Panda’s Thumbs and Irregular Verbs

Charles D. Yang
Yale University

Rules and Analogy in Acquisition, Change, and Evolution

Charles D. Yang
Yale University

A Darwinian Paradox

Irregularities and Defaults in Language

- English past tense: go-*went*, hold-*held*, lose-*lost*, walk-*walked*
- German plurals: Kind-*Kinder*, Wind-*Winde*, Ochs-*Ochsen*, Daumen-*Daumen*, Auto-*Autos*
- ...
- Chinese classifiers: yi *tou* zhu, yi *zhi* yang, ... yi *ge* igibi

Imperfection: why aren’t all words regular?

Panda’s Thumbs: Constraints

- Historical constraints
- Structural constraints
- Developmental constraints

Evolutionary backtracking

1. study children’s abilities to learn words (w/ Morris Halle)
2. extrapolate back to study sound change
3. see if any of these can be “grounded” nonlinguistically and thus, perhaps pre-linguistically in time
4. language must then be constrained by these learning abilities
5. irregularity is inevitable

The Past Tense Debate

The facts:

- In general, both adults and children add -d to novel verbs (Berko 1958).
- Overregularization errors: (*take*-*taked*) 10% out of all instances of irregular verbs (Marcus et al. 1992).
- Misregularization (*bring*-*brang*) is rare: 0.2% (Xu & Pinker 1995).

Pinker's Words and Rule Model

- regular verbs are handled by a rule (+ed)
- irregular verbs are memorized by association of stem-past pairs

  the more you hear, the better you memorize

Words by Rules

$< 2, 4 >, < 3, 4 >, < 4, 8 >, < 6, 7 >, < 7, 8 >, < 8, 16 >$

Strategy I (Pinker's): memorize each pair of numbers
2-4, 3-4, 4-8, 6-7, 7-8, 8-16

Strategy II (Ours): group numbers into two classes by rules
(3, 6, 7) $\mapsto R_{+1}$; (2, 4, 8) $\mapsto R_{\times 2}$

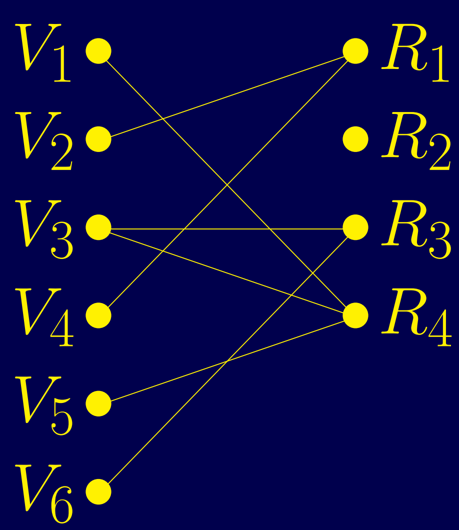
Bottom line

- Everyone agrees: irregular verbs must be memorized *somehow*
- Difference: *how* are they learned, organized, and accessed

Rules and Competitions

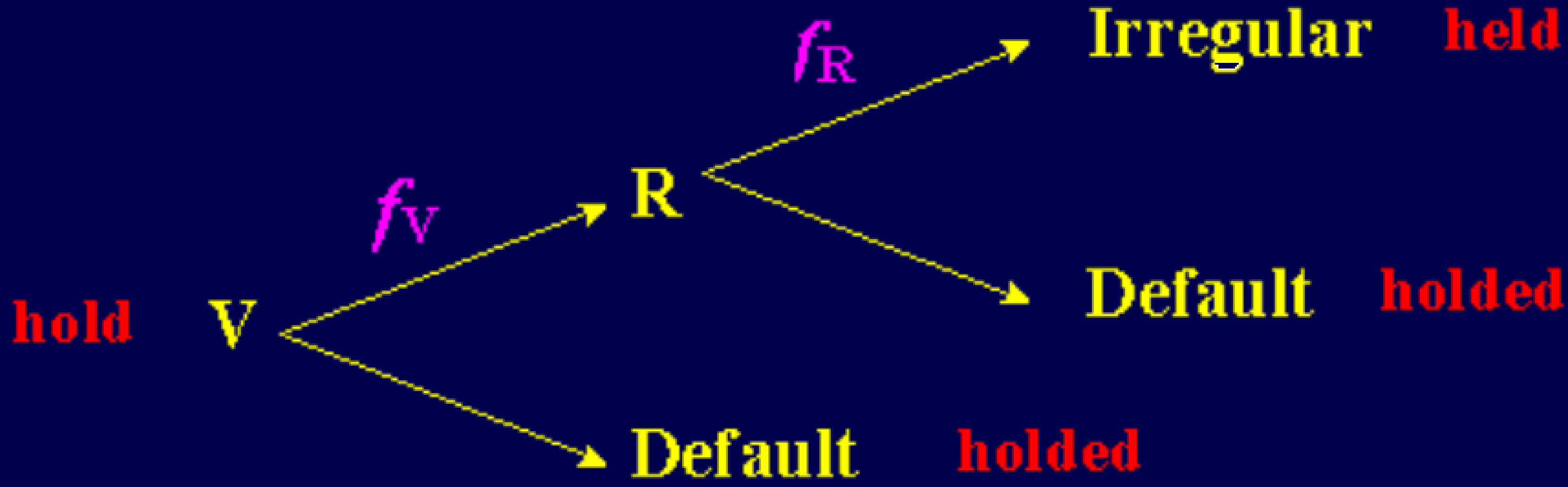
- irregular verbs are organized into *classes*, characterized by special morphophonological rules (Bloch 1947, SPE, Halle & Mohanan 1985)

  (feed, shoot) $\mapsto R_{\text{Vowel Shortening}}$, (fall, come) $\mapsto R_{\text{Umlaut}}$



- When a special rule fails, the default rule kicks in – *hold-holded*.

Two Probabilities



$$f_{\text{V correct}} \propto f_V \times f_R = f_V \times \sum_{i \to R} f_i$$

Empirical Predictions

For two verbs V_1, V_2,

- if $V_1, V_2 \rightarrow R$, then frequency determines learning performance.
- if $f_{V1} = f_{V2}$, then class size determines learning performance.

Data

- All child data from Marcus, Pinker, et al. (1992).

Adam:	2446/2491 = 98.2% correct
Eve:	285/309 = 92.2% correct
Sarah:	1717/1780 = 96.5% correct
Abe:	1786/2350 = 76% correct

- Frequency of adult usage from CHILDES (over 100,000 sentences)

Confirmation 1. Frequency Hierarchy in a Class

1. [-t & Vowel Shortening]
   lost (98%–63), left (95%–53)
2. [-t & Rime → a]
   caught (93%–36), thought (87%–363)*, brought (83%–77), bought (82%–70)
3. [-ø & No Change]
   put (95%–2248), hit (91%–66), hurt (87%–25), cut (71%–21)
4. [-ø & Vowel Shortening]
   shoot (94%–14), bite (89%–13)
5. [-ø & Backing Ablaut]
   got (96%–1511), took (90%–154), wrote (74%–28), won (56%–36)
6. [-ø & Rime → u]
   knew (74%–49), threw (32%–28)

Confirmation 2: Free-rider Effect

verbs have better performance if they belong to larger class.

1. Same frequency ($\approx$ 20), different performance

   [-ø & No Change]: hurt, cut – 80%

   [-ø & Rime $\rightarrow$ u]: draw, blow, grow, fly – 35%

2. Higher performance despite lower frequency

   [-ø & No Change]: hurt (25), cut (21) – 80%

   [-ø & Rime $\rightarrow$ u]: know (58), throw (31) – 49%

3. Lower performance despite higher frequency (Abe)

   [-ø & No Change]: hurt (25), cut (21) – 66%

   suppletion: go (557) – 64%

   [-ø & Umlaut ($\wedge \rightarrow$ ey)]: come (272) – 26%

[-ø & No Change] is HUGE class ($> 3,000$)

[-ø & Rime $\rightarrow$ u] is small (125)

Confirmation 3: Phonological Regularities

a (rare) verb may be used very well if it falls under a general phonological process of the target language

	Verb	*% Correct*	*Input Frequency*
a. [-t]	lose-lost	98%	63
	leave-left	95%	53
b. [-d]	say-said	99%	544
c. [-ø]	shoot-shot	94%	14
	bite-bit	90%	13

Vowel Shortening in English:

[ay]-[ɪ]: divine-divinity

[i]-[ε]: deep-depth

[e]-[æ]: nation-national

[o]-[a]: cone-conic

[u]-[$\wedge$]: deduce-deduction

(SPE, Myers 1987, Halle 1998)

Summary so far ...

- Words are organized in classes, most directly defined by rules.
- When irregular rule fails, the default rule is used (10% of all usage).
- Children almost never use the wrong rule (mis-irregularization): *bring-brang* type errors are extremely rare (0.2% of all usage).

Questions about Rules

- How are the rules learned?
- What's the status of the "default" rule?
- How are words associated with rules?

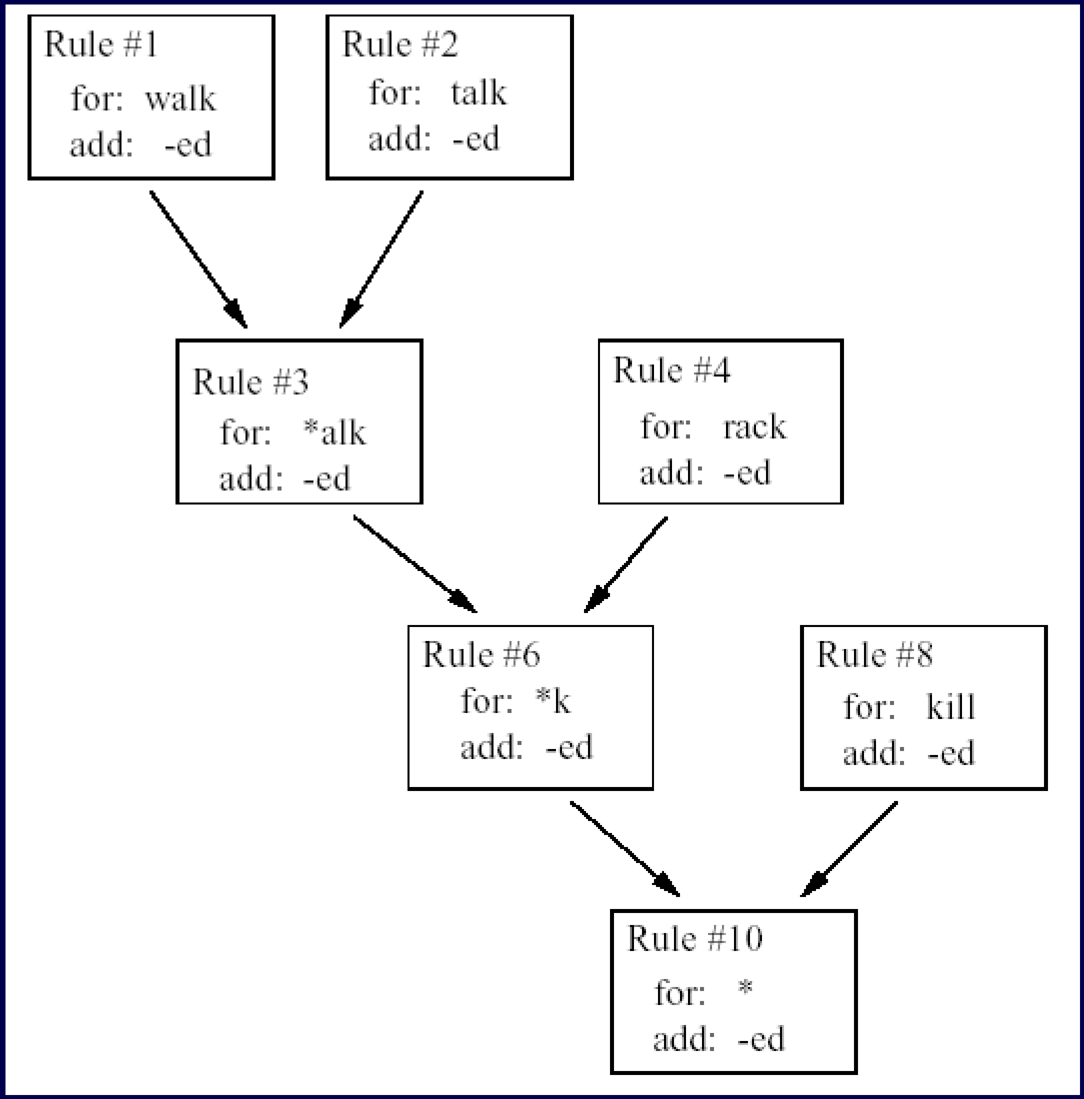
Rules result from the application of a general learning algorithm (Sussman & Yip 1997, 1998; Molnar 2001) to the specific domain of language (phonology, perhaps even more).

Rule Learning

Representations

	[ae]	[p]	[l]	[z]	"apples"
syllabic	1	0	0	0	
consonantal	0	1	1	1	
sonorant	1	0	1	0	
high	0	0	0	0	
back	0	0	0	0	
low	1	0	0	0	
round	0	0	0	0	
tense	0	1	0	1	
anterior	0	1	1	1	
coronal	0	0	1	1	
voice	1	0	1	1	
continuant	1	0	1	1	
nasal	0	0	0	0	
strident	0	0	0	1	

Learning by Generalization



Rules Learned

Regulars

1. Verbs that end in a voiced phoneme but not a `d`:

```
[*.*.[+voice,+sonorant].d]
[*.*.[+voice,-coronal].d]
[*.*.[-low,-round,-tense,+continuant],d]
```

2. Verbs that end in an unvoiced phoneme but not a `t`:

```
[*.*.[-voice,+strident].t]
[*.*.[-voice,-coronal,-continuant].t]
```

3. Verbs that end in `(d, t)`:

```
[*.(d,t).I.d]
```

Irregulars

```
[*.*.i->ae.ng]           rang, sang
[*.*.E->a.t]             forgot, got, shot
[*.E.n.->t]              bent, lent, meant
[*.*.(r,l),*->u]         blew, drew, grew, fly
[*.*.*->o,t]             bought, brought, caught, taught
[*.*.*->o.z]             chose, froze, rose
```

Rule Learning and General Learning

Generalization learning is domain neutral

- In problem solving

  If (it rains) then the ground is wet.

  If (it snows) then the ground is wet.

  $\Rightarrow$ If (things fall from the sky) then the ground is wet.

- In machine learning

  Version space learning (Mitchell 1982)

$$H_0$$

. more general

$$\Downarrow$$

Target Hypothesis

$$\Uparrow$$

. more specific

$$H_\infty$$

- Avoids the problem with minority regular rules (German)

The default

The notion of a default is domain neutral:

- In conversations (Gricean axiom)

  This is a workshop on language acquisition, change, and evolution.

  #This is a workshop on language.

- In soccer games

  (a) If one handles the ball, it's a free kick.

  (b) If one handles the ball in the penalty area, it's a penalty kick.

  If someone handles the ball in the penalty area, rule (b), the more specific one, applies.

- In phonology, the most specific rule applies

  (a) [-t & Vowel shortening] for *lose*+PAST

  (b) [*.d] for *lose*+PAST

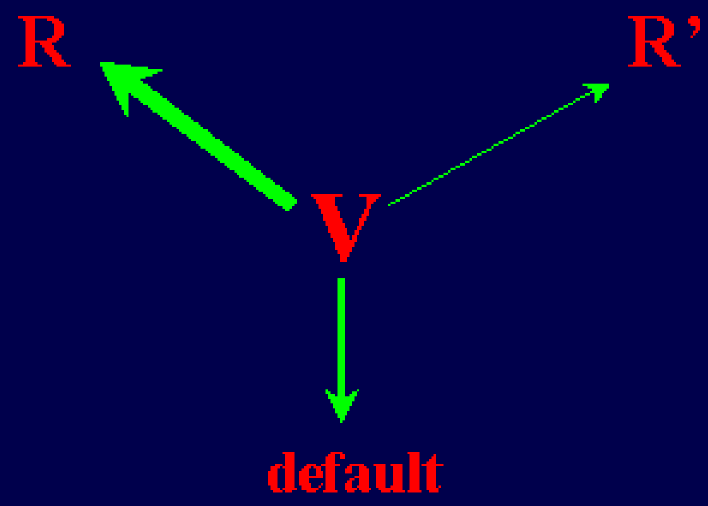
  Rule (a) wins out.

  The rule learning model treats the one with most # of *'s as the default.

Implications for Language Evolution

- If rule learning and competition are not language specific, but general to cognition, then the ability to learn rules may be a constraint that shapes the outcome of language evolution
- In order for rule learning to be applicable, evolution must have supplied the appropriate representation: distinctive features/gestures in articulation, motor movements, an atomic and combinatorial semantic space, etc.
- Must be able to perform abstract symbolic manipulation, and probabilistic learning

The Force of Analogy



- If word-rule associations are learned perfectly, then language will never change.
- Words may shift among rules by *analogy*:
  1. analogical levelling: $V \longrightarrow$ default
     strive/strove $\longrightarrow$ *strive/strived*
  2. analogical extension: $V \longrightarrow$ R’
     wear/weared $\longrightarrow$ *wear/wore*

Salvation by Volume

$$f_{V \rightarrow R} \propto f_V \times \sum_{i \rightarrow R} f_i$$

- A word can stay irregular by the virtue of its frequency and its corresponding class size
- Against Bybee & Slobin (1982) and Pinker (1999)'s frequency-based theories

Facts:

- English irregular verbs

  one huge default class, a small number of small, high-frequency irregular classes

- German noun plurals

  five classes, all of significant sizes

  four irregular classes, and one regular class (-s), which is the smallest

A Computational Model of Sound Change

1. Start with a random set of words, each with a "phonological" shape (e.g., 100 bits of 0's and 1's) as the input and with a modification on the input as its output.
2. For each generation
   (a) group words with identical I/O changes into a super rule via generalization
   (b) for each word w, with a prob. inversely correlated to $f_w f_{R:w \to R}$
      - consider the rules w might fall under
      - associate w to R', whereas R' most specifically matches w.
3. Repeat 2.

Assume an exponential decay of analogy probability wrt $f_w f_{R:w \to R}$.

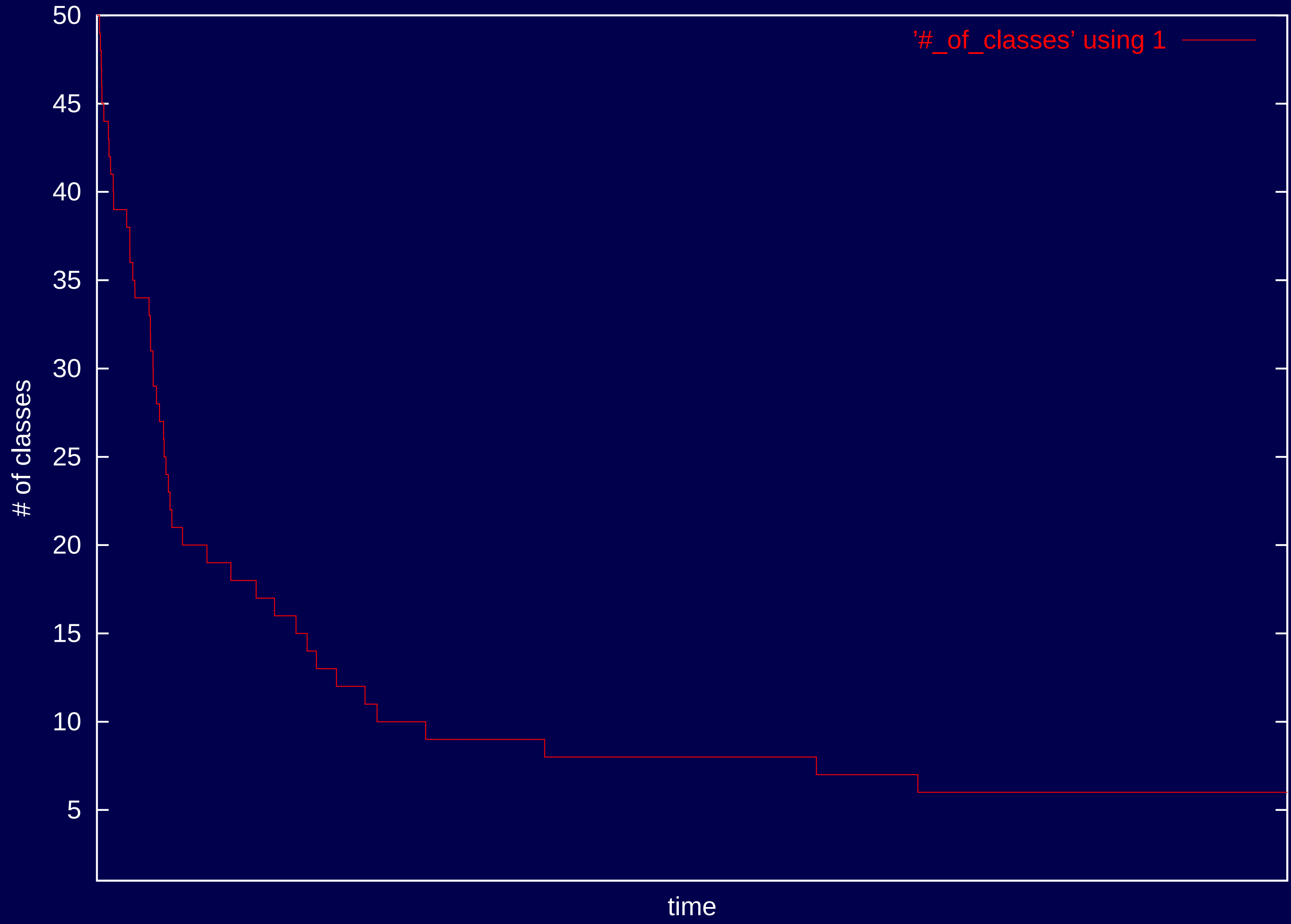
Assume a Zipfian distribution of word frequencies.

A History of Five Words

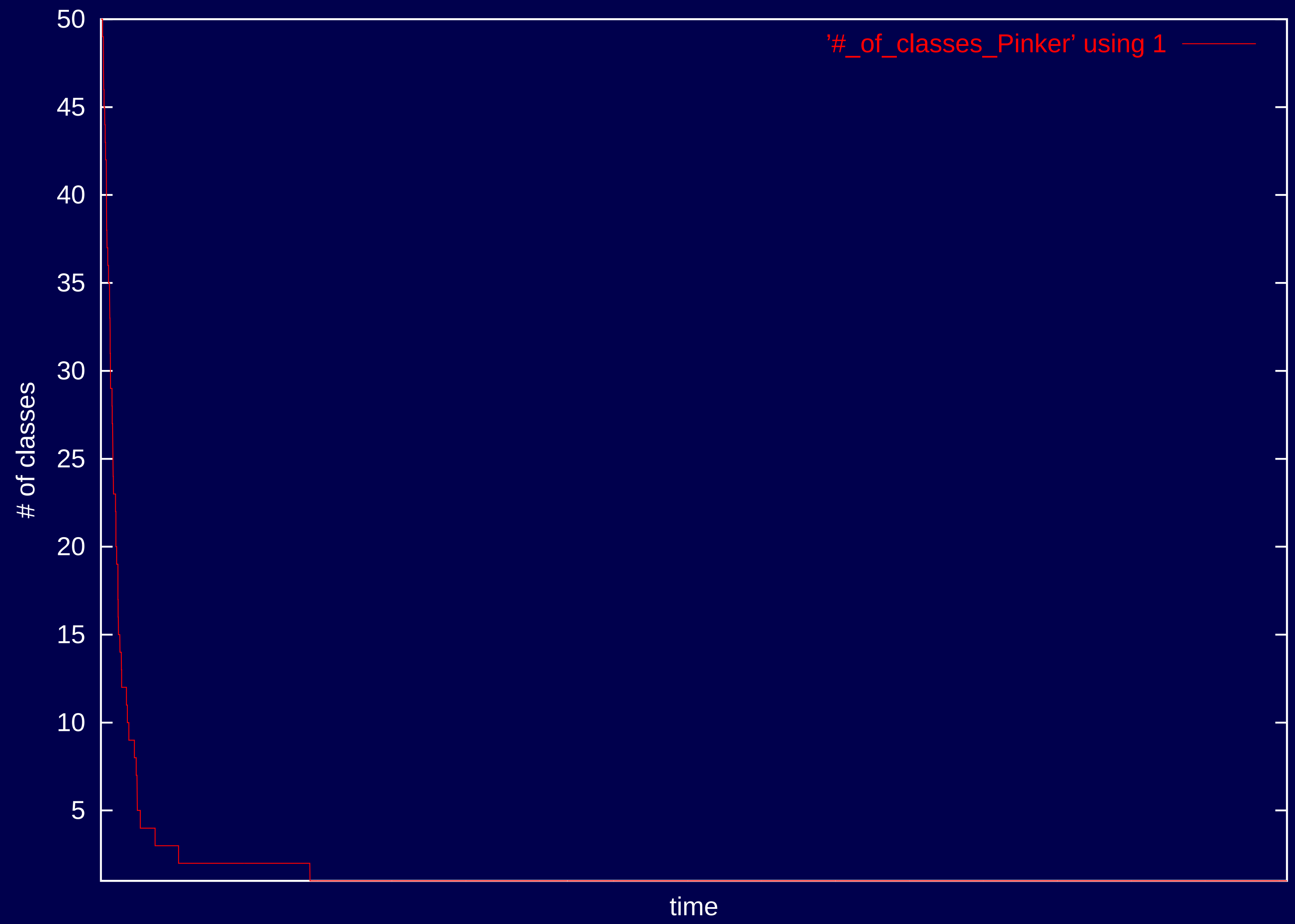
$$w_1 = [00001], w_2 = [01101], w_3 = [01001], w_4 = [11110], w_5 = [01100]$$

Generation	Rule 1	Rule 2	Change
1	(1, 2, 5) `[0 * * 0 *]`	(3, 4) `[* 1 * * * ]`	
2	(1, 5) `[0 * * 0 *]`	(2, 3, 4) `[* 1 * * *]`	$w_2 \longrightarrow R_2$
3.	(1, 3, 5) `[0 * * * 0]`	(2, 4) `[* 1 1 * *]`	$w_3 \longrightarrow R_1$

A History of 500 Words



Another History of 500 Words Under Pinker's Model



The Rules

At stable equillibrium:

Rule	Size
22222222222222222222222222222222222222222222222222	456
22222222222222222221220222222222222222222202222222	12
22222222222222222222222222220222222222222222221222	9
22221222222222222222222222222222222222222222222122	5
22222220222222222222222222222222222222222222222222	9
22222222222222221220222222222222222222222222222222	9

- many of the original rules disappeared
- more were added, and eliminated
- one rule became the default
- four irregular rules contain *relatively* high-frequency words

Future Directions

A framework for modeling sound change

- loan words (randomly replace existing words with new patterns)
  English vs. German
- systematic changes in the phonology, e.g. the Great Vowel Shift (flip some bits in all words)
- richer phonological representation and the role of phonotactics
- extension to morphological change, e.g., case

On Rules

- The phonologies of words are learned in rules/classes
- The algorithms for rule learning and use seems to be domain general
- Extrapolating learning back in time may explain the emergence of sound classes, and irregularity vs. regularity.

On Analogy

- Analogy is extremely weak in learning.
- Analogy is cumulatively strong in history.
- Analogy leads irregularity.

If language had to work with rule-based and analogical learning, then "imperfection" is a necessity.