Bruce Hayes
UCLA

Talk at Johns Hopkins
Oct. 24, 2000

A Phonologist's View of the Past Tense Controversy[1]

1. Collaborators

Adam Albright (co-PI), Argelia Andrade, Stephen Wilson

2. Origins of the Controversy

- Rumelhart and McClelland (1986): a connectionist simulation that predicts English past tense forms from present tenses, by learning from a data set
  - ➢ The system works, to the extent that it does, without symbolic rules
  - ➢ Analysis is coupled to skepticism about the need for symbolic/rule-based approaches to linguistics (e.g., generative grammar) and cognition in general.
- Pinker and Prince (1988): 121 pages of mostly critical commentary
  - ➢ The Rumelhart and McClelland system fails, most conspicuously in areas where one would need rules to succeed—forms (like *ploamph*) where there is no good model in the existing lexicon, and only a **general rule** can derive the right output.
  - ➢ But Rumelhart and McClelland have located an important new domain for investigation: the apparently analogical, "family resemblance" relationships among irregulars (*keep*/*kept, leap*/*leapt; grow*/*grew, blow*/*blew*).

3. Opposing Views that Have Emerged

- **Dual Mechanism** model (Pinker and Prince 1988, 1991, and a vast subsequent literature):
  - ➢ Regulars are generated on-line by a single default rule
  - ➢ Irregulars are memorized, and can also be formed analogically by an associative network.
  - ➢ Later versions of the model alter these views somewhat; see (54) below.
- **Connectionist replies**: technically-improved connectionist systems work better than Rumelhart and McClelland's; therefore, no fundamentally reconceived system (involving rules) is needed to model people. Among many others:
  - ➢ MacWhinney and Leinbach (1991)
  - ➢ Plunkett and Marchman (1991)
  - ➢ Daugherty and Seidenberg (1994)
  - ➢ Westermann and Goebel (1995)

[1] Research supported by grant BCS 9910686 from the National Science Foundation, by various grants from the UCLA Academic Senate, and by a NSF Graduate Fellowship award to Adam Albright.

4. Outline

- Our own approach to past tenses and morphological learning
- Implications for the Dual Mechanism theory
- Implications for connectionism

THE PROJECT

5. Original Goal

- Develop a **machine implemented algorithm** for learning phonology and morphology.
- ... in order to model morphological and phonological acquisition.
- ... in order to test learnability consequences (both good and bad) of current phonological theories (Gildea and Jurafsky 1996).

6. Narrowing the Task to Something Feasible

- Algorithm is fed data pairs, taken from two columns of a morphological paradigm; e.g.

Category 1		**Category 2**	
[dʒʌmp]	~	[dʒʌmpt]	= *jump* ~ *jumped*
[rʌb]	~	[rʌbd]	= *rub* ~ *rubbed*
[mɪs]	~	[mɪst]	= *miss* ~ *missed*
[ækt]	~	[æktəd]	= *act* ~ *acted*

- It learns a grammar from these pairs, which projects novel forms of Category 2 when given a form from Category 1.
- Grammar uses surface form of Category 1, not an abstract underlying stem—see Burzio 1994, 1996, 1997; on phonological grammars without underlying forms.

7. Empirical Testing

- **Wug test** the grammar (i.e. find how it behaves with nonce forms; Berko 1958), and Wug test native speakers with the same forms. Example from a Wug test we are currently conducting:

Next week we're going to [splɪŋ].	(*on screen*; *also voice on headphones*)
We've been waiting ages for the chance to [splɪŋ].	(*on screen*; *also voice on headphones*)
We're reading a book about___.	(*consultant reads*)
My friend ___ once before, but this will be my first time.	(*consultant reads*)

- Disagreement in productions/intuitions of people vs. model implies that modification is needed; either
  - ➢ The learning model, or
  - ➢ The theory of grammar presupposed

- We reject the easy alternative: learn on half the corpus and test on the other half.
  - ➢ Forming a model of people (not the data corpus) is the goal.
  - ➢ For *think*, a good grammar would never guess *thought*!

8. Two Strategies for the Project

(a) Solve phonology problems from textbooks: intricate, but free of lexical exceptions
☞ (b) Take on larger data problems in which exceptional forms are present.

9. What Problems do Exceptions Pose for the Learning of Rules?

- We need to "factor out" exceptions in formulating rules.
- But we don't know which forms are exceptional until we know what the rules are.
- "Exceptions" are themselves not an arbitrary collection, but fall into patterns, which must also be found (Pinker and Prince 1988). Consider:

keep	[kip]	kept	[kɛpt]	reap	[rip]	reaped	[ript]
leap	[lip]	leapt	[lɛpt]	beep	[bip]	beeped	[bipt]
sleep	[slip]	slept	[slɛpt]	seep	[sip]	seeped	[sipt]
creep	[krip]	crept	[krɛpt]				
weep	[wip]	wept	[wɛpt]				
sweep	[swip]	swept	[swɛpt]				

→ There are competing generalizations. Learn each one, despite counterexamples to both.

HOW WE DO LEARNING

10. Inspiration

Pinker and Prince (1988, 131-136), a sketch of how a rule learner might work.

11. Entertain and Test Many Hypotheses

because you never know in advance what will pan out…

12. Starting Point: Each Learning Pair Yields a Very Small Rule

- Factor out changing part from invariant part.

Category 1	**Category 2**	
[flɪŋ]	[flʌŋ]	microrule: ɪ → ʌ / [fl ___ ŋ $]_{\text{word}}$

13. Generalization

- When microrules have the same structural change, they can be combined to make more general rules.
- Method: maintain all shared elements of context; reduce the remainder to variables.

14. The Minimal Generality Principle

Never generalize beyond the point needed to match input data.

15. Example

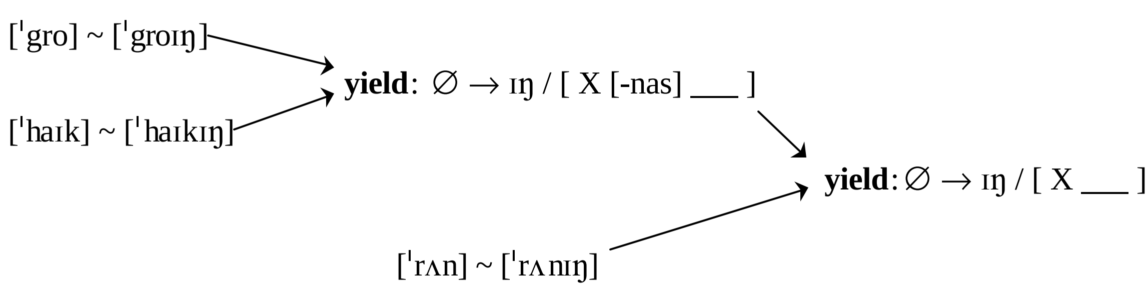
a. Forms:	[flɪŋ] ~ [flʌŋ]					
	[rɪŋ] ~ [rʌŋ] (‘*wrung*’)					
b. Microrules:	ɪ → ʌ / [	f	l	___	ŋ	]$_{word}$
	ɪ → ʌ / [		r	___	ŋ	]$_{word}$
c. Posit:	ɪ → ʌ / [	X	[liquid]	___	ŋ	]$_{word}$
not:	ɪ → ʌ / [	X		___	Y	]$_{word}$

16. Iterating Rule Creation

- Each new learning datum (pair) is compared with all existing pairs and rules, resulting ultimately in a large rule set.

17. Iteration Can Locate Highly General Hypotheses

They emerge by comparing heterogeneous data. For present participles:



18. Termination of Learning: Pinker and Prince (1988)

- Iterate the above process, until you find the rule that is most general.
- Definition of generality: number of forms the rules correctly derives. (Number of *incorrect* outcomes is ignored.)
- Every intermediate-stage rule is discarded.

19. Termination of Learning: Our Approach

- Use caution in discarding! Many of the intermediate rules could be quite useful.
- We can use them to good effect, if we find the right way to *evaluate* them.

20. An Evaluation Metric for Rules

- Rules are evaluated for how well you can **trust** them.
- Derive a statistic, the **raw reliability** of a rule:

$$\text{Raw Reliability} = \frac{\text{number of forms where rule applies successfully}}{\text{number of forms that meet rule's structural description}}$$

21. Weighing the Testimony

- 1000/1000 is more credible than 5/5.
- We use a statistical adjustment proposed by Mikheev (1995), which uses confidence limits to take this into account. We call the result **adjusted reliability**.
- Example: using a 75% lower confidence limit, 5/5 gets demoted from 1 to .825; 1000/1000 from 1 to .999

22. Taking a Wug Test with a Completed Grammar

- Input a wug form.
- Find all the rules that can apply to the wug form, apply each one.
- The learner's "well-formedness intuition" for each output is the adjusted reliability of the best rule that derives it.

FURTHER DETAILS OF THE LEARNER

23. Phonological Representations Assumed

- Currently pretty crude: *SPE*-style, with sequences of feature matrices.

24. Theory of Possible Structural Descriptions

- Also crude. Maximal template for a context is:

$$/ \; [\;\; \underset{\text{(variable)}}{X} \;\; \begin{bmatrix}\text{feature}\\\text{matrix}\end{bmatrix} \;\; \begin{matrix}\text{string}\\\text{of}\\\text{segments}\end{matrix} \;\; \_\_\_ \;\; \begin{matrix}\text{string}\\\text{of}\\\text{segments}\end{matrix} \;\; \begin{bmatrix}\text{feature}\\\text{matrix}\end{bmatrix} \;\; \underset{\text{(variable)}}{Y} \;\;]_{\text{word}}$$

- This limits us to certain problems, since {tiers, syllables, feet, phrasing} are absent.

25. Phonology: A Strategy

- Let the learner know in advance what sequences of sound are illegal. (see Jusczyk et al. 1993, 1994; Friederici and Wessels 1993 for evidence that this is a realistic assumption.)
- Try putting all the discovered affixes (= "changing parts") onto each new stem, see if you can get to the correct form by positing phonological rules.
- English: Given the knowledge that

  *[p+d] *[k+d] *[d+d]

  are ill-formed, the system will find (crude versions of) Progressive Voicing Assimilation and Schwa Epenthesis, yielding:

  [p**t**] [k**t**] [d**ə**d]

- This in turn increases the adjusted reliability of many morphological rules.

APPLICATIONS

26. English Past Tenses

- Feed learner: 2181 present-past pairs
- Learn grammar of several thousand rules
- Wug-test the grammar on the 60 forms from Prasada and Pinker's (1993) English Wug testing study
- Correlation of learner's "judgments" with subject judgments: $r = .75$, $p < .0001$

27. Sample Outputs of the Learner

		Learner's rating	Rule used	Production Probability*
[klid]	[klidəd]	.913	$\varnothing \rightarrow$ d / [X [+ant, +cor] ___]$_{word}$ + phonology	.83
	[klɛd]	.607	i $\rightarrow$ ɛ / [X [-syl, +appr] ___ d]$_{word}$	.42
	[klid]	.045	$\varnothing \rightarrow \varnothing$ / [X [+syl, +hi, -back] d ___]$_{word}$	
	[klit]	.011	d $\rightarrow$ t / [X [+son] ___]$_{word}$	
[splɪŋ]	[splɪŋd]	.913	$\varnothing \rightarrow$ d / [X [+son] ___]$_{word}$	.79
	[splʌŋ]	.786	ɪ $\rightarrow$ ʌ / [X [+vce, +cor, +ant] ___ [+vce, -cont]]$_{word}$	<.33**
	[splæŋ]	.189	ɪ $\rightarrow$ æ / [X [-syl, +vce] ___ [+nas]]$_{word}$	<.33**
	[splɔt]	.055	ɪŋ $\rightarrow$ ɔt / [X [-syl, +appr] ___]$_{word}$	

* = Prasada/Pinker human results; fraction of consultants who volunteered this form.
** V changes reported together, they sum to .33.

28. Italian Verb Conjugations (Albright 1998)

Theme vowel	Suffix	Sample Root	1sg. pres	Infinitive	
[a]	-ˈare	/rem-/	[ˈremo]	[reˈmare]	'to row'
[e]	-ere	/frem-/	[ˈfremo]	[ˈfremere]	'to shudder'
[e]	-ˈere	/tem-/	[ˈtemo]	[teˈmere]	'to fear'
[i]	-ˈire	/dorm-/	[ˈdormo]	[dorˈmire]	'to sleep'

- Task: Given a Wug root in the 1sg. pres., guess the infinitive (hence, conjugation class).
- Learning basis: 2900 Italian verbs, taken from existing corpora

29. A Wug Test

- Participants: (i) the learned grammar; (ii) 27 human native speakers of Italian
- 40 Wug verbs

"Oggi rabado con mio fratello." (*"I'm rabadating with my brother today."*)

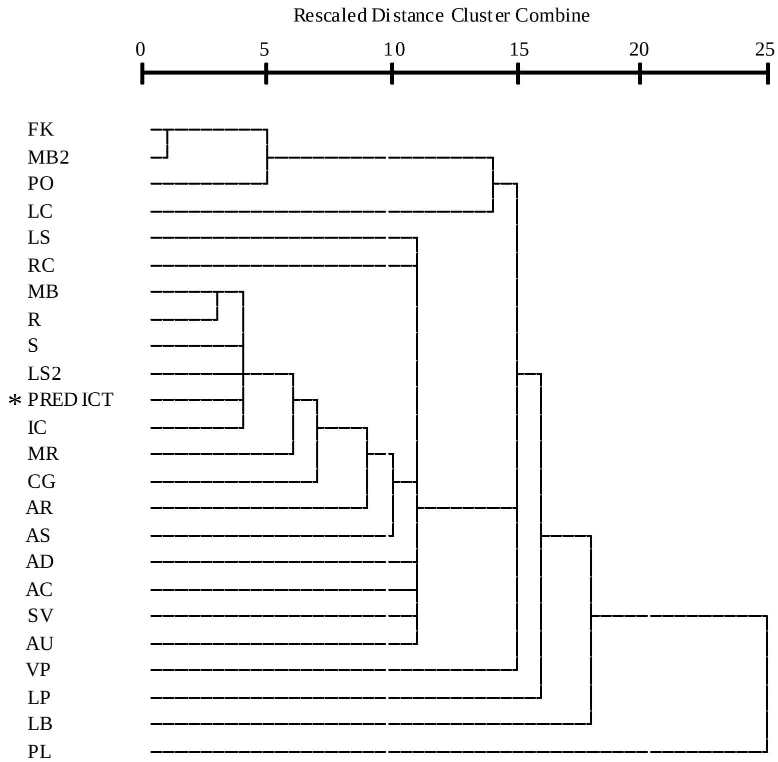
		worst						best
"Mi piace rabaˈdare"	rate:	1	2	3	4	5	6	7
"Mi piace rabaˈdere"	rate:	1	2	3	4	5	6	7
"Mi piace raˈbadere"	rate:	1	2	3	4	5	6	7
"Mi piace rabaˈdire"	rate:	1	2	3	4	5	6	7

(*"I like to rabadate."*)

30. Agreement of the Learner with the Consultants

r^2 = .792 (more details below)

31. The Learner as a Representative Italian Speaker: Cluster Analysis



32. Spanish Diphthongization

- A minority of /e, o/ become [je, we] when the stress falls on them.

Minority Pattern:			**Majority Pattern:**		
[s**e**nˈtamos] ~	[ˈs**je**nto]	'we/I sit'	[r**e**nˈtamos] ~	[ˈr**e**nto]	'we/I rent'
[p**o**ˈdemos] ~	[ˈp**we**do]	'we/I can'	[p**o**ˈdamos] ~	[ˈp**o**do]	'we/I prune'

- This is not predictable, and no published analysis attempts to predict it.
- Historically, there is no reason to expect it to be predictable. It results from a Proto-Romance phonemic distinction (diphthongizing ɛ/ɔ, non-diphthongizing e/o), lost long ago.
- Yet, the generalizations that are there by accident seems to be apprehended by speakers.

33. Wug Test for Spanish Diphthongization (Albright, Andrade, and Hayes, submitted)

Cada verano mi familia y yo _[leˈrramos]_ durante las vacaciones.
Every summer my family and I (*lerr*) while on vacation.

Hemos _________cada verano por diez años. *Check they heard it right:* [leˈrrado]
We have *lerred* every summer for ten years

Me fascina ___________. *Double check*: [leˈrrar]
I love to *lerr*

Tengo seis meses que yo no ___________. *Typical answers*: [ˈlerro, ˈljerro]
It's been six months since I not *lerr*

34. Outcome

- Correlation of:
  - ➢ Production probability observed in consultants' volunteered forms
  - ➢ Predicted production probability obtained from the learner's well-formedness scores[2]
- r(31) = .510, p = .0025

35. Rest of the Talk

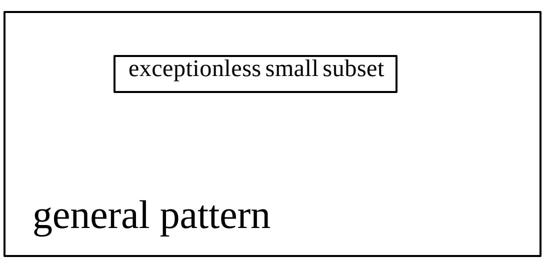
- General issues in our approach to learning: the shape of generalizations
- Remarks on the Dual Mechanism model and connectionism

THE SHAPE OF GENERALIZATIONS

36. Theme

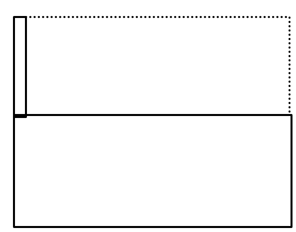
- An adequate learner should be prepared for a variety of distributional patterns in the data.
- Certain shapes can stymie inadequate approaches to learning.

37. Small, Perfect Subsets



- Example: Spanish diphthongizes in all 11 verb stems with /Xerr-/ (*aferrar/afierro* 'grasp', similarly *aserrar* 'saw', *aterrar* 'terrify', *cerrar* 'close', *desenterrar* 'disinter', *desterrar* 'banish', *encerrar* 'lock', *enterrar* 'bury', *errar* 'miss', *serrar* 'saw', and *soterrar* 'bury')
- Yet Spanish speakers are reluctant to diphthongize in /XerrY/ Wug forms—only 20% of consultants for [ˈljerro], 21% for [ˈdjerro].
- The size of the general pattern—works 918/1029 cases—overrides an exceptionless generalization. (It is also likely that a preference for non-alternation plays a role.)

[2] = (well-formedness score for diphthongized outcome)/(well-formedness score for diphthongized outcome + well-formedness score for unchanged outcome)



40. Remedy

- See Albright and Hayes (1999) for a modification of our algorithm that keeps it from falling into this trap:
    - Every rule gets a lowered adjusted reliability score if it fails to outperform all available subset generalizations.

- Wrong-overgeneralized rules thereby receive very low adjusted reliability scores, resulting in very low well-formedness ratings for forms like [lænt].

41. The General Case

- Occurs whenever an allomorph that is regular in one context occurs as an exceptional form in another context.
- Examples: Dutch plurals; French h aspiré phenomena, exceptional vowel harmony in Hungarian and Turkish

42. Implication for Pinker and Prince's (1988) Algorithm

- "Raw hit count" cannot be the correct criterion for evaluating rules.
- Such a procedure will always favor the overgeneralized rule in these cases.

BACK TO THE PAST TENSE CONTROVERSY

43. Review: Dual Mechanism Model

- There is single, typically very simple, default rule; it is used to derive regulars.
- There are no other rules, all other forms are derived by an associative network.
- Strongest version of theory:
  - Only irregulars are lexically listed.
  - The associative network only derives irregulars.

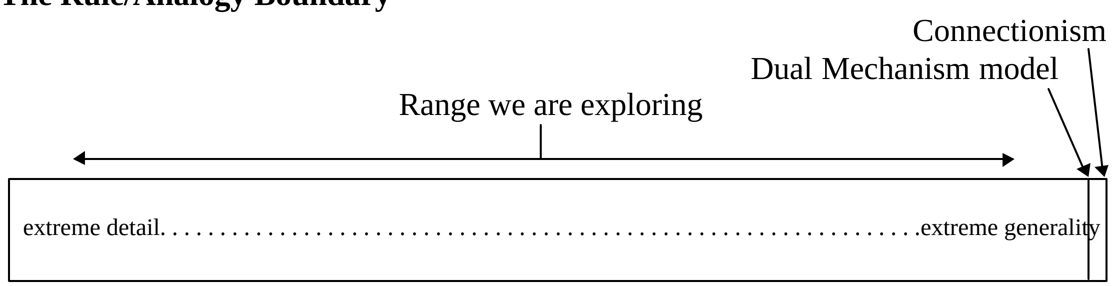
44. Review: Classical Connectionism

- Goes further: *everything* is derived in an associative network

45. Our Position

- If the problem of learning rules amidst exceptions is taken seriously, and Pinker and Prince's (1988) proposal is suitably altered, then we are led to detailed grammars, at least as intermediate phases in learning.
- Such grammars can derive, with accurate well-formedness ratings, many forms attributed in the other two theories to analogy.
- Compare Pinker and Prince's proposal: all rules of non-maximal generality should be **thrown away**; let the work they could do be done instead by an associative network.

46. The Rule/Analogy Boundary



- The rule/analogy boundary may lie further into the territory of contextual detail than other theories imagine.

DISCUSSION OF DUAL MECHANISM THEORY

47. Rules Can Be Fairly Detailed

- On occasion, dual mechanists (e.g. Clahsen 1999, 991) have come close to claiming that only simple affixation can be a rule. But the child must be prepared for more (Dressler 1999):
- Multiple Defaults, with Contexts
  - Dutch plurals (Pinker 1999, 231): [-s] in post-atonic syllables, [-ən] elsewhere
- Non-Concatenative rules
  - Samoan infixing reduplication: copy tonic CV and prepose ([aˈlofa] ~ [a-**lo**-ˈlofa] 'love-plur.', Broselow and McCarthy (1983)
  - Subtractive morphology: $VC_0 \rightarrow \varnothing$ / ___ $]_{stem}$ in Koasati plurals. [latáf] → [lát] 'kick something' (Martin 1988)
- Upshot: it probably won't work to make it easy to find defaults, simply by decreeing that defaults are formally very simple.

48. Could More Than One Rule Derive Regular Forms?

- In the purest forms of dual mechanism theory, the default rule **exhausts** the mechanisms for deriving regulars.
- Our algorithm inexorably arrives at many rules to derive regulars, each with a different context.

49. Defn. "Island of Reliability"

= a phonological structural description that isolates a set of forms in which a structural change applies with greater reliability

- All are agreed that speakers judge novel *irregulars* as reasonably good only when they occupy an island of reliability for their pattern (*spling*/*splung*).

50. Question: Are There Islands of Reliability for the Default?

Some possible candidates:

- [X [V, -long] p] for English past tenses (19/19: *stop, drop, pop, step, slip, sip, grip, top, chop, worship, tip, strip, hop, shop, mop, skip, swap, whip, chip*). Adjusted reliability = 0.967; context free default = .910.

- [Xmin-] for Italian verb stems (22/22 are first conjugation: *denominare, determinare, discriminare, disseminare, dominare, eliminare, esaminare, illuminare, incriminare, camminare, comminare, culminare, minare, nominare, predeterminare, recriminare, riesaminare, ruminare, seminare, sterminare, terminare, zgominare*). Adjusted reliability = 0.957, context free default =.762.

- [X ___ k Y] for Spanish (non-)diphthongization of /e/ (27/27: *afectar, anexar, becar, checar, colectar, conectar, defecar, desconectar, desecar, desflecar, desinfectar, desinsectar, detectar, disecar, enchalecar, hipotecar, indexar, infectar, interconectar, pecar, pretextar, recolectar, rectar, reflectar, resecar, respectar, secar*). Adjusted reliability = 0.966, context free default = .886.

51. Evidence from Albright's Italian Wug Test

Wug forms from islands of reliability			**Wug Forms Not from Islands**		
	ave. subject rating (*0-100*)	*score from Learner*		*ave. subject rating* (*0-100*)	*score from Learner*
guar**min**are	76	.96	addungare	54	.76
vegheggiare	83	.99	funadare	53	.76
fraquinare	72	.92	scincadare	46	.76

- Correlation of Learner's opinions with consultants' (*-are* forms only): r(42) = .48, p <.001
- A pure dual mechanism approach predicts identical ratings for all of these forms, since they are all derived identically.

52. Other Cases Suggesting Islands of Reliability for Default

- English: Albright's (2000) reanalysis of the Prasada and Pinker (1993) data
- Spanish (Albright, Andrade, and Hayes 2000).

53. What of the Psycholinguistic Findings?

- A large and growing dual mechanist literature shows differences between regulars and irregulars: in processing, acquisition, language disorders, brain imaging.
- But what is the causative mechanism?
- Pinker (1999, 137) may have found it: "Suppose the gatekeeper to memory follows the principle, "'Store anything that is unpredictable.'"
- Then storage is more likely for irregulars. Even when regulars are stored, they are likely to have more weakly-ensconced representations.
- But (Zuraw 2000): if this is the right story, it may suffice *alone* to explain the psycholinguistic results! Therefore, the data could well be neutral on how novel forms are generated.

54. A Possible Take on the Dual Mechanism Approach

- ~~Only those forms not derivable by the default receive lexical listing.~~ [Revised view of Pinker 1999]
- ~~There is only one default rule per morphological process; all else is analogy.~~ [removal advocated here]
- ~~The default rule serves as the only mechanism for deriving regulars~~. [removal advocated here]
- There is both a lexicon and mechanisms for generating novel forms.

- N.B. The last point is hardly trivial, is anathema to some connectionists, and continues to drive productive research…

COMMENT ON CONNECTIONIST APPROACHES

55. Performance Improves with Better Linguistics

- MacWhinney and Leinbach (1991)
  - Sensible representations, with segments, features, CV tier (cf. the notorious "Wickelphones" of Rumelhart and McClelland 1986)
  - A priori connections to enforce stem copying (i.e., "disprefer allomorphy," a widely-accepted linguistic principle)
- Westermann and Goebel (1995)
  - Adoption of a lexicon (a network segment for storing memorized forms) also improves performance

A COMMENT APPLICABLE TO BOTH APPROACHES

56. What Is Meant by a Rule/Analogy Distinction?

To generate a new form from a novel base:

- **Rules**: First, learn a grammar. I.e. look at a bunch of forms, find what they have in common, abstract from them a means of creating new forms. Then look up the relevant rules of the grammar and apply them where applicable.
- **Analogy**: find an analogue form that is *closely similar* to the novel base. Modify the base in the same way that the analogue form is modified.

57. A Claim to be Checked

- Analogy allows **single-base** operations (*bring*:*brought* = *spling*: *?*); rules do not.
- More specifically:
  - Intuitively, rules involve generalization over forms.
  - Any reasonable settings of the main parameter of our algorithm (lower confidence limit) give very low scores to any rating made based on a single example.

58. Wug Test in Progress: Choice of Wug Verbs

- Verbs that (might) rate highly as irregulars *by virtue of a semi-general pattern*
- Verbs that (might) rate highly *by virtue of close resemblance to an existing common verb*

59. Volunteered Forms from 24 Consultants: Preliminaries

- 6 were linguists, asked to volunteer **irregulars only**—a hard task.
- 18 were undergraduates, allowed to volunteer **whatever they wanted**, including regulars.
- Undergraduates volunteered few irregulars in general—9.1% of all forms.

- Consider forms volunteered by at least 3 consultants (either type):[3]

60. Forms Where Learning is Supported by a *Set* of Similar Real Verbs

Stem	Past	Set of Forms from which Generalization is learned	Learning Algorithm's "Judgment"	Total volun-teered	From "Forced Irregular" Subjects	From "Free choice" Subjects
dize	doze	rise, ride, dive	0.351	6	4	2
fleep	flept	sleep, sweep, creep, leap, weep	0.561	6	6	0
gleed	gled	read, lead, breed, bleed	0.607	5	2	3
queed	qued	read, lead, breed, bleed	0.607	5	2	3
spling	splung	dig, fling, cling, sling	0.786	3	0	3
spling	splang	begin, swim, spring, ring	0.189	5	4	1
bize	boze	rise, ride, dive	0.351	4	3	1
fro	frew	grow, throw, blow	0.396	4	3	1
murn	murnt	learn, burn	0.310	4	3	1
skride	scrid	hide, slide, light	0.214	4	3	1
skride	scrode	write, ride, shine, stride	0.607	4	2	2
teep	tept	keep, sleep, sweep, creep, leap, weep	0.537	4	4	0
drit	drat	sit, sing, swim, spring, ring	0.146	3	3	0

61. Forms Where Learning is Supported by One Frequent Analogical Model

Stem	Past	Model	Model Freq. (lemma < CELEX)	Learner's Rating (see fn.s)	Total volun-teered	From "Forced Irregular" Subjects	From "Free choice" Subjects
kive ([kɪv])	kave	give	22921	0.148[4]	1	1	0
lum	lame	come	35152	0.215[5]	0	0	0
pum	pame	come	35152	0.215	0	0	0
shee	shaw	see	36958	0.57[6]	1	1	0
zay	zed	say	76541	0	2	2	0
nold	neld	hold	8324	0.146	4	4	0

- N.B. *All* the cases that occurred, occurred in the forced-irregular condition.

[3] Some data filtration: I ignore "output oriented forms" (Bybee and Slobin 1983) like [tʃaɪnd/tʃaɪnd], (= 2.8% of volunteered forms), and three forms (*dape*/*dapt*, *chool*/*chold*, *trisk/trask*) volunteered by 3 forced-choice consultants each, for which we can find no analogical basis.

[4] Possibly this should be zero: learning data are *give* and *forgive*. If these share a stem, and only stems count as learning tokens, then this is a one-form analogue.

[5] Ditto; learning data are *come*, *become*, *overcome*; and likewise for *pum*.

[6] Ditto; learning data are *see*, *foresee*.

62. Conjecture

- Analogy (insofar as it is diagnosed by single-form cases) is perhaps not all that it's cracked up to be.
- People will do it when forced to arrive at a past tense.
- But they are perhaps less likely to do it in a maximally naturalistic condition of free-choice sentence completion.
- Minor rules are more capable of leading to the creation of novel forms.

63. Where Things Stand

- More data needed…
- But if this is right, then we may have a handle on placing the rule/analogy boundary—closer to the territory that many linguists will feel comfortable with.

64. What's Needed

- Comparative modeling of the same Wug-test results, with different theories. We are posting all our data, software, and learning sets, and encourage others to do so.

References

Adam Albright, Argelia Andrade, and Bruce Hayes (submitted) "Segmental Environments of Spanish [http://www.humnet.ucla.edu/humnet/linguistics/people/hayes/segenvspandiph/]

Albright, Adam (1998) "Phonology subregularities in productive and nonproductive inflectional classes: evidence from Italian," M.A. thesis, Dept. of Linguistics, UCLA. [http://www.humnet.ucla.edu/humnet/linguistics/people/grads/aalbrigh/papers.html]

Albright, Adam (2000) "The Lexical Bases of Morphological Well-Formedness," poster presented at the 9th International Morphology Meeting, Vienna. [http://www.humnet.ucla.edu/humnet/linguistics/people/grads/aalbrigh/papers.html]

Albright, Adam and Bruce Hayes (1999) "Distributional encroachment and its consequences for morphological learning", ms., UCLA. [http://www.linguistics.ucla.edu/people/hayes/learning/].

Berko, Jean (1958) "The child's acquisition of English morphology," *Word* 14, 150-177

Boersma, Paul and Bruce Hayes (in press) "Empirical Tests of the Gradual Learning Algorithm," to appear in *Linguistic Inquiry*, Winter 2001. [http://www.linguistics.ucla.edu/people/hayes/GLA/]

Broselow, Ellen and John McCarthy (1983) "A theory of internal reduplication," *The Linguistic Review* 3, 25-88.

Burzio, Luigi (1994) *Principles of English Stress*, Cambridge University Press, Cambridge.

Burzio, Luigi (1996) "Surface Constraints versus Underlying Representation,"in J. Durand, & B. Laks (eds.) *Current Trends in Phonology: Models and Methods*. CNRS, Paris X, and University of Salford. University of Salford Publications.

Burzio, Luigi (1997) "Cycles, Non-Derived-Environment Blocking and Correspondence," ms. Johns Hopkins University, Baltimore, MD.

Bybee, Joan L. and Daniel I. Slobin (1982) "Rules and Schemas in the Development and Use of the English Past *Language* 58, 265-289.

Clahsen, Harald (1999) "Lexical entries and rules of language: a multi-disciplinary study of German inflection," *Behavioral and Brain Sciences* 22 (6): 991-1013.

Chomsky, Noam and Halle, Morris (1968) *The Sound Pattern of English*, Harper and Row, New York.

Clahsen, Harald (1999) "Lexical entries and rules of language: a multidisciplinary study of German inflection," *Behavioral and Brain Sciences* 22, 991-1060.

Cognition 40:121-157.
Martin, Jack (1988) “Subtractive morphology as disassociation,” *Proceedings of WCCFL* 7. Stanford Linguistics Assn.
Mikheev, Andrei (1995) “Automatic rule induction for unknown-word guessing,” *Computational Linguistics* 23:405-23.
Pinker, Steven (1999) *Words and Rules.* New York: Basic Books.
Pinker, Steven and Alan S. Prince (1988) “On language and connectionism: analysis of a parallel distributed processing *Cognition* 28, 73-193.
Pinker, Steven and Alan S. Prince (1991) “Regular and irregular morphology and the psychological status of rules of grammar,” in Susan D. Lima, Roberta L. Corrigan, Gregory K. Iverson, eds., *The Reality of Linguistic Rules*, John Benjamins, Amsterdam.
Plunkett, K. and V. Marchman (1991) “U-shaped learning and frequency effects in a multilayered perceptron: *Cognition* 38, 1-60.
Prasada, Sandeep and Steven Pinker (1993) “Generalization of regular and irregular morphological patterns,” *Language and Cognitive Processes* 8, 1-56.
Prince, Alan and Paul Smolensky (1993) *Optimality Theory: Constraint Interaction in Generative Grammar*, ms., Rutgers University and University of Colorado. Available from http://ruccs.rutgers.edu/publicationsreports.html.
Rumelhart, D., & McClelland, J. (1986) “On learning the past tenses of English verbs,”. In Rumelhart, D., McClelland, J., & the PDP Research Group (Eds.), *Parallel Distributed Processing Vol 2*, pp. 216 - 271. Cambridge: MIT Press.
SPE = Chomsky and Halle (1968)
Tesar, Bruce and Alan S. Prince (1999) “Learning phonotactic distributions,” ROA 353 [http://ruccs.rutgers.edu/roa.html.]
Tesar, Bruce and Paul Smolensky (1993) “The learning of Optimality Theory: An algorithm and some basic complexity results,” ROA-52 [http://ruccs.rutgers.edu/roa.html].
Tesar, Bruce and Paul Smolensky (1998) “Learnability in Optimality Theory,” *Linguistic Inquiry* 29:229-268.
Tesar, Bruce and Paul Smolensky (2000) *Learnability in Optimality Theory*. Cambridge: MIT Press.
Vago, Robert (1976) “Theoretical implications of Hungarian vowel harmony,” *Linguistic Inquiry* 7, 243-263.
Westermann, Gert & Goebel, Rainer (1995), Connectionist rules of language, in `Proceedings of the 17th Annual Conference of the Cognitive Science Society', Erlbaum, pp. 236--241.
Zuraw, Kie (2000) *Patterned Exceptions in Phonology*, UCLA dissertation

Department of Linguistics
UCLA
Los Angeles, CA 90095-1543

bhayes@humnet.ucla.edu
http://www.linguistics.ucla.edu/people/hayes/