Productivity Learning

Driving Analogical Change in Proto-Germanic

Jordan Kodner
University of Pennsylvania

ECIEC XXXVIII
June 22, 2019, UPenn

Outline

- The Lengthened *ē-Grade Strong Verbs
- The Learning Model
- Studying Proto-Germanic Children
- Distinguishing Hypotheses
- Conclusions

Proto-Germanic Lengthened *ē-Grade Strong Verbs

The Proto-Germanic Strong Verb Paradigm

	Root	Present	Past 3sg	Past	Pptc	
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

PGmc Strong Verbs derived from PIE Ablaut Except...

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

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

Analogical Accounts

- Commonly agree that *ē analogized from Class V to Class IV
- But disagree on its origins in Class V

Analogical Accounts

- Commonly agree that ***ē* analogized from Class V to Class IV**
- But disagree on its origins in Class V

Conflicting Accounts

- **Class VI Analogy - Length analogy with Class VI **ō*-grade**
- **Eat Analogy - spread from Class V **etanǫ* ‘eat’**

Analogical Change as Productivity over Time

- It is clear that productivity plays a major role in analogical change
- But it has been unclear how[1]

[1] Hock 2003, Anderson 2015

Analogical Change as Productivity over Time

- It is clear that productivity plays a major role in analogical change
- But it has been unclear how[1]

Tendencies of analogy relate to productivity

eg, Kuryłowicz's Laws[2]

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] Hock 2003, Anderson 2015, [2] paraphrased

Analogical Change as Productivity over Time

- It is clear that productivity plays a major role in analogical change
- But it has been unclear how[1]

Tendencies of analogy relate to productivity

eg, Also[2]

1. High (token) frequency verbs are much more likely to be irregular than lower frequency verbs → low frequency verbs tend to be leveled over time[2]
2. High frequency inflectional categories are more likely to be unique than lower frequency categories → analogical leveling tends to go from high to low freq[2]
3. Languages with larger paradigms tend to have less irregular inflection[3]

[1] Hock 2003, Anderson 2015, [2] Bybee & Thompson 1997, [3] Malouf & Ackerman 2013

Analogical Change as Productivity over Time

- It is clear that productivity plays a major role in analogical change
- But it has been unclear how[1]

Productivity learning is a major subject in child language acquisition, so let's look at that!

[1] Hock 2003, Anderson 2015, [2] Bybee & Thompson 1997, [3] Malouf & Ackerman 2013

A Learning Model

The Tolerance Principle[1]

- A model for the acquisition of productivity patterns
- Should a pattern be generalized to unseen forms or not?
- How many exceptions can a pattern tolerate before we give up on it?

[1] Yang 2016

The Tolerance Principle[1]

- A model for the acquisition of productivity patterns
- Should a pattern be generalized to unseen forms or not?
- How many exceptions can a pattern tolerate before we give up on it?
- Relies on type frequencies in the data
  - Number of unique lexical items attested obeying a pattern
  - Commonly held to be the case[2]

[1] Yang 2016, [2] Aronoff 1976, MacWhinney 1978, Bybee 1985, Baayen 1993, Elman 1998, Pierrehumbert 2003, ...

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 **= number of types that should obey the generalization**

e **= number of types that do not obey the generalization**

θ = max # of exceptions that can be tolerated

Exceptions are tolerable if

$$e < \theta$$

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

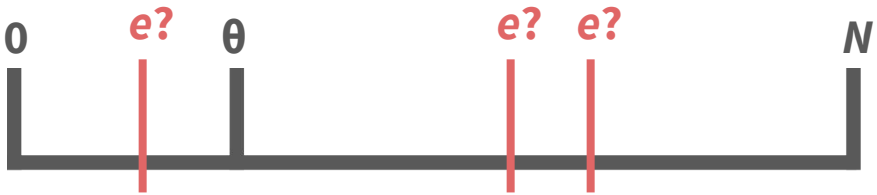
N and *e* Vary over Individual Development

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

[1] Yang 2016, [2] Xu & Pinker 1995

Visualization of the Tolerance Principle

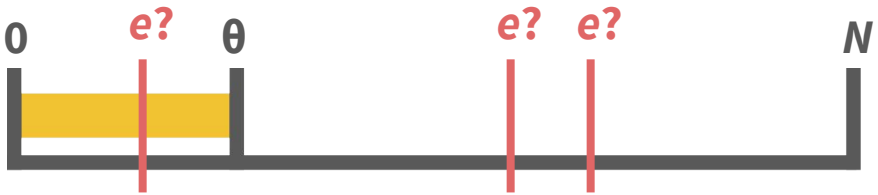
N = types it should apply to
e = types that are exceptions
θ = tolerance threshold



e falls between 0 and *N* and may be less than or greater than θ

Visualization of the Tolerance Principle

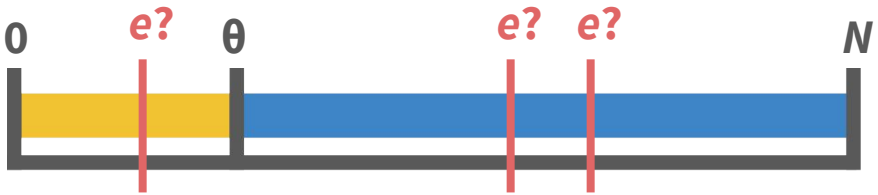
N = types it should apply to
e = types that are exceptions
θ = tolerance threshold



If *e* is below θ,
acquire generalization

Visualization of the Tolerance Principle

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



If *e* is below θ,
acquire generalization
Otherwise, do not generalize

Visualization of the Tolerance Principle

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

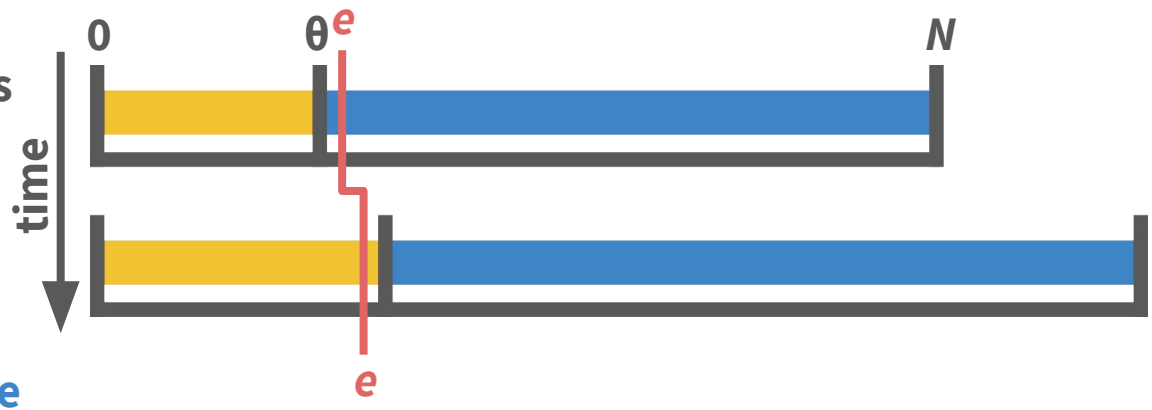
If *e* is below θ,
acquire generalization
Otherwise, do not generalize

- *N* grows over an individual's development

Visualization of the Tolerance Principle

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

If *e* is below θ,
acquire generalization
Otherwise, do not generalize

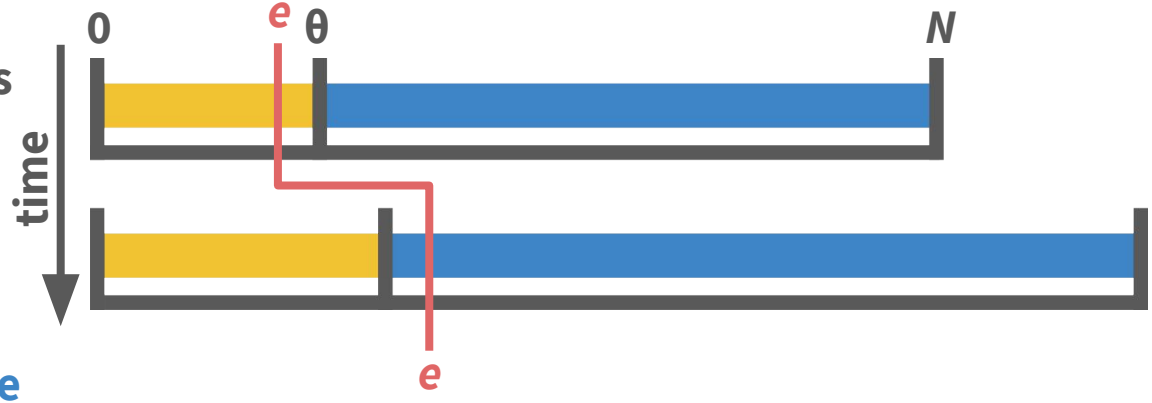


- *N* grows over an individual's development
- If θ grows faster than *e*, a generalization may fall into productivity

Visualization of the Tolerance Principle

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

If *e* is below θ,
acquire generalization
Otherwise, do not generalize



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

Child Lexical Knowledge

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

Children acquire basic morphology on the basis of surprising few items!

Language	Estimated |Vocab|
English 2;10-3;0[2]	525-1116
German 2;6[3]	$\mu = 429, \sigma > 100$

[1] Nagy & Anderson 1984, [2] Hart & Risley 2003, [3] Szagun et al 2006

What do We Know about Proto-Germanic Children?

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

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**)

And this is all we need!

The Data

- 258 “securely” reconstructed strong verbs extracted from Seebold 1979[1]
- Out of a couple thousand reconstructed items in total

We know more PGmc words than a PGmc three-year-old

Class	# Verbs
I	41
II	40
III	51
IV	16
V	28
VI	29
VII	53
Total	258

[1] Credit to Don Ringe for extracting them

Cross-Language Lexical Comparisons

- **Does the Proto-Germanic lexicon contain the same kind of vocabulary as child-directed speech?**

Cross-Language Lexical Comparisons

- **Does the Proto-Germanic lexicon contain the same kind of vocabulary as child-directed speech? Probably.**

Intuition

1. **Words can be constructed if they are attested in the daughter languages**
2. **Words that survive in the daughters tend to be common everyday terms**
3. **Common everyday terms are also the words that tend to be present in CDS**

Cross-Language Corpora

- **English CDS** - verb lemmas in CHILDES Brown (and Brent for comparison)
- **Spanish CDS** - verb lemmas in combined CHILDES FernAguado, Hess, OreaPine, Remedi, Romero, SerraSole
- **Proto-Germanic** - securely reconstructable strong verbs
- **Classical Latin** - verb lemmas in all Perseus online 3rd BC - 2nd AD (inclusive)

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

Cross-Language Comparisons

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

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

Cross-Language Comparisons

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

*Germanic Urheimat, 1st Millenium BC

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

Cross-Language Comparisons

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

*Germanic Urheimat, 1st Millenium BC

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

Accounting for the Lengthened *ē-Grade

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 pptc 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 pptc stem vowels not spread?

Avenue for Overgeneralization

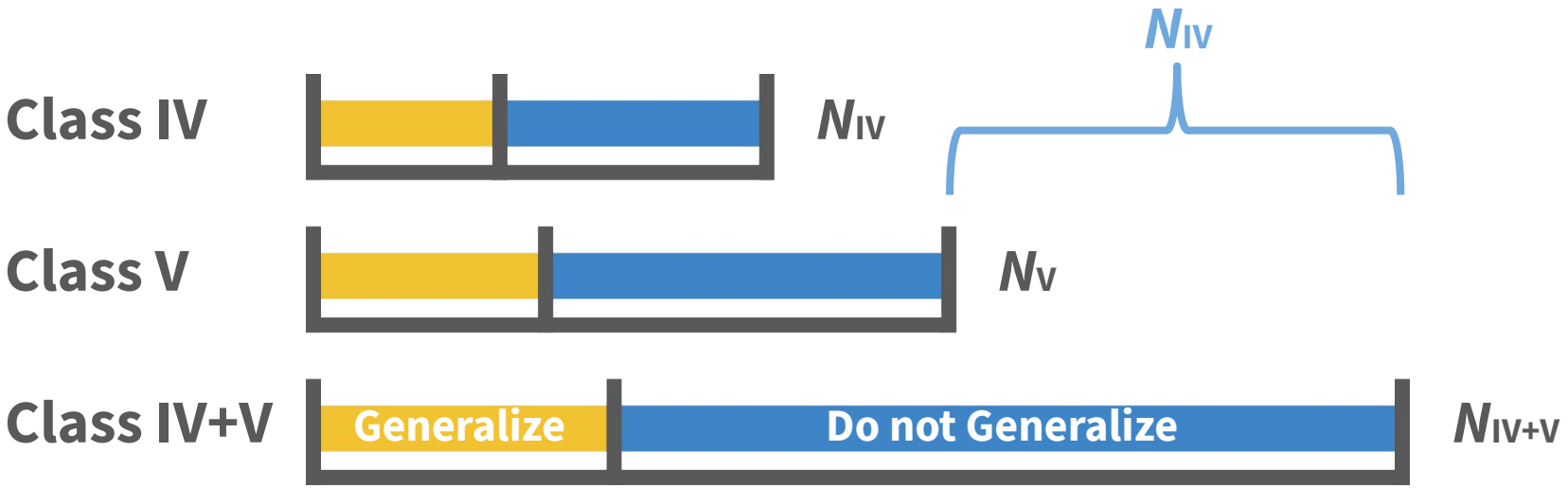
- **Analogy from Class V to IV is modeled as overgeneralization**
  - **The V past *ē becomes analyzed as the default past for Class IV**
  - **So IV verbs whose pasts are not yet attested in the input are formed in *ē**
- **Happens if a learner finds productivity over too wide a generalization**
- **There exists a generalization capturing exactly Class IV+V: *-eC-**
  - **Classes IV has a root shape *-eR- which defined the class**
  - **Class V has a root shape *-eT-**

Possible Generalizations

- ***-eR-** and ***-eT-** are non-productive
- IV's past formation applies to ***-eR-** and V's to ***-eT-** — intended
- 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

Evaluating Generalizations

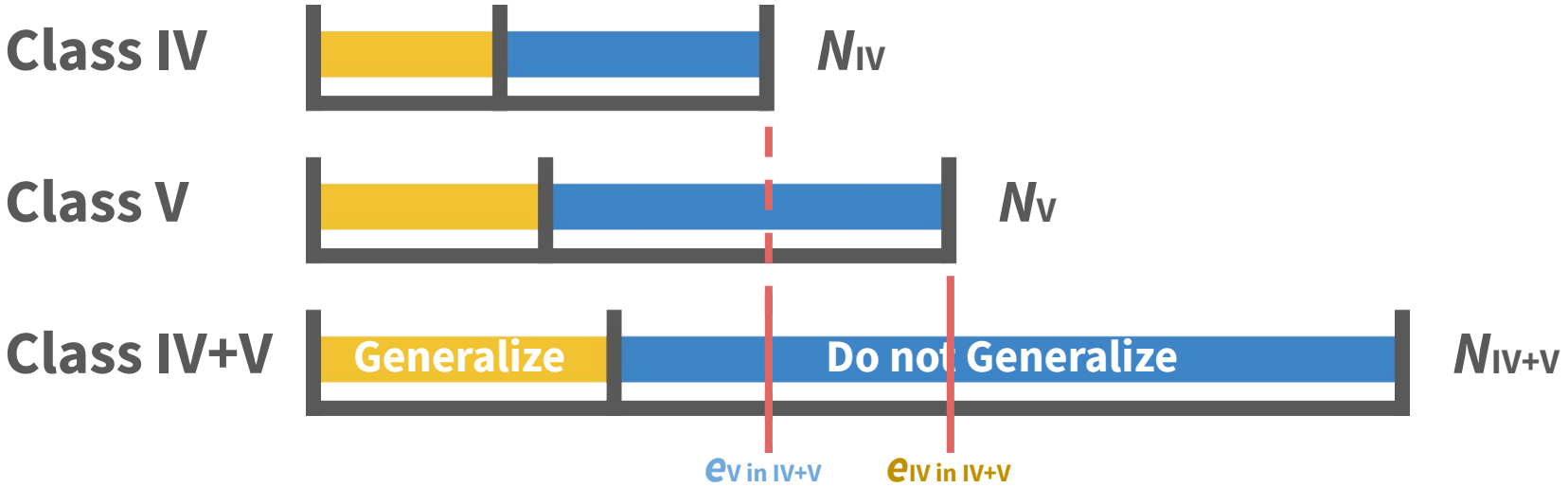
- Generalize IV's past to IV+V if few V pasts are known ($N_V = e_{IV \text{ in } IV+V} < \theta_{IV+V}$)
- Generalize V's past to IV+V if few IV pasts are known ($N_{IV} = e_{V \text{ in } IV+V} < \theta_{IV+V}$)



Evaluating Generalizations

- Generalize IV's past to IV+V if few V pasts are known ($N_V = e_{IV \text{ in } IV+V} < \theta_{IV+V}$)
- Generalize V's past to IV+V if few IV pasts are known ($N_{IV} = e_{V \text{ in } IV+V} < \theta_{IV+V}$)

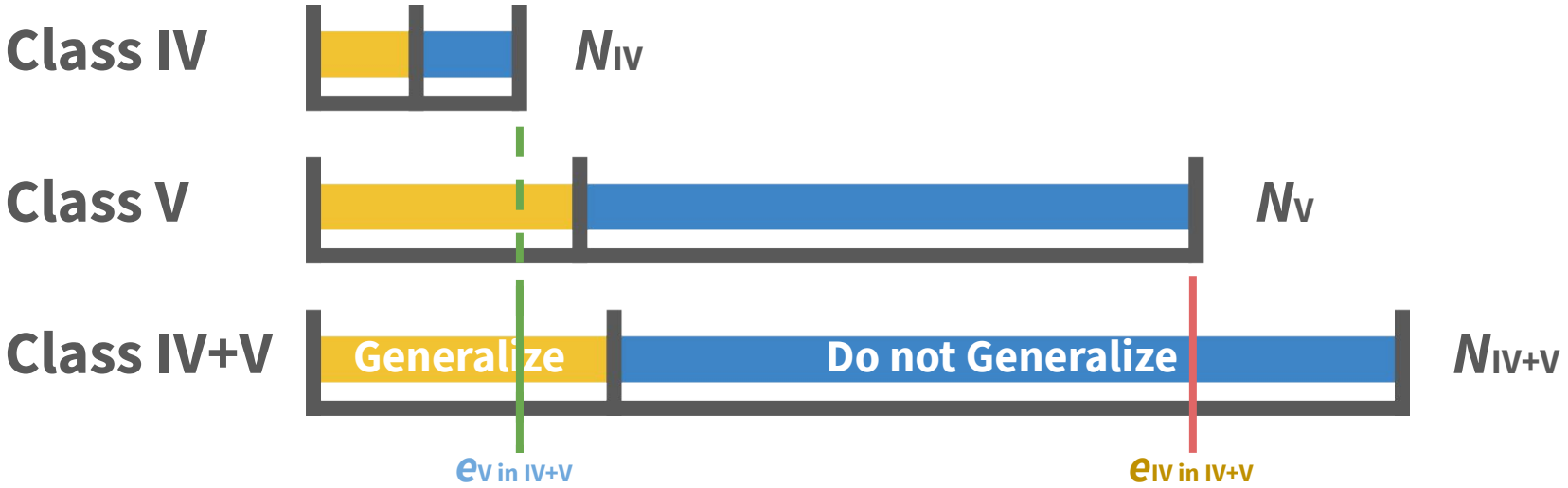
IV+V is not tolerable if *N*s are large and similarly sized



Evaluating Generalizations

- Generalize IV's past to IV+V if few V pasts are known ($N_V = e_{IV \text{ in } IV+V} < \theta_{IV+V}$)
- Generalize V's past to IV+V if few IV pasts are known ($N_{IV} = e_{V \text{ in } IV+V} < \theta_{IV+V}$)

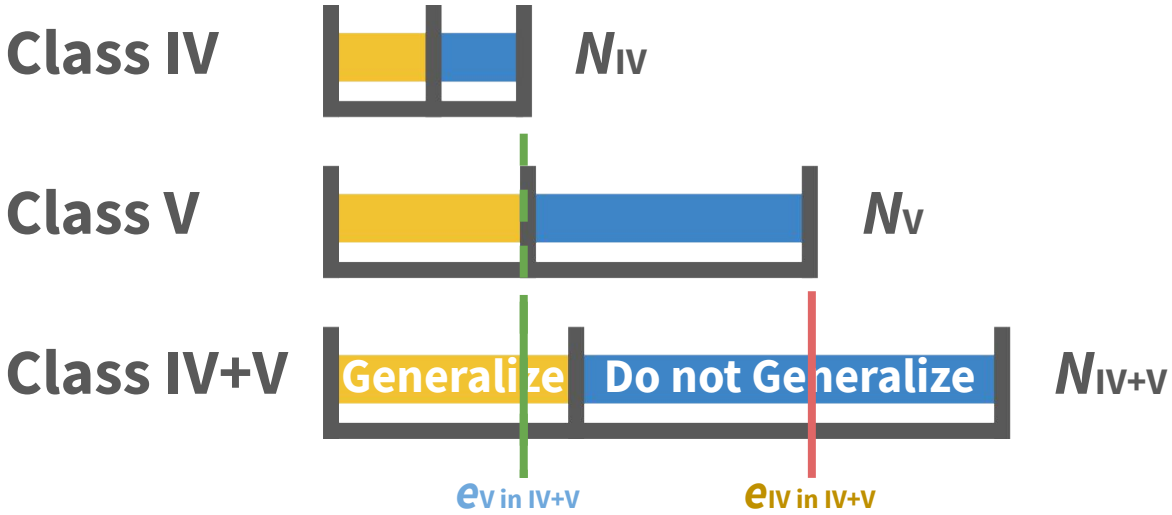
IV+V is tolerable if one *N* is much larger than the other



Evaluating Generalizations

- Generalize IV's past to IV+V if few V pasts are known ($N_{V} = e_{IV \text{ in } IV+V} < \theta_{IV+V}$)
- Generalize V's past to IV+V if few IV pasts are known ($N_{IV} = e_{V \text{ in } IV+V} < \theta_{IV+V}$)

IV+V is tolerable or if *N*s are small

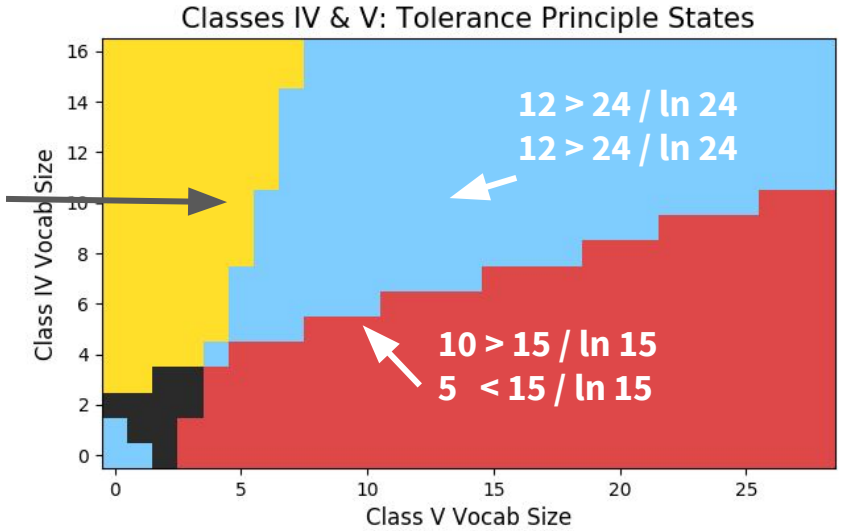


Likelihood of Overgeneralizations

Given two classes V and IV and a plausible generalization between them, there are 4 possible outcomes

- **Separate V and IV**
- **V past in IV+V**
- **IV past in IV+V**
- **Either past in IV+V**

5 < 15 / ln 15
10 > 15 / ln 15

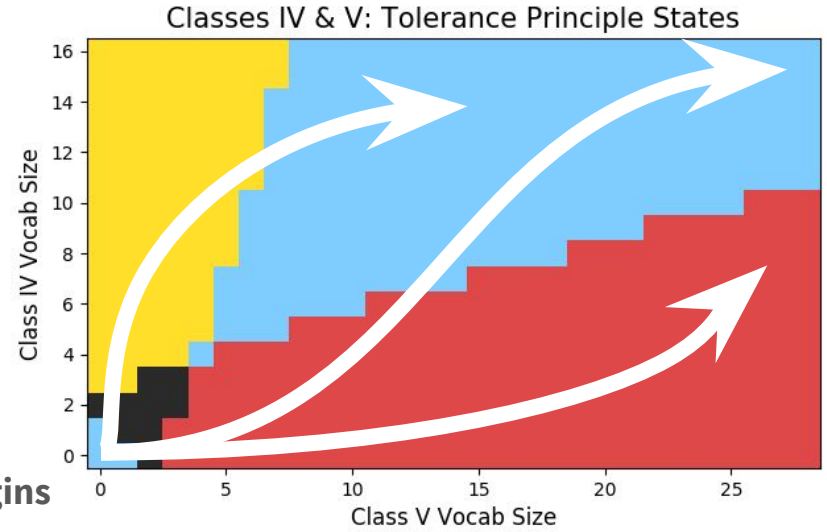


Likelihood of Overgeneralizations

Children progress along paths through this space

- **Separate V and IV**
- **V past in IV+V**
- **IV past in IV+V**
- **Either past in IV+V**

Mature learner at $N = |IV \cup V|$



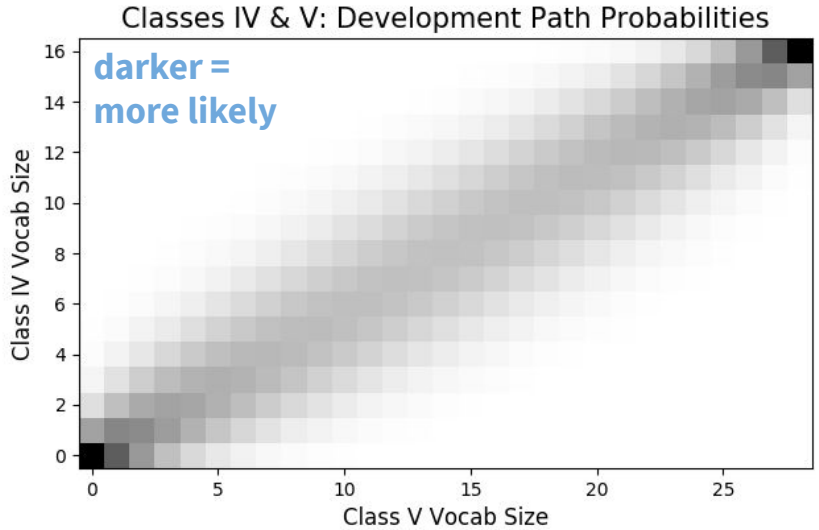
Fresh child begins at N=0

Likelihood of Overgeneralizations

Likelihood of landing in each state modeled as a hypergeometric distribution

ie drawing marbles without replacement[1]

Class	# Verbs	θ
IV	16	5.77
V	28	8.40
IV+V	44	11.62



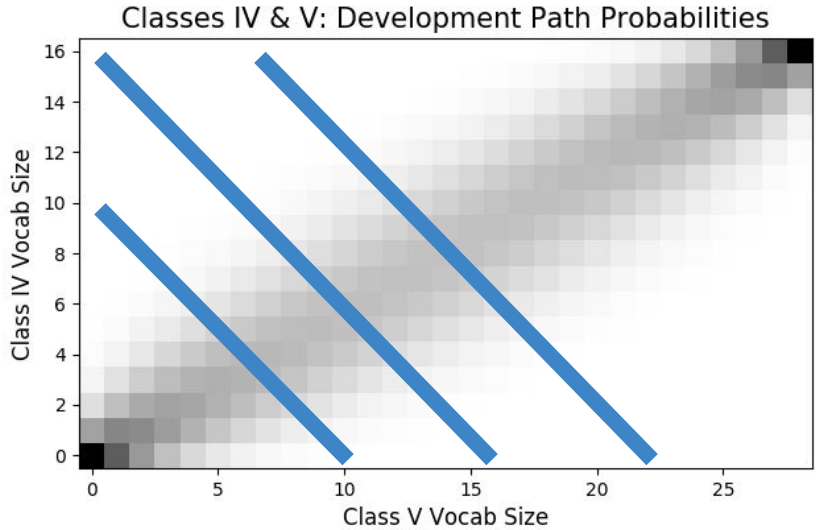
[1]If one class tends to be much more common than the other, this "line" will bow up or down

Likelihood of Overgeneralizations

Likelihood of landing in each state modeled as a hypergeometric distribution

ie drawing marbles without replacement[1]

Class	# Verbs	θ
IV	16	5.77
V	28	8.40
IV+V	44	11.62



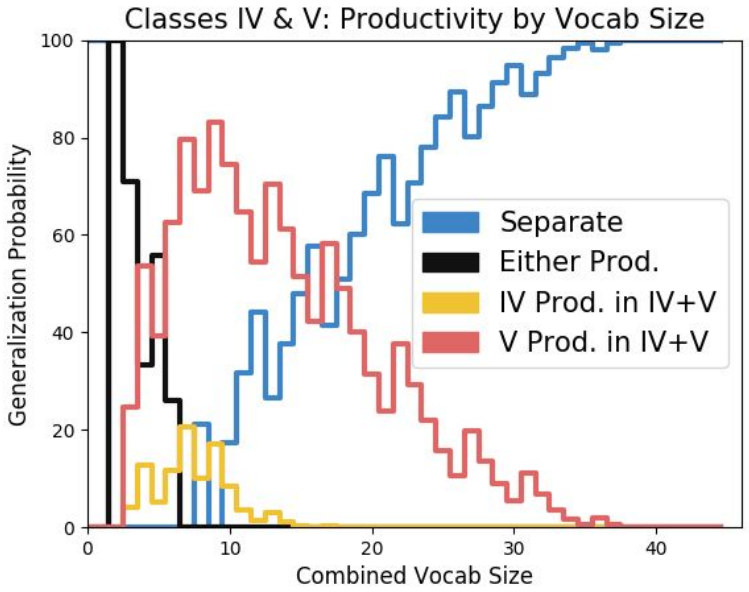
diagonals (constant *N*) each sum to 1

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

Likelihood of Overgeneralizations

Likelihood of each analogizing state by vocabulary size

- **Separate V and IV past**
- **V past in IV+V (V→IV analogy)**
- **IV past in IV+V (IV→V analogy)**
- **Either past in IV+V**



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 pptc stem vowels not spread?

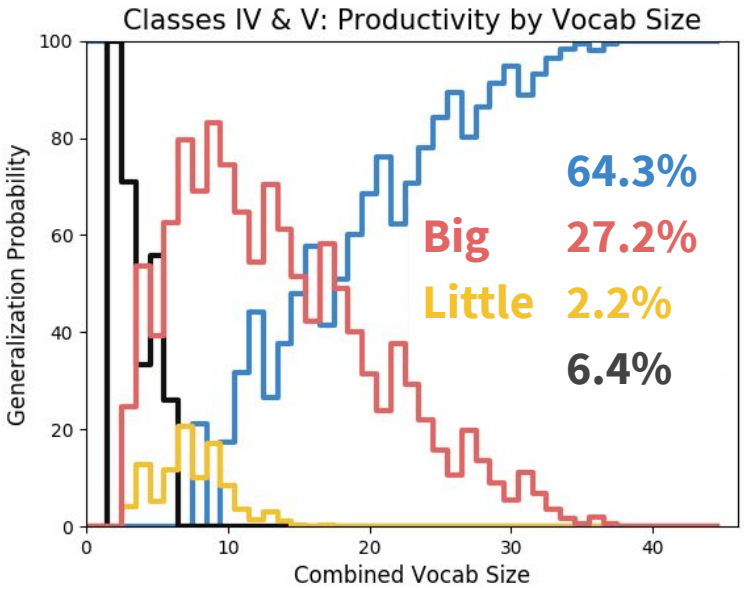
Generalization between IV+V and III

- If generalization between IV and V is tenable, analogy between III and IV+V may be tenable
- They share a unique root shape generalization
  - IV+V is defined by ***-eC-***
  - III is defined by ***-eCC-***
- There exists a generalization ***-eC(C)-*** that encompasses exactly III+IV+V
- A child can evaluate this hypothesis in the same way as the others

Comparing IV+V and III+IV+V Generalizations

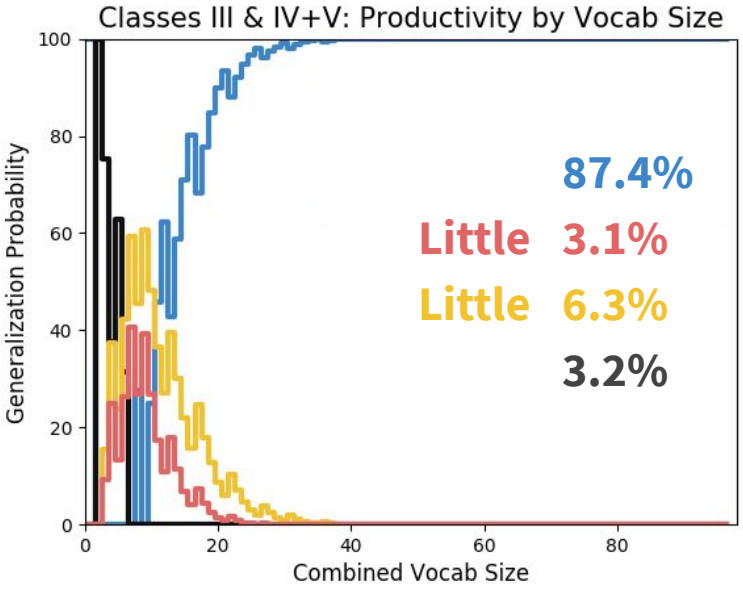
V Productive in IV+V

|IV| = 16, |V| = 28



IV+V Productive in III+IV+V

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



Comparing IV+V and III+IV+V Generalizations

Given the Proto-Germanic lexicon,

- Overgeneralization from V to IV was likely in early dev. and tenable late
- IV to V is never probable during development and not tenable late
- III to IV+V and vice-versa are never probable either

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 pptc stem vowels not spread?

The Eat Analogy

- ***etanǫ, *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-

- **By hypothesis, **ē*** **spread from this single verb to all Class V**
- **The challenge is accounting for why a form would extend from a single verb to an entire class**

Class VI Analogy

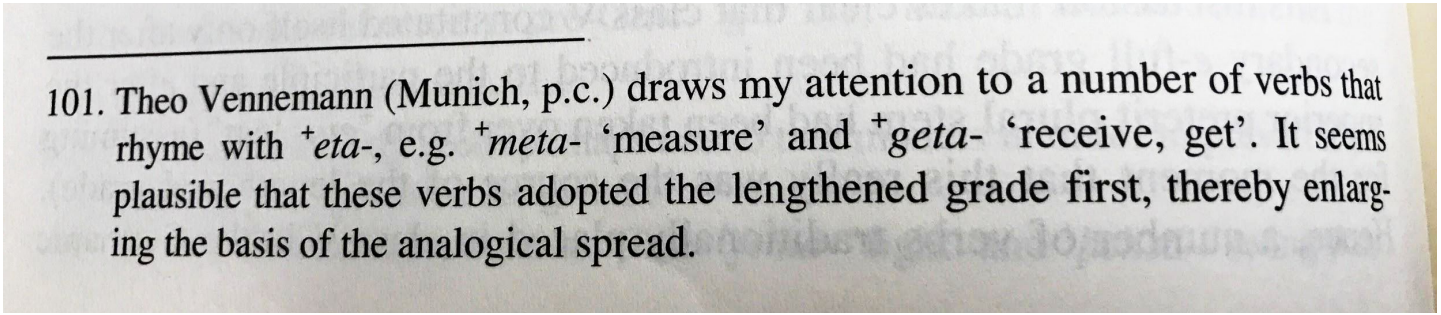
- Long vowels exist in Class VI by regular sound change
- eg, **faranǫ, *fōr, *fōrun, *faranaz* ‘travel’
- By hypothesis, the length of **ō* was analogized to Class V
- The traditional challenge is in accounting for why only length would be analogized.

The Eat Analogy - From One to Many

- This is not a job for the Tolerance Principle
  - $N / \ln N$ is invalid for $N \leq 2$
  - $N - 1 > N / \ln N$ for $N > 3$
- But some kind of generalization is likely relevant here

The Eat Analogy - 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.

[1] footnote in Mailhammer 2007

The Eat Analogy - 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 four to all V?

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!
  - **lekaną* ‘be leaky,’ **rekaną* ‘bank a fire,’ **wrekaną* ‘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	12
*-*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

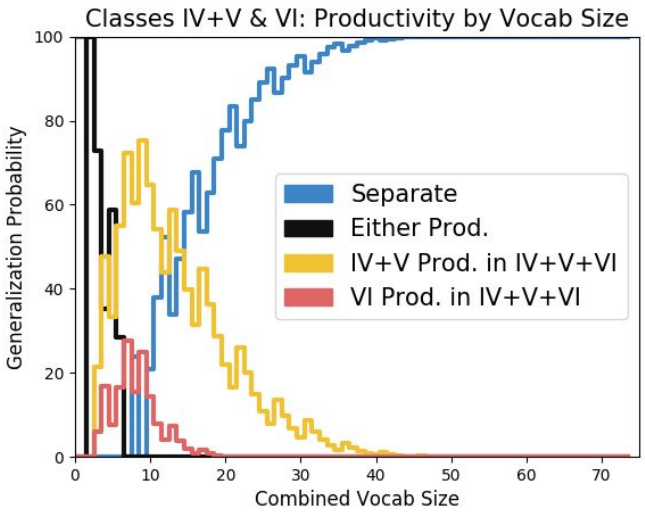
Class VI Analogy

- The TP analogy model is applied in the same way as for V+VI and III+V+VI
- No root generalization captures exclusively V (*-*eT*-) and VI (*-*aC*-)
- The closest generalization is (*-[-hi]*C*-) which captures IV+V+VI, so this model cannot predict analogy to V first then to IV

Class	# Verbs	θ
IV+V	44	11.62
VI	29	8.61
IV+V+VI	73	17.01

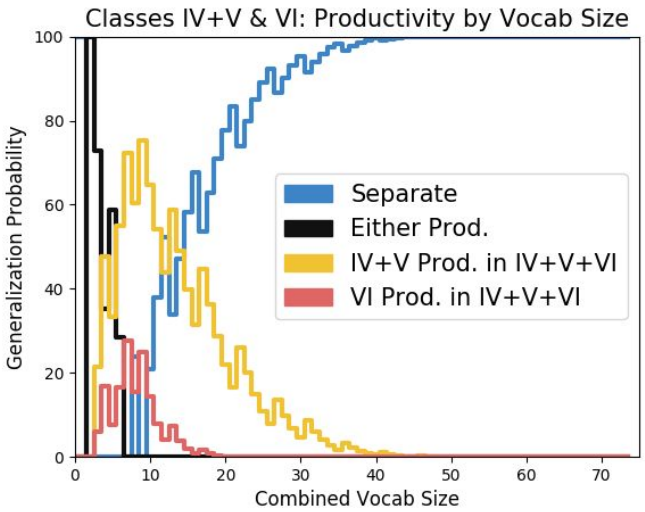
Class VI Analogy - Results

- A TP state space is constructed for analogies between IV+V and VI
- Probability of overgeneralizing VI forms, overgeneralizing IV+V forms



Opportunity for VI to IV+V Analogy

- The relevant overgeneralization is less tenable than overgeneralization in the other direction!



The Eat Analogy vs Class VI Analogy

Eat Analogy

- There exists a path for extension of *$\bar{e}$ pasts from ‘eat’ to Class V
- Establishes a course of events compatible with analogy from V to IV
- Analogical extension from a small set of verbs to all Class V is unlikely

The Eat Analogy vs Class VI Analogy

Eat Analogy

- There exists a path for extension of *ē pasts from ‘eat’ to Class V
- Establishes a course of events compatible with analogy from V to IV
- Analogical extension from a small set of verbs to all Class V is unlikely

Class VI Analogy

- Would be unlikely because of class sizes
- No way to get subsequent analogy from V to IV
- Still the question about why only vowel length was analogized

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 pptc stem vowels not spread?**

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
- **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*

Later analogies continued in both directions and involved pptc stem

Inverse Paradigm Saturation

- The proportion of attested lemmas associated with a particular inflection
- Attestation follows a long-tailed distribution

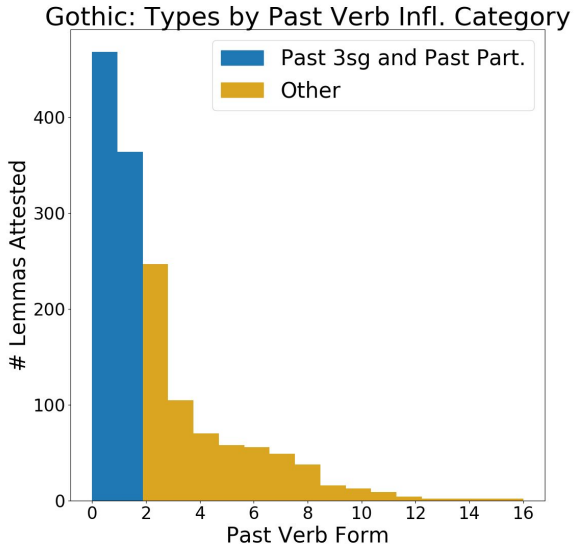
Past 3sg and past participles tend to be among the most common inflected verbs

Inverse Paradigm Saturation

- The proportion of attested lemmas associated with a particular inflection
- Attestation follows a long-tailed distribution

Gothic example

- Past 3sg and pptc are by far the best attested inflectional categories
- This works against analogy for these forms because fewer forms need to be inferred by the learner



Conclusions

Conclusions

Relationship between Analogical Change and Learning

- (Some) Analogical changes may be modeled with language acquisition
- Adopting a concrete model of learning allows us to evaluate hypotheses for change in a new light

Reconstructed Lexicons and Child Lexicons

- For Proto-Germanic, more similar to child lexicons than one might assume
- Both in size and in content

End

Acknowledgements:

- Don Ringe
- Charles Yang
- NDSEG (US ARO)

Individual-Level Variation and Change

Productivity as an Avenue for Analogy

- Analogy is overproductivity that gains a foothold in the population

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

- But children typically grow out of this

The Paradox of Language Change[1]

[1] Niyogi & Berwick 1996

The Sibling Effect

- Why might children not overcome their overgeneralizations?

Imagine big sister Alice and little brother Bob

- Alice is currently producing innovative *$\bar{e}$ pasts in Class IV
  - Plausible given how Class IV *$\bar{e}$ is tenable late
  - Bob may hear these forms
- Bob is receiving both adult conservative IV pasts and Alice's
- How does this effect Bob?

The Sibling Effect

Can Bob identify Alice's innovation?

- Bob is likely not hear adult-produced tokens for any given low frequency Class IV verb until much later
- Since Alice is mostly consistent with adults, he cannot tell if she is innovating

Will Bob adopt Alice's innovation?

- Even young children orient toward peers
- Bob may prefer Alice's forms over his parents
- He may later learn adult forms as sociolinguistic variant doublets

Tolerance Principle Background

The Tolerance Principle

- An evaluation metric[1] over linguistic hypotheses
  - an Elsewhere Condition for 'rules' and 'exceptions'[2]
  - frequency-rank correlated lexical access[3]
  - 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] Chomsky 1955, 1965, Chomsky & Halle 1968, [2] Anderson 1969, *inter alia,* [3] Murray & Forster 2004, [5] Schuler et al 2017

Language Acquisition and Change

Language Change by Language Acquisition

- Child language acquisition is one of the primary drivers of language change[1]
  - Not a new idea (Schleicher 1861, Paul 1880, etc)
  - In modern times, associated with generative linguistics[2]
- Children are both innovators and propagators of change
- Minor learning differences over successive generations

  → major population-level change

[1] Schleicher 1861, Paul 1880, Sweet 1899, Halle 1962, Kiparsky 1965, Andersen 1973, Baron 1977, Lightfoot 1979 *et seq*, Labov 1989, Niyogi 1996 *et seq*, Kroch 2005, Yang 2002 *et seq*, van Gelderen 2011, Cournane 2017, *inter alia*

[2] see Diessel 2012, Stanford 2014 for critical analyses, citing Lightfoot 1979 et seq, van Gelderen 2011, etc as generative examples