Bruce Hayes UCLA
NELS Ottawa October 27, 2007

New methods for studying UG in phonology

1. Prospect

- The search for principles of Universal Grammar and for evidence to support them has broadened to include experimental work.
- I argue for an approach that combines experimentation with the use of implemented learning models.

2. Traditional methods of studying UG

- Typological study—study multiple languages, and develop a theory that
  - encompasses what's "out there"
  - makes non-trivial predictions about what is not.
- Examples:
  - Hayes's (1995) parametric theory of stress systems, with an asymmetrical foot inventory
  - OT studies using **factorial typology** (Prince and Smolensky 1993, ch. 3): Kaun (1995), Alderete (1998), Kager and Elenbaas (1999), Gordon (2002)

3. Mixed-origin theories of typology

- What we see in the world's languages is not direct reflection of UG, but the result of history—albeit a version of history in which UG may play a significant role (Myers 2002, Wilson 2006, Koo and Cole 2006, Moreton, to appear a).
  - **Channel bias** (term from Moreton, to appear a)
    Phonological processes are more likely to be found if they can arise from reinterpretation by language learners of patterns present in natural phonetic processes like coarticulation (see Ohala 1981, Blevins 2004)
  - **Analytic bias**:
    Children construct a grammar from the ambient data, guided by UG. UG biases (either absolutely, or gradiently; Wilson 2006) what they learn.
- Thus the examination of typology is useful, but since it only indirectly reflects UG, other sources of evidence are also valuable.

4. Studying UG with experiments

- Two main strategies:
  - Have experimental subjects learn novel "mini-languages."
  - Using the subjects' native language, but examine novel forms not present in the childhood learning data (McCarthy 1981).

5. Experiment probes of phonological knowledge

- The **wug test** (Berko 1958): subjects inflect novel stems, so they must apply their rules productively ("What is the plural of [wʌg]?"). Covers **alternations**.
- The **blick test** (Chomsky and Halle 1965): ask subjects to rate novel forms ("Could [blɪk] be a word? Could [bnɪk]?"). Covers **phonotactics**.

6. History of UG experimentation in phonology

- UG experiments went through a period of interest in the 1970's (Schane et al. 1974, Pertz and Bever 1975), and have recently been vigorously revived, e.g. by

Study	*Empirical focus*	*Method*
Albright (2007)	phonotactics	project beyond native lg.
Albright and Hayes (2003)	alternations	project beyond native lg.
Becker, Ketrez and Nevins (2007)	alternations	project beyond native lg.
Berent et al. (2007)	phonotactics	project beyond native lg.
Buckley and Seidl (2005)	alternations	artificial language
Graff (2007)	alternations	artificial language
Kawahara (2006)	phonotactics	project beyond native lg.
Koo and Cole (2006)	phonotactics	artificial language
Pater and Tessier (2003)	alternations	artificial language
Pycha et al. (2003)	alternations	artificial language
Thatte (in progress)	phonotactics	artificial language
Wilson (2003, 2006)	alternations	artificial language
Zhang and Lai (2006)	alternations	project beyond native lg.
Zuraw (2007)	alternations	project beyond native lg.

7. Two kind of effects that experimenters on UG are assessing

- **Poverty of the stimulus** effects
- **Overabundance of the stimulus** effects

8. Poverty of the stimulus, example 1: Moreton (to appear a)

- Experiments compare the **ease of learning** for three artificial languages, each embodying a particular phonotactic principle:

Language	*Pattern*	*typology*	*Ease of learning in experiment*
"HH"	Vowels in consecutive syllables must agree in height	ubiquitous	easy
"VV"	Consonants separated by V must agree in voicing	rare	easy
"HV"	Vowels tend to be higher before voiceless consonants	rare	hard

- Moreton interprets the results in terms of his taxonomy (above) of transmission biases:
  - **Channel bias**: VV is rare because unlike HH and HV, it has no natural precursor in phonetics.
  - **Analytic bias**: there is some UG principle favoring the learning of HH and VV. Some possibilities:
    - "in assimilations, like assimilates to like"
    - "don't change feature values, just alter association lines"
    - Moreton (to appear b) has a more abstract approach
- Poverty of the stimulus: nothing in the stimulus (or, nothing obvious in the stimulus) tells subjects *a priori* to attend to one factor rather than another.

9. Poverty of the stimulus, example 2: Wilson (2006)

- Comparison of two artificial languages ("language games").

  I. Test if subjects generalize observed **k → tʃ / __ e** to cover cases of **k → tʃ / __ i**

  II. Test if subjects generalize observed **k → tʃ / __ i** to cover cases of **k → tʃ / __ e**

- I, not II, matches typology (Bhat 1978 et seq.).
- Subjects generalize on the lines of I, but not II.
- UG principle: "minimize the phonetic salience of alternation" (Coté 2000, 2004; Steriade 2001a,b).

10. A UG experiment based on "overabundance of the stimulus"[1]

- Becker, Ketrez, and Nevins (2007) on Turkish stem-final voicing alternations
- This is a study of "undoing final devoicing," modeled on Ernestus and Baayen's (2003) study of Dutch.
  - By using various phonological factors, you can statistically predict with greater-than-chance accuracy the presuffixal voicing of an isolation voiceless final obstruent ([ro**p**] ~ [ro**b**-u]), using phonological factors.
- Studying the TELL database (Inkelas et al., http://socrates.berkeley.edu:7037/), the authors found several factors that—in the lexicon—statistically predict the voicing alternation:
  - place of articulation of the stem-final consonant.
  - stem length
  - height and backness of the rightmost stem vowel
- But only the first two factors play a role in native speaker judgments gathered experimentally—subjects ignored vowel height and backness, even though it would have increased the accuracy of their responses.

---

[1] This term, or something like it, heard from Andrew Nevins.

- Conjectured reason: the data patterns are there, but not noticed by speakers, since UG doesn't equip them to notice. This is the "overabundance of the stimulus" effect.
- What UG principle led the subjects to ignore height and backness? We don't know, but the data are reminiscent of Moreton's "HV" case, just mentioned

THE ROLE OF MODELING IN THE STUDY OF UG

11. Some pioneers in phonological learning models

- Dresher and Kaye (1990): cue-based parameter-setting model for metrical stress theory
- Gildea and Jurafsky (1995): learning rules with a UG-bolstered inductive system
- Constraint ranking models for OT: Tesar and Smolensky (1993 et seq.), Boersma (1997)

12. The modeling process in this context

- For present purposes it is useful to have models that
    - are flexible and can be applied to all kinds of phenomena
    - handle realistic learning data: real forms (not UR-SR pairs), in numbers comparable to what the child would see.
    - include little UG at the start—if by adding in a UG principle, we improve performance, that can be informative (Gildea and Jurafsky 1995, Hayes and Wilson, in press)

13. Tying modeling to experimentation

- Give the model learning data similar to what children get.
    - It learns a grammar.
- Evaluate the grammar by giving it the same tests (wug, blick) given to people.

14. Two projects along these lines

- **Modeling alternations**
    - My collaboration with **Adam Albright** (Albright and Hayes 2002, 2003, 2006)
    - We input a list of morphological pairs: {*InflectedForm1*, *InflectedForm2*}.
    - (This presupposes Albright's theory of phonology, in which one paradigmatic form serves as the base for the others; Albright 2002 et seq.)
    - System learns morphological and phonological rules that can derive the *InflectedForm1* → *InflectedForm2* mapping.
    - The rules are stochastic and can assign a gradient predicted well-formedness value to any form they derive.
    - Model is tested with **wug test** data. Example from Albright and Hayes (2003):

Base form	*Possible past tense output*	*Rating by human subjects (1-7 scale)*	*Model prediction*
[raɪf]	[raɪft]	5.95	6.22
	[roʊf]	4.14	4.61

- You can try it; software at http://www.linguistics.ucla.edu/people/hayes/learning/

- **Modeling phonotactics**
  - My collaboration with **Colin Wilson**: Hayes and Wilson (in press)
  - Input is simply a list of forms, with a feature set for the segment inventory
  - The system induces from the data a set of phonotactic constraints.
  - Each constraint has a weight.
  - For any novel form, the model can compute a penalty score based on its constraint violations and the weights.
  - Test the model with **blick test** data, here from Scholes (1966).

Blick form	*Rating by human subjects (fraction who said "sounds ok")*	*Penalty score for learning model*
[krʌn]	1	0
[ftɪn]	.30	10.6
[ʒkip]	.03	28.0

- You can try it: http://www.linguistics.ucla.edu/people/hayes/Phonotactics/

BENEFITS OF MODELING I: INTERPRETING EXPERIMENTS MORE SECURELY

15. The Sonority Hierarchy Projection Effect

- Suppose your language has only onsets that obey sonority sequencing, like [bl], [bw], [fr], etc.
- You've never heard either [bd] (sonority violation) or [md] (extreme sonority violation), yet somehow **[md] seems worse than *[bd].
- This is experimentally documented: Pertz and Bever (1975), Berent, Steriade, Lennertz, and Vaknin (2007), Albright (2007).
- Crudest interpretation: the Sonority Hierarchy is innate.

16. Claim made here

- Modeling show that the interpretation of innate-sonority-hierarchy experiments can be very delicate, with multiple interpretations coming into play.
- Reason: *surprisingly little UG* is needed to project beyond the learning data.

17. Schematic simulation: Bwa

- The hypothetical language **Bwa** allows only the following words:

[pa]	[ta]	[ka]	[pwa]	[twa]	[kwa]	[pja]	[tja]	[kja]
[ba]	[da]	[ga]	[bwa]	[dwa]	[gwa]	[bja]	[dja]	[gja]
[fa]	[sa]							
[va]	[za]							
[ma]	[na]							
	[la]							
	[ra]							
[wa]	[ja]							

- Only stops ([ptk bdg]) may begin a cluster, only glides ([jw]) may end one.
- Question: what will speakers of Bwa think about novel onsets like [pl], or [pt], or [lp]?
- To make predictions, we will try various UG models, using the phonotactic learning model of Hayes and Wilson (in press).

18. Learning Bwa phonotactics with UG #1: tight-fisted inductivism

- This is the Hayes/Wilson model in its rawest form:
  - Find the simplest set of constraints that maximizes the probability of the learning data.
  - Constraint = *[ ]([ ])([ ]), where [ ] is a distinctive feature matrix.
  - Other than the learning model itself, UG consists merely of a feature system.[2]
  - Crucial features: the Sonority Hierarchy is formalized by aligning a feature with each cut-off point (Clements 1990), as follows:

	[consonantal]	[approximant]	[sonorant]	[continuant]
stops	+	–	–	–
fricatives	+	–	–	+
nasals	+	–	+	+
liquids	+	+	+	+
glides	–	+	+	+

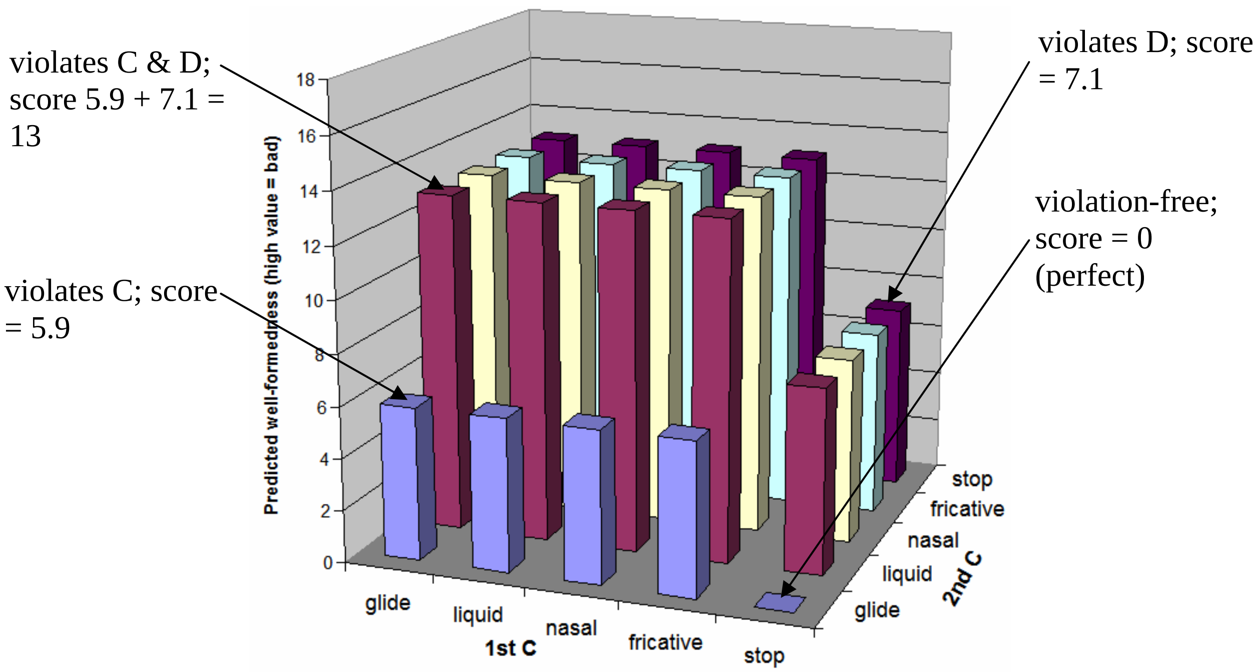
19. Constraints learned under UG #1, with commentary

- The tight-fisted model learns a very economical and accurate grammar:

	Constraint	*Weight*	*Interpretation*
A.	*C #	5.5	No final consonants
B.	*nonfinal V	5.8	All vowels final
C.	*[+continuant] C	5.9	Only stops can precede consonants
D.	*noninitial [+consonantal]	7.1	Any non-initial consonant must be a glide

[2] We don't necessarily want to attribute the feature system to UG (Mielke 2005), but this is our starting point.

- Unsurprisingly, this grammar says little beyond the data that were fed to it; all CCV word fall into one of four possible types, based on whether the first is a stop or the last is a glide:



20. UG #2: The Sonority Hierarchy is innate

- We give the system, in advance, a set of eight innate constraints.
- In the list below, the boldface ones were those learned inductively in the UG #1 simulation.

Don't put relatively nonsonorous consonants in second position:	*Don't put relatively sonorous consonants before consonants:*
***noninitial [+consonantal]**	*[–consonantal] C
*noninitial [–approximant]	*[+approximant] C
*noninitial [–sonorant]	*[+sonorant] C
*noninitial [–continuant]	***[+continuant] C**

- The model is required to include these eight in the grammar, though it can add other, inductively learned ones.
- The model can "nullify" a UG constraint by assigning it a weight of zero.

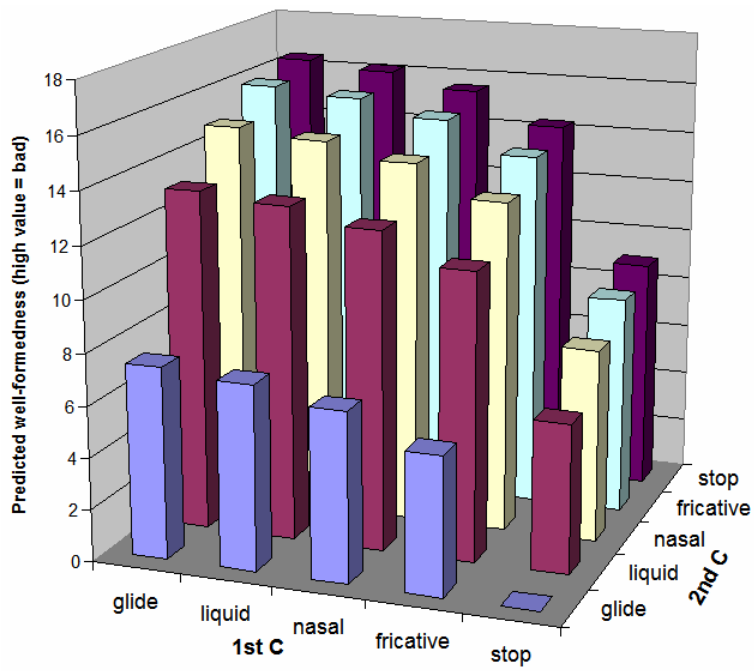
21. Results of the UG #2 simulation (innate-Sonority-Hierarchy) for Bwa

- The learning system as before, adds in the constraints that require vowels to be final and consonants to be non-final.
- Weight assignment to constraints: we get **leakage**: some of the explanatory force attributed to the constraints that were learned inductively in the "weak" UG are now

shifted to constraints of similar effect (see Martin 2007 for explanation of why this happens):

	Now	*Formerly*
*noninitial [+consonantal]	5.767	7.146
*noninitial [–approximant]	**1.728**	
*noninitial [–sonorant]	**1.010**	
*noninitial [–continuant]	**0.486**	
*[+continuant] C	5.341	5.904
*[+sonorant] C	**1.214**	
*[+approximant] C	**0.601**	
*[–consonantal] C	**0.265**	

- Result for novel well-formedness judgments: the entire sonority hierarchy comes into play, with smooth differentiation along the continuum:



22. UG #3: The Sonority Hierarchy is not innate, but special attention to sonority is

- We neutralize the sonority bias by adding to the previous UG a set of constraints with **opposite feature values** for the target feature.
- Now it has 16 UG constraints. The “nonsensical” ones added here are shown in bold.

*noninitial [+consonantal]	*[–consonantal] C
***noninitial [–consonantal]**	***[+consonantal] C**
*noninitial [–approximant]	*[+approximant] C
***noninitial [+approximant]**	***[–approximant] C**
*noninitial [–sonorant]	*[+sonorant] C
***noninitial [+sonorant]**	***[–sonorant] C**
*noninitial [–continuant]	*[+continuant] C
***noninitial [+continuant]**	***[–continuant] C**

- But this yields predictions **identical to UG #2**! Why?
  - ➢ The nonsensical constraints just added have no explanatory value, and are nullified (assigned weights of zero) by the learning algorithm.
- Upshot: a UG that merely says "care about sonority" gives same results for Bwa as a traditional-style UG.

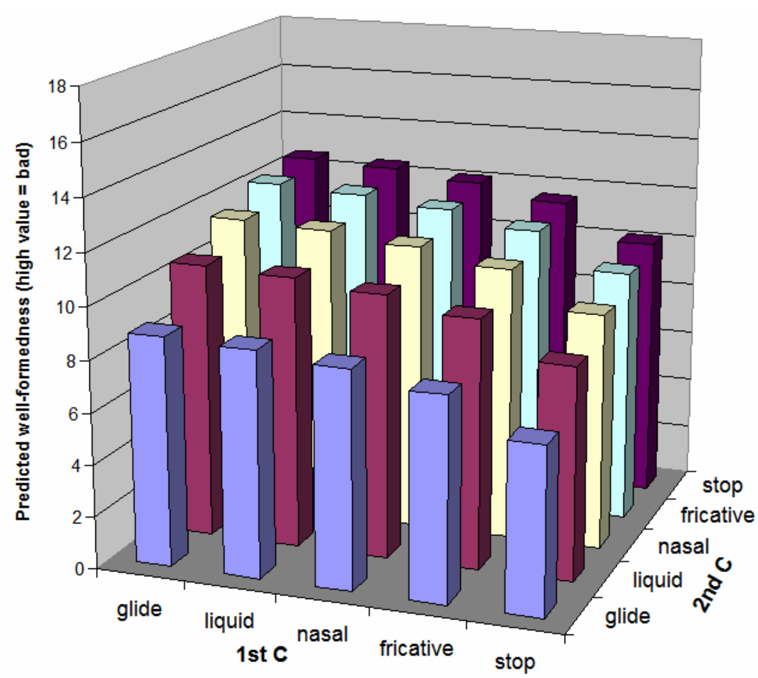
23. The C+glide onsets of Bwa are "triggers"

- They cause the constraint system to generalize the whole Hierarchy.
- Demonstration: compare Bwa with Ba, whose words are:

  [pa, ta, ka, ba, da, ga, fa, sa, va, za, ma, na, la, ra, ja, wa]

24. Ba with strong UG ("enforce sonority sequencing")

- Much weaker intuitions (though in the right direction) are projected:



- Basis of projection: the innate sonority-sequencing constraints are being used to ban onset clusters.

25. Ba with unbiased UG (as above)

- Results: all forms treated approximately alike; no sonority effect.
- This is as expected; neither UG nor the data support such an effect.

26. Simulations and UG testing for sonority sequencing: conclusions

- Simply testing your raw intuitions about "is the Sonority Hierarchy innate?" may give misleading results.
  - ➢ A weaker UG, "care about sonority", can give equivalent effects.
- Triggers in the subjects' native language (i.e. existing sonority-obeying clusters) can seriously alter the predictions.
  - ➢ It might be best to do future testing with speakers of languages that have no branching onsets at all.

27. Is English too trigger-ful to support UG experiments on Sonority?

Current work diverges on this point:

- ➢ Steriade and Wilson (2007, ms.): the ample triggers present in English mean that sonority intuitions could be projected even with primitive UG #1 (feature system only).
- ➢ Albright (2007), trying various models, suggests the opposite conclusion.

28. Generalizing a bit

- I believe that most experimental work bearing on UG:
  - ➢ does not use an explicit model of the UG principles at stake (exceptions: Wilson 2006, Albright 2007, some others)
  - ➢ Unmodeled work is vulnerable to serious reinterpretation.
- Example: Seidl and Buckley 2005
  - ➢ They make a very interesting argument that phonetic naturalness plays no role in phonotactic learning
  - ➢ But examined with the Hayes/Wilson learner, their comparison languages *differ in many ways* other than the one we care about—the learner's judgments match those of a Seidl/Buckley infant, but mostly by using constraints Seidl and Buckley may not have noticed.
- The "Dual Mechanism" model—Pinker and Prince 1988, Pinker 1999—discussed below.

BENEFITS OF MODELING II:
MAKE CONCRETE PREDICTIONS FOR INDUCTIVIST ALTERNATIVES

29. Consequences of letting inductive algorithms search databases

- Scholars who are skeptical of a rich UG (e.g. Haspelmath 1999, Blevins 2004) must perforce assume that inductive learning is used to get the grammar from the data.
- It's very intriguing actually to try this!
- One often finds generalizations that look (to linguists) unprincipled and arbitrary.
- Examined experimentally, these generalizations can have interesting consequences.

30. Example: Albright and Hayes (2003)

- We sought to predict English past tenses from their stem forms ([dʒʌmp] → [dʒʌmpt], [klɪŋ] → [klʌŋ]), using a set of rules learned by our algorithm.
- We then assessed the success of our model by carrying out a wug test.

31. Algorithm summary

- Algorithm is based on the **minimal generalization** concept given in Pinker and Prince (1988).
- Compare forms all of which undergo the same change (e.g. pairs of X → Xd verbs, pairs of ɪ → ʌ verbs), and generalize by preserving the parts of their environments that they have in common.
- Given *shine*/*shined* and *consign*/*consigned*:

  ∅ → d / [ʃaɪn ___]$_{[+past]}$ = "Insert [d] after [ʃaɪn] to form the past tense."

  ∅ → d / [kənsaɪn ___]$_{[+past]}$ = "Insert [d] after [kənˈsaɪn] to form the past tense."

- Generalization process:

change	*variable*	*shared features*	*shared segments*	*change location*		
∅ → d / [		ʃ	aɪn	___	]$_{[+past]}$	(*shine-shined*)
∅ → d / [	kən	s	aɪn	___	]$_{[+past]}$	(*consign-consigned*)
∅ → d / [	X	[+strident, +contin, –voice]	aɪn	___	]$_{[+past]}$	(generalized rule)

- When you iterate with a variety of forms (and add an elementary capacity for phonology), this rapidly converges on very general rules, of the kinds linguists are used to.
- We did this with a set of 4253 representative English stem/past pairs.

32. Important rule types discovered

- By comparing (*mis*)*read*, (*mis*)*lead*, *bleed*, *breed*:

  $i \rightarrow \varepsilon$ / [X {l, r} ___ d $]_{[+past]}$

- By comparing *face*, *wash*, *laugh*, etc.:

  $\varnothing \rightarrow t$ / [X $\begin{bmatrix} -\text{sonorant} \\ +\text{continuant} \\ -\text{voice} \end{bmatrix}$ ___ $]_{[+past]}$ "Suffix [-t] to stems ending in voiceless fricatives."

- By comparing all regular verbs and using elementary phonology:

  $\varnothing \rightarrow$ d / [X ___ $]_{[+past]}$ "Suffix /-d/ to form the past"

33. Evaluating rules

- We assign a **value** to each rule using a formula (not covered here) based on both **scope** and **accuracy**.
  - $i \rightarrow \varepsilon$ / [X {l, r} ___ d $]_{[+past]}$ is low-scope, high-accuracy (6/7, due to *plead*)
  - $\varnothing \rightarrow t$ / [X $\begin{bmatrix} -\text{sonorant} \\ +\text{continuant} \\ -\text{voice} \end{bmatrix}$ ___ $]_{[+past]}$ is medium-scope, very high accuracy (352/352)
  - $\varnothing \rightarrow d$ / [X ___ $]_{[+past]}$ is high scope, medium accuracy (4034/4253).
- In wug-testing the model, an output is scored thus: as the value of the best rule that can derive it.

34. Islands of reliability

- These are the contexts of medium-scope, very-high-accuracy rules that derive regulars, like the voiceless-fricative rule above—you can "trust" the regular mapping better in an island.
- Islands of reliability seem redundant—why not just let the general rule do its work?
- But empirically, they're not, because experiments shows that regular forms are rated as better by native speakers when they fall in an island of reliability.
- Islands of reliability, and their effect on native intuition, were discovered by Albright (2002b), and are also documented in Albright (2000, 2005) and Albright, Andrade and Hayes (2001).

35. Islands of reliability in the Albright/Hayes wug test

- Compare 10 wug verbs selected to fall within an island of reliability, vs. 10 that do not:
- Ratings task (1-7 scale): 6.1 vs. 5.3
- Production task (% of subjects): 90% vs. 75%.

36. Why we did this experiment

- Our interest in the **dual mechanism model** of morphology (Pinker and Prince 1988, Pinker 1999), which says (among other things):
    - Regular forms are derived by a single, simple rule.
- We think the Island of Reliability effect falsifies this aspect of the dual mechanism model.
- Pinker's own wug-testing work (Prasada and Pinker 1993) used human judgment, not a model, in selecting the wug words, and as a result had to use phonotactically bizarre words (e.g. "smairg") to represent the non-IOR category, contaminating the experiment and yielding (I think) a wrong conclusion.

37. More generally

- Searching for environments by machine sometimes reveals linguistic constraints not previously detected—and illustrates perhaps unexpected inductive ability in human language learners.

38. A puzzle: how to reconcile with Becker et al.'s results?

- As noted earlier, the statistically significant patterns of vowel quality, correlated with consonant voicing in Turkish, were evidently *not* learned by Turkish speakers.
- Becker et al. appeal to a rich UG:
    - "We propose that generalizations are expressed in terms of typologically-responsible OT constraints."
- I think this can't be right, because it predicts typologically-irresponsible Islands of Reliability to be unlearnable, against experimental findings (cf. (34)).
- At present, this seems a contradiction; I have no idea why the Becker et al. experiment yielded a negative result.
- Is there a principled UG theory that would predict the difference?

CONCLUSION: PROSPECTS

39. A future in modeling?

- Many fields prosper from modeling; the Chomskyan view of theoretical linguistics as an effort to solve the fundamental problem of acquisition seems ideally suited to this approach.
- Thus, we should seek the *model child*, who :
    - comes equipped with UG and appropriate learning mechanisms
    - is exposed to representative language data
    - processes the data, and learns the grammar

- Our scientific skills will be maximally tested not just in finding a good model child, but in thinking of ingenious ways to test if a given model child comes to know what real children know. Viz.
    - experiment tests of phonological knowledge
    - testing at intermediate stages of learning (same as errors of real children?)
    - comparison of errors with "permanent" errors of real children (analogical change; Albright 2002, 2006a, 20006b)

References

Albright, Adam (2000) The productivity of infixation in Lakhota. *UCLA Working Papers in Linguistics*, ed. Pamela Munro.

Albright, Adam (2002a) *The Identification of Bases in Morphological Paradigms*, UCLA Ph.D. dissertation.

Albright, Adam (2002b) Islands of reliability for regular morphology: Evidence from Italian. *Language* 78:684–709

Albright, Adam (2005) The morphological basis of paradigm leveling. In Laura Downing, Tracy Alan Hall, Renate Raffelsiefen, eds., *Paradigms in Phonological Theory*. Oxford University Press.

Albright, Adam (2006a) Base-driven leveling in Yiddish verb paradigms. Ms., Department of Linguistics and Philosophy, MIT.

Albright, Adam (2006b) Explaining universal tendencies and language particulars in analogical change, Ms., Department of Linguistics and Philosophy, MIT.

Albright, Adam (2007) Natural classes are not enough: Biased generalization in novel onset clusters. Ms., Department of Linguistics and Philosophy, MIT.

Albright, Adam, Argelia Andrade and Bruce Hayes (2001) "Segmental environments of Spanish diphthongization," *UCLA Working Papers in Linguistics* 7 (*Papers in Phonology* 5), 117-151.

Albright, Adam and Bruce Hayes (2002) "Modeling English past tense intuitions with minimal generalization," in Mike Maxwell, ed., *Proceedings of the 2002 Workshop on Morphological Learning, Association of Computational Linguistics*. Philadelphia: Association for Computational Linguistics.

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," in *Gradience in Grammar: Generative Perspectives*, ed. by Gisbert Fanselow, Caroline Fery, Ralf Vogel and Matthias Schlesewsky. Oxford: Oxford University Press, pp. 185-204.

Alderete, John (1998) Morphologically Governed Accent in Optimality Theory. Ph.D. dissertation, University of Massachusetts. Published 2002: Routledge.

Becker, Michael, Nihan Ketrez, and Andrew Nevins (2007) When and why to ignore lexical patterns in Turkish obstruent alternations. Paper given at the 2007 annual meeting of the Linguistic Society of America, Anaheim.

Berent, Iris, Donca Steriade, Tracy Lennertz, and Vered Vaknin (2007). What we know about what we have never heard: evidence from perceptual illusions. *Cognition* 104:591–630.

Berko, Jean (1958) The child's learning of English morphology. *Word* 14:150-177.

Bhat, D. (1978) A general study of palatalization. In Joseph Greenberg, ed., *Universals of Human Language* (Vol. 3, pp. 47–92). Stanford, CA: Stanford University Press.

Blevins, Juliette (2004). *Evolutionary Phonology*. Cambridge: Cambridge University Press.

Boersma, Paul (1997) How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences* 21: 43–58.
Buckley, Eugene and Amanda Seidl (2005) On the learning of arbitrary phonological rules. L*anguage Learning and Development* 1(3&4): 289-316.
Chomsky, Noam, and Morris Halle (1965) Some controversial questions in phonological theory. *Journal of Linguistics* 1:97–138.
Clements, G. N. (1990). The role of the sonority cycle in core syllabification. In John Kingston and Mary Beckman, eds., *Papers in Laboratory Phonology I*, Cambridge University Press.
Côté, Marie-Hélène (2004) Syntagmatic distinctness in consonant deletion. *Phonology* 21:1-41.
Côté, Marie-Hélène (2000) *Consonant cluster phonotactics: A perception-based approach*, Ph.D. dissertation, MIT.
Dresher, B. Elan and Jonathan D. Kaye (1990) A computational learning model for metrical phonology. *Cognition* 34:137-195.
Ernestus, M. and R. Harald Baayen, R. H. (2003) Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language* 79:5–38.
Gildea, Daniel, and Daniel Jurafsky (1996) Learning bias and phonological rule induction. *Computational Linguistics* 22:497–530.
Gordon, Matthew (2002) A factorial typology of quantity insensitive stress. *Natural Language and Linguistic Theory* 20:491-552
Graff, Peter (2007) The Nature of Systematic Sound Change – An Experiment. M.A. thesis, University of Edinburgh.
Haspelmath, Martin (1999) Optimality and diachronic adaptation. *Zeitschrift für Sprachwissenschaft* 18:180-205.
Hayes, Bruce (1995) *Metrical Stress Theory: Principles and Case Studies*. Chicago: University of Chicago Press.
Hayes, Bruce and Colin Wilson (in press) A maximum entropy model of phonotactics and phonotactic learning. To appear in *Linguistic Inquiry.*
Kager, René and Nine Elenbaas (1999) Ternary rhythm and the lapse constraint. *Phonology* 16:273-329.
Kaun, Abigail (1995) The Typology of Rounding Harmony: An Optimality Theoretic Approach. Ph.D. dissertation, UCLA.
Kawahara, Shigeto (2006) Mimetic gemination in Japanese: A challenge for Evolutionary Phonology. *Theoretical Linguistics* 32.3: 411-424.
Koo, Hahn and Jennifer Cole (2006) "On learnability and naturalness as constraints on phonological grammar. *Proceedings of ISCA Tutorial and Research Workshop on Experiemental Linguistics*, 28-30 August, Athens.
Martin, Andrew (2007) *The Evolving Lexicon*. Ph.D. dissertation, UCLA. http://www.linguistics.ucla.edu/general/dissertations/Martin_dissertationUCLA2007.pdf
McCarthy, John (1981) Prosodic Structure and Expletive Infixation. *Language* 58:574-590.
Mielke, Jeff (2005) Modeling distinctive feature emergence. *WCCFL* 24:281-289.
Moreton, Elliott (to appear b). Underphonologization and modularity bias. To appear in: Steve Parker (ed.), *Phonological Argumentation: Essays on Evidence and Motivation.* London: Equinox.
Moreton, Elliott (to appear a). Learning bias as a factor in phonological typology. To appear in: Charles Chang and Anna Haynie (eds.), *Proceedings of the 26th Meeting of the West Coast Conference on Formal Linguistics (WCCFL)*.
Myers, Scott (2002). Gaps in factorial typology: The case of voicing in consonant clusters. MS, University of Texas at Austin.
Ohala, John J. (1981) The listener as the source of sound change. In C. S. Masek, R. A. Hendrick,&M. F. Miller (Eds.) *Papers from the parasession on language and behavior* (pp. 178–203).
Pater, Joe and Anne-Michelle Tessier (2003) Phonotactic knowledge and the acquisition of Alternations. In M.J. Solé, D. Recasens, and J. Romero (eds.) *Proceedings of the 15th International Congress on Phonetic Sciences*, Barcelona. 1177-1180.

Pertz, D. L. and T. G. Bever (1975). Sensitivity to phonological universals in children and adults. *Language* 51:149–162.
Philip, Victoria (in progress) Phonetic naturalness as a learning bias in phonological acquisition: an experimental study. Almost-finished M.A. thesis, Department of Linguistics, UCLA.
Pinker, Steven (1999). *Words and rules: the ingredients of language*. New York: Basic Books.
Pinker, Steven and Alan Prince (1988) On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, 28:73-193.
Prasada, Sandeep and Steven Pinker (1993) Generalisation 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*. Published 2004: Blackwell, Oxford.
Pycha, Anne, Pawel Nowak, Eurie Shin, and Ryan Shosted (2003). Phonological rule learning and its implications for a theory of vowel harmony. In: G. Garding and M. Tsujimura (eds.), *WCCFL* 22, 423–435. Somerville: Cascadilla Press.
Schane, Sanford, Bernard Tranel, and Harlan Lane (1974). On the psychological reality of a natural rule of syllable structure. *Cognition* 3(4): 351–358.
Scholes, Robert (1966) *Phonotactic grammaticality*. The Hague: Mouton.
Seidl, Amanda, and Eugene Buckley (2005). On the learning of arbitrary phonological rules. *Language Learning and Development* (1):289–316.
Steriade, Donca (2001a). The phonology of perceptibility effects: the P-map and its consequences for constraint organization. MS., Department of Linguistics, University of California, Los Angeles.
Steriade, Donca (2001b). Directional asymmetries in place assimilation: a perceptual account. In: Hume & Johnson 2001: 219–250. San Diego: Academic Press.
Steriade, Donca and Colin Wilson (2007) Learning the grammar of English word-initial clusters. Handout, LSA Institute, Stanford University.
Tesar, Bruce and Paul Smolensky. 1993. The learnability of Optimality Theory: an algorithm and some basic complexity results. ROA.
Thatte, Victoria (in progress) Phonetic naturalness as a learning bias in phonological acquisition. MA thesis, UCLA.
Wilson, Colin (2003) Experimental investigation of phonological naturalness. In: G. Garding and M. Tsujimura (eds.), *WCCFL* 22, 533–546. Somerville: Cascadilla Press.
Wilson, Colin (2006). Learning phonology with substantive bias: an experimental and computational study of velar palatalization. *Cognitive Science* 30 (5): 945–982.
Zhang, Jie and Yuwen Lai (2006). Testing the role of phonetic naturalness in Mandarin tone sandhi. *Kansas Working Papers in Linguistics* 28:65-126.
Zuraw, Kie (2007). The role of phonetic knowledge in phonological patterning: Corpus and survey evidence from Tagalog. *Language* 83:277-316.