Investigating Acquisition

in Unattested Dead Languages

Jordan Kodner
University of Pennsylvania

GLAC 24, May 12, 2018
Pennsylvania State University

Goals

Analogical Change

Begin to develop a mechanism grounded in native acquisition and cog sci

Plausibility Testing in Historical Linguistics

Concretely state the assumptions that go into historical linguistic hypotheses

Application to Proto-Germanic

Understand the PGmc strong verb’s lengthened **ē*-grade as a case study

Proto-Germanic Strong Verbs

PGmc Strong Verbs Overview

- Verbs that undergo a stem change (ablaut)
- Overall similar 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
- Most ablaut is derived from PIE by regular sound change
- A few hundred roots are securely reconstructable
  - Common, but not quite as common as weak verbs

The Strong Verb Paradigm

	Root	Present	Past 3sg	Past Default	PParticiple	Trans
I	**-ĩC-*	**bītanq̨*	**bait*	**bitun*	**bitanaz*	‘bite’
II	**-euC-*	**teuhanq̨*	**tauh*	**tugun*	**tuganaz*	‘pull’
III	**-eCC-*	**helpanq̨*	**halp*	**hulpun*	**hulpanaz*	‘help’
IV	**-eR-*	**beranq̨*	**bar*	**bērun*	**buranaz*	‘carry’
V	**-eT-*	**gebanaz*	**gab*	**gēbun*	**gebanaz*	‘give’
VI	**-aC-*	**faranq̨*	**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

Ancestral PIE Ablaut Grades

A Long-Standing Problem!

	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

[1]Summary of Mailhammer 2007

Previous Accounts[1]

A Long-Standing Problem!

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 **etanq* '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)

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 **etaną* ‘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)

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 **etaną* '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)

Analogical Change

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 Analogy

Four-Part Analogy[1]

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

(Morphological leveling and extension)

[1]Based on Hock 2003. There are many similar typologies.

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]
- Determining productivity is an issue within the scope of child language acquisition

So...

- Let's look at analogy from the perspective of language acquisition

[1]Hock 2003

Why We can Study Unattested Dead Children

Reconstructed Lexicons as Child Lexicons

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

What we Cannot Do

- Run laboratory experiments (no access to children)
- Use token frequency information (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 accept 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 information (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 (with wide variation)[1]
- There are 358 frequent verbs (lemmas occurring ≥10 times) in Brown (CHILDES) CDS

PGmc Reconstruction

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

[1]Hart & Risley 2003
[1]Ringe (p.c.) following Seebold 1979

Lexicon Size

Children

- 3-year-olds know a couple thousand lemmas (with wide variation)[1]
- There are 358 frequent verbs (lemmas occurring ≥10 times) in Brown (CHILDES) CDS

PGmc Reconstruction

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

[1]Hart & Risley 2003
[1]Ringe (p.c.) following Seebold 1979

Semantic Content

Children

- Extracted 358 verb lemmas occurring ≥10 times in Brown
- CDS from CHILDES is often used to approximate child knowledge

PGmc Reconstruction

- Tabulated 258 securely reconstructable strong verb lemmas with English translations

Semantic Content

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

Strict Translations

- **bītaną* ‘bite’ bite
- **grētaną* ‘weep’ cry

Good-Enough Translations

- **wringaną* ‘twist’ turn, roll, screw
- **dragană* ‘haul’ pull, carry
- **fanhaną* ‘seize’ take, steal

English CHILDES → *Proto-Germanic

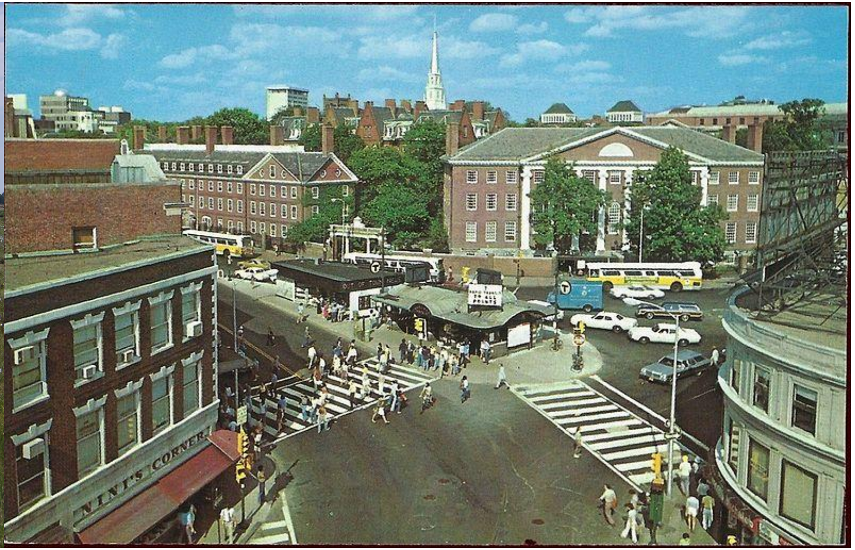
	#PGmc	#EN→PGmc	%
I	41	30	73.2
II	40	29	72.5
III	51	45	88.2
IV	16	13	81.3
V	28	21	75.0
VI	29	23	79.3
VII	53	41	77.4
Total	258	202	78.3

Why isn't this number higher?

And is it good?

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

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
- Then tabulated how many of the Spanish verbs have English translations
  - Careful to look at the numbers only after doing the matching
  - Tried to be equally “liberal” in the matches:
    - *prende* ‘ignite’ light
    - *frei* ‘fry’ cook

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**

Acquiring Paradigms

Learning Rules vs Exceptions

- Given some pairs that appear to follow a pattern, and some that violate the pattern
- Is it better to learn
  - one general rule that has exceptions?
  - multiple more specific rules with fewer exceptions?

How Many English Past Rules?

- +*ed* is obvious
- What about *-iN(C)* → *-aN(C)*?
  - *sing~sang*, *swim~swam*, *drink~drank*, etc.
  - but not *wing~winged*, *sting~stung*, *bring~brought*, etc.
- (Ignoring other small classes), two options:
  - **One rule:**
    - +*ed* with *-iN(C)*→*-aN(C)* as exceptions
  - **Two rules:**
    - +*ed* with no exceptions
    - *-iN(C)*→*-aN(C)* with exceptions

The Tolerance Principle[1]

- Model of productivity learning
- Based on economy of lexical access - Is it more efficient to assume some pattern is productive?
- Calculated over type frequencies (counts in a lexicon), not token frequencies (counts in a corpus)

[1]Yang, 2016

The Tolerance Principle[1]

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

$N = |C|$

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

[1]Yang, 2016

The Tolerance Principle[1]

- Given a hypothesized rule R operating over a class C, quantitatively defines the number of exceptions below which the rule is tenable

$N = |C|$

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

Exceptions are tolerable if

$$e < N / \ln N$$

[1]Yang, 2016

The Tolerance Principle[1]

- Given a hypothesized rule R operating over a class C, quantitatively defines the number of exceptions below which the rule is tenable

$N = |C|$

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

Exceptions are tolerable if

$$e < N / \ln N$$

- If there are too many exceptions, the child can try narrower generalization (smaller C)
- **If all attempts fail, give up and memorize everything**

[1]Yang, 2016

Tolerance Principle and Representation

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

How Many English Past Rules?

- One-Rule is tantamount to deciding that *-iN(C)* → *-aN(C)* is non-productive
  - ie, the child should assume +*ed* for new *-iN(C)* words
- Two-Rules is equivalent to deciding that *-iN(C)* → *-aN(C)* is productive
  - ie one should assume *-iN(C)* → *-aN(C)* for new *-iN(C)* words

N over Time

- *N* is the number of class members learned *so far*
- *N* and *e* grow as the vocabulary grows

N over Time

- *N* is the number of class members learned *so far*
- *N* and *e* grow as the vocabulary grows
- Children fall into and out of productivity during development
- Which explains overgeneralization errors attested in child speech
- Successfully applied to a wide range of problems
  - Modern English strong verbs
  - English diatones
  - German noun plurals
  - Russian and Polish genitives
  - English and Mandarin numeracy
  - etc.
- And psychological backing from artificial language learning experiments[1]

Accounting for the *ē-Grade with Acquisition

“The Eat Analogy”

- **etaną, *ēt, *ētun, *etanaz* ‘*eat*’ is the only Class V verb with **ē* by regular sound change

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

    PIE **h_1e-h_1d-´* > PGmc **ēt-*

- This is uncontroversially *a* source for the **ē*-grade. By hypothesis, it is *the* source.

Steps of the Eat Analogy

- The **ē*-grade spread from *eat* to the rest of Class V
- Then the **ē*-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)

Explicanda

Positives

- Why did **ē* spread from ‘eat’ to V?
- 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 past participle stem vowels not spread?

Explicanda

Positives

- Why did **ē* spread from ‘eat’ to V?
- 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 past participle 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 between these: **-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 between these: **-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 V verbs are exceptions (IV forms in V)
    - V's rule applies to **-eC-*, and IV verbs 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 between these: *-*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 V verbs are exceptions (IV forms in V)**
  - **V's rule applies to *-*eC*-, and IV verbs 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 verbs and Class V verbs will learn two distinct classes**
- **One class contributes too many exceptions to the other's rule**

V to IV+V

- According to the TP, a child who knows all Class IV verbs and Class V verbs will learn two distinct classes
- One class contributes too many exceptions to the other's rule

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

V to IV+V

- According to the TP, a child who knows all Class IV verbs and Class V verbs will learn two distinct classes
- One class contributes too many exceptions to the other's rule

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

$e = \min(|\mathrm{IV}|, |\mathrm{V}|) = 18$

V to IV+V

- According to the TP, a child who knows all Class IV verbs and Class V verbs will learn two distinct classes
- One class contributes too many exceptions to the other's rule

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

$e = \min(|\mathrm{IV}|,|\mathrm{V}|) = 18$

$N / \ln N = 11.6$

V to IV+V

- According to the TP, a child who knows all Class IV verbs and Class V verbs will learn two distinct classes
- One class contributes too many exceptions to the other's rule

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

$e = \min(|IV|,|V|) =$ 18

$N / \ln N =$ 11.6

18 > 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 probable is this state?

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

$5 < 5.3$.

IV+V SUCEEDS!

Likelihood of Overgeneralizations

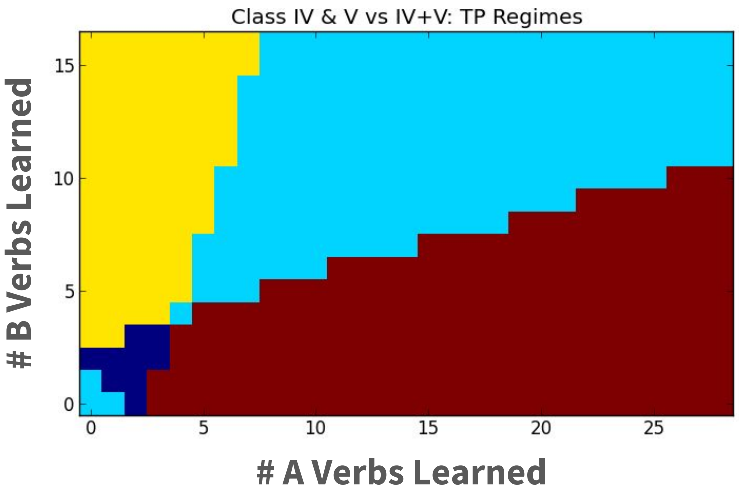
Given two classes A and B of sizes *K* and *N-K* and a plausible generalization between them, there are 4 possible outcomes

- Separate rules for A and B
- Rule A for A+B (Analogy A→B)
- Rule B for A+B (Analogy B→A)
- Rule A or B for A+B (Either)

Likelihood of Overgeneralizations

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

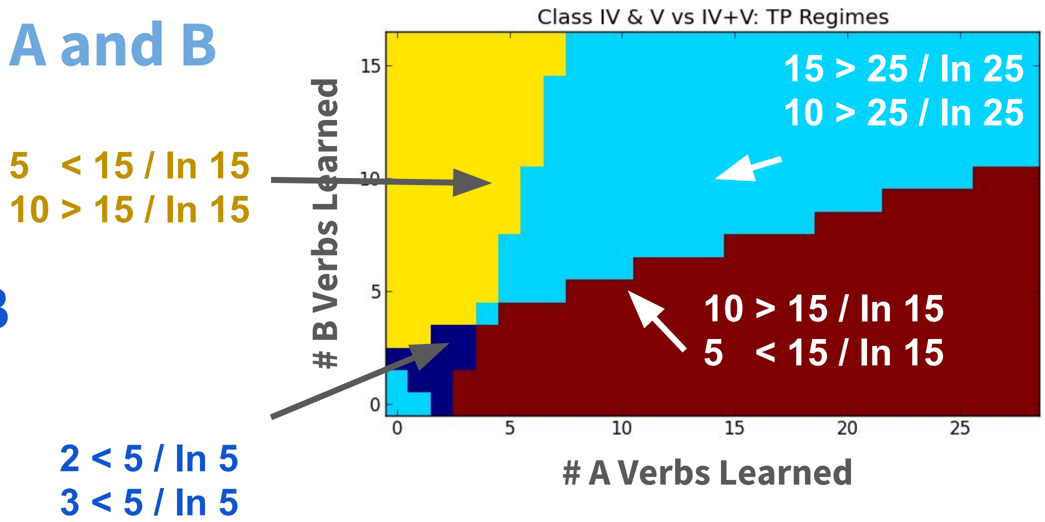
- Separate rules for A and B
- Rule A for A+B (Analogy A→B)
- Rule B for A+B (Analogy B→A)
- Rule A or B for A+B (Either)



Likelihood of Overgeneralizations

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

- Separate rules for A and B
- Rule A for A+B
- Rule B for A+B
- Rule A or B for A+B



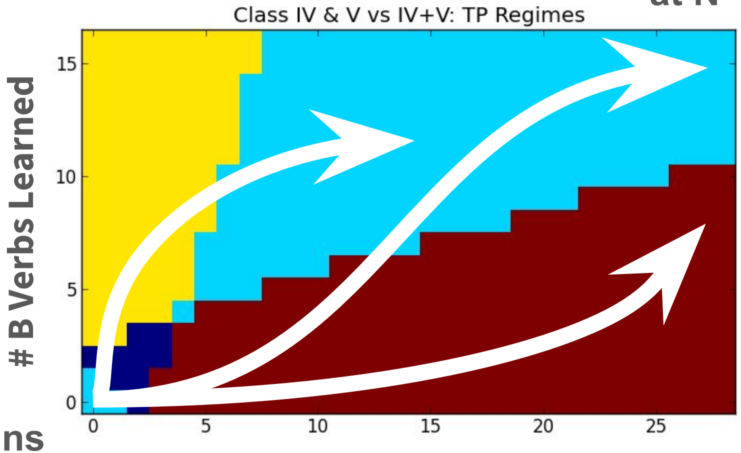
Likelihood of Overgeneralizations

Children progress along paths through this space

but not all paths are equally likely!

- Separate rules for A and B
- Rule A for A+B (Analogy A→B)
- Rule B for A+B (Analogy B→A)
- Rule A or B for A+B (Either)

Mature learner at N = |A ∪ B|



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]

$$P(X = k) = f(k; N, K, n) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

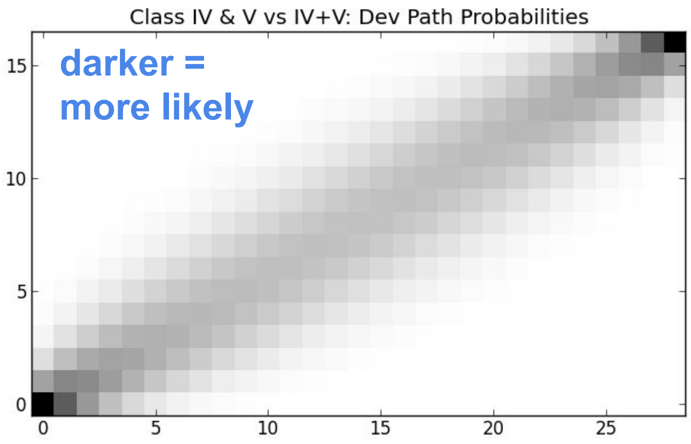
- $N = |A \cup B|$
- $K = |A|$
- $n = |\subseteq A \cup B \text{ learned so far}|$
- $k = |\subseteq A \text{ learned so far}|$
- $n\text{-}k = |\subseteq B \text{ learned so far}|$

[1]Unweighted marbles 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]

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

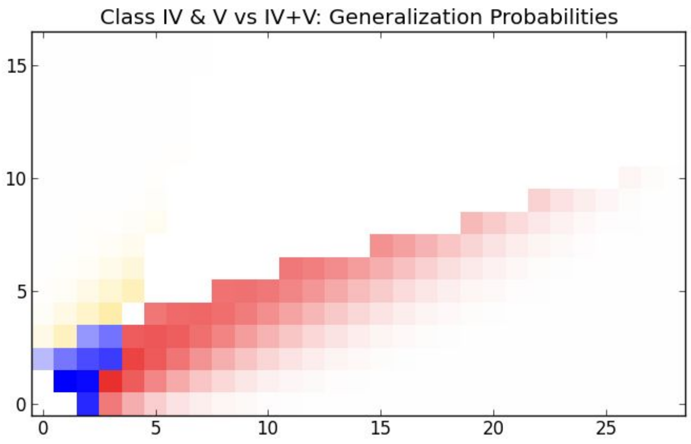


[1]If one class tends to be much more common than the other, this "line" will bow up or down

Likelihood of Overgeneralizations

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

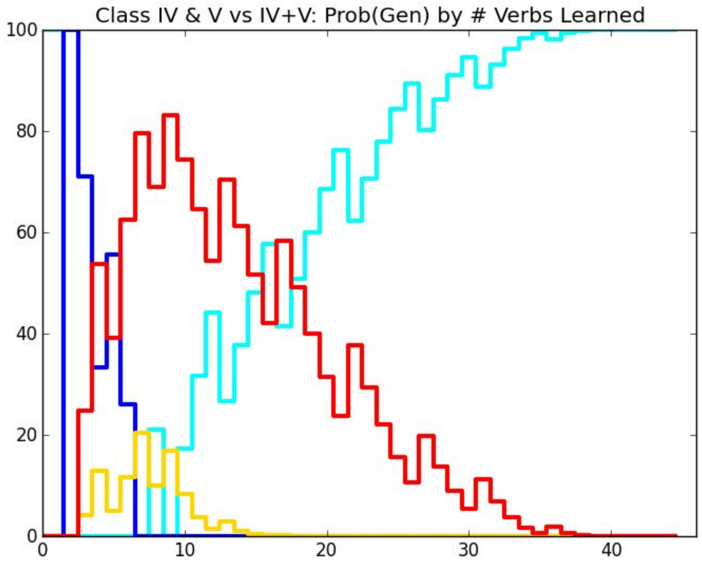
- Rule V for IV+V (Analogy V→IV)
- Rule IV for IV+V (Analogy IV→V)
- Rule V or IV for IV+V (Either)



Likelihood of Overgeneralizations

Plotting likelihood of each state by # of lemmas learned so far

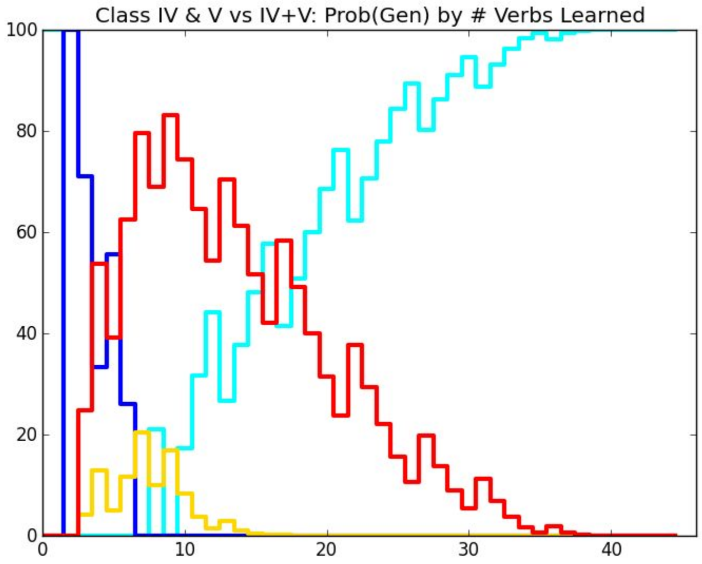
- Separate rules for V and IV
- Rule V for IV+V (Analogy V→IV)
- Rule IV for IV+V (Analogy IV→V)
- Rule V or IV for IV+V (Either)



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

Post-PGmc IV/V confusions

- **Shift from V to IV in Old High German**
  - eg OHG *gisprohhan* 'spoken' vs OE *sprecen*
  - After OHG and OE diverged, so this was late
- ***brekaną* '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*

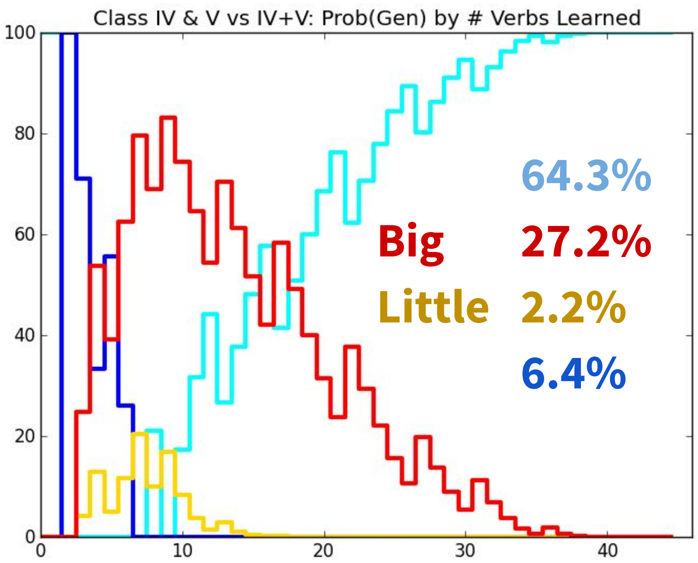
Generalization between IV+V and III

- 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

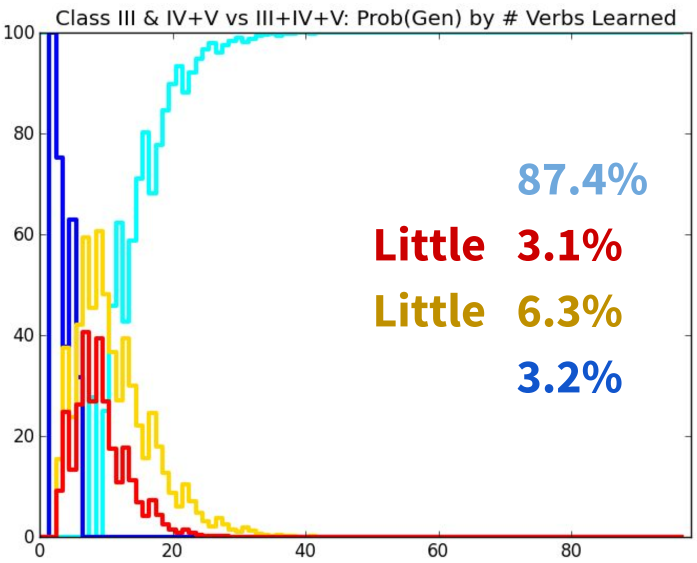
V → IV+V

|IV| = 16, |V| = 28



IV+V → III+IV+V

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



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

- Overgeneralizations 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)
  - Why the analogy was from V to IV rather than vice-versa
- IV+V→III+IV+V and III→III+IV+V were unlikely too (3.1, 6.3%)
  - Why further generalization did not happen

Explicanda

Positives

- Why did *ē spread from ‘eat’ to V?
- 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 past participle stem vowels not spread?

Explicanda

Positives

- Why did **ē* spread from ‘eat’ to V?
- 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 past participle 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

My idea is not much better than the classic stories

- Are the past 3sg and pparticiple stems more or less frequent than past?
  - Influences how early forms are heard/learned
  - Could affect the TP and the Sibling Effect

Explicanda

Positives

- Why did *$\bar{e}$* spread from ‘eat’ to V?
- 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?

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-*
  - **etaną* 'eat,' **fetaną* 'fall,' **getaną* 'get,' **metaną* '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**

Initial extension

- **Are there any subclasses of V to which **ē* 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!**
- **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**

End

Acknowledgements:

- Don Ringe
- NDSEG Fellowship

Contact:

- jkodner@sas.upenn.edu
- ling.upenn.edu/~jkodner

The Paradox of Language Change[1]

[1] 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]

[1]Niyogi & Berwick 1995
[2]A 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]

We've shown which overgeneralizations are more likely to occur, but we haven't explained why they persisted in adult speakers

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

Learner Errors

Blame the Child

- The learner does not act correctly on its input
- "A buggy algorithm"
- Hard-coded ε parameter (cf Griffiths, Kirby, etc)

Learner Errors

Blame the Child

- The learner does not act correctly on its input
- "A buggy algorithm"
- Hard-coded ε parameter (cf Griffiths, Kirby, etc)

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 trivial variation (cf Niyogi & Berwick)

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, can 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?
- Likelihood of ***not*** hearing an adult token varies by domain
  - Not hearing (many tokens of) a phone?

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?
- Likelihood of ***not*** hearing an adult token varies by domain
  - Not hearing (many tokens of) a phone? **~impossible**

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?
- Likelihood of ***not*** hearing an adult token varies by domain
  - Not hearing (many tokens of) a phone? **~impossible**
  - Not hearing some corner of verbal inflection?

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?
- Likelihood of ***not*** hearing an adult token varies by domain
  - Not hearing (many tokens of) a phone? ~impossible
  - Not hearing some corner of verbal inflection? sure

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?**
- **Likelihood of *not* hearing an adult token varies by domain**
  - **Not hearing (many tokens of) a phone?** **~impossible**
  - **Not hearing some corner of verbal inflection?** **sure**
  - **Not hearing some obtuse syntactic construction?**

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?
- **Likelihood of *not* hearing an adult token varies by domain**
  - Not hearing (many tokens of) a phone? ~impossible
  - Not hearing some corner of verbal inflection? sure
  - Not hearing some obtuse syntactic construction? yes.

Can Bob Ignore Alice?

- I don’t know…

Can 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?

Can 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"

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 *$\bar{e}$?
- If so, next generation will receive competent IV *$\bar{e}$ inputs