Investigating Acquisition

in Unattested Dead Languages

Jordan Kodner
University of Pennsylvania

LSA, New York
Jan. 5, 2019

Goals

Analogical Change

Begin to develop a mechanism grounded in child language acquisition

Plausibility Testing in Historical Linguistics

Concretely state the assumptions that go into historical linguistic hypotheses

A Concrete Application

Understand the Proto-Germanic strong verb's lengthened *ē-grade as a case study

Analogical Change

- Change by analogy
- A classic mode of language change
- Erratic and irregular, contrary sound change
- Hard to explain, but often easy to identify

What I Mean By Analogical Change

Four-Part Analogy

dog : *dog-s*

cat : *cat-s*

cow : X=*cow-s* (replacing earlier *kine*)

Tendencies of Analogy

Quantitative and descriptive work has focused on cataloguing typological tendencies

Some Examples: Kuryłowicz's Laws[1]

1. Bipartite markers replace simpler ones
2. Analogy is from the "basic" to the "subordinate" within their sphere of usage
3. Basic+subordinate structures serve as the basis for later basic ones
4. When a new (analogical) and older form coexist, the new one is productive
5. Marginal distinctions are eliminated in favor of more significant ones
6. Analogized forms may be borrowed from prestige dialects

[1]Paraphrased

Tendencies of Analogy

But tendencies are often violated, they do not explain analogy, and they do not account for individual cases

Some Examples: Kuryłowicz's Laws[1]

1. Bipartite markers replace simpler ones
2. Analogy is from the "basic" to the "subordinate" within their sphere of usage
3. Basic+subordinate structures serve as the basis for later basic ones
4. When a new (analogical) and older form coexist, the new one is productive
5. Marginal distinctions are eliminated in favor of more significant ones
6. Analogized forms may be borrowed from prestige dialects

[1]Paraphrased

Analogical Change as Productivity

- It is clear that productivity plays a major role in analogical change, but it is unclear how[1]

[1]Hock 2003

Analogical Change as Productivity

- It is clear that productivity plays a major role in analogical change, but it is unclear how[1]
- Productivity learning is an issue within the scope of child language acquisition

[1]Hock 2003

Proto-Germanic Strong Verbs

PGmc Strong Verbs Overview

- Overall comparable to modern Germanic languages’
- Four principle parts:
  - present, past 3sg, past default, past participle
- Seven classes (I-VII)
- I-VI are transparently defined by root shape
- A few hundred roots are securely reconstructable
  - Common, but not quite as common as weak verbs

The Strong Verb Paradigm

	Root	Present	Past 3sg	Pastdefault	PParticiple	Trans
I	**-īC-*	**bītanǫ*	**bait*	**bitun*	**bitanaz*	‘bite’
II	**-euC-*	**teuhanǫ*	**tauh*	**tugun*	**tuganaz*	‘pull’
III	**-eCC-*	**helpanǫ*	**halp*	**hulpun*	**hulpanaz*	‘help’
IV	**-eR-*	**beranǫ*	**bar*	**bērun*	**buranaz*	‘carry’
V	**-eT-*	**gebanaz*	**gab*	**gēbun*	**gebanaz*	‘give’
VI	**-aC-*	**faranǫ*	**fōr*	**fōrun*	**faranaz*	‘travel’

C = Consonant; R = Sonorant; T = Obstruent

Ancestral PIE Ablaut Grades

	Present	Past 3sg	Past	PParticiple
I	*e*-grade	*o*-grade	zero-grade	zero-grade
II	*e*-grade	*o*-grade	zero-grade	zero-grade
III	*e*-grade	*o*-grade	zero-grade	zero-grade
IV	*e*-grade	*o*-grade	*ē*-grade	zero-grade
V	*e*-grade	*o*-grade	*ē*-grade	*e*-grade

A Long-Standing Problem!

[1]Summary of Mailhammer 2007

Previous Accounts[1]

Phonological Accounts

- Rectifying stems after reduplication was lost (eg **g^heg^hb-* → **gb-*) (Streitberg 1896, Schumacher 2005)
- Compensatory lengthening (Hirt 1931)

Analogical Accounts

- Some kind of old aorist (Sverdrup 1927, Prokosch 1939, Cowgill 1957)
- Length analogy with Class VI *ō*-grade (eg Kuryłowicz 1968, Meid 1971, Bammesberger 1986)
- Analogical spread from **etanǫ* 'eat' (Kortlandt 1992, Schumacher 1998, 2005, Mottausch 2000, Ringe 2006, Mailhammer 2007)
- From the nominal system (Bammesberger 1994, 1996)

Other Accounts

- Brugmann 1913's second perfect formation (Matzel 1970, Meid 1971)

Why We can Study Unattested Dead Learners

Reconstructed Lexicons as Child Lexicons

For this enterprise to work, we need to use reconstructed lexicons as stand-ins for child lexicons

Reconstructed Lexicons as Child Lexicons

For this enterprise to work, we need to use reconstructed lexicons as stand-ins for child lexicons

What we Cannot Do

- **Run laboratory experiments (no access to children)**
- **Use token frequency info (no access to corpora)**

Reconstructed Lexicons as Child Lexicons

For this enterprise to work, we need to use reconstructed lexicons as stand-ins for child lexicons

What we Cannot Do

- Run laboratory experiments (no access to children)
- Use token frequency info (no access to corpora)

What we Can Do

- Use type counts (what lexicons have)
- Use rough translations (can be reconstructed)

Reconstructed Lexicons as Child Lexicons

For this enterprise to work, we need to use reconstructed lexicons as stand-ins for child lexicons

We're good to go if we can show that the size and semantic composition of the PGmc lexicon is similar to (our approximations of) child lexicons

What we Cannot Do

- Run laboratory experiments (no access to children)
- Use token frequency info (no access to corpora)

What we Can Do

- Use type counts (what lexicons have)
- Use rough translations (can be reconstructed)

Lexicon Size

Children

- 3-year-olds know a **couple thousand lemmas** ***at most***[1]

PGmc Reconstruction

- There are a **couple thousand** "securely" reconstructable lemmas (your mileage may vary)

[1]Hart & Risley 2003

Lexicon Size

Children

- 3-year-olds know a **couple thousand lemmas** ***at most***[1]
- There are **358** frequent verbs (lemmas occurring ≥10 times) in Brown (CHILDES) child-directed speech
- **CDS from CHILDES is often used to approximate child knowledge**

PGmc Reconstruction

- There are a **couple thousand** "securely" reconstructable lemmas (your mileage may vary)
- There are **~258** securely reconstructable strong verb lemmas[2]

[1]Hart & Risley 2003
[2]More accurately, Don Ringe extracted a superset from Seebold 1979 and I sorted through that

Lexicon Size

Children

- 3-year-olds know a couple thousand lemmas *at most*[1]
- There are 358 frequent verbs (lemmas occurring ≥10 times) in Brown (CHILDES) child-directed speech
- CDS from CHILDES is often used to approximate child knowledge

PGmc Reconstruction

- There are a couple thousand "securely" reconstructable lemmas (your mileage may vary)
- There are ~258 securely reconstructable strong verb lemmas[2]

[1]Hart & Risley 2003
[2]More accurately, Don Ringe extracted a superset from Seebold 1979 and I sorted through that

Semantic Content

- **Tabulated the number of PGmc strong verb with translations among the 358 CHILDES verbs**

Example Matches

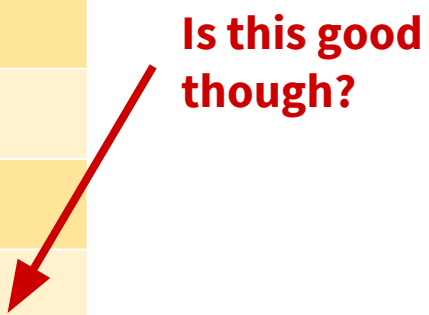
bītaną	**‘bite’**	**bite**
grētaną	**‘weep’**	**cry**
wringaną	**‘twist’**	**turn, roll, screw**
draganą	**‘haul’**	**pull, carry**
fanhaną	**‘seize’**	**take, steal**

English CHILDES → *Proto-Germanic

	#PGmc	#EN→PGmc	%
I	41	30	73.2
II	40	29	72.5
III	51	45	88.2
IV	16	13	81.3
V	28	21	75.0
VI	29	23	79.3
VII	53	41	77.4
Total	258	202	78.3

English CHILDES → *Proto-Germanic

	#PGmc	#EN→PGmc	%
I	41	30	73.2
II	40	29	72.5
III	51	45	88.2
IV	16	13	81.3
V	28	21	75.0
VI	29	23	79.3
VII	53	41	77.4
Total	258	202	78.3



Why that Number is Good Enough

- **Reran the experiment with 300 verbs occurring ≥10 times in Spanish FernAguado+PineOrea+Hess+Remedi+Romero+SerraSole (CHILDES)**
  - **Used translations provided by the corpora**
  - **Compared English → PGmc, Spanish → PGmc, English → Spanish**

English, Spanish, and Proto-Germanic

	#PGmc	#EN→PGmc	%	#ES→PGmc	%	#EN→ES	%
I	41	30	73.2	30	73.2		
II	40	29	72.5	33	82.5		
III	51	45	88.2	35	68.6		
IV	16	13	81.3	12	75.0		
V	28	21	75.0	21	75.0		
VI	29	23	79.3	21	72.4		
VII	53	41	77.4	34	64.2		
Total	258	202	78.3	186	72.1	234	77.8

English, Spanish, and Proto-Germanic

	#PGmc	#EN→PGmc	%	#ES→PGmc	%	#EN→ES	%
I	41	30	73.2	30	73.2		
II	40	29	72.5	33	82.5		
III	51	45	88.2	35	68.6		
IV	16	13	81.3	12	75.0		
V	28	21	75.0	21	75.0		
VI	29	23	79.3	21	72.4		
VII	53	41	77.4	34	64.2		
Total	258	202	78.3	186	72.1	234	77.8

The PGmc lexicon is “in the space” of child lexicons for our purposes

English, Spanish, and Proto-Germanic

	#PGmc	#EN→PGmc	%	#ES→PGmc	%	#EN→ES	%
I	41	30	73.2	30	73.2		
II	40	29	72.5	33	82.5		
III	51	45	88.2	35	68.6		
IV	16	13	81.3	12	75.0		
V	28	21	75.0	21	75.0		
VI	29	23	79.3	21	72.4		
VII	53	41	77.4	34	64.2		
Total	258	202	78.3	186	72.1	234	77.8

The Learning Model

Learning Productivity in Morphology

Need a model for productivity learning

- One that operates on type frequencies
- And is motivated by acquisition research
- And has found synchronic empirical success

Learning Productivity in Morphology

Need a model for productivity learning

- One that operates on type frequencies
- And is motivated by acquisition research
- And has found synchronic empirical success

The Tolerance Principle

The Tolerance Principle[1]

- An evaluation metric over linguistic hypotheses

[1]Yang, 2016

The Tolerance Principle[1]

- An evaluation metric over linguistic hypotheses
- Derived from
  - an Elsewhere Condition for 'rules' and 'exceptions'[2] (Anderson 1969 inter alia)
  - frequency-rank correlated lexical access[3] (Murray & Forster 2004)
  - Generally Zipfian input distributions

[1]Yang, 2016 [2]Anderson, 1969 *inter alia* [3]Murray & Forster 2004

The Tolerance Principle[1]

- An evaluation metric over linguistic hypotheses
- Derived from
  - an Elsewhere Condition for 'rules' and 'exceptions'[2] (Anderson 1969 inter alia)
  - frequency-rank correlated lexical access[3] (Murray & Forster 2004)
  - Generally Zipfian input distributions
- Successfully applied to a wide range of problems
  - Modern English strong verbs, German noun plurals, Russian and Polish genitives
  - English diatones, American sociolinguistic variables
  - English and Mandarin numeracy, etc.
- And psychological backing from artificial language learning experiments[4]

[1]Yang, 2016 [2]Anderson, 1969 *inter alia* [3]Murray & Forster 2004 [4]Schuler et al, 2016

The Tolerance Principle

- **Given a hypothesized generalization *R* operating over a class *C*, quantitatively define the number of exceptions below which the generalization is tenable**

The Tolerance Principle

- Given a hypothesized generalization *R* operating over a class *C*, quantitatively define the number of exceptions below which the generalization is tenable

$N = |C|$

$e = |\text{exceptions to } R \in C|$

Exceptions are tolerable if

$$e < N / \ln N$$

The Tolerance Principle

- Given a hypothesized generalization *R* operating over a class *C*, quantitatively define the number of exceptions below which the generalization is tenable

$N = |C|$

$e = |\text{exceptions to } R \in C|$

Exceptions are tolerable if

$$e < N / \ln N$$

- If it holds, the child can try a broader generalization (larger *C*)

Tolerance Principle and Representation

- Words can be associated with generalizations governing their derivations or memorized as word-derivation pairs
- Rule = productive; memorization = non-productive
- So learning a rule is tantamount to hypothesizing productivity

Productive generalizations will be extended to unseen forms

N Varies during Individual Development

- ***N*** **is the number of class members a child has learned *so far***
- ***N*** **and *e* grow as the learner's vocabulary grows**

N Varies during Individual Development

- *N* is the number of class members a child has learned *so far*
- *N* and *e* grow as the learner's vocabulary grows
- Children fall into and out of productivity during development
- Which explains overgeneralization errors attested in child speech
- This is an avenue for historical analogy

Accounting for the *ē-Grade with Acquisition

Explicanda

Positives

- Where is **ē* from originally?
- Why did **ē* spread from V to IV?

Negatives

- Why did ***u* not spread from IV to V?
- Why did **ē* not spread from IV+V to III?
- Why did **u* not spread from III to IV or V?
- Why did the past 3sg and pparticiple stem vowels not spread?

Explicanda

Positives

- Where is **ē* from originally?
- **Why did **ē* spread from V to IV?**

Negatives

- **Why did ***u* not spread from IV to V?**
- **Why did **ē* not spread from IV+V to III?**
- **Why did **u* not spread from III to IV or V?**
- Why did the past 3sg and pparticiple stem vowels not spread?

Avenue for Overgeneralization

- Classes IV has a root shape ***-eR-*** which defined the class for the purposes of the Tolerance Principle
- Class V has a root shape ***-eT-***
- There exists a generalization covering exactly IV+V: ***-eC-***

Avenue for Overgeneralization

- Classes IV has a root shape ***-eR-** which defined the class for the purposes of the Tolerance Principle
- Class V has a root shape ***-eT-**
- There exists a generalization covering exactly IV+V: ***-eC-**
- **The learner needs to pick the appropriate generalization:**
  - IV's rule applies to ***-eR-** and V's to ***-eT-**
  - IV's rule applies to ***-eC-**, and learned V pasts are exceptions (**IV forms in V**)
  - V's rule applies to ***-eC-**, and learned IV pasts are exceptions (**V forms in IV**)

Avenue for Overgeneralization

- Classes IV has a root shape ***-eR-** which defined the class for the purposes of the Tolerance Principle
- Class V has a root shape ***-eT-**
- There exists a generalization covering exactly IV+V: ***-eC-**
- **The learner needs to pick the appropriate generalization:**
  - IV's rule applies to ***-eR-** and V's to ***-eT-**
  - IV's rule applies to ***-eC-**, and learned V pasts are exceptions (IV forms in V)
  - V's rule applies to ***-eC-**, and learned IV pasts are exceptions (V forms in IV)
- Kuryłowicz's 2nd Law "within their 'sphere of usage'"

V to IV+V

- According to the TP, a child who knows all **Class IV** and **V** verbs will learn **two distinct classes**
- The hypothesis that one class's learned pasts are just exceptional cases to the other's rule cannot hold

V to IV+V

- According to the TP, a child who knows all **Class IV** and **V** verbs will learn **two distinct classes**
- The hypothesis that one class's learned pasts are just exceptional cases to the other's rule cannot hold

$N = |\text{IV+V}| =$ 44

V to IV+V

- According to the TP, a child who knows all **Class IV** and **V** verbs will learn **two distinct classes**
- The hypothesis that one class's learned pasts are just exceptional cases to the other's rule cannot hold

$N = |\text{IV+V}| = 44$

$e = \min(|\text{IV}|,|\text{V}|) = 16$

V to IV+V

- According to the TP, a child who knows all **Class IV** and **V** verbs will learn **two distinct classes**
- The hypothesis that one class's learned pasts are just exceptional cases to the other's rule cannot hold

$N = |\text{IV+V}| = 44$

$e = \min(|\text{IV}|,|\text{V}|) = 16$

$N / \ln N = 11.6$

V to IV+V

- According to the TP, a child who knows all Class IV and V verbs will learn two distinct classes
- The hypothesis that one class's learned pasts are just exceptional cases to the other's rule cannot hold

$N = |\text{IV+V}| = 44$

$e = \min(|\text{IV}|,|\text{V}|) = 16$

$N / \ln N = 11.6$

16 > 11.6.

IV+V FAILS!

V to IV+V

- Imagine a younger child
- Say, one who knows 5 Class IV verbs and 9 Class V verbs

V to IV+V

- Imagine a younger child
- Say, one who knows 5 Class IV verbs and 9 Class V verbs

$N = 5+9 =$	14
$e =$	5
$N / \ln N =$	5.3

5 < 5.3.

IV+V SUCEEDS!

V to IV+V

- Imagine a younger child
- Say, one who knows 5 Class IV verbs and 9 Class V verbs

Great, but how plausible is this state?

$N = 5+9 =$	14
$e =$	5
$N / \ln N =$	5.3

5 < 5.3.

IV+V SUCEEDS!

Likelihood of Overgeneralizations

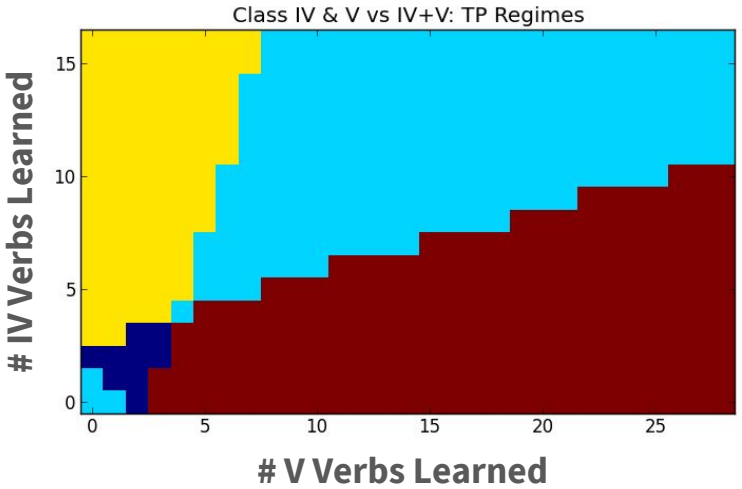
Given two classes V and IV of sizes *K* and *N-K* and a plausible generalization between them, there are 4 possible outcomes

- **Separate rules for V and IV**
- **Rule V for IV+V**
- **Rule IV for IV+V**
- **Rule V or IV for IV+V**

Likelihood of Overgeneralizations

Given two classes V and IV of sizes *K* and *N-K* and a plausible generalization between them, there are 4 possible outcomes

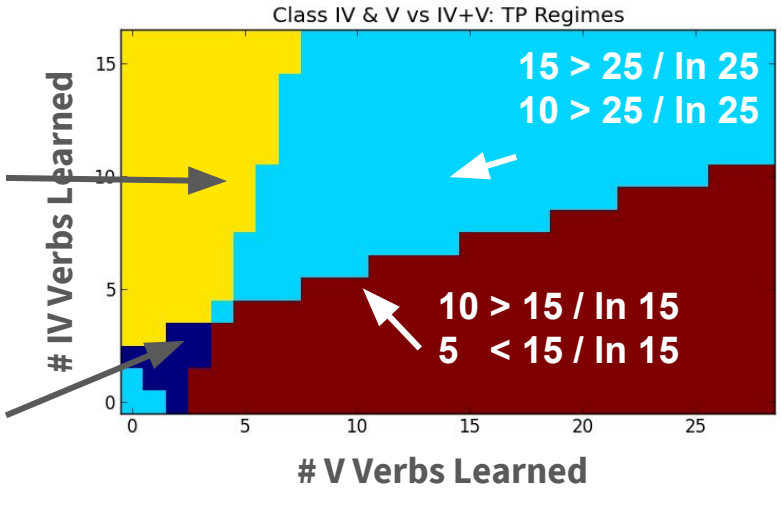
- **Separate rules for V and IV**
- **Rule V for IV+V**
- **Rule IV for IV+V**
- **Rule V or IV for IV+V**



Likelihood of Overgeneralizations

Given two classes V and IV of sizes *K* and *N-K* and a plausible generalization between them, there are 4 possible outcomes

- **Separate rules for V and IV**
- **Rule V for IV+V**
- **Rule IV for IV+V**
- **Rule V or IV for IV+V**

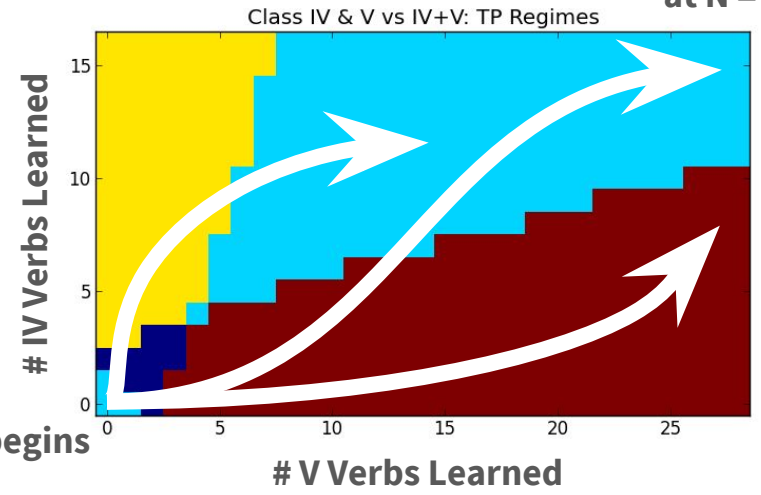


Likelihood of Overgeneralizations

Children progress along paths through this space

- Separate rules for V and IV
- Rule V for IV+V
- Rule IV for IV+V
- Rule V or IV for IV+V

Mature learner
at N = |IV ∪ V|

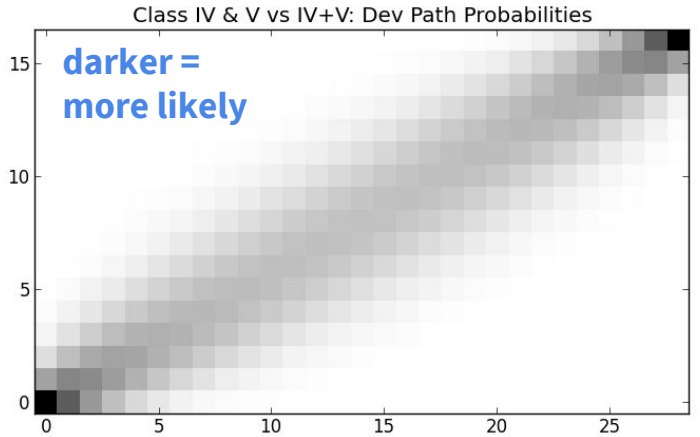


Fresh child begins
at N=0

Likelihood of Overgeneralizations

Likelihood of landing in each state modeled as a hypergeometric distribution ie drawing marbles without replacement[1]

- $N_{total} = |V \cup IV| = 44$
- $|IV| = 16$
- $|V| = 28$

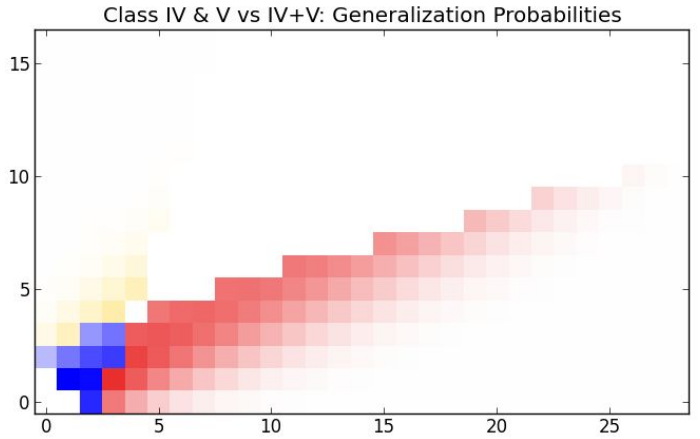


[1]If one class tends to be much more common than the other, this "line" will bow up or down

Likelihood of Overgeneralizations

Composing the previous two plots visualizes likelihood of each kind of overgeneralization

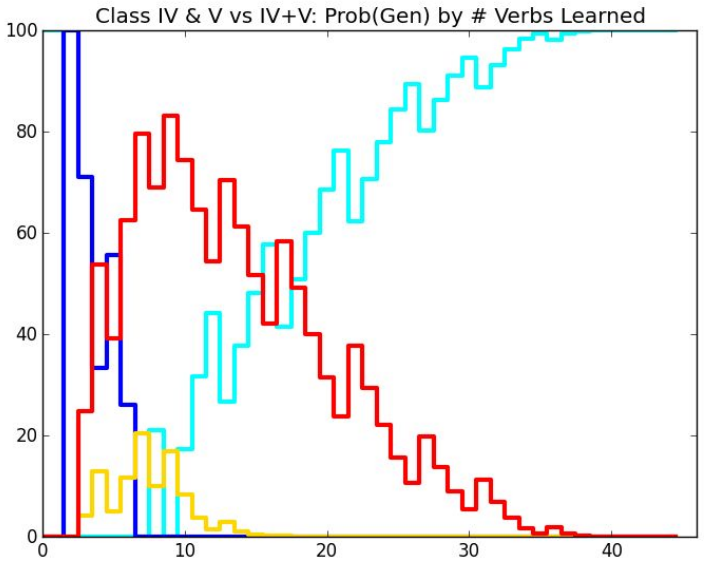
- **Rule V for IV+V (V→IV analogy)**
- **Rule IV for IV+V (IV→V analogy)**
- **Rule V or IV for IV+V (either)**



Likelihood of Overgeneralizations

Plotting *N* by likelihood of each state

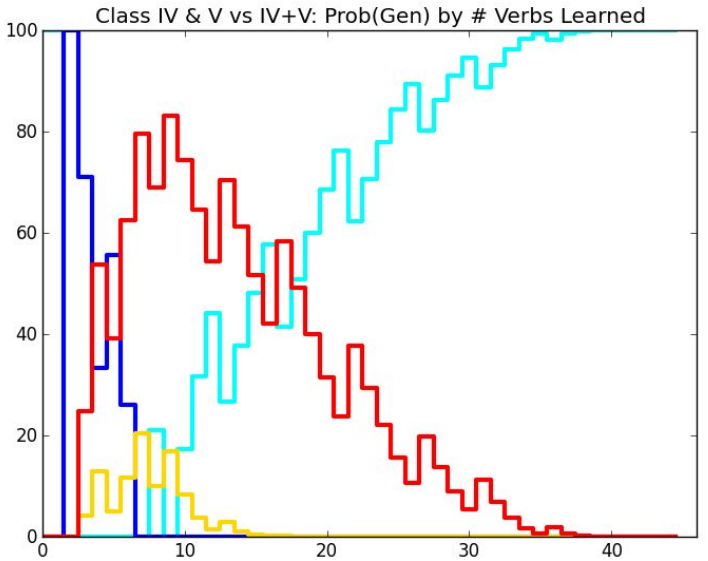
- Separate rules for V and IV
- Rule V for IV+V (V→IV analogy)
- Rule IV for IV+V (IV→V analogy)
- Rule V or IV for IV+V (either)



Likelihood of Overgeneralizations

Area under the curves ≈ proportion of time spent in state[1] ≈ proportion of learners in state[2]

- **64.3% (wins by the end)**
- **27.2% (dominant early, trails)**
- **2.2% (present early only)**
- **6.4% (dominant very early)**



[1]Related to learning rate
[2]Related to population structure

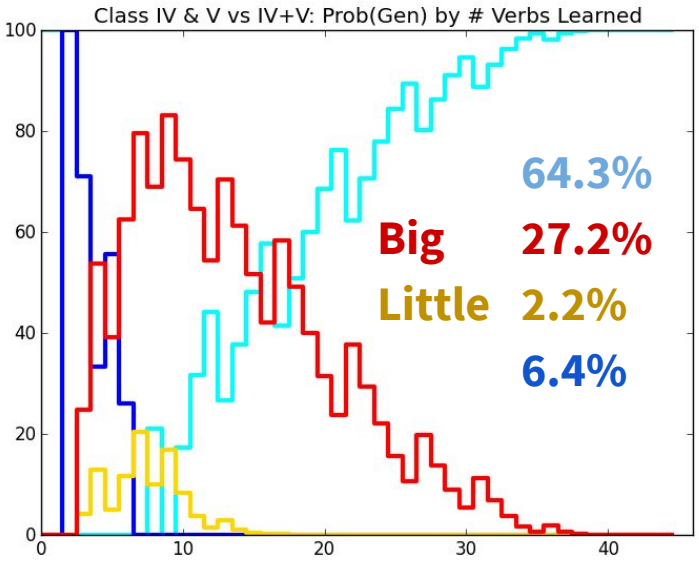
Generalization between IV+V and III

- IV+V is defined by ***-eC-***
- III is defined by ***-eCC-***
- There exists a generalization ***-eC(C)-*** that encompases exactly III+IV+V

Comparing V → IV+V and IV+V → III+IV+V

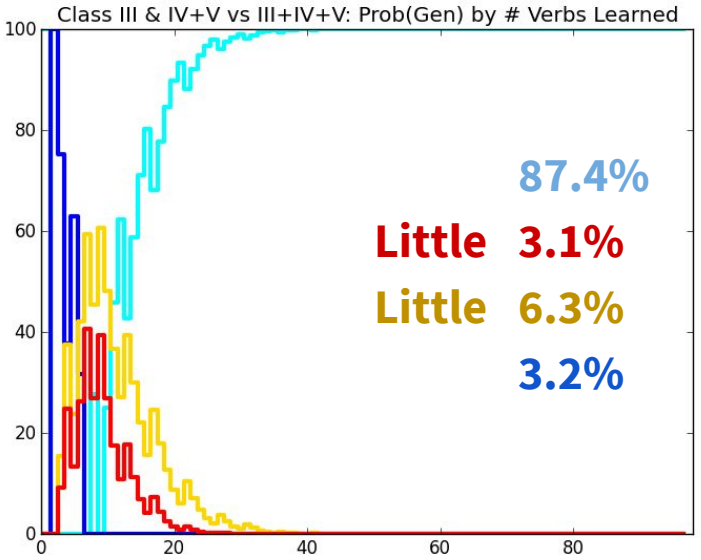
V → IV+V

|IV| = 16, |V| = 28



IV+V → III+IV+V

|III| = 52, |IV+V| = 44



Comparing V → IV+V and IV+V → III+IV+V

- Productivity provides the avenue for analogy
- Some overgeneralizations are more likely than others

Kuryłowicz's 4th Law "the newer option is productive"

Comparing V → IV+V and IV+V → III+IV+V

- **Productivity provides the avenue for analogy**
- **Some overgeneralizations are more likely than others**

Given the Proto-Germanic lexicon,

- **V→IV+V is much more likely than IV→IV+V (27.2 vs 2.2%) and more persistent (some late learners could make it)**
- **IV+V→III+IV+V and III→III+IV+V were unlikely too (3.1, 6.3%)**

Child Errors → Change

The Paradox of Language Change[1]

If children are so good at acquiring language, why are they so bad at it?[2]

[1]Niyogi & Berwick 1995
[2]My paraphrase of Niyogi & Berwick 1995

The Paradox of Language Change[1]

If children are so good at acquiring language, why are they so bad at it?[2]

- **A common criticism of child-driven models of change**
- **The solution is multi-part:**
  - **Sociolinguistic factors:** variation in the input...
  - **Psycholinguistic factors:** some things are actually hard to learn...
  - **The input data itself:** it is sparse in key ways...

[1]Niyogi & Berwick 1995
[2]My paraphrase of Niyogi & Berwick 1995

The Input Data Sparsity Problem

- **Paradigm Saturation**[1] - The proportion of a verb's paradigm that is actually attested
- **Zipfian distribution** - very low average

Spanish (~1mil; UDT)

- 1st (*ir*): 54.2%
- mean: 4.7%
- median: 1.4%

Spanish(~1mil; CHILDES)[2]

1st (*decir*): 72.2%
mean: 7.9%

[1]Chan, 2008 [2]Lignos & Yang, 2018

The Input Data Sparsity Problem

- **Paradigm Saturation**[1] - The proportion of a verb's paradigm that is actually attested
- **Zipfian distribution** - very low average

Spanish (~1mil; UDT)

- 1st (*ir*): 54.2%
- mean: 4.7%
- median: 1.4%

Spanish(~1mil; CHILDES)[2]

- 1st (*decir*): 72.2%
- mean: 7.9%

The Zipfian distr. scales, so more data cannot fix. You must rely on productivity for what you haven't heard!

[1]Chan, 2008 [2]Lignos & Yang, 2018

Child Learner Analogy

Input-driven

- Assumes poorly attested, highly incomplete paradigms
- Contra Skousen 1989 et seq, Albright 2005 et seq, Kirov et al 2018...

*A*functional

- The result of the learning algorithm and the learning environment
- Functional factors are not invoked. Correlations are emergent, not causal

Empirically Grounded

- Explained in terms of a learning process that we can study today
- Both the learning model (TP) and the input data (saturation)

End

Acknowledgements:

- Don Ringe
- NDSEG (US ARO)

Implementation:

`github.com/jkodner05/PGmcTP`

The Sibling Effect

The Paradox of Language Change[1]

If children are so good at acquiring language, why are they so bad at it?[2]

- A common criticism of child-driven models of change
- The solution is multi-part:
  - Sociolinguistic factors: variation in the input...
  - Psycholinguistic factors: some things are actually hard to learn...
  - The input data itself: it is sparse in key ways...

[1]Niyogi & Berwick 1995
[2]My paraphrase of Niyogi & Berwick 1995

The Sibling Effect

- Children rarely receive input from a single source grammar
- Trivial variation is ever-present in the input

The Sibling Effect

- Children rarely receive input from a single source grammar
- Trivial variation is ever-present in the input

Imagine two incompetent peers Alice & Bob

- Maybe Alice is an older sibling to Bob?
- Alice is currently overgeneralizing and Bob is listening
- Bob receives “correct” adult tokens + Alice’s tokens
- What does Bob do?

The Sibling Effect

Is Bob Skeptical?

- Can Bob recognize Alice's incompetence?
- If so, will Bob ignore her?

The answers to these predict different behaviors

Can Bob Recognize Alice's Incompetence?

- Only if Bob has heard an adult-produced token

Can Bob Recognize Alice's Incompetence?

- **Only if Bob has heard an adult-produced token**
  - **Alice is only somewhat untrustworthy**
  - **Without an adult reference, when can he assume that Alice is wrong?**

Can Bob Recognize Alice's Incompetence?

- Only if Bob has heard an adult-produced token
  - Alice is only somewhat untrustworthy
  - Without an adult reference, when can he assume that Alice is wrong?
- **Less often than you would think!** (cf paradigm saturation)

Will Bob Ignore Alice?

- I don’t know…

Will Bob Ignore Alice?

- I don’t know…
  - How “bad” do Alice’s mistakes have to be?
  - Does relative age matter? Are 3yo’s cool to 2yo’s?

Will Bob Ignore Alice?

- I don’t know…
  - How “bad” do Alice’s mistakes have to be?
  - Does relative age matter? Are 3yo’s cool to 2yo’s?
- Likely dependent on the domain again
  - Morphological doublets

Germanic Inflectional Doublets

A persistent feature of the family

- Post-PGmc IV/V confusions
- Weak Verbs in Old/Middle English
- Modern English
    - dived/dove, sneaked/snuck, brought/brang, saw/seen…

Kuryłowicz's 4th Law "the newer option is productive"

Post-PGmc IV/V confusions

- **Shift from V to IV in Old High German**
  - **eg OHG *giprohhan* ‘spoken’ vs OE *precen***
  - **After OHG and OE diverged, so this was late**
- ****brekanǫ* ‘break’**
  - **Goth *gabrukano*, OE *brocen*, (ModE *broken*)**
- **Old English**
  - **Beowulf 2981 *dropen* ‘smitten’ vs usual *drepen* < PGmc **drepanaz* (V)**
- **E and N Gmc with IV’s pparticiple vowel in the present**
  - **eg Goth *trudan* ‘step’, ON *troða* vs OE *treden*, OHG *gitretan***

The Sibling Effect Effect

- If Bob accepts Alice’s overgeneralized tokens of IV+V,

Short-term

Long-term

The Sibling Effect Effect

- If Bob accepts Alice's overgeneralized tokens of IV+V,

Short-term

- Do these decrease the number of exceptions *e*?
- If anything, these work in favor of IV+V

Long-term

The Sibling Effect Effect

- If Bob accepts Alice's overgeneralized tokens of IV+V,

Short-term

- Do these decrease the number of exceptions *e*?
- If anything, these work in favor of IV+V

Long-term

- Even if Bob matures into separate IV and V, will adult Bob occasionally produce IV verbs with V's **ē*?
- If so, next generation will receive **competent** IV **ē* inputs

The Other Explicanda

Explicanda

Positives

- Where is **ē* from originally?
- Why did **ē* spread from V to IV?

Negatives

- Why did ***u* not spread not IV to V?
- Why did **ē* not spread from IV+V to III?
- Why did **u* not spread from III to IV or V?
- **Why did the past 3sg and pparticiple stem vowels not spread?**

Why did only Class V’s past stem form spread?

- The other stems *could* spread and have (cf WGmc)
  - So the real question is not why they did not spread, it’s **why they did not stick at the PGmc stage**

Why did only Class V's past stem form spread?

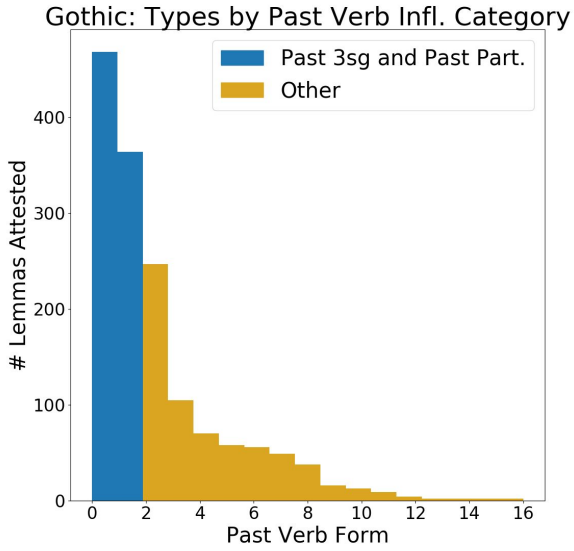
- The other stems *could* spread and have (cf WGmc)
  - So the real question is not why they did not spread, it's why they did not stick at the PGmc stage
- A matter of (type) attestation
  - Inflected forms that are not attested must be hypothesized by productive generalization
  - Ones that are well attested can be memorized whatever their forms

Why did only Class V's past stem form spread?

- The other stems *could* spread and have (cf WGmc)
  - So the real question is not why they did not spread, it's why they did not stick at the PGmc stage
- A matter of (type) attestation
  - Inflected forms that are not attested must be hypothesized by productive generalization
  - Ones that are well attested can be memorized whatever their forms
- Past 3sg and past participles tend to be among the most common inflected verbs

Why did only Class V's past stem form spread?

- The other stems *could* spread and have (cf WGmc)
  - So the real question is not why they did not spread, it's **why they did not stick at the PGmc stage**
- **Past 3sg and past participle are much better attested than other pasts in Gothic**
- **Works against analogical change in these forms**



Explicanda

Positives

- **Where is *$\bar{e}$ from originally?**
- Why did *$\bar{e}$ spread from V to IV?

Negatives

- Why did **u not spread not IV to V?
- Why did *$\bar{e}$ not spread from IV+V to III?
- Why did *u not spread from III to IV or V?
- Why did the past 3sg and pparticiple stem vowels not spread?

“The Eat Analogy”

- ***etanǫ, *ēt, *ētun, *etanaz*** **‘*eat*’ is the only Class V verb with *$\bar{e}$ by regular sound change**

  PIE *h_1e-$h_1$ód- > *ēt- > PGmc *ēt-
  PIE *h_1e-h_1d-ˊ > PGmc *ēt-

By hypothesis, it is the source for the *$\bar{e}$-grade in Classes IV and V.

Steps of the Eat Analogy

1. The *$\bar{e}$-grade spread from *eat* to the rest of Class V
2. Then the *$\bar{e}$-grade spread from Class V to Class IV

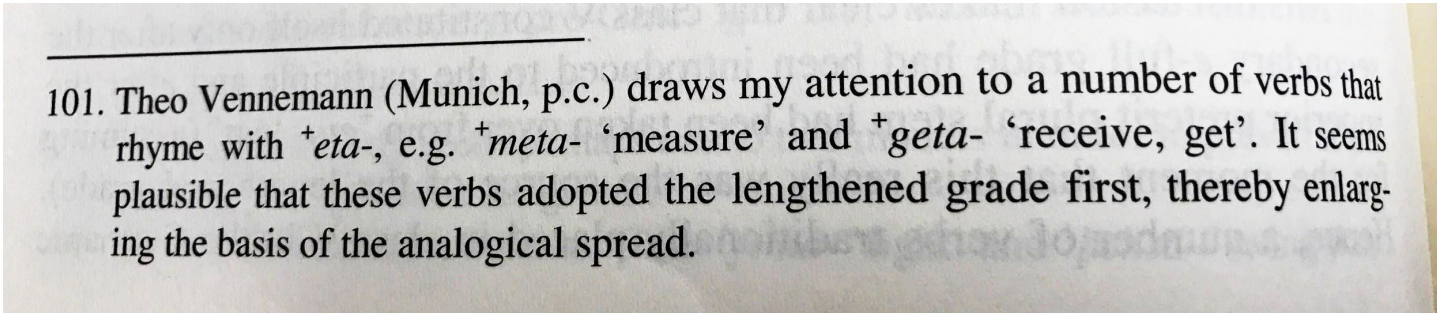
The latter point is well accepted and not specific to the Eat Analogy (eg Matzel 1970, Bammesberger 1986, Mottausch 2000, Ringe 2006)

From One to Many

- This is **not** a job for the Tolerance Principle
- But some kind of generalization is likely relevant here

From One to Many

- This is **not** a job for the Tolerance Principle
- But some kind of generalization is likely relevant here



101. Theo Vennemann (Munich, p.c.) draws my attention to a number of verbs that rhyme with +*eta*-, e.g. +*meta*- 'measure' and +*geta*- 'receive, get'. It seems plausible that these verbs adopted the lengthened grade first, thereby enlarging the basis of the analogical spread.

(Mailhammer, 2007)

From One to Many

- This is **not** a job for the Tolerance Principle
- But some kind of generalization is likely relevant here

From Four to Many

- There are 4 Class V verbs of the shape **-et-*
  - **etanǫ* ‘eat,’ **fetanǫ* ‘fall,’ **getanǫ* ‘get,’ **metanǫ* ‘measure’
- What would have to happen to spread **ē* from these to V?
- An application of **analogical extension**

Sequences of Overgeneralization

- Modeling the extension of *ē as a series of increasingly general overgeneralizations

Sequences of Overgeneralization

- Modeling the extension of *ē as a series of increasingly general overgeneralizations

Initial extension

- Are there any subclasses of V to which *ē could extend from 4 *-*et*- verbs?

Sequences of Overgeneralization

- Modeling the extension of *$\bar{e}$ as a series of increasingly general overgeneralizations

Initial extension

- Are there any subclasses of V to which *$\bar{e}$ could extend from 4 *-*et*- verbs?

Generalization	*N*	*N* / ln *N*	*e* = *N*-4
*-*e*[-voi -cont -son]-	7	3.59	3
*-*e*[-voi -son]-	19	6.45	15
*-*e*[-voi COR]-	11	4.58	7
*-*e*[-cont -son]-	12	4.83	8
*-*e*[-son COR]-	12	4.83	8

Sequences of Overgeneralization

- An extension to *-e[voiceless stop]- works!
  - **lekanǫ* ‘be leaky,’ **rekanǫ* ‘bank a fire,’ **wrekanǫ* ‘drive out’
- Nothing else quite works, but some come close

Generalization	*N*	*N* / ln *N*	*e* = *N*-4	*e* = *N*-7
*-*e*[-voi -cont -son]-	7	3.59	3	*NA*
*-*e*[-voi -son]-	19	6.45	15	13
*-*e*[-voi COR]-	11	4.58	7	*NA*
*-*e*[-cont -son]-	12	4.83	8	5
*-*e*[-son COR]-	12	4.83	8	*NA*

Sequences of Overgeneralization

- An extension to *-e[voiceless stop]- works!
- Nothing else quite works, but some come close
  - If PGmc had one extra verb, **plausible but untestable**, it would work as-is
- The same process could not facilitate spread between III and IV+V because there are no (obvious) intermediate generalizations between IV+V's *-*eC*- and III's *-*eCC*- and their joint *-*eC(C)*-
- As expected, extension is tenuous but not impossible

Hypergeometric Distribution

Likelihood of Overgeneralizations

Likelihood of landing in each state modeled as a hypergeometric distribution ie drawing marbles without replacement[1]

[1]Unweighted marbles are approximated when both classes have similar frequency distributions

Likelihood of Overgeneralizations

Likelihood of landing in each state modeled as a hypergeometric distribution ie drawing marbles without replacement[1]

- $N = |V \cup IV|$
- $K = |V|$
- $n = |\subseteq V \cup IV \text{ learned so far}|$
- $k = |\subseteq V \text{ learned so far}|$
- $n\text{-}k = |\subseteq IV \text{ learned so far}|$

[1]Unweighted marbles are approximated when both classes have similar frequency distributions

Likelihood of Overgeneralizations

Likelihood of landing in each state modeled as a hypergeometric distribution ie drawing marbles without replacement[1]

$$P(X = k) = f(k; N, K, n)$$

- $N = |V \cup IV|$
- $K = |V|$
- $n = |\subseteq V \cup IV \text{ learned so far}|$
- $k = |\subseteq V \text{ learned so far}|$
- $n\text{-}k = |\subseteq IV \text{ learned so far}|$

$$= \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

[1]Unweighted marbles are approximated when both classes have similar frequency distributions

What We Know about Unattested Dead Languages

More than You Might Think

Some unattested languages exist by logical necessity

- Related languages must have had a common ancestor
- The ancestor is family X is often called **Proto-X**

More than You Might Think

Some unattested languages exist by logical necessity

- Related languages must have had a common ancestor
- The ancestor is family X is often called **Proto-X**

Examples

- The ancestor of the Germanic languages (English, Swedish, Gothic, etc) is called **Proto-Germanic**
- The ancestor of the Indo-European languages (Proto-Germanic, Latin, Sanskrit, Hittite, Proto-Balto-Slavic, etc) is called **Proto-Indo-European**

Comparative Reconstruction

We can figure out a lot about the sounds and vocabularies of proto-languages with comparative reconstruction

Comparative Reconstruction

We can figure out a lot about the sounds and vocabularies of proto-languages with comparative reconstruction

- Methodical process that determines sound correspondences between related languages
- These correspondences define a partial ordering of sound changes
- "Unravelling" the sound changes yields the ancestral forms

Comparative Reconstruction

- Depends on the observation that sound change is overwhelmingly regular
- The more data that is available, the more secure the results will be
  - In terms of attested branches per family
  - And cognates per language

Comparative Reconstruction

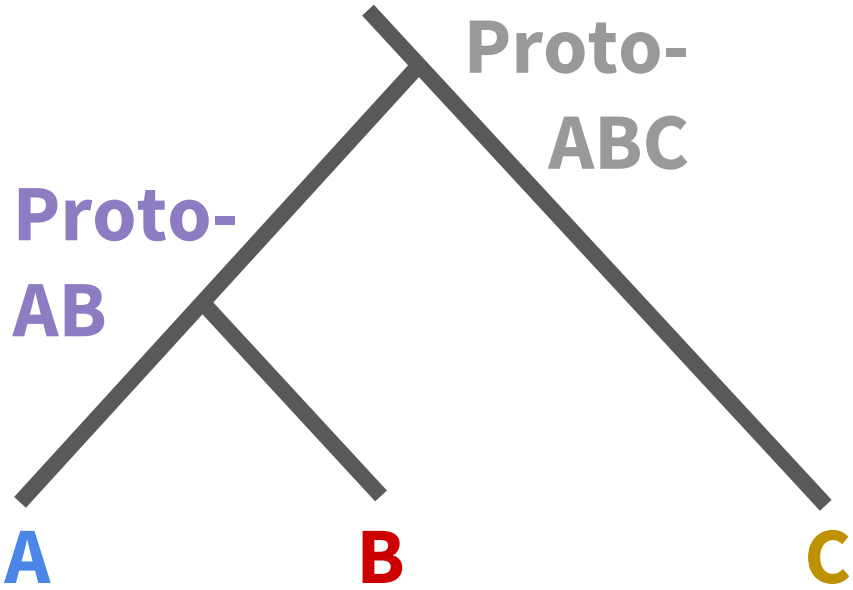
- Depends on the observation that sound change is overwhelmingly regular
- The more data that is available, the more secure the results will be
  - In terms of attested branches per family
  - And cognates per language
- Without enough data, the outcome is too unconstrained to be confident about anything

Comparative Reconstruction

Depends on comparison, given cognates from two branches of a family, forms can only be reconstructed back to their common ancestor

Comparative Reconstruction

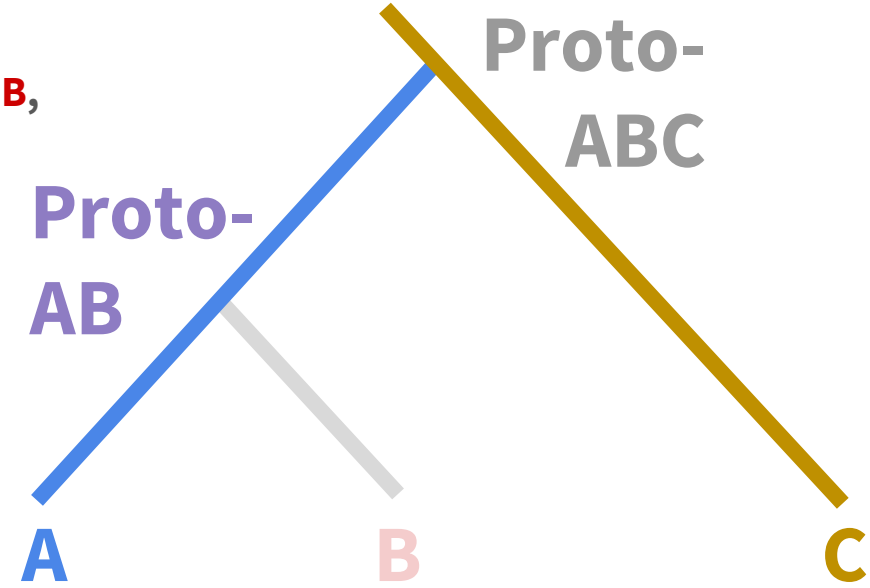
Depends on comparison, given cognates from two branches of a family, forms can only be reconstructed back to their common ancestor



Comparative Reconstruction

Depends on comparison, given cognates from two branches of a family, forms can only be reconstructed back to their common ancestor

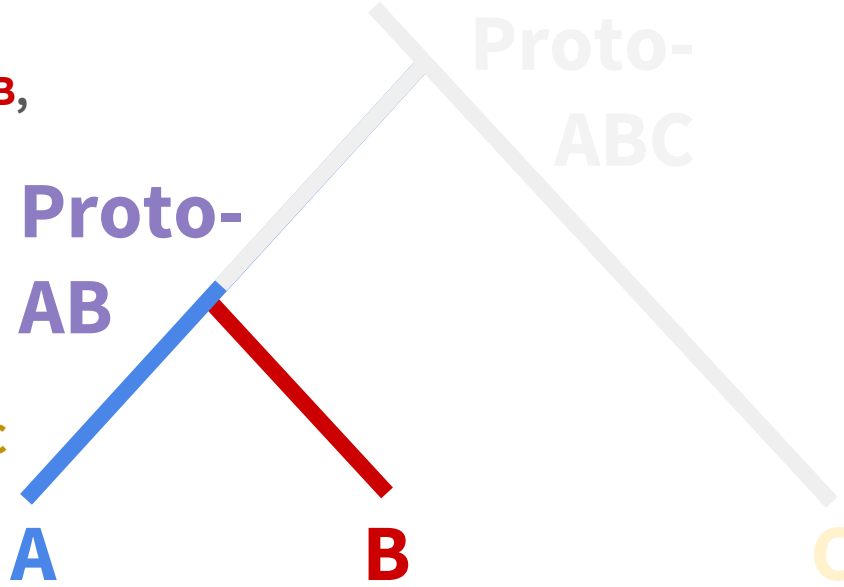
- If a form is attested in C and either A or B, it can be reconstructed to Proto-ABC



Comparative Reconstruction

Depends on comparison, given cognates from two branches of a family, forms can only be reconstructed back to their common ancestor

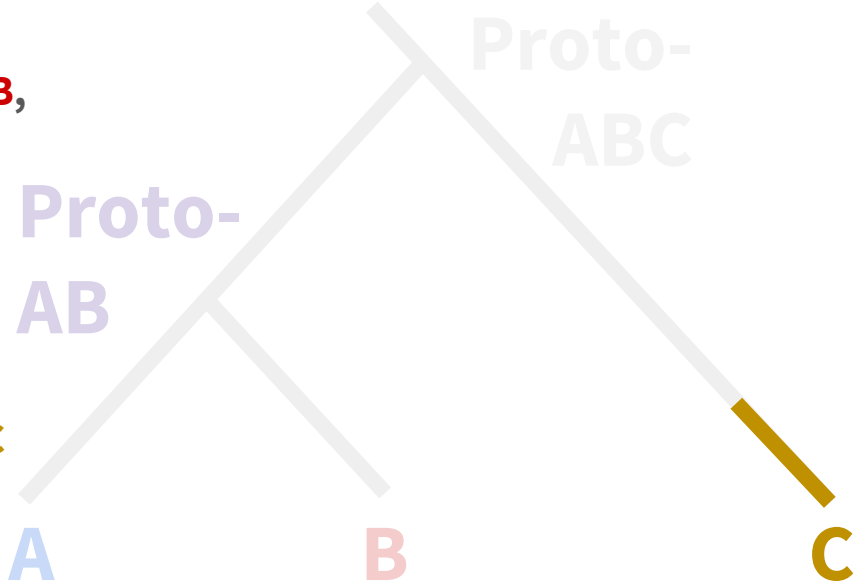
- If a form is attested in C and either A or B, it can be reconstructed to Proto-ABC
- If a form is attested in A and B, it can be reconstructed to Proto-AB but not Proto-ABC
  - Either it was innovated in AB
  - or it existed in ABC but was lost in C



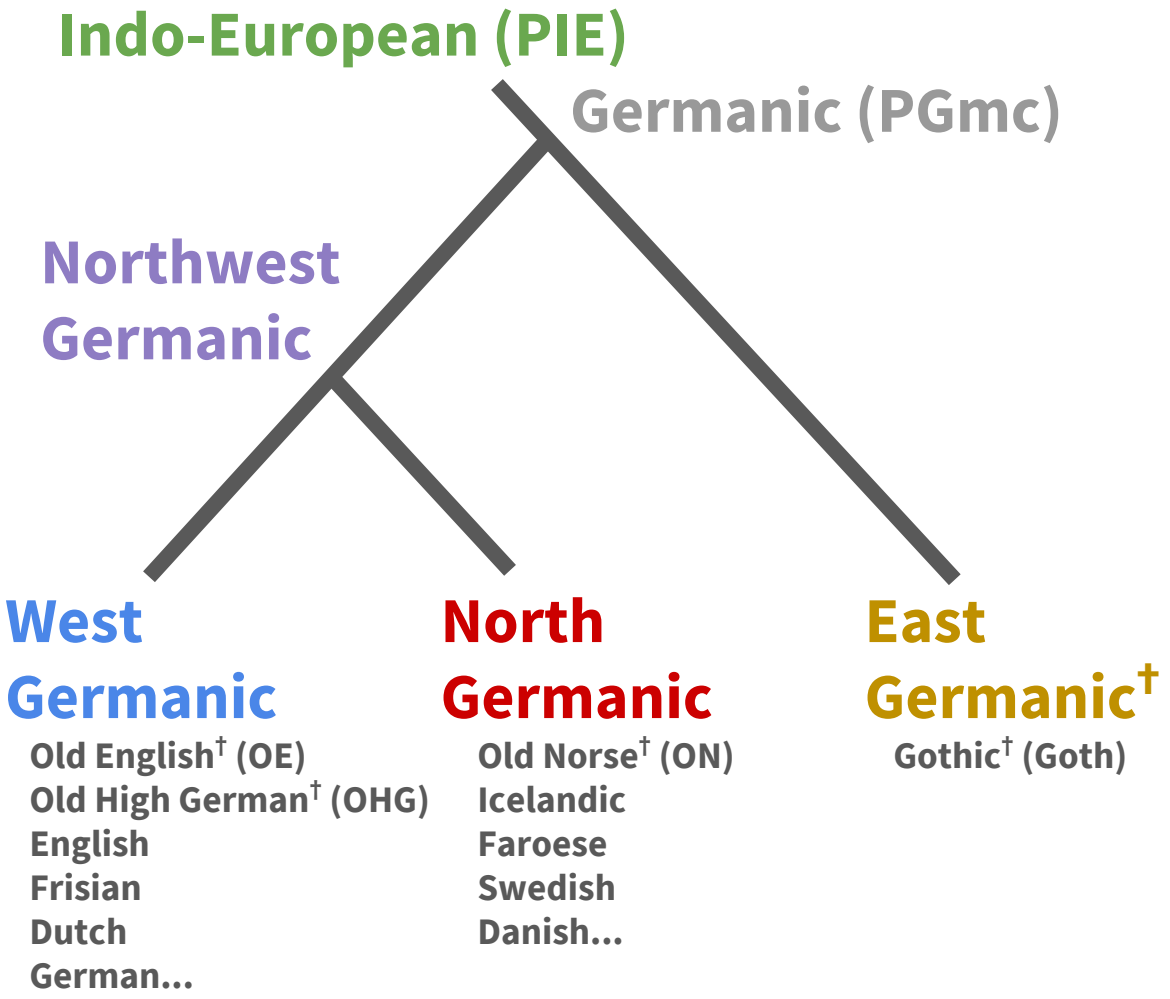
Comparative Reconstruction

Depends on comparison, given cognates from two branches of a family, forms can only be reconstructed back to their common ancestor

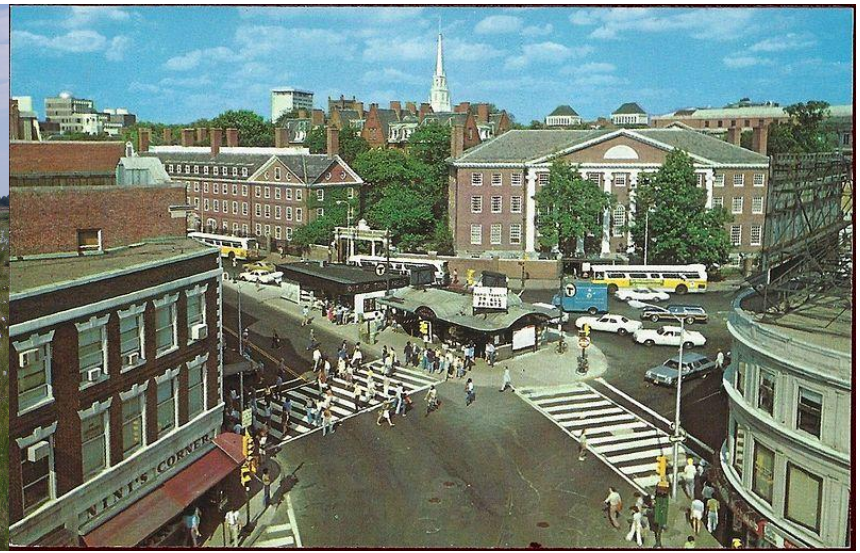
- If a form is attested in C and either A or B, it can be reconstructed to Proto-ABC
- If a form is attested in A and B, it can be reconstructed to Proto-AB but not Proto-ABC
  - Either it was innovated in AB
  - or it existed in ABC but was lost in C
- If a form it attested in only A, B, or C it cannot reconstructed



The Germanic Family



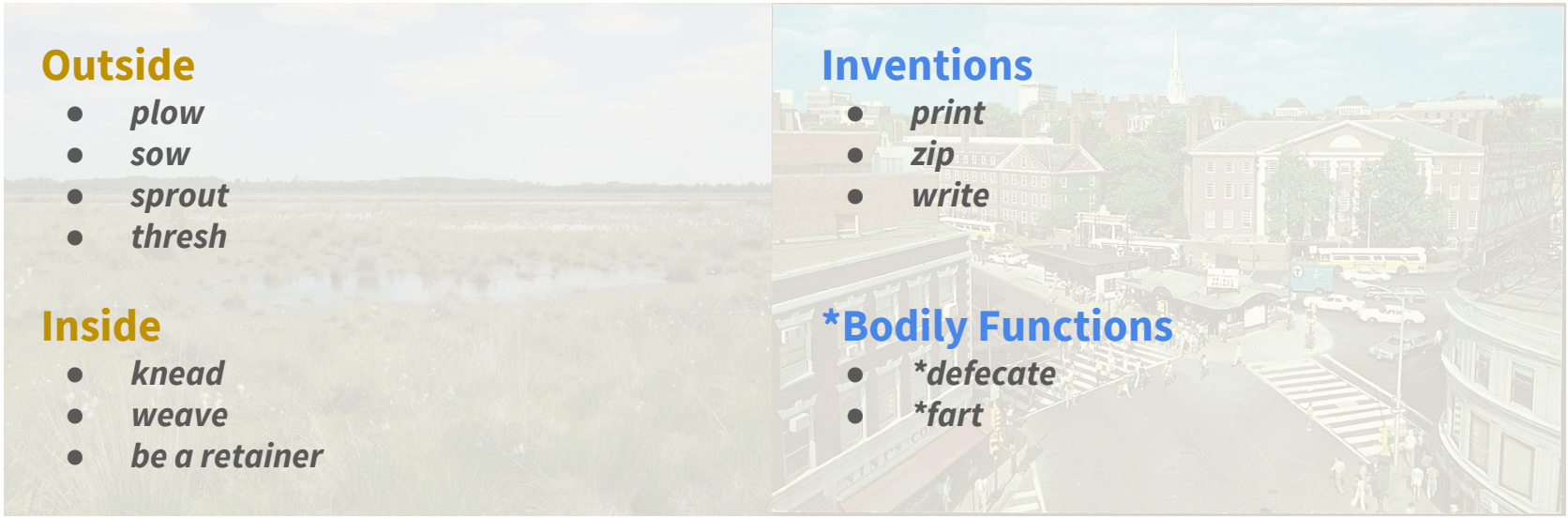
Why that Number isn't Higher



*Germanic Urheimat, 1st Millenium BC

Cambridge, MA, c. 1970

Why that Number isn't Higher



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