Exceptions

Class 2: Finding predictability

Adam Albright

CreteLing 2024

creteling2024.phonology.party

Taking stock

- Reorientation: goal of analysis is not to minimize exceptions, but to maximize predictiveness
    - Concentrate probability mass on the data
- We were rather informal about what a predictive grammar looks like and how it's identified
- Today: examine a specific model of rule-learning
    - How it discovers predictive rules
    - Evaluate performance
    - Consequences for what is law-governed and what is exceptional

Strategy

- Goal is to assess whether we can compare analyses based on which forms they consider exceptional
- Makes sense to start with cases where there's widespread agreement about which forms are exceptional: irregular morphology
  - Numerous correlates of exceptionality
- Game plan
  - Irregular morphology: English past tenses
  - Irregular morphophonology: Spanish diphthongization
  - Phonology: Dutch, Turkish

Morphological exceptions

Irregular morphology: English past tenses

- Regular/default suffix, with contextual allomorphy
  - *wait*[əd], *need*[əd]
  - *jump*[t], *kick*[t], *laugh*[t], *brush*[t]
  - *rub*[d], *sag*[d], *starve*[d], *use*[d]
- Irregular suffixes
  - Wrong allomorph: *burn*[t]
  - Null suffix: *lay*-∅
  - Vowel changes (usually with null suffix): *slid*, *wrote*, *ran*, *sang*, *clung*
- More drastic changes
  - *caught*, *was*

The model

- Minimal Generalization Model (MGL) (Albright & Hayes, 2002)
- Simple model of rule induction, designed to find properties that predict distribution of competing outcomes
- In this case, allomorphy (almost perfectly predictable) and irregularity (less predictable)

Inducing morphological rules from data

- Words that show common morphological behavior (e.g., past in *-ed*, past with [aɪ] → [oʊ]) are compared to see what they have in common
- This common material forms the structural descriptions for rules
  - E.g., aɪ → oʊ / [+cor] ___ $\begin{bmatrix} +\text{cor} \\ +\text{cont} \end{bmatrix}$
- These rules can then be applied to produce new (possibly innovative) forms
- For concreteness, we start with an approach sketched by Pinker & Prince (1988), developed and implemented in the form of the Minimal Generalization Learner (Albright & Hayes, 2002)

Step 1: Input data

- Assume that the learner has available some (present, past) pairs, as input data for learning the present→past mapping

a.	mɪs	mɪst	'*miss(ed)*'
b.	prɛs	prɛst	'*press(ed)*'
c.	læf	læft	'*laugh(ed)*'
d.	hʌg	hʌgd	'*hug(ged)*'
e.	rʌb	rʌbd	'*rub(bed)*'
f.	nid	nidəd	'*need(ed)*'
g.	dʒʌmp	dʒʌmpt	'*jump(ed)*'
h.	plæn	plænd	'*plan(ned)*'
i	sɪŋ	sæŋ	'*sing/sang*'
j.	drɪŋk	dræŋk	'*drink/drank*'

Step 2: Discovering changes and contexts

Comparison of individual data pairs (Pinker 1984; **?**; Ling & Marinov, 1993; Albright & Hayes, 2002)

E.g., Pinker & Prince (1988), p. 130:

1. "Candidates for rules are hypothesized by comparing base and past tense versions of a word, and factoring apart the changing phonetic portion, which serves as the rule operation, from certain morphologically-relevant phonological components of the stem, which serve to define the class of stems over which the operation can apply." (= the *context*)

Step 2: Discovering changes and contexts (*cont.*)

- Factor out the changing portion: some alignment is necessary to find minimal change
- Edge-in alignment: start from both left and right, aligning until there's a mismatch; combined leftward and rightward alignments to align as many segments as possible

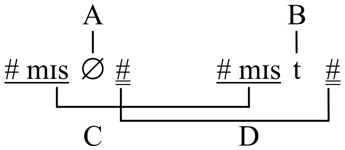
$$\begin{array}{ccc||c} \text{m}_1 & \text{ɪ}_2 & \text{s}_3 & \\ \text{m}_1 & \text{ɪ}_2 & \text{s}_3 & \text{t} \end{array} \quad + \quad \begin{array}{cccc||} & \text{m} & \text{ɪ} & \text{s} \\ \text{m} & \text{ɪ} & \text{s} & \text{t} \end{array} \quad = \quad \begin{array}{ccc||c||} \text{m}_1 & \text{ɪ}_2 & \text{s}_3 & \\ \text{m}_1 & \text{ɪ}_2 & \text{s}_3 & \text{t} \end{array}$$

- Medial changes: leftward and rightward scans are both able to align some segments; residue left in the middle

$$\begin{array}{c||cc} \text{s}_1 & \text{ɪ} & \text{ŋ} \\ \text{s}_1 & \text{æ} & \text{ŋ} \end{array} \quad + \quad \begin{array}{cc||c} \text{s} & \text{ɪ} & \text{ŋ}_1 \\ \text{s} & \text{æ} & \text{ŋ}_1 \end{array} \quad = \quad \begin{array}{c||c||c} \text{s}_1 & \text{ɪ} & \text{ŋ}_2 \\ \text{s}_1 & \text{æ} & \text{ŋ}_2 \end{array}$$

Step 2: Discovering changes and contexts (*cont.*)

- This simple alignment strategy is not adequate for all morphological processes, but will be sufficient for all the examples we discuss here
- Mismatched portions = change (A → B); aligned portions = context (C ___ D)



- Rule: ∅ → t / mɪs ___

Step 2: Discovering changes and contexts (*cont.*)

- Result: a set of word-specific candidate rules

a. $\emptyset \rightarrow$ t / #mɪs ___ #

b. $\emptyset \rightarrow$ t / #prɛs ___ #

c. $\emptyset \rightarrow$ t / #læf ___ #

d.. $\emptyset \rightarrow$ d / #hʌg ___ #

e. $\emptyset \rightarrow$ d / #rʌb ___ #

f. $\emptyset \rightarrow$ əd / #nid ___ #

g. $\emptyset \rightarrow$ t / #dʒʌmp ___ #

h. $\emptyset \rightarrow$ d / #plæn ___ #

i. ɪ $\rightarrow$ æ / s ___ ŋ#

j. ɪ $\rightarrow$ æ / dr ___ ŋk#

Step 2: Discovering changes and contexts (*cont.*)

2. "If two rule candidates have been coined that have the same change operation, a single collapsed version is created, in which the phonological feaures distinguishing their class definitions are eliminated."
    - E.g., *miss* and *press* both take $\varnothing \rightarrow t$. Assume that this is because they have some crucial property in common
        - In both cases, the change is word-final
        - In both cases, the segment before the change is an [s]
        - In both cases, the segment before [s] is a non-low lax front V
        - etc…??? (sonorant before the vowel, monosyllabic, etc.)
    - Albright & Hayes (2002): *Minimal generalization* approach
        - Conservative collapsing: new rule keeps *everything* that the pair has in common

Step 2: Discovering changes and contexts (*cont.*)

- Once again, some alignment scheme is needed. Assume *locality*

	m	ɪ	s	___	#
p	r	ɛ	s	___	#

  Precludes many possible generalizations, such as:
  - First segment is labial in both cases
- A pragmatic assumption: *myopic generalization*
  - Once a mismatch is encountered and featural generalization is needed, any shared similarities that are even farther away from the change location are not likely to be crucial; convert more distant segments to a free variable (X)
  - See Albright & Hayes (2006) for a procedure that permits one to relax this assumption

Step 2: Discovering changes and contexts (*cont.*)

- Comparing *miss* and *press*: $\varnothing \rightarrow$ t / …

	Residue	Shared Features	Shared Segments	Change Location	Shared Segments	Shared Features	Residue
A.	#m	ɪ	s	____	#		
B.	#pr	ɛ	s	____	#		
C.	X	$\begin{bmatrix} +\text{syllabic} \\ -\text{low} \\ -\text{back} \\ -\text{tense} \\ -\text{round} \end{bmatrix}$	s	____	#		

Step 2: Discovering changes and contexts (*cont.*)

- Minimal generalization of features: all shared feature values are retained
  - Prevents generalization to unseen feature values
  - Does permit generalization to unseen feature *combinations*, however
  - Example: comparing *b* and *n*

$$
\begin{array}{c}
\text{b} \\
\begin{bmatrix} -\text{syllabic} \\ +\text{cons} \\ -\text{contin} \\ +\text{voice} \\ -\text{nasal} \\ +\text{labial} \\ -\text{coronal} \\ -\text{velar} \end{bmatrix}
\end{array}
+
\begin{array}{c}
\text{n} \\
\begin{bmatrix} -\text{syllabic} \\ +\text{cons} \\ -\text{contin} \\ +\text{voice} \\ +\text{nasal} \\ -\text{labial} \\ +\text{coronal} \\ -\text{velar} \end{bmatrix}
\end{array}
\mapsto
\begin{array}{c}
\{\text{ b, m, d, n}\} \\
\begin{bmatrix} -\text{syllabic} \\ +\text{cons} \\ -\text{contin} \\ +\text{voice} \\ \\ \\ \\ -\text{velar} \end{bmatrix}
\end{array}
$$

Step 2: Discovering changes and contexts (*cont.*)

- Comparing *hug* and *plan*: $\varnothing \rightarrow$ d / …

	Residue	Shared Features	Shared Segments	Change Location	Shared Segments	Shared Features	Residue
A.	#hʌ	g		___	#		
B.	#plæ	n		___	#		
C.	X	$\begin{bmatrix} -\text{syllabic} \\ -\text{continuant} \\ -\text{labial} \\ -\text{lateral} \\ +\text{voice} \\ \ldots \end{bmatrix}$		___	#		
And generalizing further with *rub*:							
D.	#rʌ	b		___	#		
E.	X	$\begin{bmatrix} -\text{syllabic} \\ -\text{continuant} \\ -\text{lateral} \\ +\text{voice} \\ \ldots \end{bmatrix}$		___	#		

Step 2: Discovering changes and contexts (*cont.*)

- Comparing *sing* and *drink*: ɪ → æ / …

	Residue	Shared Features	Shared Segments	Change Location	Shared Segments	Shared Features	Residue
A.	#	s		___	ŋ		#
B.	#d	r		___	ŋ		k#
C.	X	$\begin{bmatrix} -\text{syllabic} \\ +\text{coronal} \\ +\text{continuant} \end{bmatrix}$		___	ŋ		Y

Step 2: Discovering changes and contexts (*cont.*)

3. “Rule candidates increase in strength each time they have been exemplified by an input pair”

	Rule	Examples covered
1.	∅ → t / #mɪs ___ #	1
2.	∅ → t / #prɛs ___ #	1
3.	∅ → t / X [+syl, −low, −bk, −tns, −rnd] s ___ #	2
4.	∅ → d / #hʌg ___ #	1
5.	∅ → d / #plæn ___ #	1
6.	∅ → d / ___ X [−syl, −cont, −lab, −lat, −del.rel.]	2
7.	ɪ → æ / #s ___ ŋ#	1
8.	ɪ → æ / #dr ___ ŋk#	1
9.	ɪ → æ / X [−syl, +cor, +cont] ___ ŋ Y#	2

Step 2: Discovering changes and contexts (*cont.*)

4. "When an input stem has to be processed by the system …[it] is matched in parallel against all existing rule candidates, and if it falls into several classes, several past tense forms may be generated."
    - Example: novel verb *spling* [splɪŋ] meets criteria for rules 6 ($\varnothing$ $\rightarrow$ d) and 9 (ɪ $\rightarrow$ æ)
    - Thus, both *splinged* and *splang* are generated

Step 2: Discovering changes and contexts (*cont.*)

5. "The outcome of a competition among the past tense forms is determined by the strength of the relevant rule and the proportion of a word's features that were matched by that rule."
    - In this unrealistically small example, both rules have strength 2, predicting tie.
    - In real English, steady stream of new $\varnothing \rightarrow d$ words quickly boosts strength of *splinged*

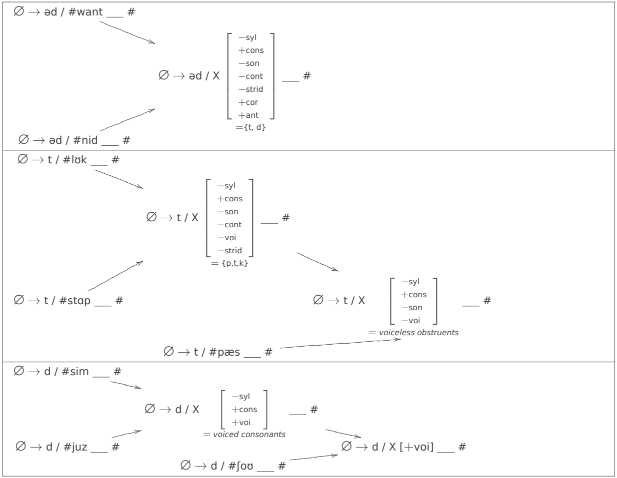
Pinker and Prince's final step

(p. 134): "[by various means], some regularities can be enshrined as permanent productive rules whereas others can be discarded or treated differently."

- Rather vague on what should be "enshrined", but larger context implies that final grammar will consist mainly of the "regular" rule(s)
- As we'll see below, there's reason to suppose that learners retain a substantial set of regularities

What this model does well

- Generalizes quickly over input pairs to discover most general environment for each pattern



What this model does well (*cont.*)

- Generalization proceeds rapidly, given sufficiently diverse stems
- The pathways shown here can be found using verbs that are among the 75 most frequent verbs of English (according to CELEX) (of which the majority are actually irregular)
- By counting the STRENGTH of each rule, we find the hypothesis that covers as much of the data as possible

Rule creation not limited to overtly suffixed forms

ɪ → æ / #s ___ ŋ#

$$\text{ɪ} \rightarrow \text{æ} \;/\; \text{X} \begin{bmatrix} -\text{syl} \\ +\text{cons} \\ +\text{cont} \\ +\text{cor} \end{bmatrix} \;\_\_\_\; \text{ŋ}\#$$

= θ, ð, s, z, ʃ, ʒ, l, r

Strength = 2

ɪ → æ / #r ___ ŋ#

$$\text{ɪ} \rightarrow \text{æ} \;/\; \text{X} \begin{bmatrix} -\text{syl} \\ +\text{cons} \\ +\text{cont} \\ +\text{cor} \end{bmatrix} \;\_\_\_\; \text{ŋ}\;\text{Y}\#$$

= θ, ð, s, z, ʃ, ʒ, l, r

Strength = 3

ɪ → æ / #dr ___ ŋk#

- Irregular patterns (at least in English) tend to cover relatively few forms, which are similar to one another ≈ share a set of phonological properties

Rule creation not limited to overtly suffixed forms (*cont.*)

- Result: rules that characterize them are more specific, and weaker strength
- These patterns are not productive, except possibly for inputs that fit a very particular phonological shape ($\approx$ prototypical forms)

Where this version of the model fails quickly

Consider hypothesized rules after exposure to *sing*, *ring*, *drink*, *sit*

ɪ → æ / #s ___ ŋ#

ɪ → æ / #r ___ ŋ#

$$\text{ɪ} \rightarrow \text{æ} \;/\; X \underbrace{\begin{bmatrix} -\text{syl} \\ +\text{cons} \\ +\text{cont} \\ +\text{cor} \end{bmatrix}}_{=\ \theta,\ \eth,\ \text{s, z, ʃ, ʒ, l, r}} \text{\_\_\_ ŋ\#}$$

Strength = 2

ɪ → æ / #dr ___ ŋk#

$$\text{ɪ} \rightarrow \text{æ} \;/\; X \underbrace{\begin{bmatrix} -\text{syl} \\ +\text{cons} \\ +\text{cont} \\ +\text{cor} \end{bmatrix}}_{=\ \theta,\ \eth,\ \text{s, z, ʃ, ʒ, l, r}} \text{\_\_\_ ŋ}\ Y\#$$

Strength = 3

ɪ → æ / #s ___ t#

$$\text{ɪ} \rightarrow \text{æ} \;/\; X \underbrace{\begin{bmatrix} -\text{syl} \\ +\text{cons} \\ +\text{cont} \\ +\text{cor} \end{bmatrix}}_{=\ \theta,\ \eth,\ \text{s, z, ʃ, ʒ, l, r}} \text{\_\_\_} \underbrace{\begin{bmatrix} -\text{syl} \\ +\text{cons} \\ -\text{cont} \end{bmatrix}}_{=\ \text{p,t,k,b,d,g,m,n,ŋ,tʃ,dʒ}} \#$$

Strength = 4

- “Prototypical” forms (*sing*, *ring*, *drink*) and “outlier” (*sit*) combine to yield more general rule

Where this version of the model fails quickly (*cont.*)

- More general rule encompasses more forms, so has higher strength
- Prediction: ɪ→æ change may apply equally well to any form within this broader context
  - Hypothetical *zick* $\sim$ *zack* just as well supported as *spling* $\sim$ *splang*

Diagnosis:

- Strength, measured purely in terms of number of examples covered, favors the broadest possible generalizations
- Under this view, "compatible with the data" $=$ "encompasses all the examples"

Where this version of the model fails quickly (*cont.*)

- Broad generalizations are not always virtuous; sometimes, they incur numerous exceptions.
- A more intuitive notion of “compatible with the data” must incorporate not only generality, but also accuracy

Reliability of a hypothesized rule

$$= \frac{\text{\# of items that the rule works for (“\textit{hits}”)}}{\text{\# of items that meet the structural description (CAD) (“\textit{scope}”)}}$$

- Also known as ACCURACY, or COVERAGE

Rule	Hits	Exceptions
$\text{ɪ} \rightarrow \text{æ} \;/\; \text{X} \underbrace{\begin{bmatrix} -\text{syl} \\ +\text{cons} \\ +\text{cont} \\ +\text{cor} \end{bmatrix}}_{= \text{θ, ð, s, z, ʃ, ʒ, l, r}} \_\_\_ \;\text{ŋ}\;\; \text{Y\#}$	*ring*, *sing*, *drink* $= n$	*wring*, *fling*, ... $= m$
$\text{ɪ} \rightarrow \text{æ} \;/\; \text{X} \underbrace{\begin{bmatrix} -\text{syl} \\ +\text{cons} \\ +\text{cont} \\ +\text{cor} \end{bmatrix}}_{= \text{θ, ð, s, z, ʃ, ʒ, l, r}} \_\_\_ \underbrace{\begin{bmatrix} -\text{syl} \\ +\text{cons} \\ -\text{cont} \end{bmatrix}}_{= \text{p,t,k,b,d,g,m,n,ŋ,tʃ,dʒ}} \;\text{Y\#}$	*ring*, *sing*, *drink*, *sat* $= n+1$	*wring*, *fling*, *lick*, *rig*, *ship*, *rid* ... $= m+\text{lots}$

Reliability of a hypothesized rule (*cont.*)

- Island of reliability: context in which a particular change is especially likely to apply
  - "especially likely" = more likely than in its most general context (Albright, 2002)

The intended effect

- Intuitively, favoring rules based on reliability should let the model find specific features which (tend to) *uniquely* characterize members of a particular class
- Generalizing to overly broad contexts is discouraged, because of the cost of adding exceptions
  - I.e., number of positive hits gained (in numerator) must exceed number of exceptions added to scope (in denominator)
  - For small, “irregular” class, optimal contexts will be quite narrow
  - For the “regular” class, there are (by hypothesis) no distinguishing features that uniquely characterize the bulk of regular verbs, so no benefit to staying small; model zooms to full generality, incurring a few exceptions along the way
- Guides model to the right level of generality for each class

How this plays out in practice

- Example: ran the model on the 200 most frequent verbs in CELEX
    - 134 regular (80 -*d*, 30 -*t*, 24 -*əd*); 66 irregular, of various types

Islands of reliability: irregulars

- The model discovers narrow environments where irregular changes are especially common
  - 4/5: i → ɛ / [+cons] ____ [−son,−cont,+cor] (*read*, *lead*, *meet*, *feed*)
  - 2/4: ɪ → æ / [−syl,+voi] ____ ŋ Y (*begin*, *drink*)
  - 3/6: aɪ → oʊ / X ɹ ____ [+cor] (*write*, *drive*, *rise*)

Islands of reliability: regulars

- The model does discover general rules for allomorphs of the regular/default suffix
  - 24/45: $\varnothing \rightarrow$ əd / [−son,−contin,+cor] ___
  - 30/60: $\varnothing \rightarrow$ t / [−voi] ___
  - 80/134: $\varnothing \rightarrow$ d / [+voi] ___
- However, it finds many more reliable specific contexts
- E.g., for $\varnothing \rightarrow$d, there are 170(!) more reliable rules
  - 11/11: $\varnothing \rightarrow$ d / ɚ ___
  - 15/16: $\varnothing \rightarrow$ d / [−high,−low] ___
  - 13/14: $\varnothing \rightarrow$ d / [−high] n ___
  - 18/21: $\varnothing \rightarrow$ d / [−ant,+voi] ___ (= ɹ,ʒ,dʒ)
- Similarly for other allomorphs
  - 10/10: $\varnothing \rightarrow$ əd / C t ___
  - 8/11: $\varnothing \rightarrow$ əd / V d ___
  - 14/14: $\varnothing \rightarrow$ t / [+contin,−voi] ___
  - 4/4: $\varnothing \rightarrow$ t / [−high,−low] k ___

An unintended consequence!

- This maneuver has something of the desired effect (finds consists clusters of similar words, and tells us that it is advantageous to exclude more distant words), but it takes small-scale generalizations too seriously
- By favoring reliability instead of generality, we end up finding a long list of very accurate, but seemingly quite accidental contexts

Why are the more general contexts so unreliable?

- Genuine exceptions: almost half of *t*, *d*-final verbs in this sample are irregular
  - We would expect this proportion to go down in a larger sample (general contexts improve, and we may encounter exceptions to accidentally consistent subcontexts)
  - Larger sample will never make the problem disappear completely, however, since some subcontexts are 100% regular (so will always be more reliable than the corresponding general context)
    - Verbs ending in voiceless fricatives, verbs ending in [ɚ], etc.
  - In any system with irregularity, broad generalizations will necessarily involve some exceptions (decreased reliability)
- Inadequate characterization of rule interaction:
  - *t*, *d*-final verbs act as exceptions for more general -*d*, -*t* suffixes

Tackling the generality vs. reliability problem

- Intuitively, it appears that we need some way to pay less attention to the small-scale generalizations, and to reward broader generalizations
- What is the correct trade-off?
  - How many exceptions can a simple, broad generalization have before we stop trusting it, and opt for a more specific generalization?
  - How many examples must a consistent subcontext have in order for us to take notice?
- The guiding intuition thus far

Tackling the generality vs. reliability problem (*cont.*)

- The “regular” past tense class is well-enough instantiated/has few enough exceptions that it is most appropriately stated in terms of unified, broad contexts
- “Irregular” classes (like ɪ→æ) are restricted to rather particular phonological contexts, and should be stated in these terms, even if they happen to have a few members outside the “core” contexts

- The model fails to capture this intuition
  - It finds subgeneralizations (“islands of reliability”) for the regular suffix as well as for irregular changes

But what is the basis for this intuition? Are we holding the model to an unrealistic standard?

Empirical interlude

Commonly held belief about English past tenses (Pinker & Prince, 1988)

- Irregular classes have high degree of internal similarity
  - Members tend to resemble one another to a significant extent
  - Characteristic "prototypes" or "schemas" (Bybee & Moder, 1983)
  - Novel ("wug") words only classified as irregular when they are similar to prototypical members of the relevant class
- Regulars, on the other hand, can be of any phonological shape
  - No such thing as a "prototypical regular"
  - Wug words treated as regular even if not particularly similar to any existing words

Prasada & Pinker (1993)

- Constructed wug words of varying degrees of similarity to existing regular and irregular verbs

	Prototypical	Intermediate	Distant
Pseudo-irregular	spling, cleed, froe	ning, cleef, voa	trisp, blip, goav, flape
Pseudo-regular	plip, proke, slace	brilth, ploab, smeeg	trilb, smairg, smeenth

- Asked experimental participants to volunteer past tense forms, or to rate goodness of suffixed and vowel-change pasts (“John likes to *smeenth*. Yesterday, he ___.”)
- Crucial comparison: does distance (phonological context) play a greater role for irregulars than for regulars?

☞ On the face of it, not really!

  - Wug forms like *plipped* do sound better than forms like *trilbed*.
  - Ambiguous result: have speakers have retained a subgeneralization that specifically covers verbs ending in *p* (or *ip*, or...)? Or just because *trilbed* is phonotactically awkward?

Albright & Hayes (2003) *Rules vs. analogy*

- Attempted to de-confound phonological well-formedness (likelihood of the word) and morphological well-formedness (likelihood of process in that phonological context)
- Morphological well-formedness is measured as reliability in different phonological contexts
- As a result, it is possible to construct both low- and high-likelihood regulars that are phonotactically ordinary

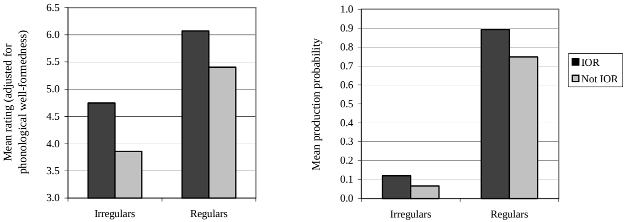
Albright & Hayes (2003) *Rules vs. analogy* (*cont.*)

Four-way comparison:

Island of reliability for regulars only (tɛʃ, brɛdʒ, wɪs, spæk)	Island of reliability for both regs & irregs (daɪz, raɪf, froʊ)
Island of reliability for neither (priːk, trɪsk, nʌŋ, pæŋk)	Island of reliability for irregs only (splɪŋ, blɪg, glid, flip)

- Pre-test to gather phonotactic ratings (how plausible does X sound as an English verb?)
- Wug test: participants produce or rate past tense forms
- Any small differences in phonotactic well-formedness are controlled for by partialling out phonotactic ratings from past tense ratings

Results



(a) IOR Effect on Ratings (adjusted) (b) IOR Effect on Production Probabilities

- Unsurprisingly, a general preference for regulars
    - Corresponds to the fact that rules for regulars are very reliable, and supported by a large number of forms

Results (*cont.*)

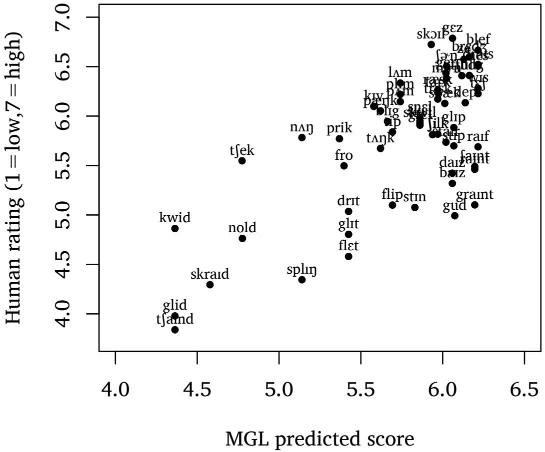
- Rules for irregulars are much less reliable, since it is difficult to isolate phonological contexts in which irregular verbs are dominant (ɪŋ $\rightarrow$ ʌŋ, iːp $\rightarrow$ ɛpt, etc.)

Results (*cont.*)

- Within irregular responses, preference for those involving high reliability contexts
  - ɪŋ → ʌŋ, iːp → ɛpt, iːd → ɛd
  - Supports the intuitive claim that patterns compete according to their strength in the lexicon
  - Connection to change: occasional over*ir*regularization
- Significant effect of phonological context (IOR's), both for regulars & irregulars
  - Irregulars sound better/are produced more often when the verb is in a "strongly irregular context" (*fleep/flept* > *preak/prekt*, *blig/blug* > *trisk/trusk*)
  - Regulars also sound better/are produced more often when the verb is in a "strongly regular context" (*wiss/wissed* > *trisk/trisked*, *spack/spacked* > *pank/panked*)

Results (*cont.*)

- The effect for regulars, in more detail ($r = .71$)



Taking stock: what we have seen so far

- Generalization of morphological patterns involves competition, according to strength among existing words
- Pattern strength depends not just on the number of forms involved, but also on the number of exceptions (accuracy, or reliability)
- A ‘single morphological process’ may have difference strengths in different phonological contexts (islands of reliability)

Type vs. token frequency

- So far, we have estimated pattern strength/reliability in terms of the number of words involved (type frequency)
- In principle, the fact that some verbs occur very frequently could boost the strength of their patterns (type frequency)
- Empirically, type frequency appears to be the primary (or only?) determinant of pattern strength (Bybee, 1995; Hay et al., 2004; Buchwald, 2007)

Unimportance of token frequency: regular defaults

- 200 most frequent verbs in CELEX, regulars outnumber irregulars in types by nearly 2:1 (126 vs. 69), but reverse holds of tokens (571,934 vs. 1,007,737, excluding *have*, *be*)
- Top 1000 verbs, regulars outnumber irregulars in types by more than 6:1 (856 vs. 138), but by tokens, regulars and irregulars are very closely matched (1,033,467 vs. 1,063,395)
- All 4271 verbs with token freq $>$10 in CELEX, there is not a strong majority of regular tokens (1,239,162 vs. 1,069,258)
- If pattern strength depended on token frequency, preference for *-ed* should be weak, and emerging as the lexicon grows
- In fact, the preference for regulars emerges early and is strongly asymmetrical

Unimportance of token frequency: single-form analogy

- If pattern strength could be supported by either many types or many tokens, then a limiting case is patterns with just one type by high token frequency
- Examples: *see* $\sim$ *saw*, *say* $\sim$ *said*, *come* $\sim$ *came*, *give* $\sim$ *gave*, *get* $\sim$ *got*, *choose* $\sim$ *chose*, *tell, sell* $\sim$ *told, sold*
- Albright & Hayes (2003): tested nonce verbs that were very close to these existing words: *shee*, *zay*, *pum*, *lum*, *kive*, *gez*, *choole*, *snell*,
- Result: generalizations based on single models were occasionally volunteered, but extremely rarely (2/44 *kave*; 2/88 *pame*, *lame*; 0/44 *goz*, *zed*, *snold*, etc.)
- Conclusion: high token frequency does not appear to encourage generalization

Where this leaves us

- Synchronic generalization (i.e., responses to wug verbs in experimental settings) is modeled well by a grammar of rules that produce competing morphological patterns, with different strengths (reliabilities) in different phonological contexts
- Rules for ‘regular’ (-*ed*) patterns tend to be the most reliable in English, since most verbs are regular (high type frequency)
  - Most wug reponses are regular -*ed*
- Rules for ‘irregular’ patterns can be fairly reliable in specific phonological contexts (ɪ → æ / [+son] ___ ŋ, iː → ɛ / [liquid] ___ d)
  - Wug verbs like *spling*, *gleed* often inflected *splang/splung*, *gled*
  - Wug verbs like *trisk*, *preak* rarely inflected *trask/trusk*, *preck*

Identifying the exceptions

- With this model in place, we can now ask whether it makes sensible predictions about morphological exceptions?
- A simple way of defining exceptions
  - Law-governed = forms for which the rules give highest probability to the attested output
  - Exceptions = forms for which the rules give highest probability to an incorrect output
- An interesting possibility we might come back to later: gradient exceptionality (Moore-Cantwell)

Examining the exceptions

- Used MGL to learn a grammar for 4,243 verbs most frequent verbs in CELEX (lemma frequent ≥ 10)
  - 4,032 regular $=$ 95%
- 52 different changes, most involving just a few verbs
- Used the grammar to derive outputs for all 4,243 trained words
- Calculated accuracy as defined above:
  - Correct $=$ model assigns highest probability to trained form
  - Incorrect $=$ model assigns highest probability to a different form

Which forms are exceptional?

- Overall, grammar gets 95% of trained words correct
- Unsurprising! Grammar prefers ‘regular’ outputs, so prob correct $\approx$ prob regulars
- Don’t forget: we’re using this as a sanity check, to test the model on a case where we have a pretty good sense of what the exceptions are

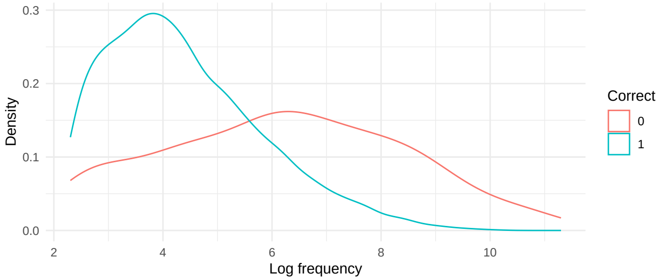
Producing existing exceptions

- In real life, human speakers (usually) produce the exceptional forms correctly
- Two main approaches
    - Exceptions bypass the grammar (dual route)
    - Exceptions undergo a different grammar (lexical indexation/co-phonologies)
- In the current model, “word-specific rules” could be thought of in either way (memory, or lexically-specific grammar)

Generalization: wug words

- (Albright & Hayes, 2003) show that this model provides a fairly good match to human ratings/production probabilities in a wug task
- Various details
  - Contextual allomorphs: phonology, or separate rules?
- Relation between reliability and ratings/probabilities

Which forms are exceptional?



- Exceptions (irregulars) strongly skewed towards higher frequency items
- Density plot provides a quick and dirty way of visualizing this

Acquisition: overregularization

- Ervin and Miller (1963) and many subsequent authors: characterize development of irregular verbs in three stages
  - Early correct use of irregulars (*ate*, *held*, etc.)
  - Overregularization: period in which irregulars are incorrectly produced as regulars (☆*eated*, ☆*holded*)
  - Recovery: child eventually resumes correct use of irregulars
- Claimed pattern: U-shaped development
  - Plotting accuracy vs. age shows temporary decrease in accuracy
- If this is right, then it could be a challenge to view that overregularization is ‘innocent’ (using grammar to project unknown forms)

‘U-shaped’ development

Marcus et al. (1992): *Overregularization in Language Acquisition*

- Systematic study of available longitudinal data from children passing through age range in which verb errors tend to be greatest (generally ages 2–5)
- Finding 1: overregularization indeed often (though not always) preceded by correct irregular productions of the same verb
- E.g., productions by Abe of four irregular verbs (Figs 12–15, pp. 54–55)
    - Wild fluctuations probably due at least in part to sampling issues

‘U-shaped’ development (*cont.*)

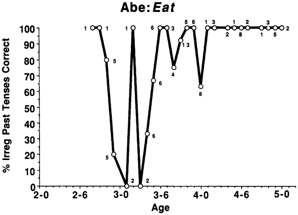


FIG. 12.—Example of a verb that resists overregularization as the child gets older

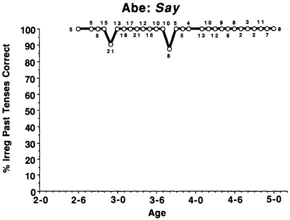


FIG. 13.—Example of a verb that is rarely overregularized at any age

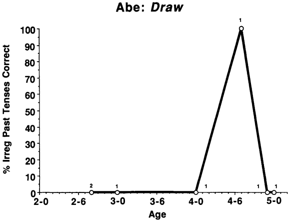


FIG. 14.—Example of a verb that is overregularized throughout development

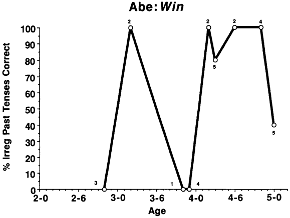


FIG. 15.—Example of a verb with a chaotic developmental pattern

‘U-shaped’ development (*cont.*)

- However, some important amendments to the classic characterization
  - For most children and most verbs, overregularization rate is quite low ($<$10%)
  - Overregularized productions generally coexist with correct irregular productions
  - Overregularization persists (at gradually lower rates) for a long time, even into adulthood
  - Marcus et al., using data from Stemberger, suggest a rate of $\approx$.00004 (p.45)
  - Rate of overregularization increases under time pressure and fatigue in a speeded naming task (Bybee and Slobin 1982)

Do kids forget something they once knew well?

Probably not...

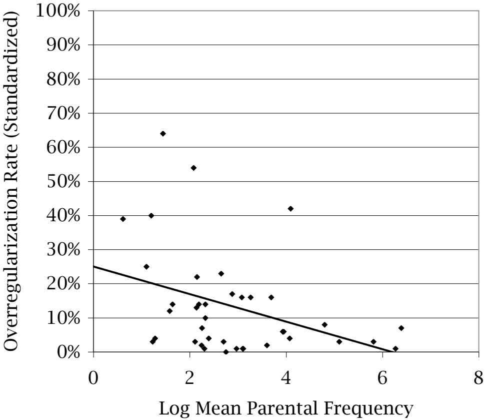
- Plots above obscure the fact that several other things are changing as well
- Change 1: kids know more words and are starting to speak more $\rightarrow$ using more low-freq words
  - This has led some (e.g., Rumelhart and McClelland 1987) to hypothesize that onset of overregularization is tied to a large increase in vocabulary size
  - Marcus et al. try to estimate vocabulary size in various ways (tally of all verbs used so far, and fancier estimation techniques)
  - Counts don't support idea of a rapid growth spurt, but do show that vocabulary prior to overregularization is smaller and includes higher proportion of irregulars

Do kids forget something they once knew well? (*cont.*)

- Change 2: past tense marking starts to be employed more consistently
  - Early productions frequently omit tense marking altogether
  - Onset of overregularization is at approximately same time that tense marking becomes much more common in the child's speech[1] (See Figs 33–38, pp. 108–110)
  - More consistent use of tense marking in obligatory past contexts creates more opportunities for erroneous marking
- Conclusion: more morphologically marked past tense productions $\rightarrow$ more errors

[1] Potential for circularity here: do children omit past tense marking because they don't have the syntax right to enforce it, or because they don't know how to form it morphologically? More careful verb-by-verb data is needed.

Overregularization and frequency



- Lower frequency verbs more subject to overregularization
- Lower frequency verb ⇒ lower frequency past tense ⇒ greater chance of not knowing

Language change: loss of exceptions

Present	Past (older)	Past (modern)
laugh	†lough[2]	laughed
starve	†storve[3]	starved
walk	†welk[4]	walked
wash	†wosh[5]	washed
climb	†clamb/clomb[6]	climbed
spill	spilt	spilled

[2] c1000 Ælfric, Gen. xviii. 15 *Þa ætsoc Sarra: Ne **hloh** ic na…God cwæð þa…ac þu **hloȝe**.* ‘Then Sara denied ‘I did not laugh’…and God said…‘but you did laugh’.’

[3] 1526 Pilgr. Perf. (W. de W. 1531) 140 *All bathed in rayne & frosen with yce, & nere **storuen** for colde.*

[4] a1425 St. Lucy l. 121 in C. Horstmann Altengl. Legenden (1881) 2nd Ser. 18/1 *Whils he **welk** in þis werld here, He said to his appostels…*

[5] c1300 Judas 125 in E.E.P. (1862) 110 *His fet heo **wosch** wiþ hire teres.* ‘his feet she washed with her tears’

[6] 1697 Dryden, Æneid VIII. 293 *He **clomb**, with eager haste, th’ Aerial height.*

Psycholinguistic and neuroimaging evidence

- Large body of experimental evidence documenting processing differences between regular and irregular past tenses
  - Lexical decision time
  - Priming
  - Localization
  - Selective impairment in special populations
- Most studies have assumed coarse ‘regular’ $=$ *-ed*, ‘irregular’ (exceptional) $=$ other
- More on fine-grained regularities in a moment

Summary for English past tenses

- Syndrome of properties associated with exceptionality
  - Not generalized to new items
  - Subject to regularization in acquisition
  - Subject to regularization over time
  - Protected by higher token frequency
  - Perhaps: processed differently, once learned
- In this case, converging evidence for which forms are exceptional, for the most part
- Fine-grained grammar does distinguish contexts and degrees of exceptionality, though…

Subregularities

- Limited investigation of regularities among irregulars (*flung*, *hid*), or *-ed* exceptions to these regularities (*blinked*, *chided*), but some indications of “exceptionality inversion”
- Pinker & Prince (1994) cite Ullman and Pinker (1991): *-ed* verbs that rhyme with irregular verbs (e.g., *blinked*) show frequency effects on naturalness ratings of pasts
  - I.e., *blinked* acts like an irregular
- Fruchter et al. (2013) MEG study: masked priming activation sensitive to reliability of pattern for “regulars” and “irregulars”
  - Higher reliability pattern = more ‘regular’
- Consistent with more nuanced grammar of past tense formation

References

Albright, Adam. 2002. Islands of reliability for regular morphology: Evidence from Italian. *Language* 78.684–709.

Albright, Adam, and Bruce Hayes. 2002. Modeling English past tense intuitions with minimal generalization. *Sigphon 6: Proceedings of the sixth meeting of the acl special interest group in computational phonology*, 58–69. ACL.

Albright, Adam, and Bruce Hayes. 2003. Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90.119–161.

Albright, Adam, and Bruce Hayes. 2006. Modeling productivity with the Gradual Learning Algorithm: The problem of accidentally exceptionless generalizations. *Gradience in grammar: Generative perspectives*, ed. by Gisbert Fanselow, Caroline Féry, Ralf Vogel, and Matthias Schlesewsky, 185–204. Oxford University Press.

References (*cont.*)

Buchwald, Adam. 2007. Determining well-formedness in phonology: Type vs. token frequency. Paper presented at the Annual Meeting of the Linguistic Society of America, Anaheim, CA, January 4-7.

Bybee, Joan. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes* 10.425–455.

Bybee, Joan, and Carol Moder. 1983. Morphological classes as natural categories. *Language* 59.251–270.

Fruchter, Joseph; Linnaea Stockall; and Alec Marantz. 2013. Meg masked priming evidence for form-based decomposition of irregular verbs. *Frontiers in Human Neuroscience* 7.1–16.

References (*cont.*)

Hay, Jennifer; Janet Pierrehumbert; and Mary Beckman. 2004. Speech perception, well-formedness and the statistics of the lexicon. *Phonetic interpretation: Papers in laboratory phonology vi*, ed. by J. Local, R. Ogden, and R. Temple, 58–74. Cambridge: Cambridge University Press.

Ling, Charles, and Marin Marinov. 1993. Answering the connectionist challenge: a symbolic model of learning the past tenses of English verbs. *Cognition* 49.235–290.

Marcus, G.; S. Pinker; M. Ullman; M. Hollander; T.J. Rosen; and F. Xu. 1992. Overregularization in language acquisition. *Monographs of the Society for Research in Child Development*.

References (*cont.*)

Pinker, Steven, and Alan Prince. 1988. On language and connectionism: Analysis of a Parallel Distributed Processing model of language acquisition. *Cognition* 28.73–193.

Pinker, Steven, and Alan Prince. 1994. Regular and irregular morphology and the psychological status of rules of grammar. *The reality of linguistic rules*, ed. by S. D. Lima, R. L. Corrigan, and G. K. Iverson, 321–351. Amsterdam: J. Benjamins.

Prasada, Sandeep, and Steven Pinker. 1993. Generalization of regular and irregular morphological patterns. *Language and Cognitive Processes* 8.1–56.