Transparent /aɪ/-Raising

as a Contact Phenomenon

Jordan Kodner
Caitlin Richter

NWAV 47
New York, NY

Penn

Outline

- "Transparent" /aɪ/-raising
- Computational model for the acquisition of raising
- Learning from mixed input
- Transparent raising over time
- Implications and future directions

“Transparent” /aɪ/-Raising

“Canonical” /aɪ/-Raising

- “Canadian” Raising of front diphthongs before (underlyingly) voiceless segments, eg

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

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

/raɪɾɚ/“rider”	/rʌɪɾɚ/“writer”

“Transparent” /aɪ/-Raising

- **Raising before surface voiceless segments only**

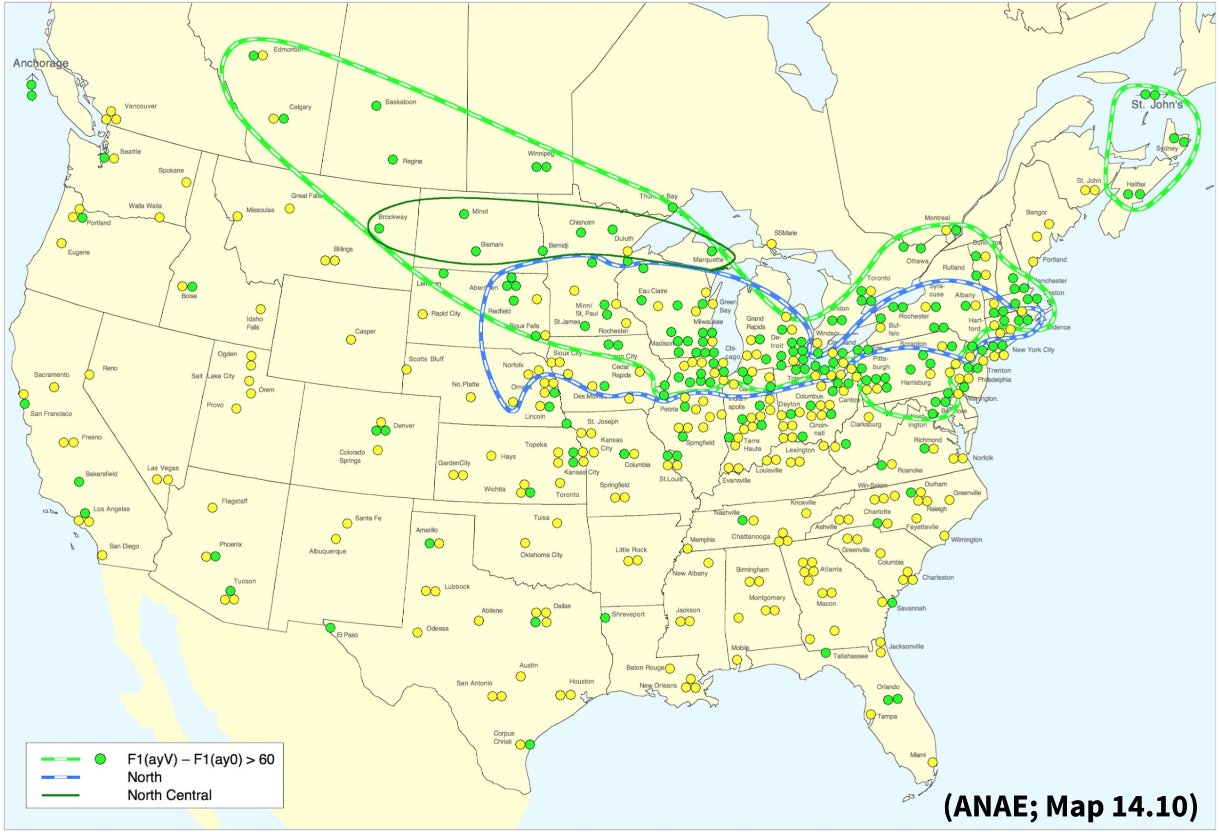
Canonical

/raɪd/ “ride”	/rʌɪt/ “write”
/raɪɾɚ/ “rider”	/rʌɪɾɚ/ “writer”

Transparent

/raɪd/ “ride”	/rʌɪt/ “write”
/raɪɾɚ/ “rider”	/raɪɾɚ/ “writer”

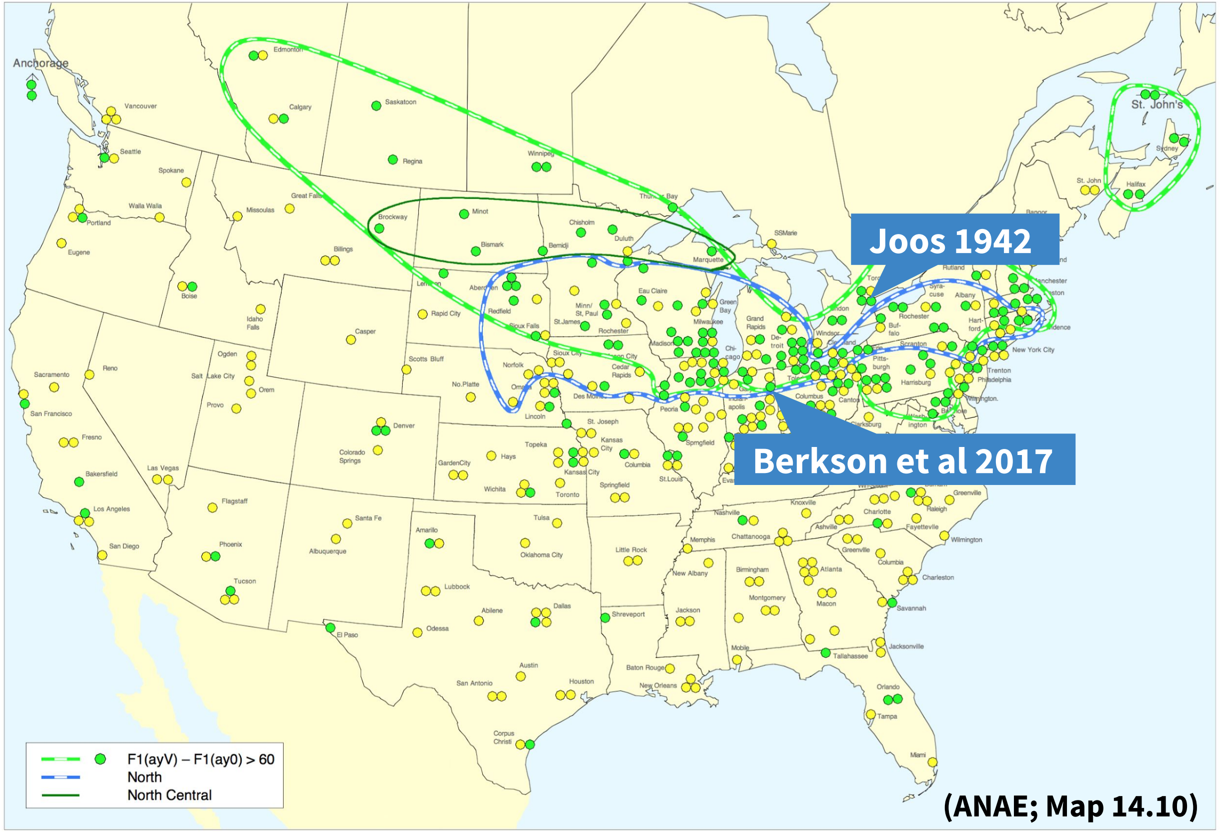
“Transparent” /aɪ/-Raising in the Wild



(ANAE; Map 14.10)

“Transparent” /aɪ/-Raising in the Wild

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



(ANAE; Map 14.10)

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

A phonetically-driven precursor to canonical raising

- **Hypocorrection (Ohala 1981) before surface /t/ spread to flapped /t/**
- **Offglide peripheralization (Moreton & Thomas 2007), pre-voiceless shortening (Joos 1942)**
- **Berkson et al 2017 find evidence for the former**

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

A phonetically-driven precursor to canonical raising

- **Hypocorrection (Ohala 1981) before surface /t/ spread to flapped /t/**
- **Offglide peripheralization (Moreton & Thomas 2007), pre-voiceless shortening (Joos 1942)**
- **Berkson et al 2017 find evidence for the former**

But...

- **Fruehwald 2016 finds that /aɪ/-raising was always conditioned by the underlying consonant in Philadelphia**

An Alternative Account

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

An Alternative Account

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

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

An Alternative Account

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

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

We will model this with the Tolerance Principle (Yang 2016)

An Alternative Account

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

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

We will model this with the Tolerance Principle (Yang 2016)

Accounts for Berkson et al's findings but remains consistent with Fruehwald

Predictions

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

Expectation - No incrementing phonetic precursor: raising is in phonology

- **Fruehwald finds no articulatorily-motivated gradient, and immediate phonological conditioning of raising**
- **Berkson et al's community snapshot observes progression of raising environments, not phonetic targets**

Predictions

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

Expectation - No incrementing phonetic precursor: raising is in phonology

- **Fruehwald finds no articulatorily-motivated gradient, and immediate phonological conditioning of raising**
- **Berkson et al's community snapshot observes progression of raising environments, not phonetic targets**

Incrementation accounts expect a phonetic peak in late adolescence

- **Not observed - raised phonetic target essentially flat across lifetime**

Early Acquisition of Phonology

Early in acquisition, children identify inventory of surface segments

- **Stable system of contrasts emerges at ~6-12 months (Kuhl et al 1992, Werker & Tees 1984)**

Can learn allophones underlyingly relating some segments, like aspirated and unaspirated English /p/ (Pierrehumbert 2003)

- **Influence of learned allophones evident in perception ~8 months (Pegg & Werker 1997)**

Modeling Acquisition

Tolerance Principle

A productive generalisation exists in a grammar when the number of exceptions to that generalisation do not exceed the tolerance threshold

Tolerance Principle

A productive generalisation exists in a grammar when the number of exceptions to that generalisation do not exceed the tolerance threshold

- e **exceptions to the generalisation**
- θ_N **tolerance threshold**
- N **number of linguistic items (types) that the generalisation pertains to**

$$e \leq \theta_N$$

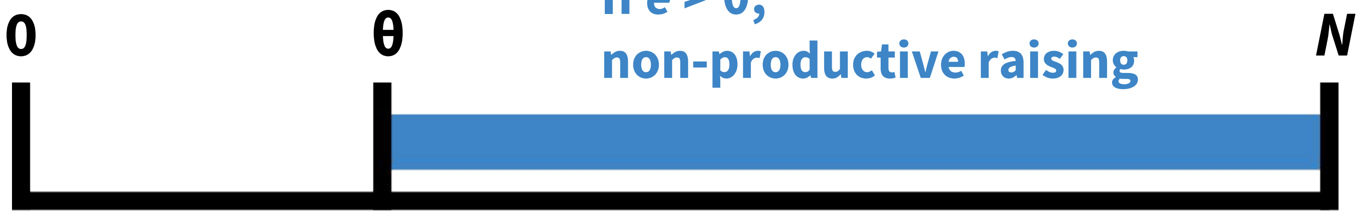
$$\text{where} \quad \theta_N := \frac{N}{\ln(N)}$$

Threshold for Canonical /aɪ/-Raising

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

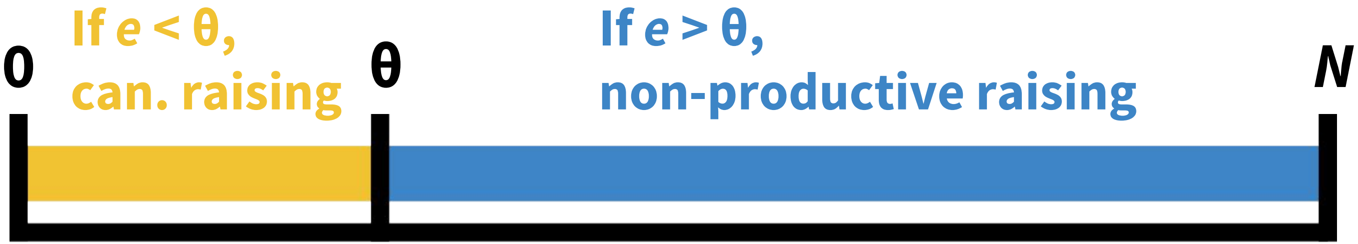
Threshold for Canonical /aɪ/-Raising

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



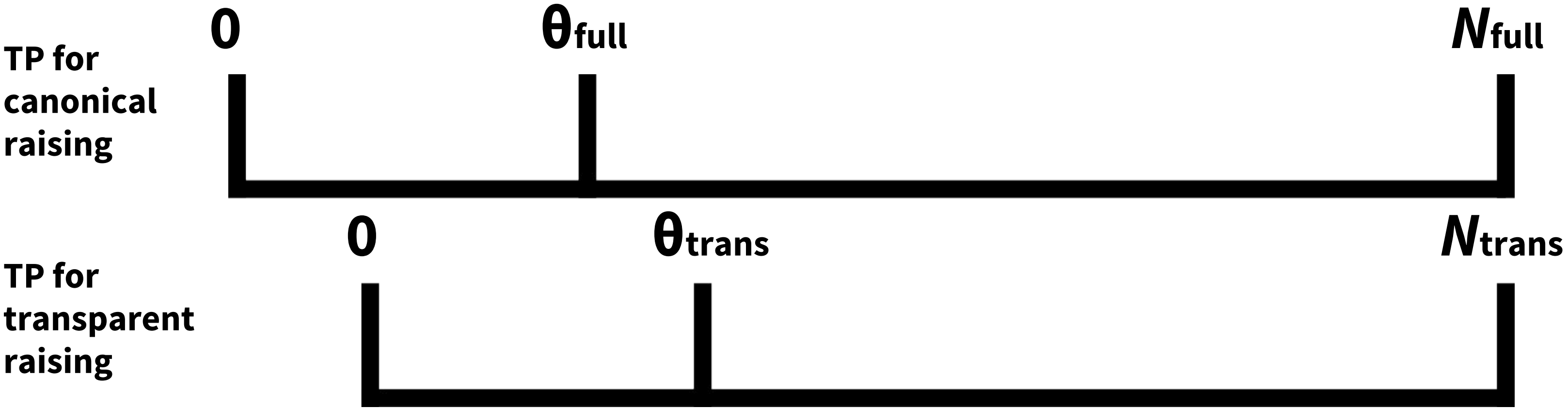
Threshold for Canonical /aɪ/-Raising

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



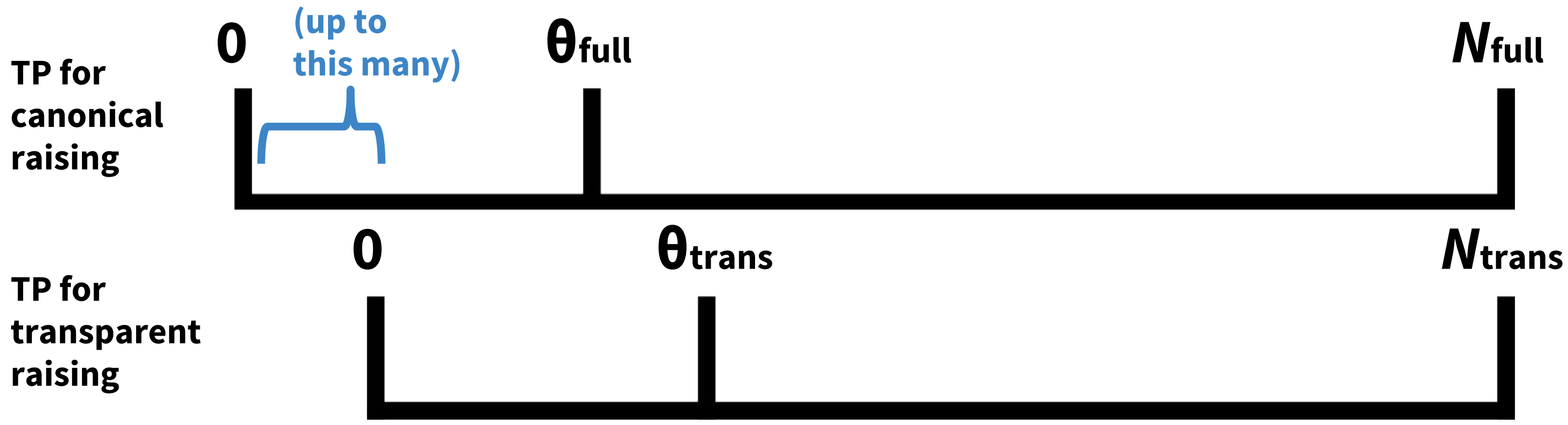
Threshold for Transparent /aɪ/-Raising

- **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

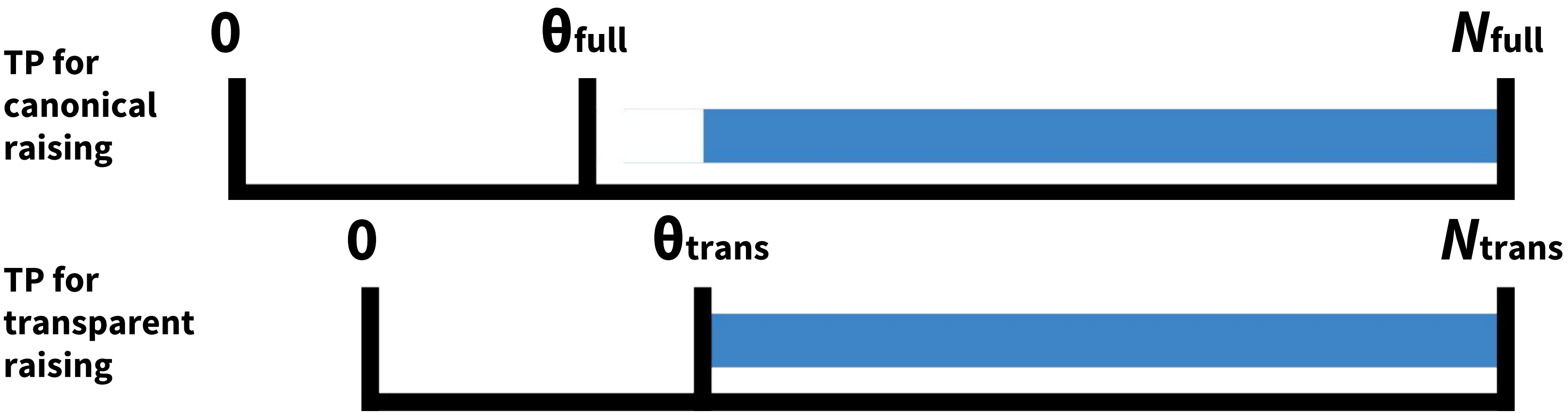
- The same calculation but with its own N, e, θ
- N_{full} = # of surface /t/ /aɪt/ words + # of flapped /aɪt/ words
- N_{trans} = # of surface /t/ /aɪt/ words
- So canonical can have extra exceptions that are irrelevant to transparent



Threshold for Transparent /aɪ/-Raising

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

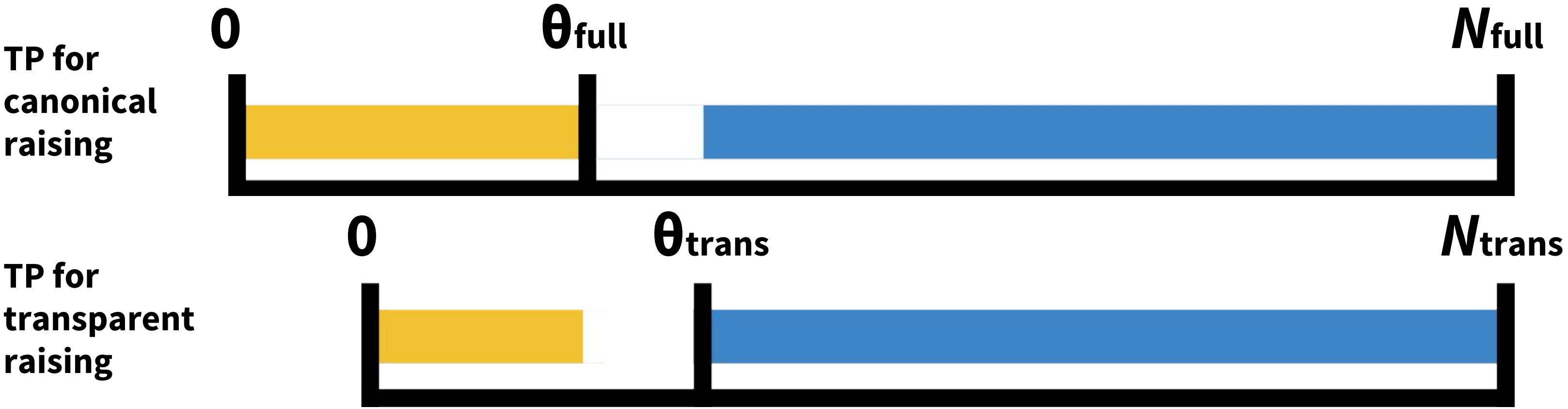
If both thresholds are exceeded, no productive raising



Threshold for Transparent /aɪ/-Raising

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

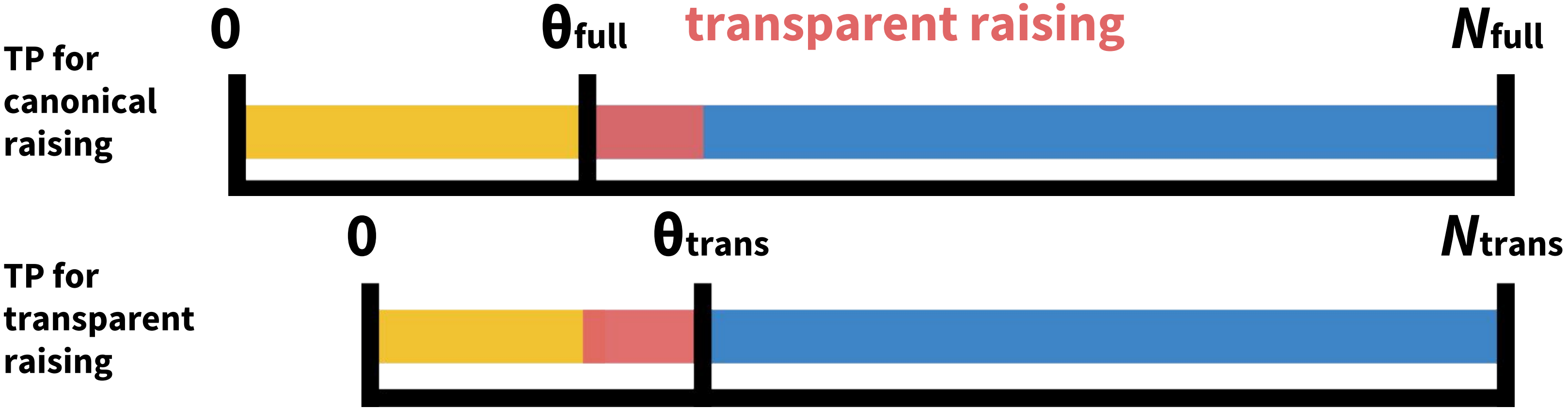
If the broader generalization is upheld, canonical raising



Threshold for Transparent /aɪ/-Raising

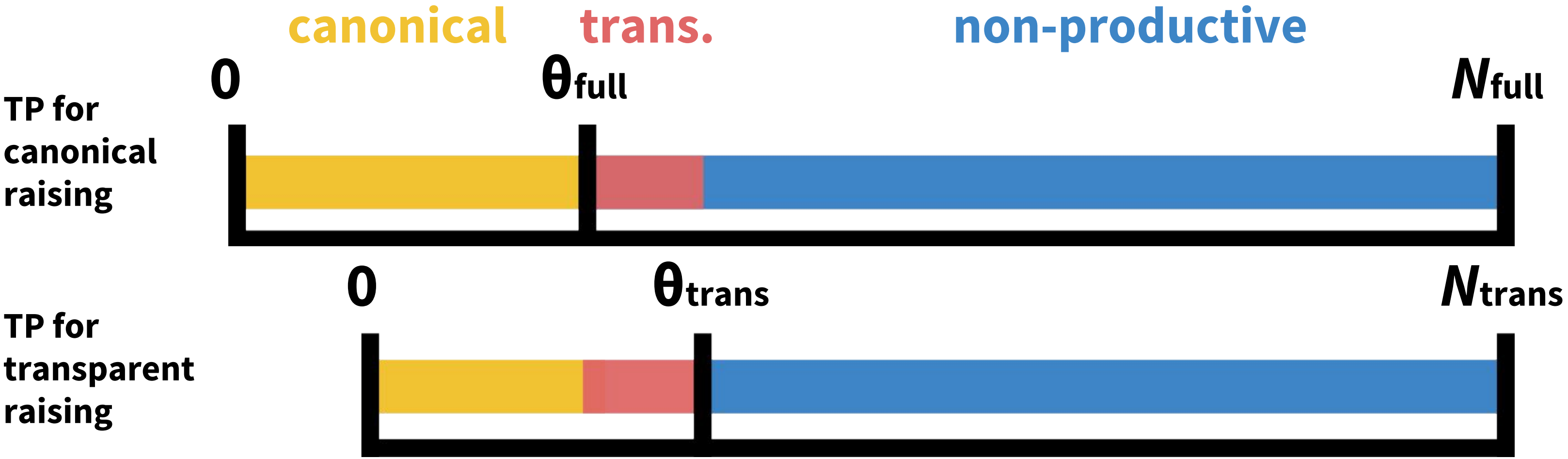
- The same calculation but with its own N, e, θ
- $N_{full} = N_{trans} + N_{flap}$

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 but not canonical raising. How likely is this?



Learning from Mixed Input

Children Regularise Inconsistent Input

Children whose input varies inconsistently at token level impose systematicity by regularising to most frequent variant - not probability matching

- **Productivity of single prevalent form emerges when learning inconsistent artificial language (Schuler et al 2016)**
- **Natively signing children learning from only non-native signers “clean up” errors, are more systematic than their input (Singleton & Newport 2004)**

Children Regularise Inconsistent Input

Children whose input varies inconsistently at token level impose systematicity by regularising to most frequent variant - not probability matching

- **Productivity of single prevalent form emerges when learning inconsistent artificial language (Schuler et al 2016)**
- **Natively signing children learning from only non-native signers "clean up" errors, are more systematic than their input (Singleton & Newport 2004)**

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

Tolerance of generalisations is based on type frequency

Mixed Input Learning is Probabilistic

- **What are the values for *N* and *e*?**
  - **The TP applies to individuals**
  - **so the answer varies child-to-child**

Mixed Input Learning is Probabilistic

- **What are the values for *N* and *e*?**
  - **The TP applies to individuals**
  - **so the answer varies child-to-child**
- **Children maturing in a mixed raising/non-raising population may hear both raised and non-raised versions of a given word**
  - **They learn the more frequent variant (Sneller et al 2018)**
  - ***e* is probabilistic ⇨ Learning outcome is probabilistic**
  - **Depends on ratio of raising and non-raising input**

Mixed Input Learning is Probabilistic

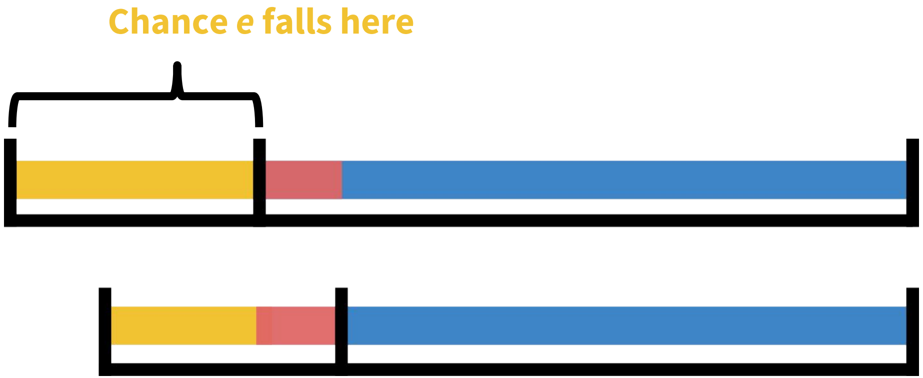
- **What are the values for *N* and *e*?**
  - **The TP applies to individuals**
  - **so the answer varies child-to-child**
- **Children maturing in a mixed raising/non-raising population may hear both raised and non-raised versions of a given word**
  - **They learn the more frequent variant (Sneller et al 2018)**
  - ***e* is probabilistic ➪ Learning outcome is probabilistic**
  - **Depends on ratio of raising and non-raising input**
- **Individuals' *e* values can be modelled by a binomial distribution (weighted coin tosses)**

Calculating Learner Outcomes

Probability of learning canonical raising

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

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



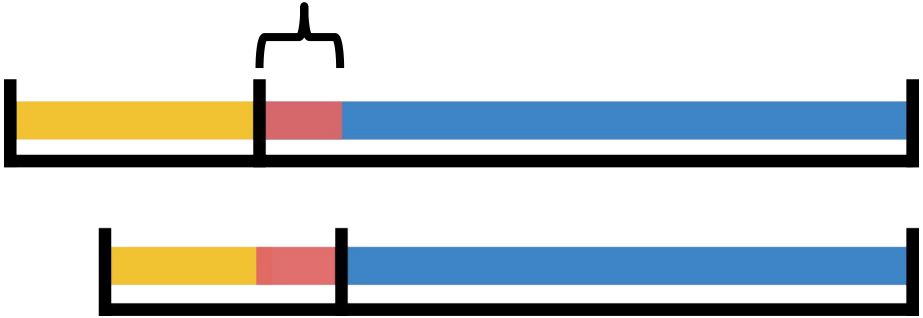
Calculating Learner Outcomes

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

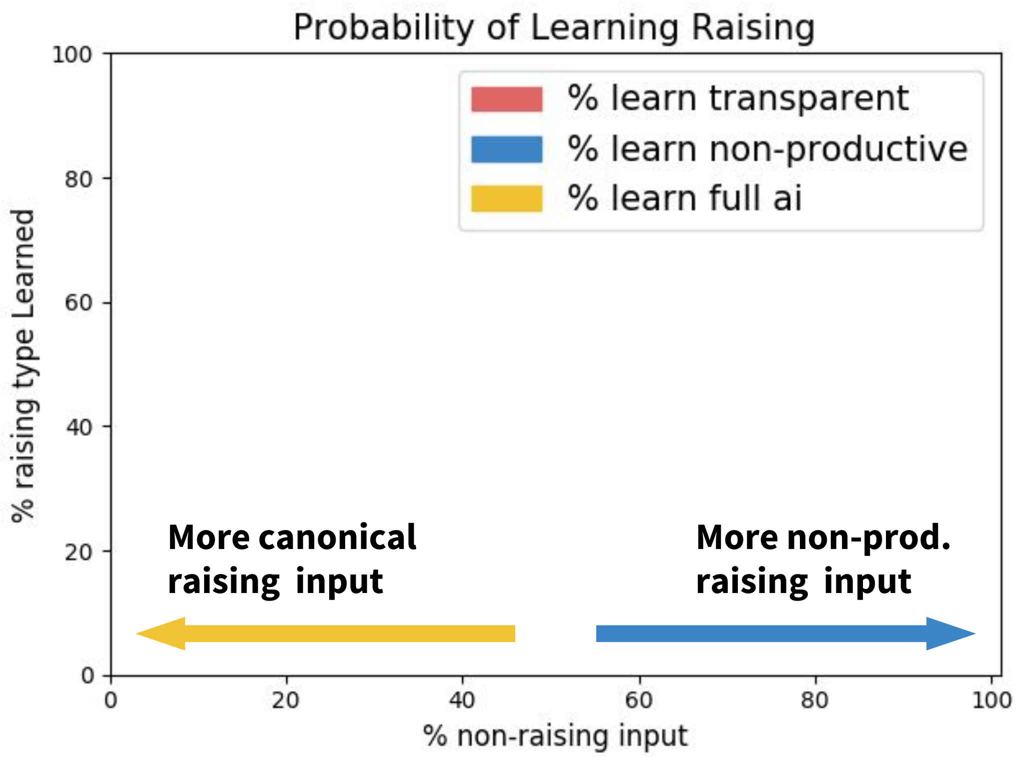
$$= \sum_{e_{\text{trans}}=\theta_{\text{full}}-N_{\text{flap}}}^{\lfloor \theta_{\text{trans}} \rfloor} \left(\binom{N_{\text{trans}}}{e_{\text{trans}}} p_{\text{none}}^{e_{\text{trans}}} \, p_{\text{full}}^{N_{\text{trans}}-e_{\text{trans}}} \sum_{e_{\text{flap}}=\theta_{\text{full}}-e_{\text{trans}}}^{N_{\text{flap}}}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

Estimating N_{full} and N_{trans}

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

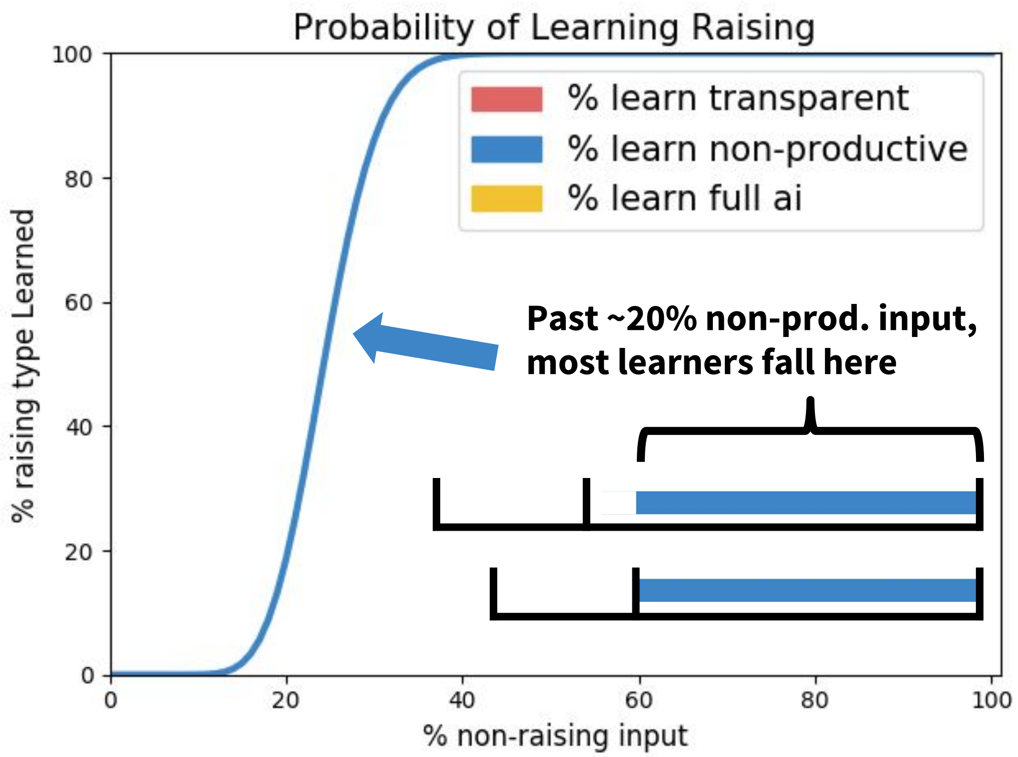
Estimate	Size in tokens	N_{full} (# types)	N_{trans} (# types)
Brown (freq ≥ 5)	356,959	53	45
B+B (freq ≥ 5)	883,698	82	69
Brown (all)	364,267	122	103
B+B (all)	895,501	182	155

Transparent Raising Emerging *De Novo*



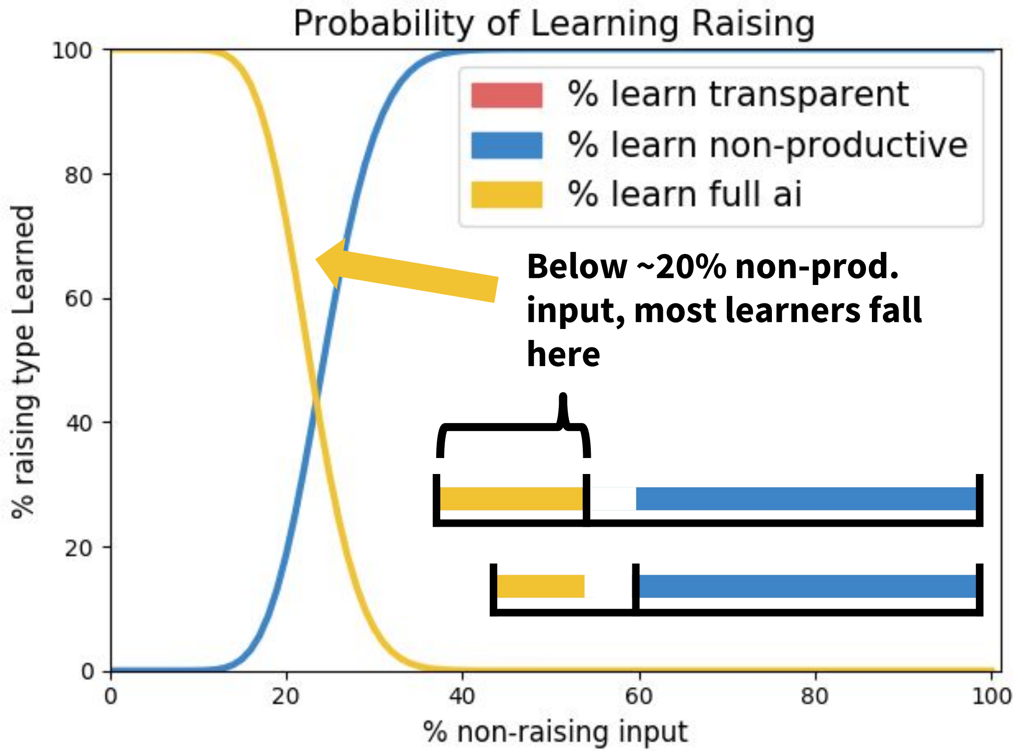
- *x*-axis **initial rate of non-raising vs canonical raising in community**
- *y*-axis **proportion of learners learning each raising type**

Transparent Raising Emerging *De Novo*



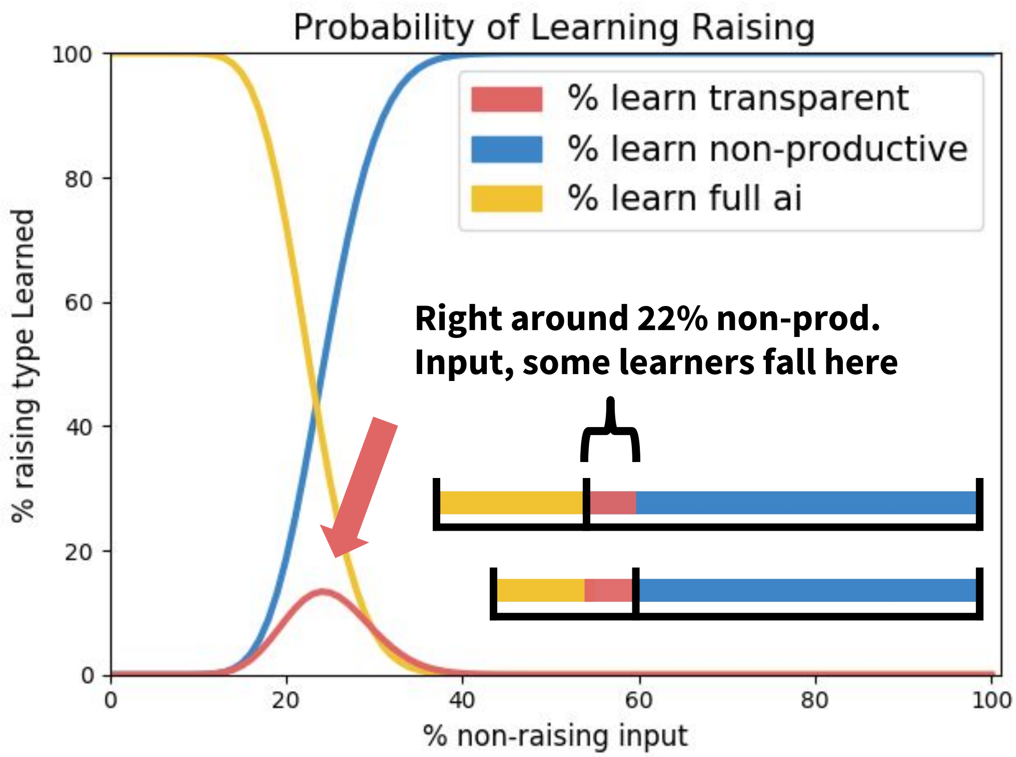
- *x*-axis **initial rate of non-raising vs canonical raising in community**
- *y*-axis **proportion of learners learning each raising type**

Transparent Raising Emerging *De Novo*



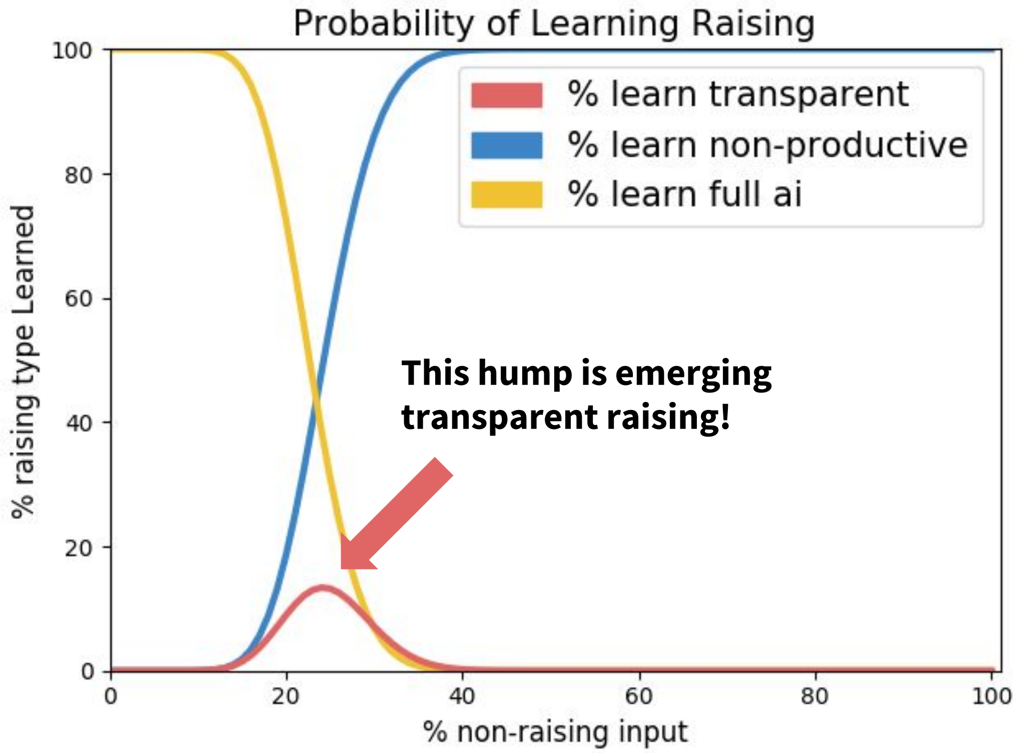
- *x*-axis **initial rate of non-raising vs canonical raising in community**
- *y*-axis **proportion of learners learning each raising type**

Transparent Raising Emerging *De Novo*



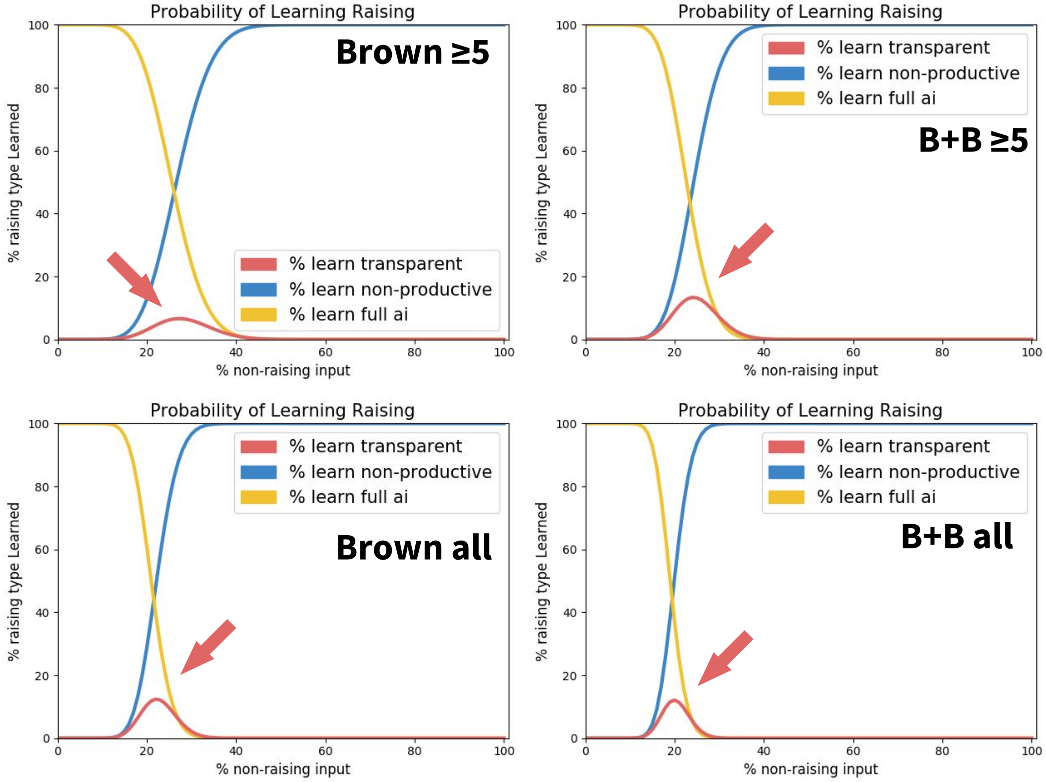
- *x*-axis initial rate of non-raising vs canonical raising in community
- *y*-axis proportion of learners learning each raising type

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

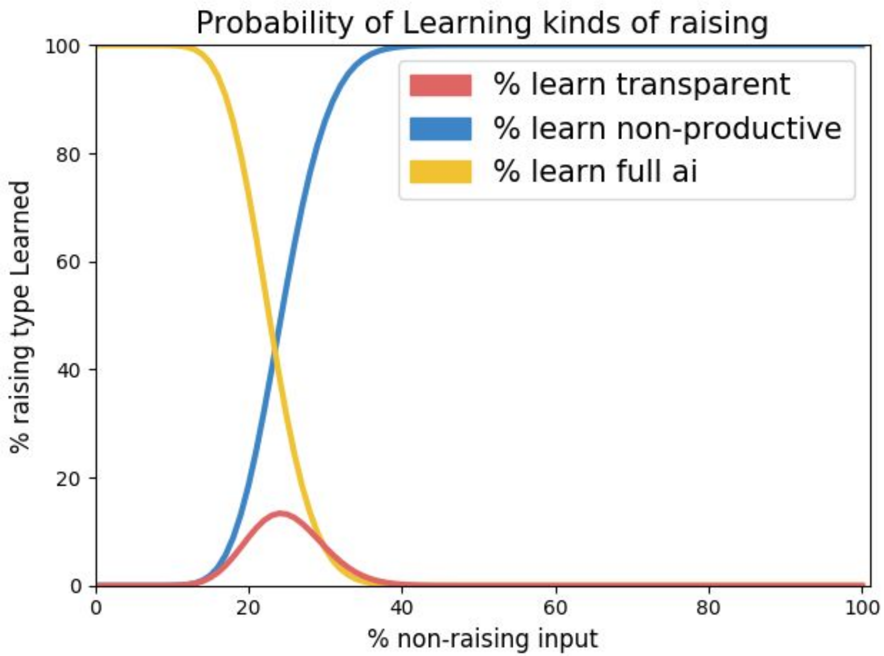


- **Transparent peaks occur at >20% non-raiser communities**
- **Transparent peaks reach <20% max**
- **This works because N_{full} tends to be just slightly larger than N_{trans}**

Interim Summary

This casts transparent raising as a contact effect, not as incipient raising

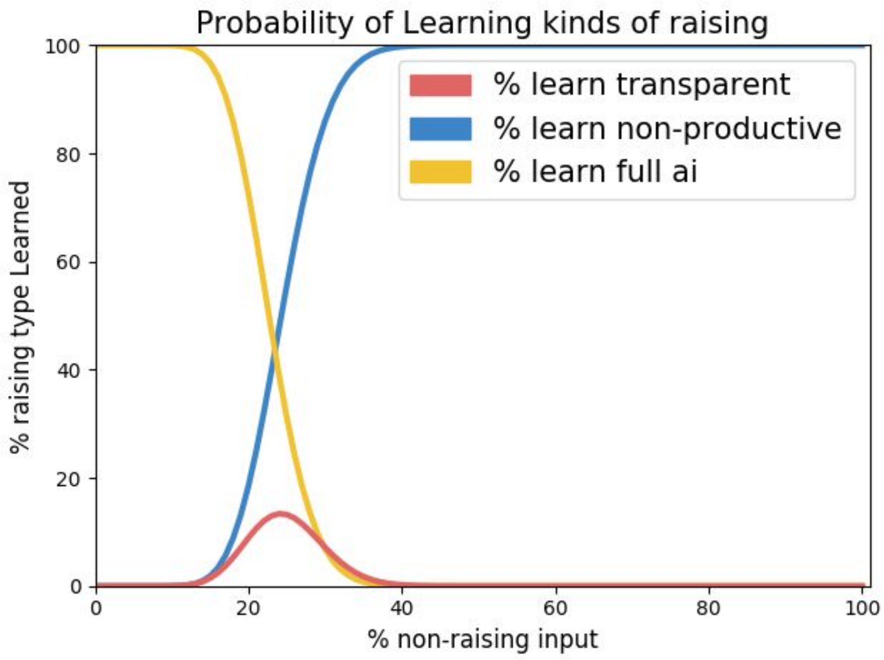
- **We predict that transparent /aɪ/-raising should appear sporadically among talkers in mixed raising/non-raising communities (eg 1940s ON and 2010s Fort Wayne)**



Interim Summary

This casts transparent raising as a contact effect, not as incipient raising

- **We predict that transparent /aɪ/-raising should appear sporadically among talkers in mixed raising/non-raising communities (eg 1940s ON and 2010s Fort Wayne)**
- **And do not expect more transparent raising to be more common early on (cf Philadephia study)**



Transparent Raising over Time

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

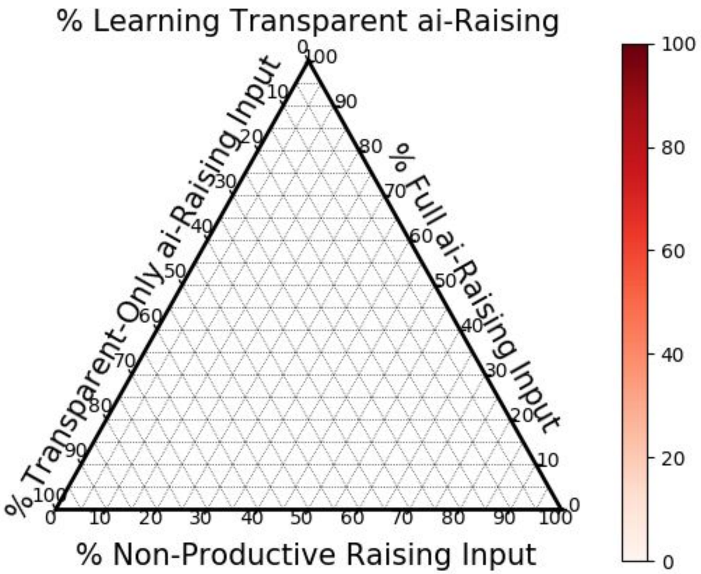
- They trend toward either non-raising or canonical raising over time
- Transparent raising dies out rapidly

Learning in a 3-way Mixed Setting

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

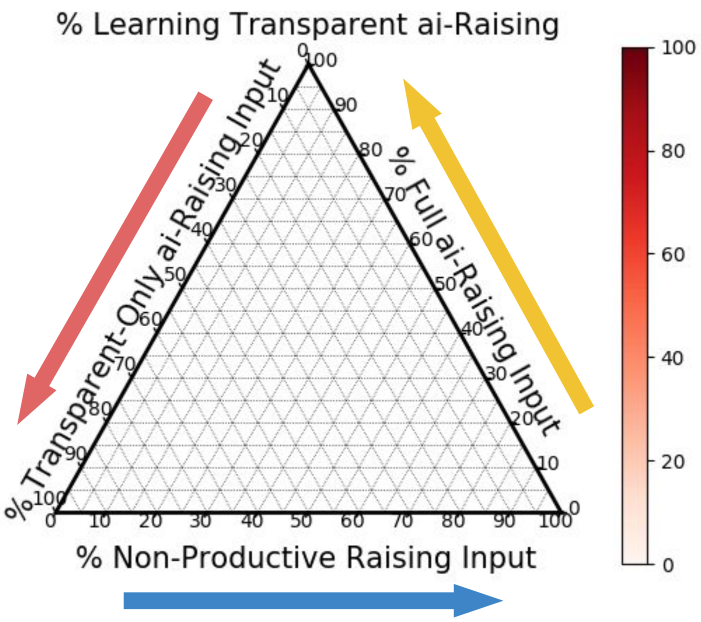
Learning in a 3-way Mixed Setting

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



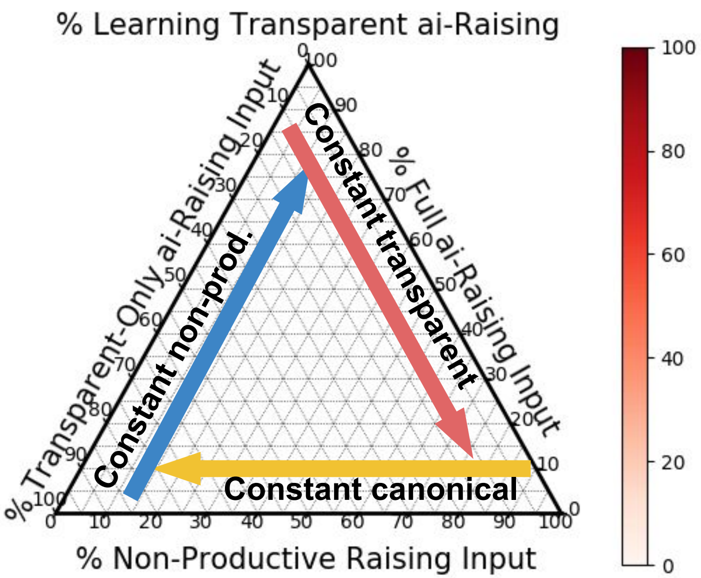
Learning in a 3-way Mixed Setting

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



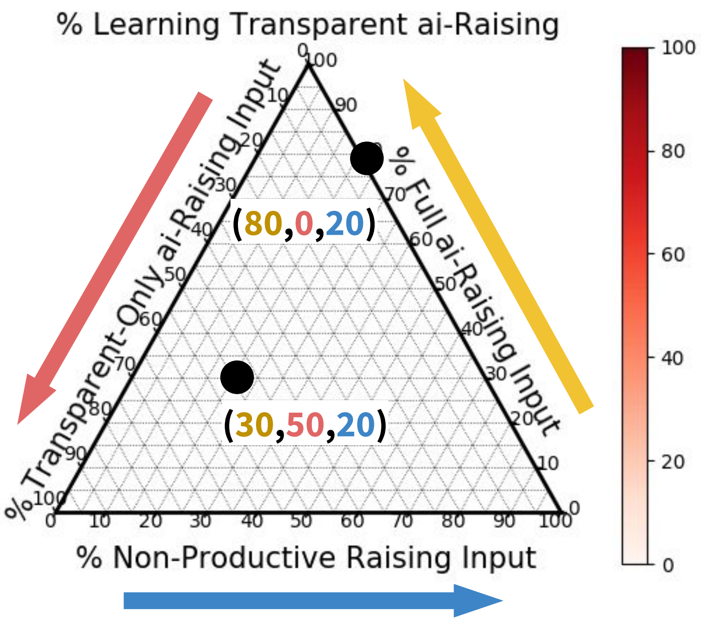
Learning in a 3-way Mixed Setting

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



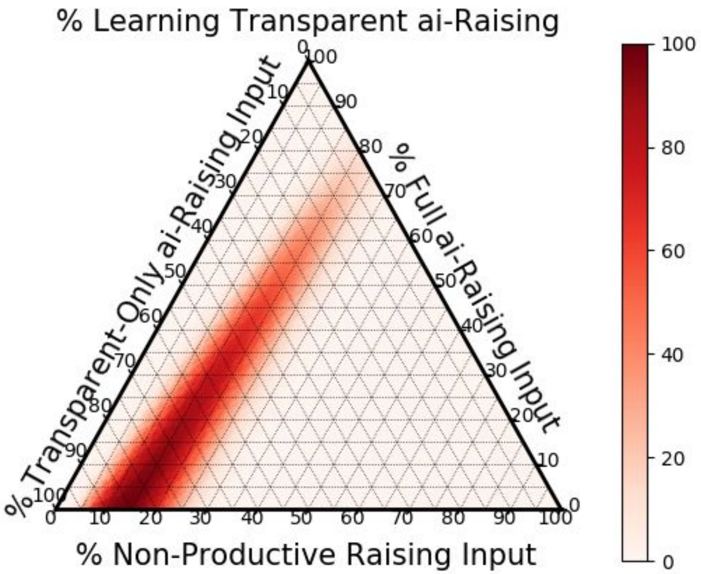
Learning in a 3-way Mixed Setting

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



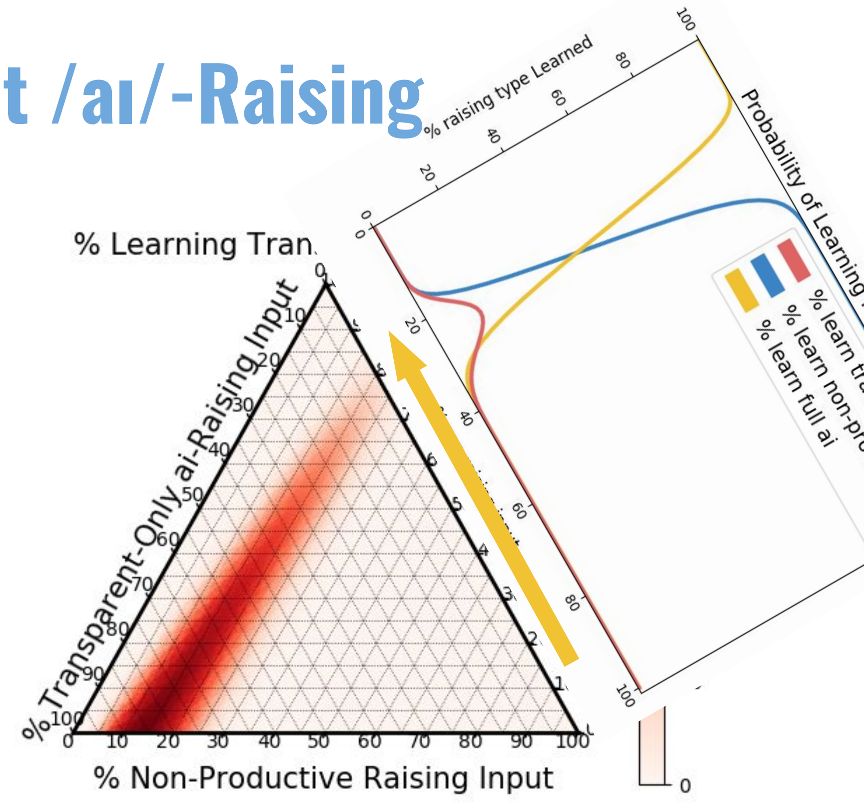
Proportion Learning Transparent /aɪ/-Raising

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



Proportion Learning Transparent /aɪ/-Raising

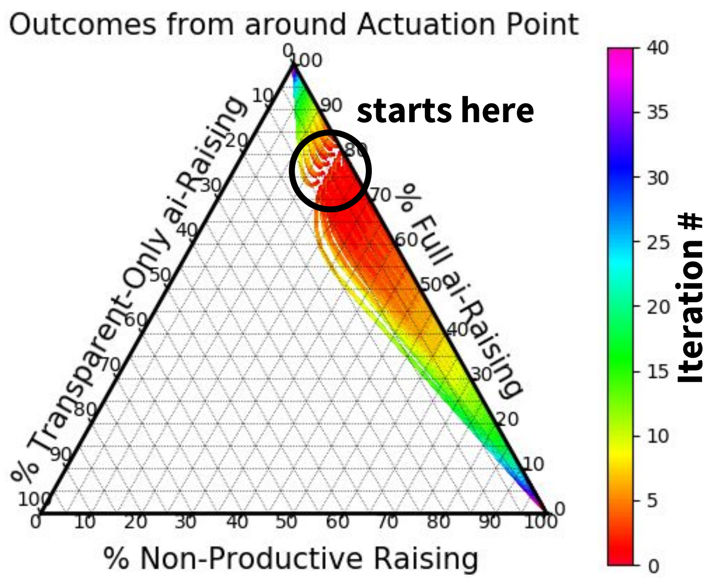
- **Transparent raising has a narrow band of viability**
- **Previous plots were on the top right side of the triangle where $p_{trans} = 0$ (no transparent raising input)**



Transparent /aɪ/-Raising over Time

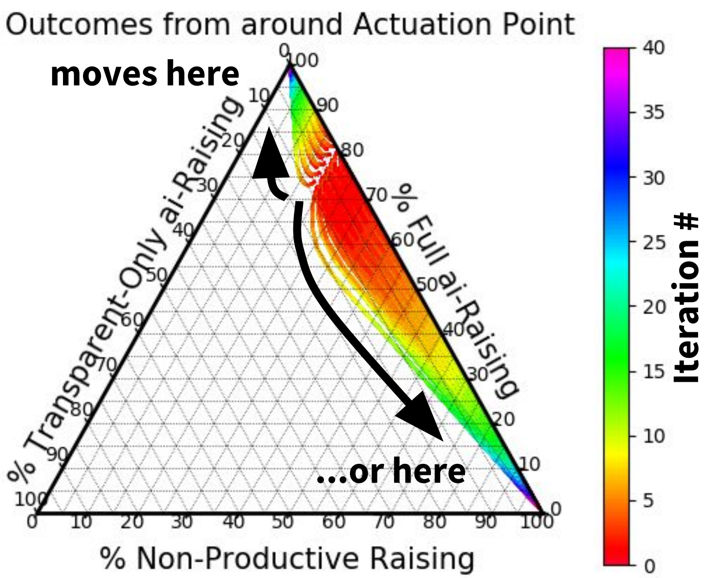
Simulate the population over time

- 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

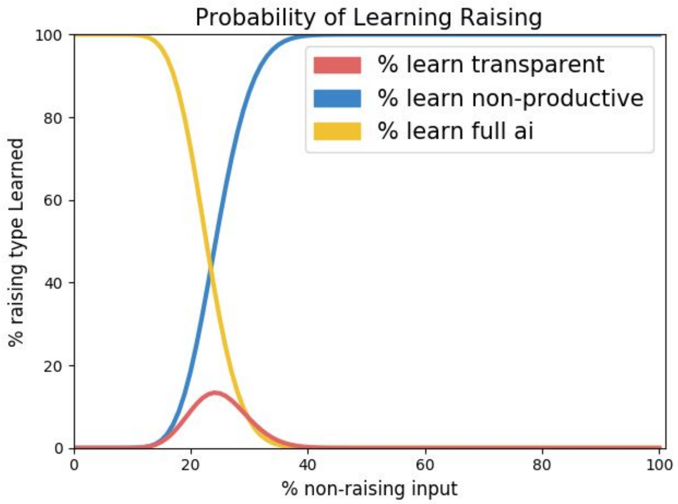
- As it evolves, it rapidly falls off the band of viability depending on the exact starting condition
- Transparent raising dies out and never becomes common



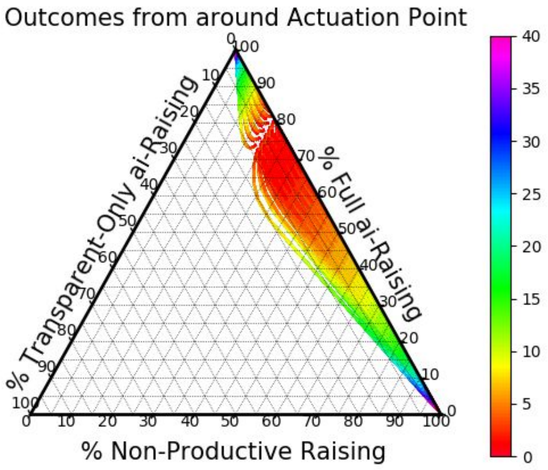
Summary

Transparent /aɪ/-raising as a contact phenomenon

- **Can emerge from mixed raising/non-raising input**
- **Attested at boundaries of the raising region**



Summary



Transparent /aɪ/-raising as a contact phenomenon

- **Can emerge from mixed raising/non-raising input**
- **Attested at boundaries of the raising region**

Transparent /aɪ/-raising is ephemeral

- **Transparent raising populations should rapidly transition away**
- **Consistent with rarity of attestation**
- **Transparent raisers are not expected in the earliest phases of change**

Future Directions

Future Directions

Directed search for transparent raisers

- **There is no aggregate transparent raising in Fruehwald’s data. Are there transparent raiser individuals in the PNC?**
- **Lab-based methods (Berkson et al’s contribution) may prove critical for finding transparent raisers at large**

Future Directions

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**
- **In Fort Wayne for the /aɪ/-raising case or elsewhere for other problems**

Future Directions

Implications for phonological change

- **Can the emergence of other phonological patterns be explained this way?**
- **The development of “simpler” short-*a* tensing systems across the US may be a good case study (building on Sneller et al)**

End

Acknowledgements:

- Meredith Tamminga
- Charles Yang

- NDSEG Fellowship (US Army Research Office)
- NSF GRFP Fellowship DGE-1321851

Extra Slides

Lexical Exceptions

Learning Lexical Exceptions

- **The Tolerance Principle was initially developed to handle questions like this**
- **If the tolerance threshold is exceeded, evidence for the would-be generalization is still learned, but lexically as exceptions**

For example,

Someone who hears *high school* as /hʌɪskul/ but no other raising is expected to learn the raised variant of *high school* regardless

It would be surprising under the TP if non-raisers never exhibited lexical raising

Hence the use of "non-productive raising" instead of "non-raising" in this talk

Extra Slides

More on Phonological Acquisition

Child Language Acquisition and Phonology

Child must accumulate enough evidence in input

- **Requires cognitive ability and morpho/syntactic/semantic knowledge to recognise it as evidence (Pierrehumbert 2003)**
- **e.g. learn English is not stress-initial ~2 years (Legate & Yang 2012), acquire Finnish vowel harmony 2;6-3 years (Kulju & Savinainen-Makkonen 2008)**

Child Language Acquisition and Phonology

Child must accumulate enough evidence in input

- **Requires cognitive ability and morpho/syntactic/semantic knowledge to recognise it as evidence (Pierrehumbert 2003)**
- **e.g. learn English is not stress-initial ~2 years (Legate & Yang 2012), acquire Finnish vowel harmony 2;6-3 years (Kulju & Savinainen-Makkonen 2008)**

/aɪ/-raising not learnable very early

- **Raisable and unraised forms both generally absent from initial small (50-word, 100-word,...) vocabularies**
- **By the age (non)raising enters lexicon, assume lower-level phonology already in place**

Extra Slides

More on the Tolerance Principle

The Tolerance Principle

Children generalise about their language

- **even with exceptions to the generalisations**

V[PAST] = V + *ed*

run[PAST] = r*a*n

Tolerance Principle - when are generalisations learnable?

The Tolerance Principle

Children generalise about their language — **V[PAST] = V + *ed***

- **even with exceptions to the generalisations** — **run[PAST] = r*a*n**

Tolerance Principle - when are generalisations learnable?

Processing efficiency! Generalise if that speeds processing. Don't if prevalence of exceptions would slow processing - more exceptions than tolerable (Yang 2016)

- **Frequency-correlated lexical access**
- **Elsewhere condition - exceptional forms processed before generalisations**
- **Language use is characterised by Zipf's law**

Modelling assumptions

The Tolerance Principle (Yang 2016) is derived from the following observations on language processing and use:

- **Frequency-correlated lexical access**
  - **relative frequency/rank (Murray & Forster 2004)**
  - **Exceptions to generalisations tend towards higher frequencies (*be, fish*)**
- **Elsewhere condition: exceptional forms processed first (by frequency rank) before broader generalisations are otherwise applied**
- **Language use characterised by Zipf's Law - many tokens belong to very frequent types; many other types are very infrequent**

Role in acquisition

Tolerance Principle is an evaluation metric on linguistic hypotheses

- **The grammars subject to evaluation must enter hypothesis space via input data, internal factors, etc....**
- **Active during course of child acquisition; tolerability of a generalisation can change as the lexicon it applies to changes**

Favours learning generalisations with small *N* (young child lexicon), before acquiring adults' potential long tail of infrequent exceptions

Empirical Support

- **Case studies spanning phonology, morphology, and syntax (Yang 2016)**
- **Counting ability emerges when vocabulary supports productive successor function - differs predictably by language**
- **English lexical stress compatible with stress-initial grammar until ~2 yrs; change reflected in children's productions (Legate & Yang 2012)**
- **Artificial language learning predicts children's categorical generalisation behaviour from individuals' unique lexicon (Schuler 2017)**

Extra Slides
Rough Comparison with Berkson et al

Comparing to Berkson et al's Experimental Sample

These numbers are not directly comparable and cannot be construed as anything more than a sanity check

Berkson et al's sample *N*=27			Our Model (*de novo*; B+B > 5)
Pattern 0*			
Pattern 1	37% (Patterns 0* and 1)	non-productive	48%
Pattern 2	33%	transparent	13% (@ max)
Pattern 3	30%	canonical	39%

*Patterns 0 and 1 are virtually identical in our model and CDS corpus