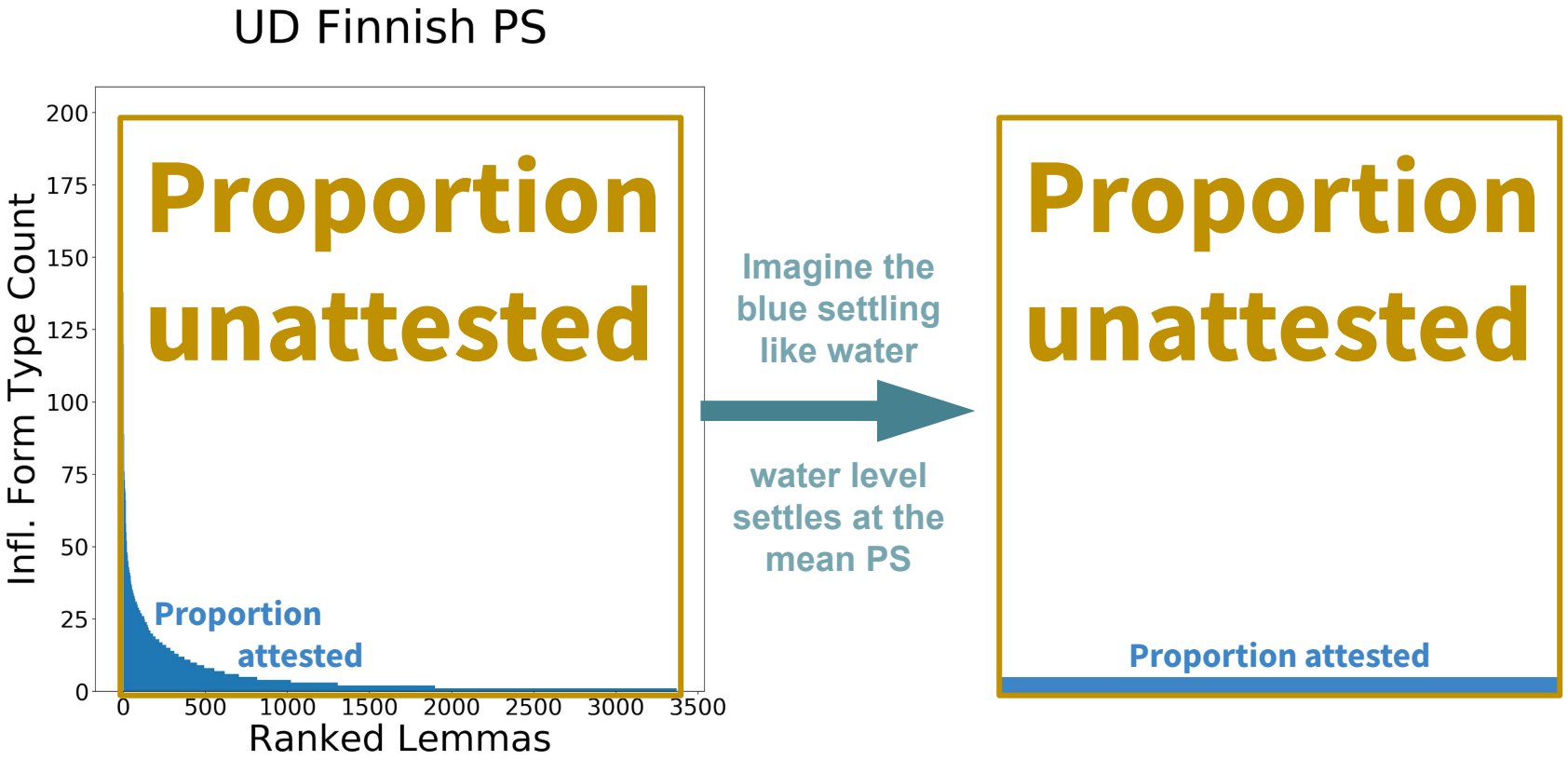
Historical corpora behave just like any other in this respect



A different way to read these plots



A different way to read these plots



Five Studies

1. Trimming infrequent vocabulary from Mod. English CDS and non-CDS corpora
2. Morphophonological and syn-sem type freqs across ModE CDS and non-CDS
3. Sem. overlap between ModE and Spanish, Latin, and PGmc lexicons
4. Morphological sparsity in Modern CDS, adult and historical corpora
5. Outcome of learning model applied to Modern English CDS and non-CDS data

Study 5 - Deploying a Learning Model

- A learning algorithm applied to high frequency items should yield similar trajectories and final outcomes regardless of genre
- I apply the Tolerance Principle[1] to two problems of linguistic generalization

Acquisition Problems

1. Default past inflection in English

[1] Yang 2016

Study 5 - Deploying a Learning Model

- A learning algorithm applied to high frequency items should yield similar trajectories and final outcomes regardless of genre
- I apply the **Tolerance Principle**[1] to two problems of linguistic generalization

Acquisition Problems

1. Default past inflection in English
2. **Subset of past inflection in Old vs. Modern Icelandic**
   - pre-1400 vs. post-1900 in IcePaHC Corpus[2]
   - Can strong inflection be applied productively to monosyllabic verbs?
   - eg, should they be inflected like English *see* ~ *saw* (*sé* ~ *sá*)
     or suffixed like English *flay* ~ *flay-ed* (*flá* ~ *flá-ði*)?

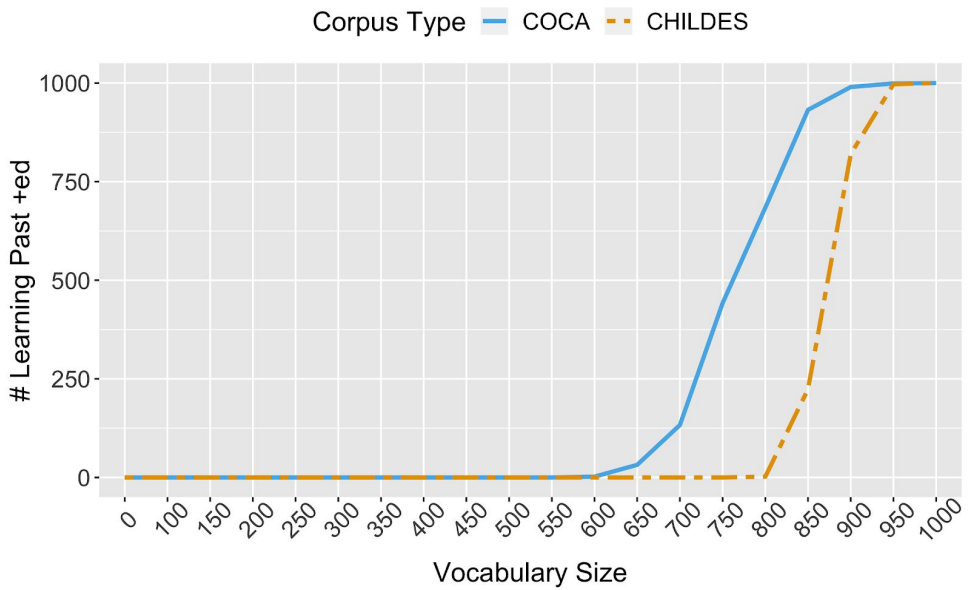
[1] Yang 2016, [2] Wallenberg et al. 2011

Procedure

- **Sample 1,000 lexicons of size *N* by frequency weight from each corpus**
- **Compute the learning outcome over each lexicon at size *N***
- **Plot learning trajectories as average learning outcomes as *N* increases**

English *-ed*

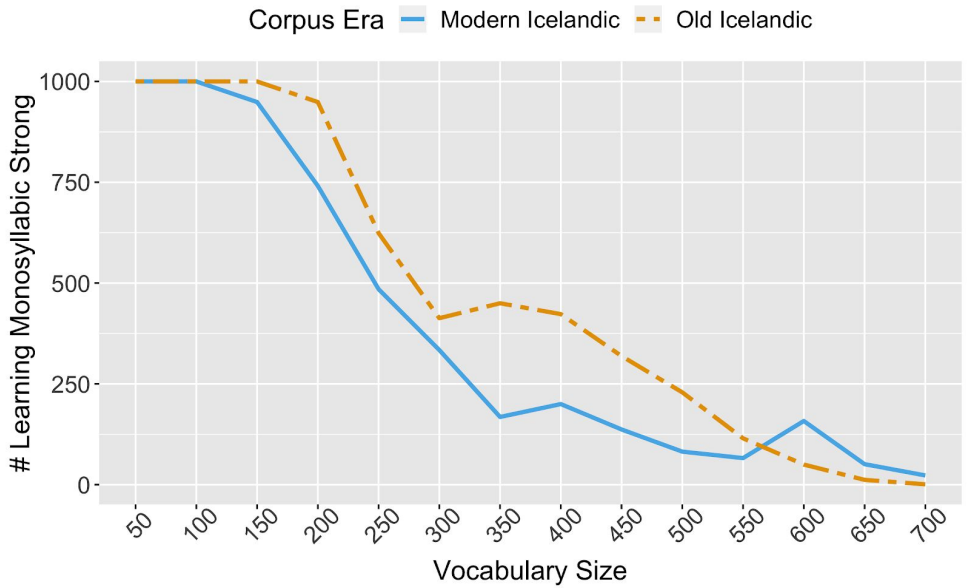
- All 1,000 lexicons yield same final learning outcome
- Follow similar trajectories
- COCA is slightly shifted
- Reasonable observed relative developmental variation[1]



[1] Marcus et al 1992 observing the Brown Corpus children, Maratsos 2000, Yang 2002

Icelandic Monosyllabic Stem Mutation

- Nearly all lexicons yield same final learning outcome
- Follow similar trajectories



As expected, similar type frequencies in the data correspond to similar learning outcomes

Conclusions

Though CDS-derived and non-CDS derived lexicons differ in terms of exact lexical makeup and other superficial corpus stats

- They are quantitatively similar or indistinguishable over linguistic dimensions
- When frequency-trimmed to approximate learner vocabulary sizes

With appropriate pre-processing, historical and modern adult-derived corpora may be reasonably used to approximate child linguistic experience

A Learners' Perspective on the Rise of the *to*-Dative

The Dative Constructions in Modern English

- A classic syntax-semantics mapping problem
- Ditransitives with recipient/goal indirect objects

Double Object

- Alice gave Bob the book
- Alice told Bob a story

to-Dative

- Alice gave the book to Bob
- Alice told a story to Bob

The Dative Constructions in Modern English

- A classic syntax-semantics mapping problem
- Ditransitives with recipient/goal indirect objects

Double Object

- Alice gave Bob the book
- Alice told Bob a story

But...

- *Alice donated Bob the book
- Alice guaranteed Bob the win

to-Dative

- Alice gave the book to Bob
- Alice told a story to Bob

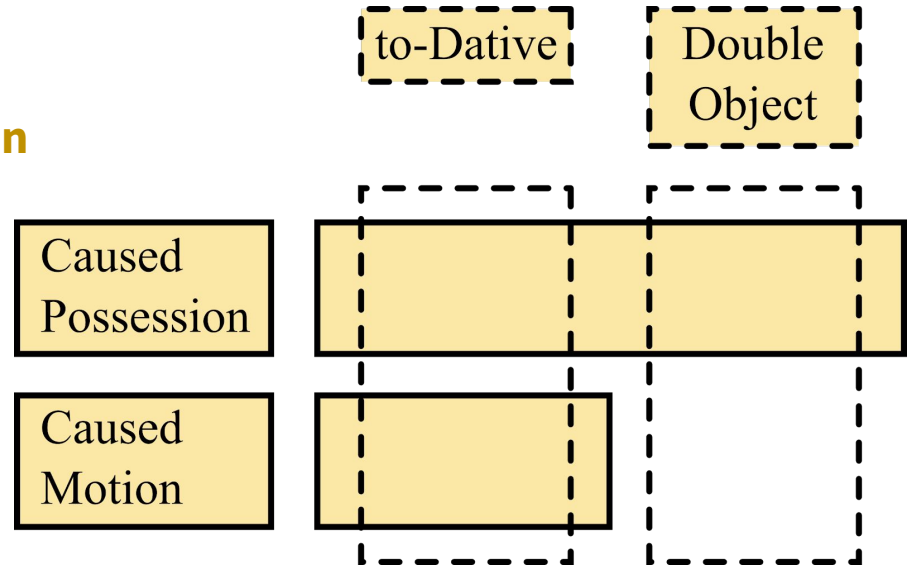
- Alice donated the book to Bob
- ? Alice guaranteed the win to Bob

Broad-Range Semantic Classes

- Semantics provides restrictions on the constructions' distributions
- A verb-sensitive approach[1]

The constructions require

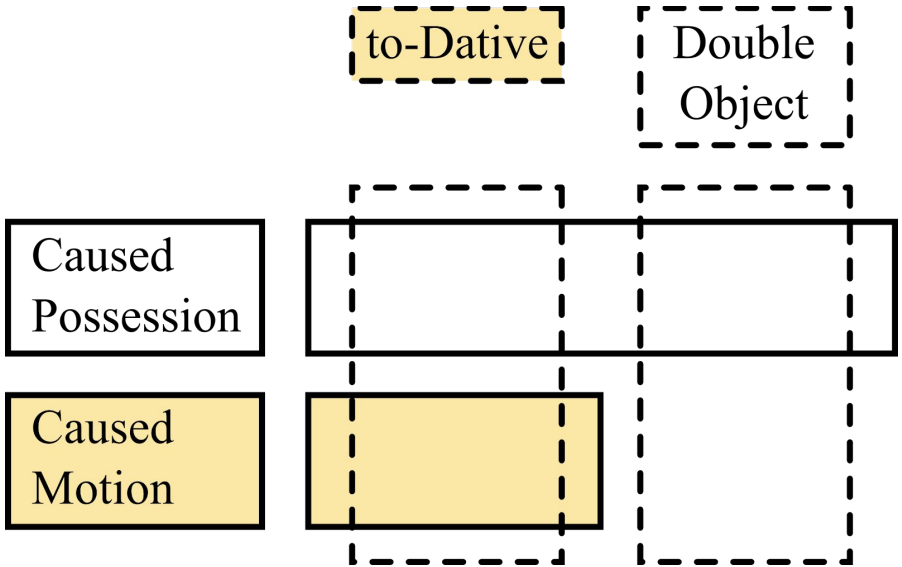
- Verbs expressing caused possession or caused motion meanings
  - Double objects require CP
  - to-Datives require CP or CM



[1] Levin 2008, Rappaport Hovav & Levin 2008, other approaches: Pinker 1989, Goldberg 1995

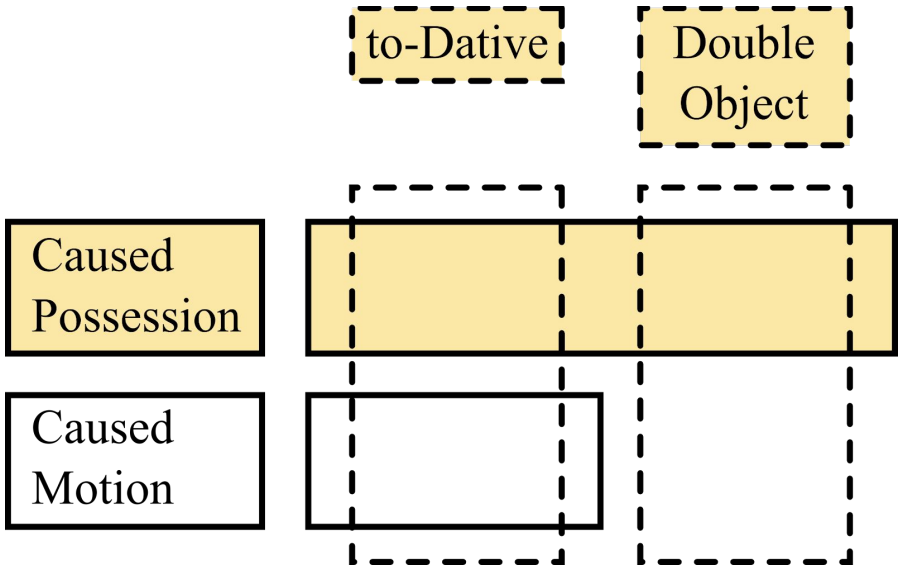
Caused Motion Verbs

- Direct physical transfer or abstract transfer (such as of messages)
- ***Say*-type** verbs is caused motion-only so it is to-dative only



Caused Possession Verbs

- Some such as *give*-type verbs, lack a path argument
- Others may be either such as *send*-type (eg, throw)

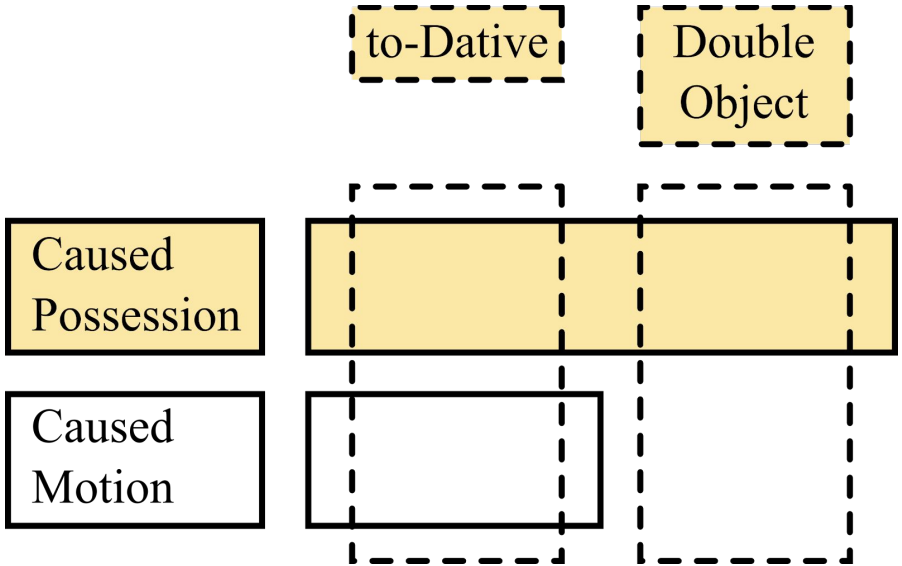


Caused Possession Verbs

- Some such as *give*-type verbs, lack a path argument
- Others may be either such as *send*-type (eg, throw)

Identified by diagnostics[1]

- * Where did you give the ball?
- Where did you throw the ball?
- I gave the package to Maria/*London
- I sent the package to Maria/London
- * Susan gave the ball all the way/halfway to Bill
- Jake threw the ball all the way/halfway to Bill
- and others...



[1] Rappaport Hovav & Levin 2008, emphasis mine

Broad-Range Classes are not Sufficient

Not all caused possession verbs can support both constructions

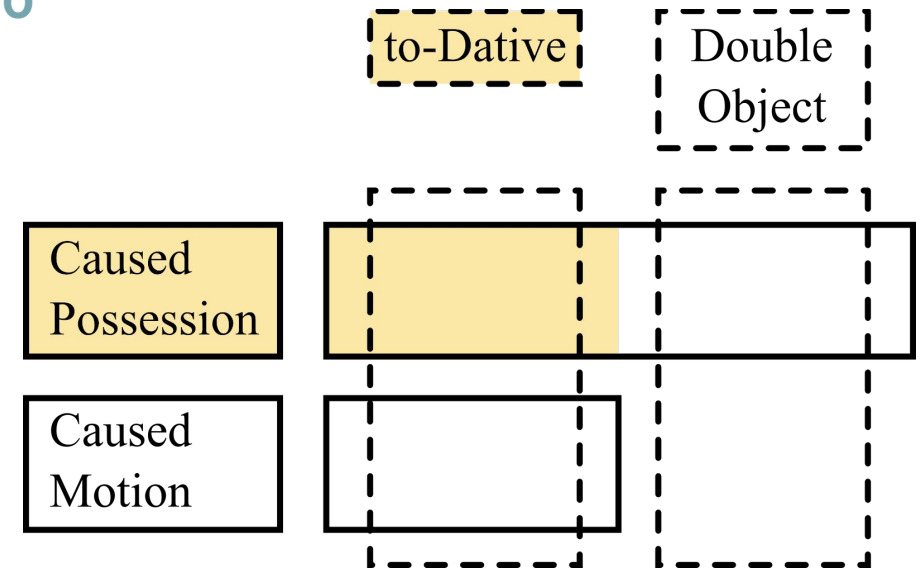
Famously, Latinate CP verbs prohibit DO[1]

John told/*reported Bill the news

Kate showed/*demonstrated Alan the technique

But these exceptions have exceptions[2]

Either	advance, refund, extend, etc
DO-only	nominate, refuse, suppose, etc



[1]Storm 1977, Pinker 1989, Gropen et al 1989, [2]Levin 1993

The Constructions and their Distributions

There are three factors at play

- The grammar behind the constructions
- How they are acquired
- The history of the language

An acquisition-driven diachronic account connects the three

The Constructions over Time

- **The double object is attested throughout Old English**
- **The to-dative arose during the Middle English Period**
  It was actuated in Early Middle English at the latest, could be much older
  It rapidly rose in token frequency
  It rapidly expanded in its semantic range (type frequency)

A historical account of the *to*-dative should cover its innovation and spread through the lexicon

Old English[1]

- The double object was symmetric (IO-DO and DO-IO both licit)
- There was (probably) no to-dative[2]
- There was an overt dative-accusative (DAT-ACC) distinction

DO-IO (* in Modern English)[3]

… þæt he forgeafe godne willan þam seocan hæðenan

… that he would grant good will.ACC the sick heathen.DAT

IO-DO (*ok* in Modern English)

… gif þu geoffrast Gode ænige lac æt his weofode.

… if you offer God.DAT any sacrifice.ACC at his altar

[1] Visser 1963, Mitchell 1985, Allen 1995, [2] Mitchell 1985, [3] Polo 2002

Old English[1]

- The double object was symmetric (IO-DO and DO-IO both licit)
- There was (probably) no to-dative[2]
- There was an overt dative-accusative (DAT-ACC) distinction
- Dative and accusative mark IO and DO respectively for most verbs
- But there were plenty of exceptions[3]
  - Genitive and Dative themes
  - Accusative recipients
  - Optionality

Example	Theme	Recipient
giefan 'give'	ACC	DAT
forwyrnan 'forbid'	GEN	DAT
bereafian 'deprive'	GEN	ACC
læran 'teach'	ACC	ACC
bereafian 'deprive'	DAT	ACC

[1] Visser 1963, Mitchell 1985, Allen 1995
[2] Mitchell 1985
[3] Allen 1995 pg 29

Old English *to*-ditransitives

- ***to*** **could indicate goals[1]: *bringan*, *niman* ‘take,’ *lætan* ‘permit,’ *sendan*…**
- **Including abstract goals: *secgan* ‘say, speak,’ *cweþan* ‘speak, name, declare,’ *sprecan* ‘speak,’ *cleopian* ‘cry, call’…**
- **Similar pattern to Old Norse and Icelandic**

Old English "*to*-datives"

- ***to*** **is attested a few times with goals which are plausible recipients[1]**
  eg '*agifan to a monastery*,' '*(ge)sellan to a church*'
- **And dubiously a couple examples with human goal-like recipients**
- **There is some disagreement about whether these count as *to*-datives[2]**

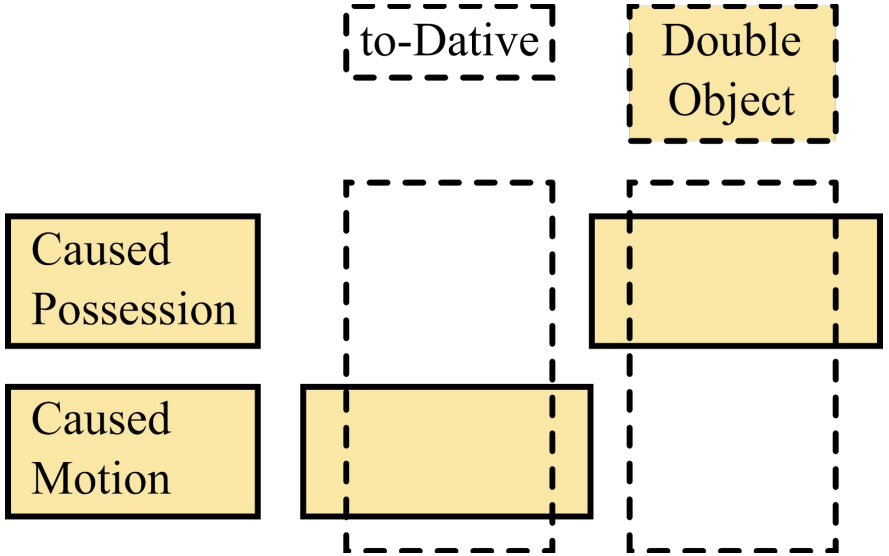
Icelandic has these too, but they typically aren't described as *to*-datievs[3]
Recipient reading of *til* 'to' illicit except with metonymy reading

Ég gaf bækurnar til Háskólabókasafnsins
'I gave the books to the University Library'

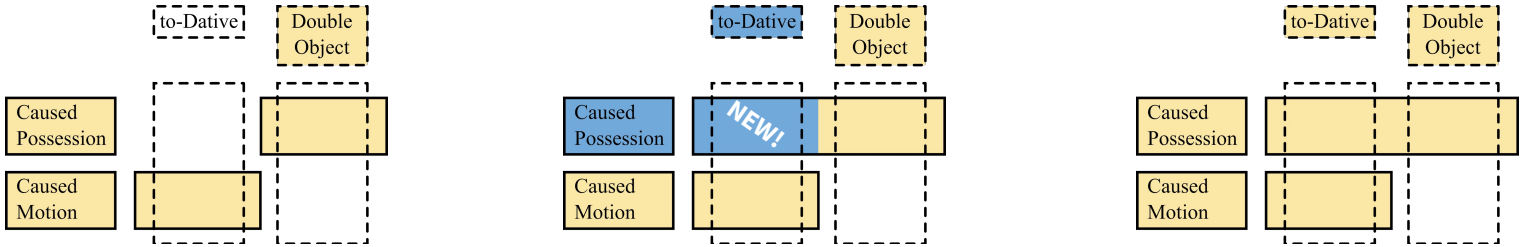
[1] Mitchell 1985, [2] Mitchell 1985 argues *no*, Visser 1963 argues *yes*, De Cuypere 2015 emphatically argues *yes*, but includes "say to"-type examples

Old English Broad-Range Alignment

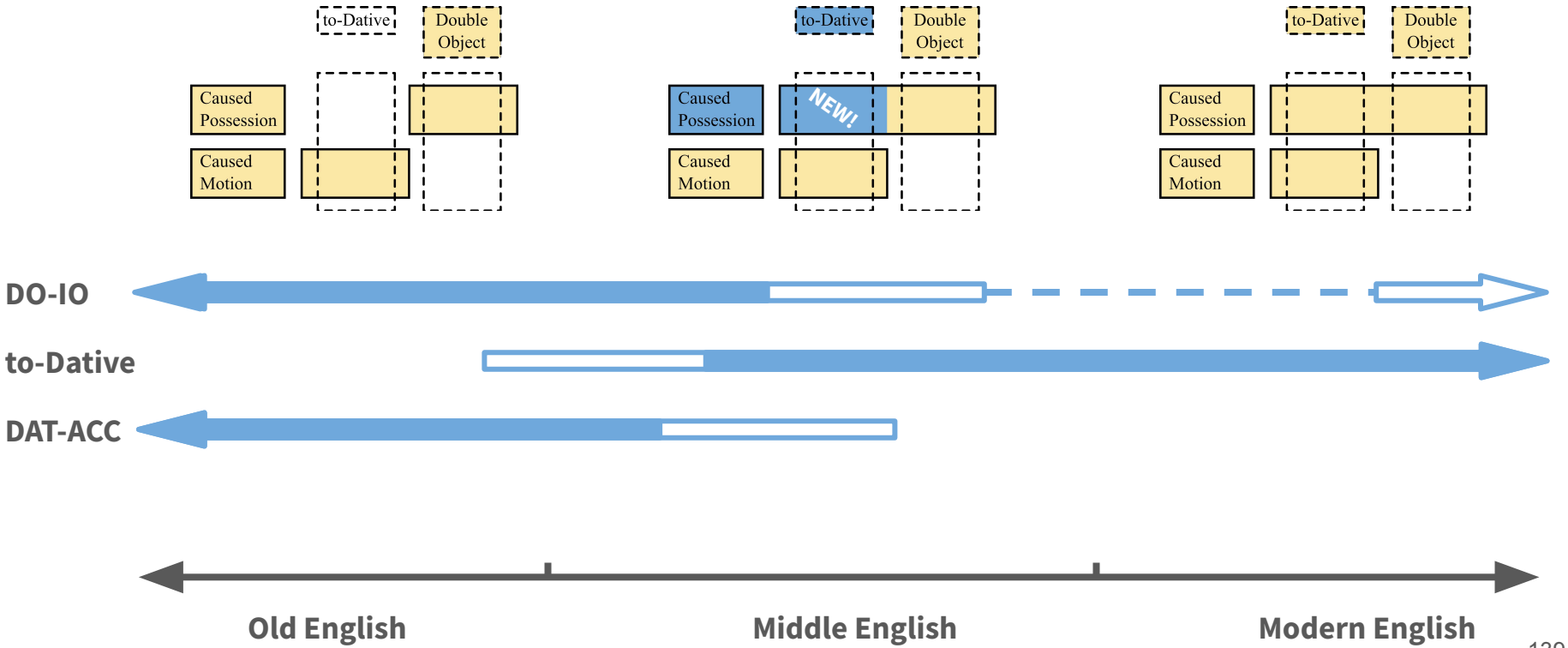
- **Caused possession did not allow for the prepositional construction**
- **ie, no to-dative**



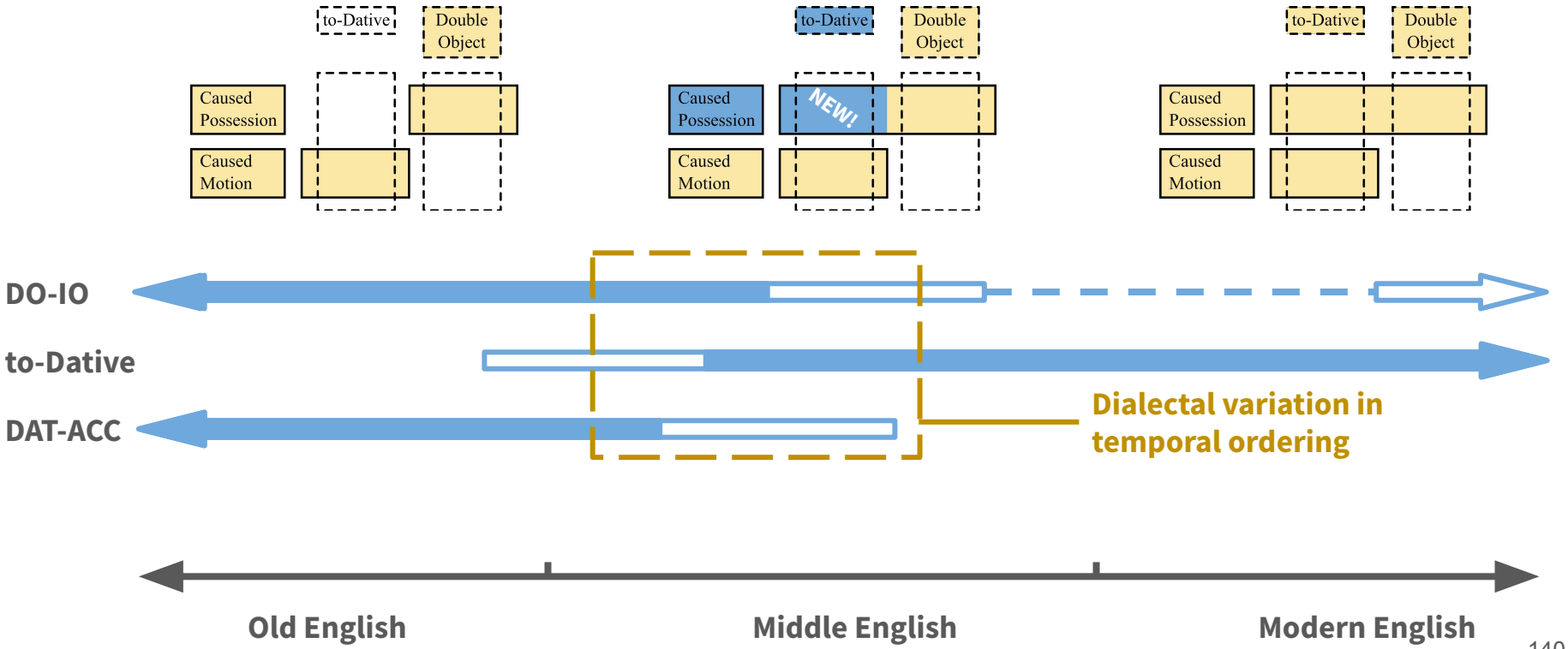
Timeline of the English to-Dative



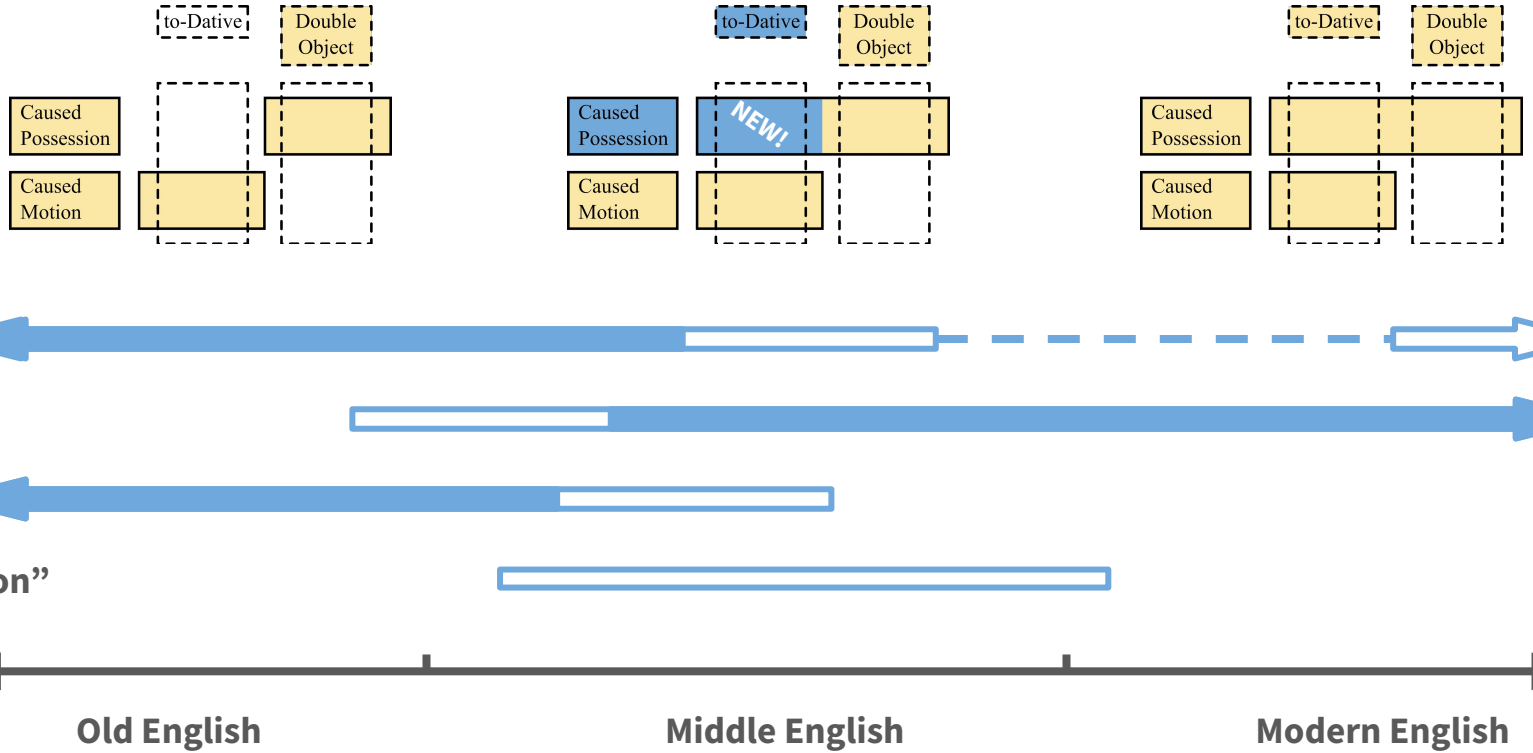
Timeline of the English to-Dative



Timeline of the English to-Dative



Timeline of the English to-Dative



Accounting for the Middle English to-Dative

Four Classes of Hypothesis

- **Borrowing**[1] - the *to*-dative entered and spread from Old French
- **Morphological erosion** - the to-dative replaced DO-IO in response to ambiguity introduced by the loss of a DAT-ACC distinction
- **Semantic expansion** - gradual expansion of scope from the attested OE to-dative-like *agiefan to* constructions
- **Learner Overgeneralization** - a common but usually transient phase of the acquisition process happened in the right time and place to gain a foothold

I find evidence in support of learner overgeneralization initiated following individuals' semantic expansion and against morphological erosion

[1] Stein & Trips 2008, Elter 2018

Change Following Morphological Erosion

Strong Hypothesis

When overt case marking was lost, DO-IO becomes ineffable or otherwise problematic because of ambiguity. Overtly marking the goal/recipient with *to* fixed this.[1]

- **Consistent with the observed trade-off between syntactic and morphological complexity**
- **Not dependent on a specific theory of Case[2]**
- **Can be integrated into a competing grammars account[3]**
- **Essentially functional in nature**

[1] Allen 1995, Polo 2002, McFadden 2002, [2] Contrast Allen 1995 and Bacovcin 2017, [3] McFadden 2002

Predictions of the Strong Account

If morphological erosion were the primary driver of this change,

1. **The *to*-dative should replace DO-IO around the time that overt DAT-ACC is lost**
2. **The DO-IO double object should be rare when DAT-ACC is lost**
3. **The *to*-dative should be rare where overt DAT-ACC is maintained**

Prediction #1

The *to*-dative should replace DO-IO around the time that overt DAT-ACC is lost

- **Should hold if overt syntactic marking of the indirect object (to) is innovated as a synchronic repair to the grammar, but**
- **The overt DAT-ACC was lost on nouns well before DO-IO was lost in the SE Midlands**
- **And if it did exist in OE,[1] the temporal disconnect is even more serious**

The temporal correlation between the loss of DAT-ACC on pronouns and DO-IO is closer, so perhaps pronouns provided sufficient evidence to learn DO-IO?[1]
This turns out to be problematic…

[1] De Cuypere 2015, [2] Polo 2002

Prediction #2

The DO-IO double object should be rare when DAT-ACC is lost

- It should be absent, modulo competing grammars, but
- Swedish, which also lost DAT-ACC, retains it lexically with a few particle verbs[1]

DO-IO

Stevie Wonder tillägnade	*konserten*	*sin*	*hustru*
Stevie Wonder dedicated	consert.DEF	his	wife

IO-DO

Stevie Wonder tillägnade	*sin*	*hustru*	*konserten*
Stevie Wonder dedicated	his	wife	consert.DEF

'*Stevie Wonder dedicated the concert to his wife.*'

[1] Lundquist 2014, Garbacz 2010 fn. 85 *tillskriva* 'ascribe'

Prediction #2

The DO-IO double object should be rare when DAT-ACC is lost

- It should be absent, modulo competing grammars, but
- Modern Liverpool English has surface DO-IO[1]
- Much of the English North and Midlands have pronominal DO-IO[2]

DO-IO

Mary gave the book the teacher

Mary sent the package her nan's

[1] Biggs 2015, [1] McKenzie & Bailey 2016

Prediction #2

The DO-IO double object should be rare when DAT-ACC is lost

- It should be absent, modulo competing grammars, but
- Swedish retains it lexically with a few particle verbs
- Northern English have DO-IO

There exist syntactic structures that render surface DO-IO is learnable in these languages without overt case marking

Prediction #3

The *to*-dative should be rare where overt DAT-ACC is maintained

- There would be no motivation to innovate it, but
- Faroese: overt DAT-ACC distinction, *to*-dative, but no DO-IO[1]

*** DO-IO, *to*-Dative**

* *Hon gaf troyggiuna *(till) Mariu*
She gave sweater.DEF.ACC to Maria.DAT
'*She gave Maria the sweater / the sweater to Maria.*'

[1] Lundquist 2014

Prediction #3

The *to*-dative should be rare where overt DAT-ACC is maintained

- There would be no motivation to innovate it, but
- Halsa Norwegian: overt DAT-ACC on pronouns and definite nouns, *to*-dative, DO-IO pronouns only[1]

DO-IO

Ho	*ga*	*det*	*'nå*
She	gave	it	him.DAT

IO-DO

Ho	*ga*	*'nå*	*det*
She	gave	him.DAT	it

'She gave him it'

***to*-Dative**

Ho	*ga*	*det*	*åt*	*'nå*
She	gave	it	to	him.DAT

'She gave it to him'

[1] Åfarli & Fjøsne 2012, pc.

Prediction #3

The *to*-dative should be rare where overt DAT-ACC is maintained

- There would be no motivation to innovate it, but
- Halsa Norwegian: overt DAT-ACC on pronouns and definite nouns, *to*-dative, DO-IO pronouns only[1]

DO-IO

* *Ho* *ga* *mat* *kattåinn*
She gave food cat.DEF.DAT

IO-DO

Ho *ga* *kattåinn* *mat*
She gave cat.DEF.DAT food
'*She gave the cat food*'

[1] Åfarli & Fjøsne 2012, pc.

Prediction #3

The *to*-dative should be rare where overt DAT-ACC is maintained

- There would be no motivation to innovate it, but
- Elfdalian/Älvdalian: overt DAT-ACC, *to*-dative, no DO-IO (at least for nouns)[1]

DO-IO[2]

* *Ig* *gav* *dukkur* *kullum*
I gave dolls girls.DAT

IO-DO

Ig *gav* *kullum* *dukkur*
I gave girls.DAT dolls
'*I gave the girls dolls*'

[1] Dahl 2009, Garbacz 2010, [2] Garbacz 2010 (20)

Prediction #3

The *to*-dative should be rare where overt DAT-ACC is maintained

- There would be no motivation to innovate it, but
- Elfdalian/Älvdalian: overt DAT-ACC, *to*-dative, no DO-IO (at least for nouns)[1]

to-Dative[2]

dier	*åvå*	*selt*	*gardn*	*að*	*buälaę*
they.NOM	have.PRS.3P	sell.PAP.N	farm.DEF.ACC.SG	to	company.DAT.DEF.SG

‘*They have sold the farm to the company*’

[1] Dahl 2009, Garbacz 2010, [2] Dahl 2009 (7) reproducing Levander 1909

Prediction #3

The *to*-dative should be rare where overt DAT-ACC is maintained

- **Faroese, Elfdalian: overt DAT-ACC distinctions, to-datives, but no DO-IO**
- **Norwegian: overt DAT-ACC on pronouns + definite nouns, to-dative, DO-IO pronouns only**

The *to*-dative arises even when there is no pressure from morphological ambiguity.

The Norwegian examples show that case marking on pronouns (and def nouns) does not maintain DO-IO on nouns, so the English temporal gap remains unaccounted for[1]

[1] contra Polo 2002

Change Following Morphological Erosion

Weak Hypothesis

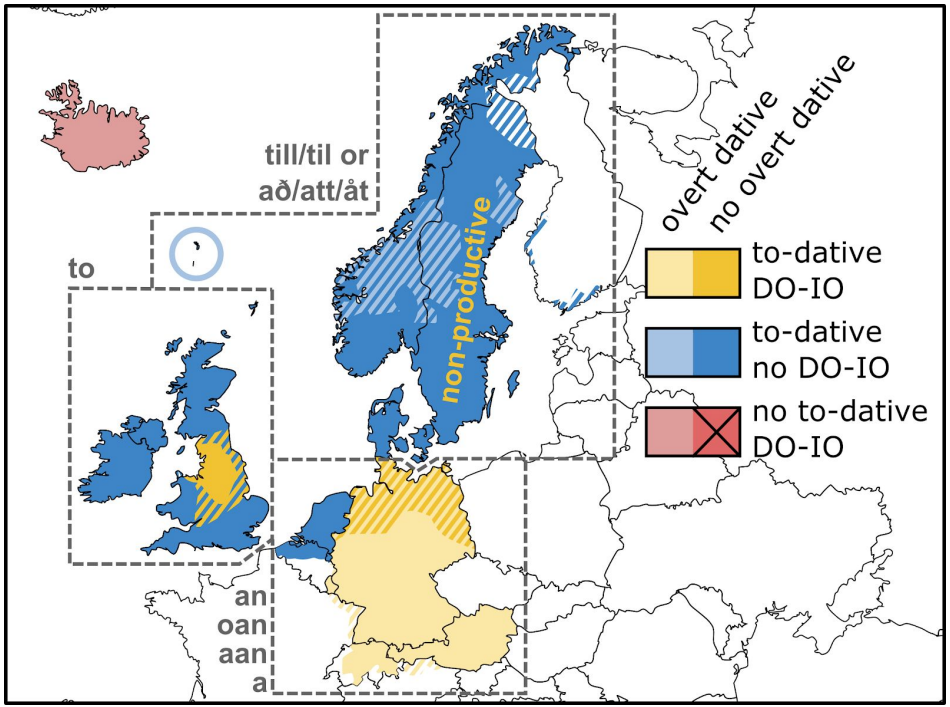
Morphological erosion did not force the to-dative to replace DO-IO. Rather, ambiguity created a pressure in favor of the former over the latter[1]

- **Supposes a functional trade-off between syntactic and morpho. complexity**
- **But it doesn't actually answer the relevant questions**
  How did categorical changes to the grammar occur?
  Why did *to* become a recipient marker?
  How did the *to*-dative achieve its modern lexical distribution?
- **And the correlation is just not good!**

[1] Allen 2006, De Cuypere 2015

Summarizing the Modern Germanic Distribution

- Only Icelandic lacks any kind of to-dative
- DO-IO exists without overt dative marking in English, non-productively in Swedish, and in Low German
- To-datives exist in overt dative varieties without DO-IO in Faroese, Norwegian (e.g., Halsa), in Sweden (e.g., Elfdalian)
- To-datives are expressed with non-cognate prepositions in different regions

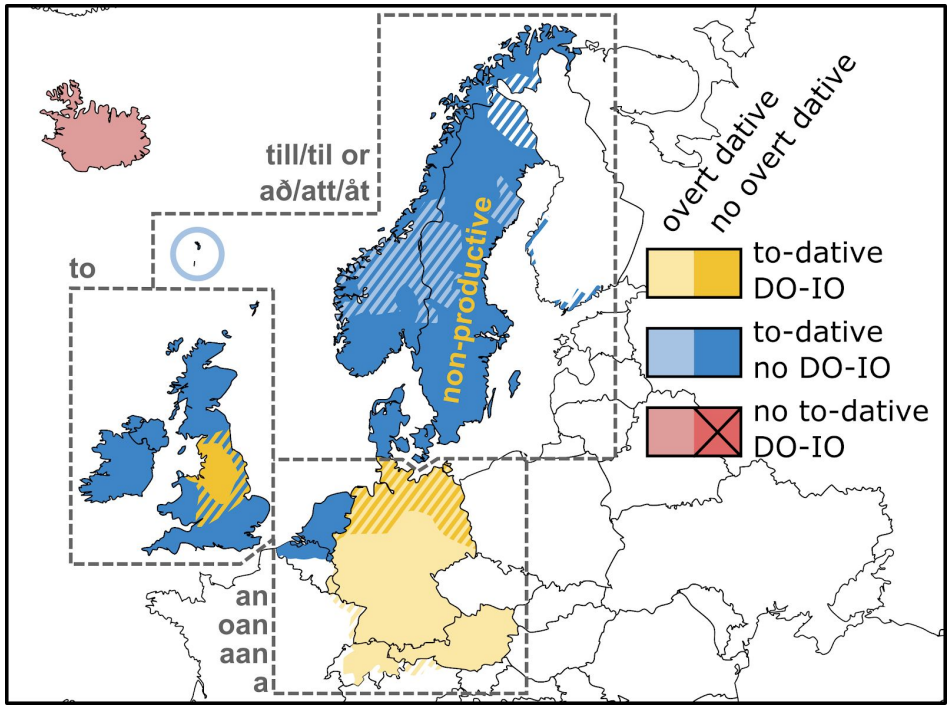


* Map based on data from Dahl 2004, 2009, Garbacz 2010, Åfarli & Fjøsne 2012, pc., Lundquist 2014, Biggs 2015, MacKensie & Bailey 2016, Bacovcin 2017, pc. with Danielle Turton and George Walkden

Summarizing the Modern Germanic Distribution

- Strong morphological erosion predicts pink and dark blue with transitional dark gold, non-existence of light blue and red
- Weak morphological erosion prefers pink and dark blue, strongly disprefers light blue and red
- Erosion does not predicts dark blue → gold in England
- Borrowing might predict to-datives with cognate prepositions, not three zones

Light blue across North Germanic poses a serious synchronic problem for morpho. erosion



* Map based on data from Dahl 2004, 2009, Garbacz 2010, Åfarli & Fjøsne 2012, pc., Lundquist 2014, Biggs 2015, MacKensie & Bailey 2016, Bacovcin 2017, pc. with Danielle Turton and George Walkden

Semantic Extension

Semantic Extension

Speakers reinterpreted recipient-like goal constructions such that the object of *to* gained a possible recipiency reading and became the to-dative

- **Allative → recipient shift is common cross-linguistically[1]**
- **Attested in child language acquisition**
- **Predicts smooth expansions in both semantic scope and token frequency, but the former does not necessarily bear out[2]**

[1] Cuychens & Verspoor 1998, Haspelmath 2003, [2] Elter 2018

Attested “Overextensions”

- **Semantic expansion may explain the attested “overextended” Middle English to-datives *Commaunde to the peuple*, *saued to hym*, *acsy to his uader*, or *forbed…to Roboam***
- **But none of these hypotheses explains why these were later lost**
- **We would still need a second mechanism to account for this even if morphological erosion bore out.**

How did the *to*-dative go from Old to Modern English?

I present a two-part model

1. ***to*-Dative innovated as naive recipiency analysis of certain goal constructions[1]**
2. ***to*-Dative expanded rapidly as typical learner overgeneralization[2]**

Accounts for

- **Typological disconnect between morphological erosion and *to*-dative/DO-IO**
- **Rapid attestation of broadly applicable *to*-dative**
- **Possibility for parallel evolution across Germanic**

[1] cf De Cuypere 2015, [2] cf Yang 2016

Innovation

Semantic Extension

Speakers reinterpreted recipient-like goal constructions→
the object of *to* gained a possible recipiency reading and became the to-dative

- **Allative → recipient shift is common cross-linguistically[1]**
- **Attested in child language acquisition**

[1] Cuychens & Verspoor 1998, Haspelmath 2003

Old English "*to*-datives"

- ***to* is attested a few times with goals which are plausible recipients[1,2]**
  eg '*agifan to a monastery*,' '*(ge)sellan to a church*'
- **And maybe[1] a couple examples with human goal-like recipients**

If these already were true *to*-datives,
then this part of the analysis is not needed

[1] Mitchell 1985, [2] De Cuypere 2015

Ambiguous Allative-*to*

Alice threw the ball to Bob **recipient-like goal or goal-like recipient?**

Alice said something to Bob **abstract goal or abstract recipient?**

- **Their semantic interpretations (conservative, innovative) may be formally distinct, but they are practically the same in use**
- **Language-specific broad-range mappings must be learned**

Ambiguous Allative-*to*

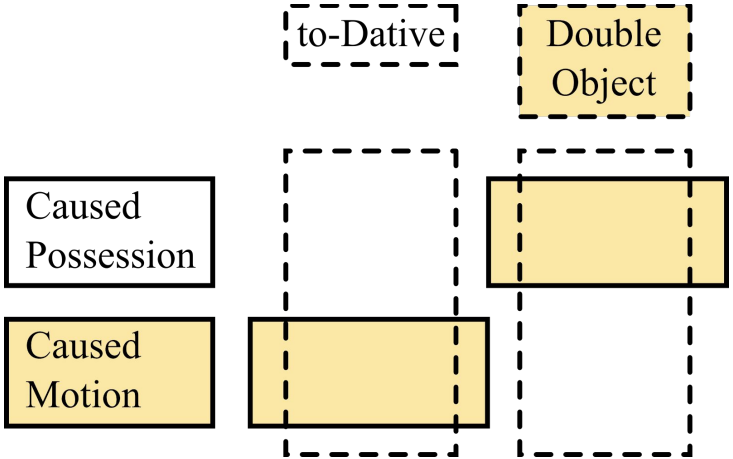
Alice threw the ball to Bob — recipient-like goal or goal-like recipient?

Alice said something to Bob — abstract goal or abstract recipient?

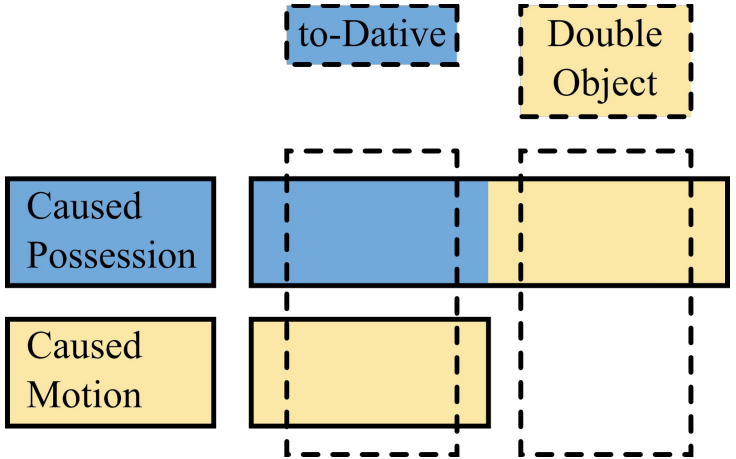
- Their semantic interpretations (conservative, innovative) may be formally distinct, but they are practically the same in use
- Language-specific broad-range mappings must be learned

At issue here is not what the proper semantic analysis these sentences is. It's whether a child could get away pragmatically with interpreting these as recipient sentences

Asymptomatic Innovation

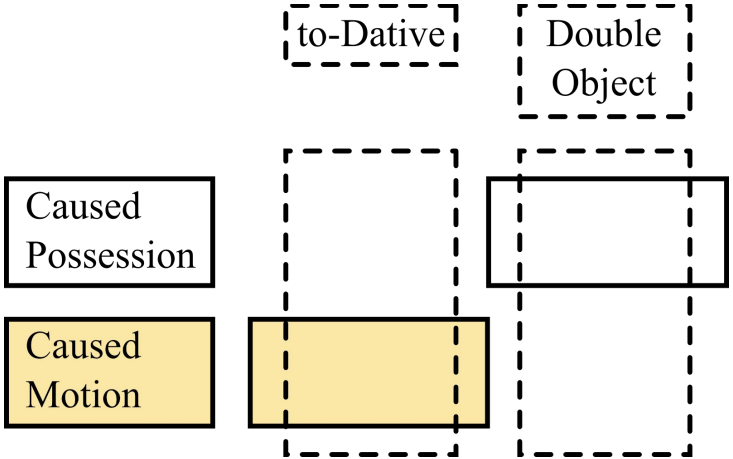


Old English *throw*

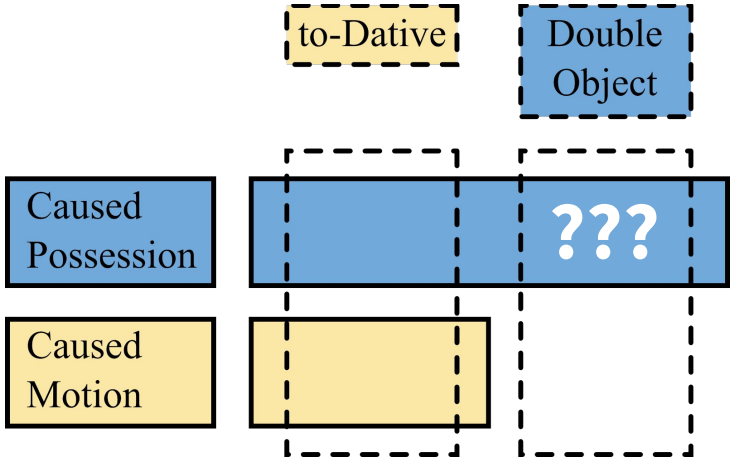


Naive realigned *throw*

Symptomatic Innovation?



Old English *say*



Naive realigned *say*

The Modern Analogue

- **Modern children occasionally overgeneralize the dative alternations**

Overgeneralized to-dative

‘*I asked this to you*’[1]

Overgeneralized double object

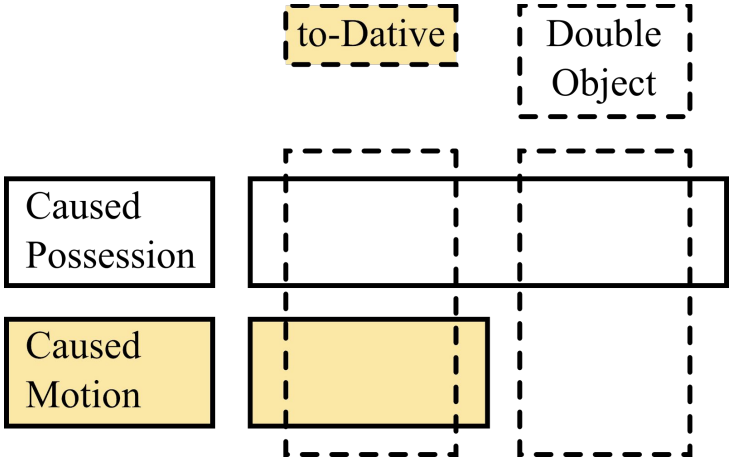
‘*Jay said me no*’[2]

‘*Mattia demonstrated me that yesterday*’[3]

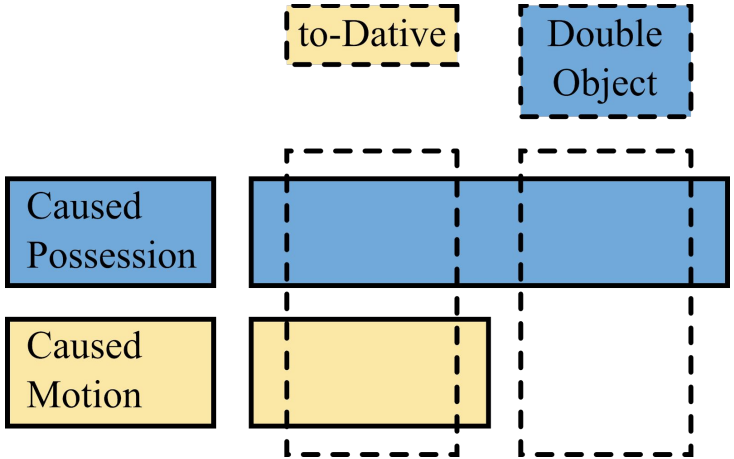
[1] 4;8 Hall et al 1984, [2] ROSS 2;8 CHILDES, [3] 3;8 CHILDES

Child Overgeneralization

Realigning *say*-type verbs yields "*Jay said me no*"



Adult *say*



Naive learner *say*

Child Language Acquisition

- Argument structure acquisition is a classic research topic[1]
- Focused on how children learn the partially arbitrary lexical mappings between verbs and the double object and to-dative
- Children must generalize past their input,
  but they cannot generalize too much. This is **Baker's Paradox**[2]

[1] Bowerman 1983, Fodor 1985, Gropen et al. 1989, Pinker 1989, Ambridge et al. 2008, Yang 2016, etc., [2] Baker 1979

Child Language Acquisition

Children need enough innate knowledge to render the constructions learnable but enough inductive learning to explain variation

Child Language Acquisition

Children need enough innate knowledge to render the constructions learnable but enough inductive learning to explain variation

Lexical conservatism is misleading[1]

- **Even three-year-olds use the constructions where unattested to them[1]**
- **Including "I asked this to you" innovations**
- **Even frequent verbs may not be attested in a construction in large corpora**
  - **Eg, *Throw* is attested 146 in Brown CDS, but only 3 times in the to-dative and 0 times in the DOC**

[2] Pinker 1989, [2] Naigles 1990, Conwell & Demuth 2007, Bowerman & Croft 2008 Table 13.1

Broad- and Narrow-Range Semantic Classes

- Verbs are given broad-range[1] and narrow-range[2] semantic classification
- Broad-range classes provide necessary conditions for each construction
- Narrow-range account for more specific patterns and exceptions
- The two work together in describing the constructions' lexical distribution

[1] Mazurkewich & White 1984, Pinker et al 1987, Gropen et al 1989, Rappaport Hovav & Levin 2008, [2] Gropen et al 1989, Levin 1993

Narrow-Range Classes

Finer-grained classifications can be specified to describe grammaticality better than broad-range classes[1]

Double Obj & to-Dative:

GIVE, TRANSFER OF MESSAGE, FUTURE HAVING, CARRY, BRING/TAKE, THROWING, SEND, DRIVE

Double Object Only:

DO ONLY, DUB, APPOINT, BILL, DECLARE

to-Dative Only:

SAY, MANNER OF SPEAKING, FULFILLING, LATINATE, PUTTING IN SPEC DIRECTION

[1] Gropen et al 1989, Levin 1993. Levin's are presented

Narrow-Range Classes

- May be learned distributionally[1]
- There is cross-linguisitic variation
  eg, Norwegian THROWING is to-dative-only, unlike English[2]
- There is diachronic and individual variation in English[3]
- The classes are violable

These are a useful descriptive tool, but how do children leverage them to learn the dative constructions?

[1] Pinker 1989, Braine & Brooks 1995, [2] Barðdal et al 2011, [3] Fellbaum 2005

Sufficient Evidence

Is there enough evidence that a given narrow-range class productively supports the to-dative?

- If a learner has experienced the construction with enough verbs in a class, it is reasonable to treat it as productive and apply it to all members of that class
- Otherwise, the learner can memorize which members it applied to and use it only with those

The Sufficiency Principle[1]

- A corollary to the Tolerance Principle
- Asks whether the learner has received enough evidence for a generalization
- Calculated over *m* (yet) unattested forms rather than *e*

N = number of types that should obey the generalization

m = number of types not attested obeying the generalization

θ = max # of exceptions that can be tolerated

Exceptions are sufficent if

$$m < \theta$$

$$\theta = N / \ln N$$

[1] Yang 2016

Acquiring the Modern Dative Alternation

Consider narrow generalizations: one for each narrow-range class

- Each class has its own *N*, *m*, θ according to that child's experience

- These numbers are estimated from text corpora for a "typical" child
- A frequency cutoff gives a child-like lexicon size and composition

Acquiring the Modern Dative Alternation

Consider narrow generalizations: one for each narrow-range class

- Each class has its own *N*, *m*, θ according to that child's experience

- These numbers are estimated from text corpora for a "typical" child
- A frequency cutoff gives a child-like lexicon size and composition

Acquiring the Modern Dative Alternation

Consider narrow generalizations: one for each narrow-range class

- Each class has its own *N*, *m*, θ according to that child's experience

- These numbers are estimated from text corpora for a "typical" child
- A frequency cutoff gives a child-like lexicon size and composition

Modeling Middle English Learners

- Extracted and lemmatized verbs from Penn Parsed Corpus of Middle English 2[1]
- About 1.2 million tokens total
- All verb lemmas occuring > 2x in DOC or to-PP constructions were extracted

Yields 75 lemmas sorted into Levin's narrow-range classes

- 39 of 75 have possible ambiguous allative-to readings
- Represent the nascent to-dative's distribution for a new innovator

[1] Kroch & Taylor 2000

Extending from Innovation

The children who innovated the *to*-dative would be "asymptomatic" if they did not extend it from ambiguous allative-*to*. Could they extend it?

- **The SP only needs counts of relevant verb types**
- **For each narrow-range class,**
  ***N* = number verb types in that class**
  ***m* = number without possible ambiguous reading**

If they could extend it, younger learners would receive unambiguous evidence from "symptomatic" peers

Ambiguous Directional-to by Class

Doub Obj + to-Dat	*N*	*m*
TRANS. MESSAGE	10	8
GIVE	5	1
FUTURE HAVING	14	4
CARRY	0	-
BRING/TAKE	4	0
THROWING	1	0
SEND	1	0

to-Dative Only	*N*	*m*
DRIVE	1	0
SAY	2	0
MANN. OF SPEAK	2	2
FULFILLING	3	1
PUT SPEC. DIR.	7	3
LATINATE	9	4

Doub Object Only	*N*	*m*
DO ONLY	6	6
DUB	4	4
APPOINT	3	3
BILL	0	-
DECLARE	3	3

Would an innovator be symptomatic?

Initial Expansion via Productivity

Doub Obj + to-Dat	Generalize
TRANS. MESSAGE	no
GIVE	YES
FUTURE HAVING	YES
CARRY	-
BRING/TAKE	YES
THROWING	YES
SEND	YES

to-Dative Only	Generalize
DRIVE	YES
SAY	YES
MANN. OF SPEAK	no
FULFILLING	YES
PUT SPEC. DIR.	YES
LATINATE	YES

Doub Object Only	Generalize
DO ONLY	no
DUB	no
APPOINT	no
BILL	-
DECLARE	no

Already almost the modern distribution

Further Expansion

- Most narrow-range classes could support a productive *to*-dative
- Innovators are predicted to be symptomatic

Following the initial expansion,

- Older learners producing novel to-datives could pass them onto their peers
- **Younger learners receive unambiguous evidence** for the construction
- Sparse input makes it hard to identify errant behavior among peers

Further Expansion

- ME learners who heard the new unambiguous *to*-datives from older peers had a broader basis for generalization
- The Sufficiency Principle works up to broader generalizations

An example broader classification:

1. TRANSFER OF MESSAGE, GIVE, FUTURE HAVING
2. CARRY, BRING/TAKE, THROWING, SEND
3. DRIVE, SAY, MANNER OF SPEAKING, FULFILLING, PUT SPEC DIR
4. LATINATE
5. DO ONLY, DUB, APPOINT, BILL, DECLARE

Sufficiency Result by Broader Class

Doub Obj + to-Dat	Generalize
CLASS 1	YES
CLASS 2	YES

to-Dative Only	Generalize
CLASS 3	YES
CLASS 4	YES

Doub Object Only	Generalize
CLASS 5	no

- This is the modern distribution

Sufficiency Result by Broader Class

Doub Obj + to-Dat	Generalize
CLASS 1	YES
CLASS 2	YES

to-Dative Only	Generalize
CLASS 3	YES
CLASS 4	YES

Doub Object Only	Generalize
CLASS 5	no

- This is the modern distribution
- But because of the relative sizes of the classes in Middle English, a further generalization is possible

Classes 1-4 provide enough evidence to extend the *to*-dative to all caused possession/motion verbs despite Class 5.

The attested "overgeneralization"

Empirical Predictions

Broad semantic range from the earliest attestation

- **This recursive application of the SP only characterizes its initial innovation in a single speech community over a few cohorts**
- **It should already have a wide range by the time it is attested**

Empirical Predictions

Broad semantic range from the earliest attestation

- This recursive application of the SP only characterizes its initial innovation in a single speech community over a few cohorts
- It should already have a wide range by the time it is attested

Contingency on the lexicon

- The broad Middle English overgeneralization is predicated on the lexicon
- If the lexicon were very different, the change may not have happened

Modeling Retreat

- "Over-generalized" *to*-datives disappeared in the 16th century[1]
- The presence of these *to*-datives was predicated on the composition of the Middle English lexicon
- A change to the lexicon had the power to upset it
- English underwent significant lexical change in the 16th century

More Latin borrowings in the 16th Century than French in the previous centuries combined.

[1] Visser 1963

Modeling Retreat

- I consider lexical change in English by counting lemmas in the PPCEME[1]
- Same methodology as before
- Lemmas carried over from ME are assumed to support the to-dative
- New lemmas are assumed not to

118 lemmas (57 carried over), 44 ambig-to lemmas (27 carried over)

29 Latinate verbs compared to 9 previously.
Many are attested in modern CDS: *administer*, *convey*, *mention*, *return*, *submit*...

[1] Kroch, Santorini, & Delfs 2004

EME Broader Classes

Doub Obj + to-Dat	*N*	*m*
CLASS 1	27	0
CLASS 2	8	0

to-Dative Only	*N*	*m*
CLASS 3	29	8
CLASS 4	29	15

Doub Object Only	*N*	*m*
CLASS 5	25	16

- **Middle English holdovers + ambig-to verbs present substantial evidence for the *to*-dative in all classes**

EME Broader Classes

Doub Obj + to-Dat	Generalize
CLASS 1	YES
CLASS 2	YES

to-Dative Only	Generalize
CLASS 3	YES
CLASS 4	no

Doub Object Only	Generalize
CLASS 5	no

- The broadest generalization no longer works
- Neither does generalization in Classes 4 and 5

This brings Class 5 into line with modern grammar but incorrectly predicts that the to-dative is unproductive in Latinate Class 4

Implications

Reason for change

- **The change was not primarily a functional response to morphological erosion**
  Too many disconnects exist between the two
  Learners can learn DO-IO without overt case marking
  Learners can acquire to-datives even with overt case marking
- **Then whatever caused its increased usage frequency was sociological in origin**
  Such as influence from French

Change is "*afunctional*"

Implications

Innovation was “the easy part”

- A naïve analysis of adult productions
- Could have happened multiple times in multiple places
- Gaining a population foothold is the hard part of change

Following Labov et al. (1972)

Actuation = Innovation + Initial Propagation

The focus of this research

This is the limiting factor in the changes discussed today

Some Thoughts on Acquisition and Propagation

Actuation[1] and the Paradox of Language Change[2]

If children are so good at acquiring language,
why are they so bad at it?

Helps to have a precise definition of actuation

Actuation = Innovation + uptake into the speech community
(The hand-off from an individual-level process to a population-level one)

[1] Definition paraphrased from Labov, Yager & Steiner 1972, [2] term coined by Niyogi & Berwick 1997

Transmission is not strictly linear and generational

- Children mature in communities and receive input from multiple speakers
- Community input is formally necessary for attested dynamics of change[1]
- Young children learn sociolinguistic variables[2]
- Children attend to input from older children[3] who are not linguistically mature
- Multiple competing targets may be present in the input

Everybody receives input from multiple grammars

"Monolingual" "Multilingual"

Multi-idiolect multi-dialectal traditional multilingual

[1] Niyogi & Berwick 2009, [2] Labov 1989, Anderson 1990, [3] Manly 1930, Weinreich, Labov & Herzog 1968 p 145, Roberts and Labov 1995, Labov 2001 p449, Nardy, Chevrot & Barbu 2014

How do we get from innovation to change?

- Need a way to get from individual innovation to population-level change

Solution to the Paradox of Language Change

- Acquisition is hard!
- Learning targets are obscured by
  Abject poverty in the input
  Interpersonal variation
- So even a “perfect” learner can initiate change - “blame the environment”

A thought experiment: “Sibling-Induced Change”

“Sibling-Induced Change”

Imagine two young children, Alice is slightly older than Bob

- Alice is currently producing innovative forms
- Bob is receiving both conservative adult input and Alice’s
- How does this effect Bob?

“Sibling-Induced Change”

Can Bob identify Alice’s innovation?

- Alice is mostly consistent with adults
- Bob may rarely if ever hear a conservative token corresponding Alice’s
- If Bob never hears a conservative token, he cannot know if Alice is innovating

“Sibling-Induced Change”

Can Bob identify Alice’s innovation?

- Alice is mostly consistent with adults
- Bob may rarely if ever hear a conservative token corresponding Alice’s
- If Bob never hears a conservative token, he cannot know if Alice is innovating

Will Bob adopt Alice’s innovation?

- In cases of severe sparsity, yes. What choice does he have?
- In other cases,

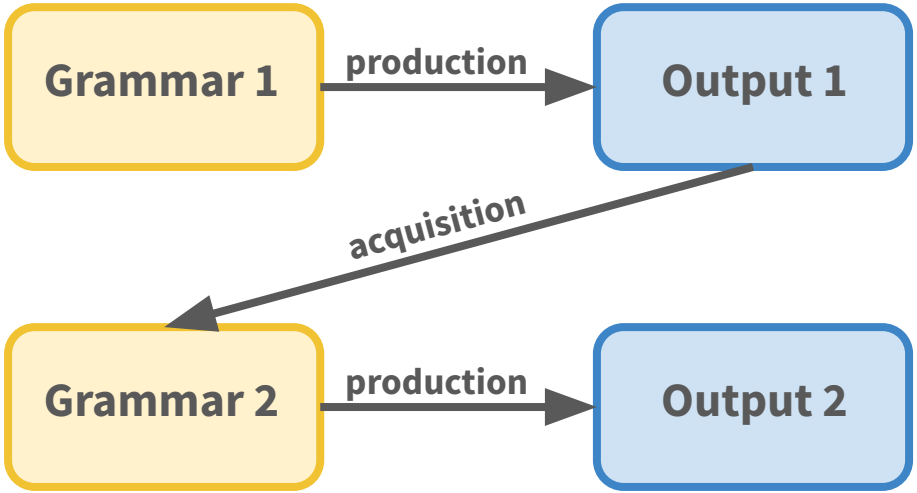
  Even young children orient toward peers

  Bob may prefer Alice’s forms over his parents

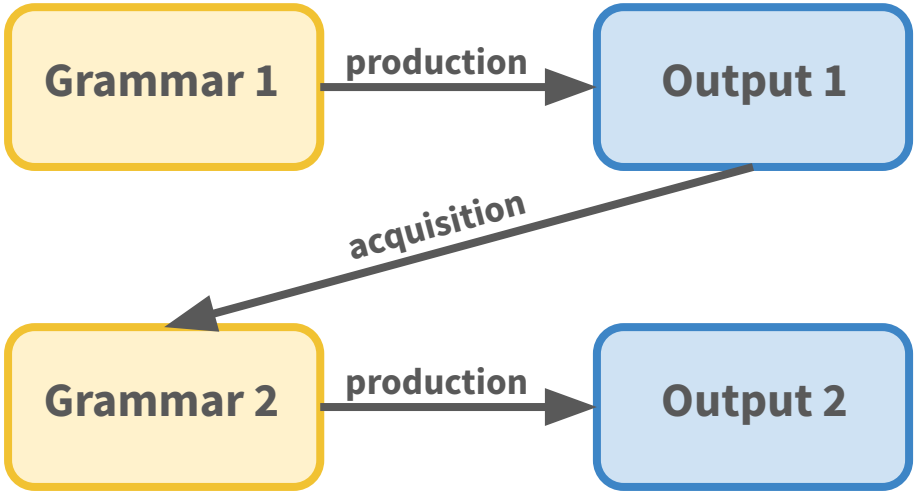
  He could learn both! (Competing grammars and sociolinguistic variation)

Z-Model of Language Acquisition and Change

- 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

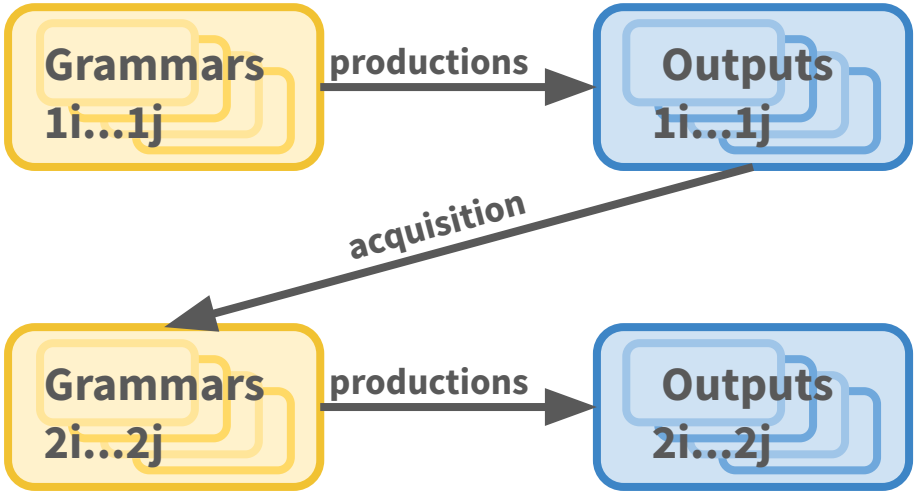


Insufficiency of the Z-Model



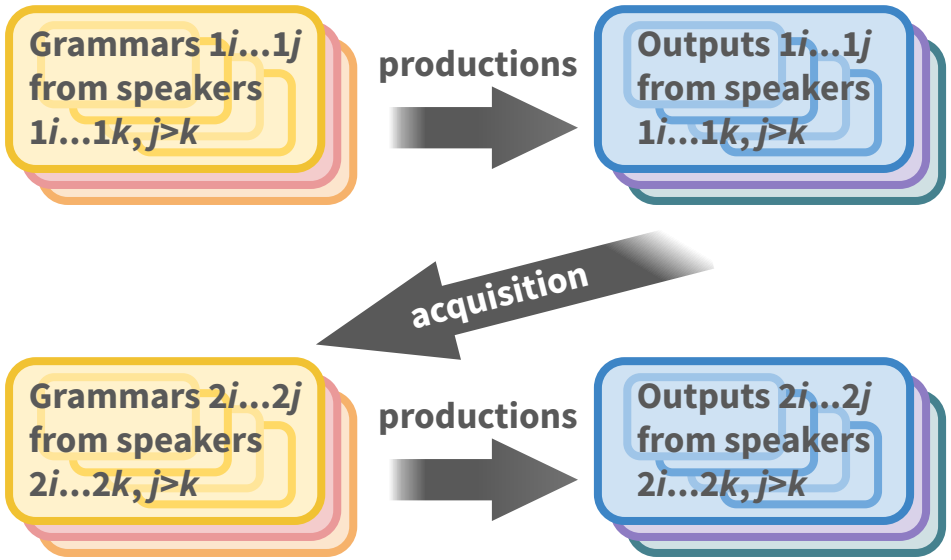
Insufficiency of the Z-Model

- **Individual production**
  - Variation across social settings
  - Variation over lifetimes



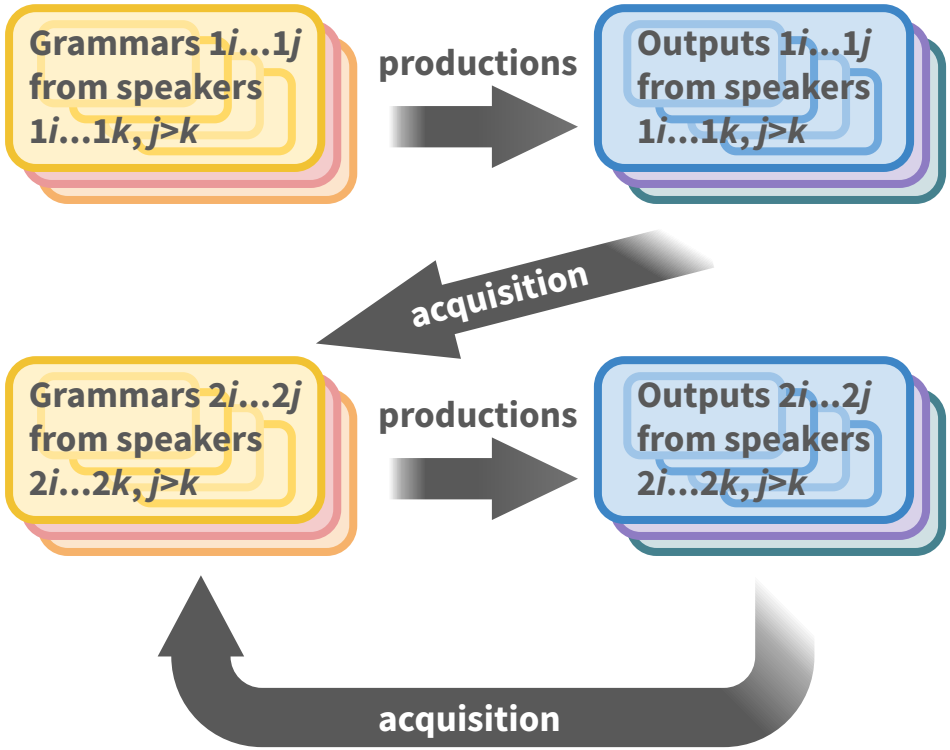
Insufficiency of the Z-Model

- **Individual production**
  - Variation across social settings
  - Variation over lifetimes
- **Community Embedding**
  - Variation across people
  - Everyone receives many inputs



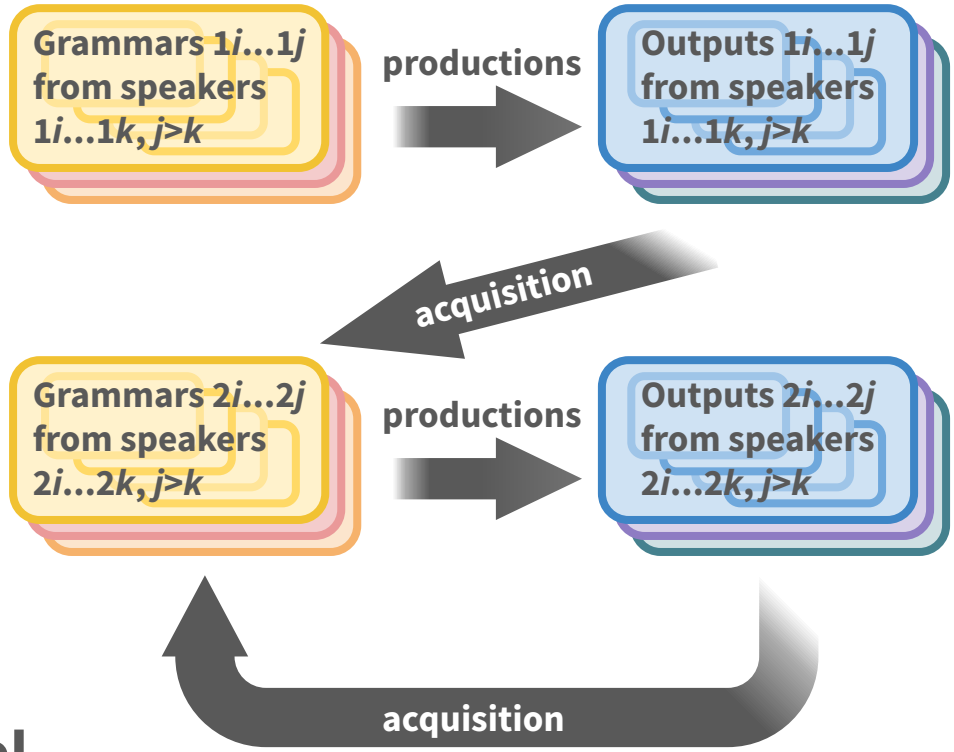
Insufficiency of the Z-Model

- **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



Insufficiency of the Z-Model

- **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

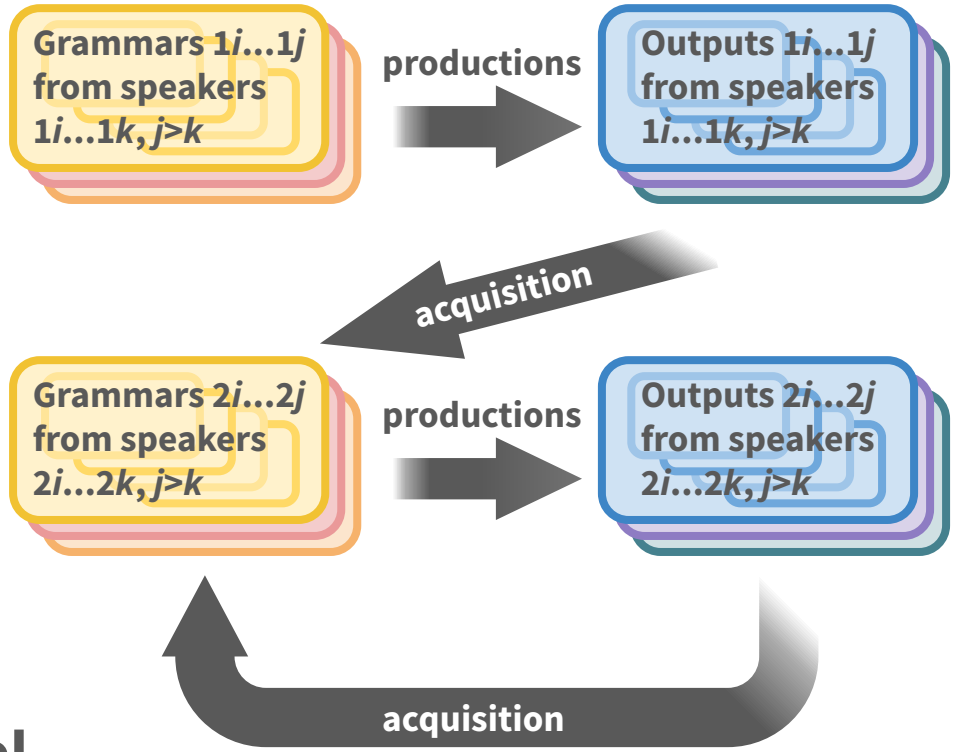


More of a

"Cyclic multi-multi-Z" model

Insufficiency of the Z-Model

- **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



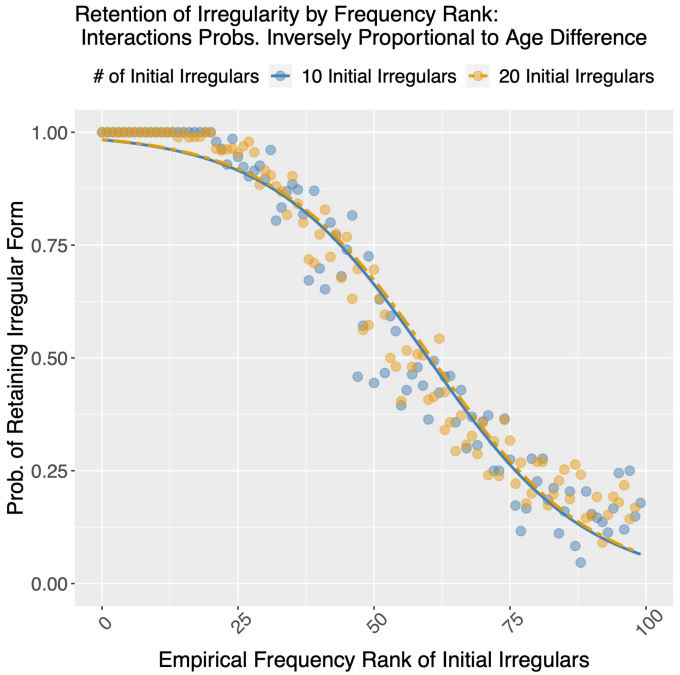
Sibling-Induced Change

~~"Cyclic multi-multi-Z" model~~

Proof-of-Concept “Sibling-Induced Change”

Sibling-Induced Change as a Baseline

- **It is sufficient on its own to reproduce**
  - Correlations between token frequency and irregularity
  - Correlations between paradigm size and irregularity
- **A much richer model than iterated learning**
  - Includes a population ← **change is population-level!**
  - Does not privilege generational transmission



The End.

Thank you!

- Charles Yang
- Mitch Marcus
- Don Ringe
- George Walkden
- Caitlin Richter
- Tony Kroch
- Carola Trips
- Betsy Sneller