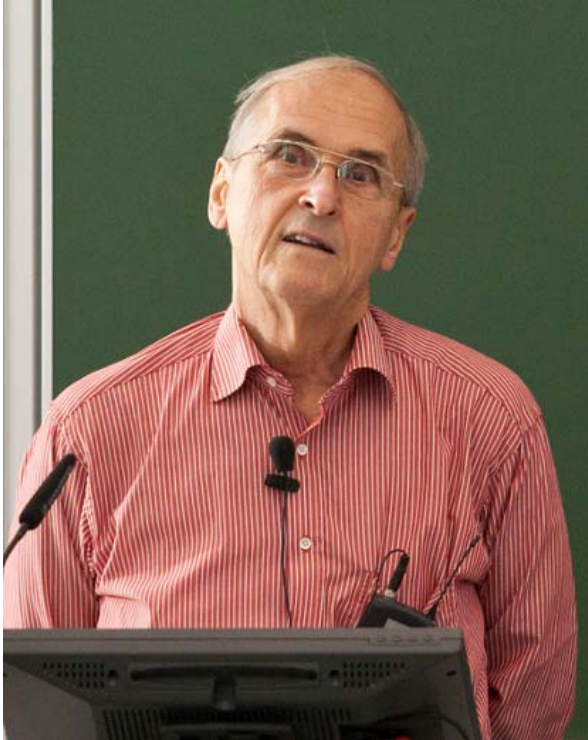


“Phonological acquisition is not always accurate”: extending the Kiparskyan research program

Bruce Hayes
Department of Linguistics
UCLA

Linguistic Society of America
Portland, OR
January 8, 2015

Credits up front



Paul Kiparsky
(Stanford
University)



Dustin Bowers
Ph.D. UCLA
forthcoming June 2015



James White
(UCL)
Ph.D. UCLA
2013

Overview of the talk

- In an important research program during the 1960's and 1970's, Paul Kiparsky put forth a criterion for phonological theories:
 - *Not* just explain how children effectively acquire the ambient phonological system.
 - *but also* explain the cases where they fail and acquire something different.
- Goals: reemphasize Kiparsky's original point, and pursue it in some novel directions

Preliminary background: the classical theory of phonological change

- This is textbook material; see Hayes and White (forthcoming) for a recent summary.
- The phonological grammar is somehow *bifurcated* — this idea has been put forth in multiple ways.

<i>Type I</i>	<i>Type II</i>	<i>Source</i>
phonological processes	phonetic processes	Keating (1985)
lexical processes	postlexical processes	Mohanan (1986)
analogy	sound change	19th century

Phonetic processes as creator of puzzles for language acquirers

- Phonetic processes evolve over time, often becoming **more extreme**.
- They evolve with relative **independence** from the “deeper” phonological grammar (Labov 1994)
- After a certain point they phonetically **no longer clearly manifest the original phonological pattern** — confronting a new generation with an acquisition conundrum ...
- ... leading, sometimes, to **misacquisition**
- i.e. the drift of phonetic change ultimately serves up to the next generation of children a data pattern on which they place a radically different interpretation.

Why is this interesting from the viewpoint of phonological theory?

- Whenever it happens, it forms a **real-life phonological experiment**.
- And so the record of phonological change, taken from
 - philology
 - the Comparative Method
 - internal reconstruction

becomes a trove of data that can bear on how children learn phonology.

- This was the research program launched by Kiparsky in the 1960's and 1970's.

Legacy of the Kiparskyan research program

- Many elements of our current thinking arose from this program:
 - Opacity
 - Rule ordering typology (feeding, bleeding, etc.)
 - Paradigm uniformity
 - Learning biases

A useful place to read everything

- Kiparsky, Paul (1982) *Explanation in phonology*.
Dordrecht: Foris.

Foci for this talk

- A vivid case where acquisition was indeed imperfect, and its consequences for theory
- Pursuing the Kiparskyan paradigm with new tools in the 21st century

I. A case of imperfect phonological acquisition

Source and affiliation

- Bowers, Dustin (submitted) Phonological restructuring in Odawa, ms. Department of Linguistics, UCLA;
<https://sites.google.com/site/dustinbowerslinguist/papers>
 - Bowers draws heavily on Rhodes (1985a, 1985b)
- Odawa is Algonquian, spoken in the Great Lakes region

Historical evolution, earliest stage: iambic stress assignment, left to right

- In a sequence of short-voweled syllables, this places stress on all even ones; also on final and V: syllables.

(gʊtí)(gʊmí)(nʌgí)(bíná:) ‘he rolls someone’

(nɪ-gú)(tígú)(mɪnʌ)(gɪbí)(ná:) ‘I roll someone’

- This stress pattern is widely found in Algonquian languages and is likely ancient; see e.g. Hayes (1995).
- Note that the existence of short-voweled prefixes like *nɪ-* makes possible stress **alternations** in the paradigm.

- Data like these are not a historical conjecture; they appear in 19th century studies by Baraga.

Next stage of evolution: phonetic change in stressless syllables

- Iambic stress systems are prone to **vowel reduction** (Hayes 1995).
- This happened in Odawa: the stressless vowels become **steadily shorter and more reduced**.

Shorten:

(gǔtí)(gǔmí)(nǎgí)(bǎná:)

‘he rolls someone’

(nǎ-gú)(tǐgú)(mǎnǎ)(gǐbǐ)(ná:)

‘I roll someone’

Reduce:

(gǎtí)(gǎmí)(nǎgí)(bǎná:) ‘he rolls someone’

(nǎ-gú)(tǎgú)(mǎnÁ)(gǎbí)(ná:) ‘I roll someone’

- This stage was heard in the 1930’s by **Leonard Bloomfield** (publ. 1957), who reported the reduced vowels as:

“rapidly spoken and often whispered
or entirely omitted”

Step 3: a new generation of children hears the degraded data, in the late 1930's

- What for Mom and Dad is a quick and lazy way of pronouncing a vowel that is phonologically there, is now simply no vowel at all.
- For these data see Rhodes (1985a,b), based on speakers born around this time.

Consequences of taking reduction to its logical conclusion (deletion)

- Stress is no longer relevant (all stressless vowels are gone!) — so I won't transcribe it.
- What was originally a **vowel-reduction** alternation was heard by the new generation as a — potential — **syncope** alternation.

gtɪgmingibna:

‘he rolls someone’

ngʊtgʊmnʌgbina:

‘I roll someone’

The correct textbook-style analysis for the data late-1930's Odawa children heard

- Recapitulate diachrony; i.e.
- Assume “**etymological**” underlying representations — all vowels in their correct historical places.
- Assume **abstract left-to-right iambic stress**, followed by **categorical syncope** of stressless vowels.

- This is not what the kids did...

What actually happened I: new underlying representations

- For each stem, roughly, the **isolation form** is now the underlying form.
 - This oversimplifies — visit Bowers’s poster Fri. 10:30 for the more interesting version.
- Prefixation is to this form, with relatively little phonology:

gtigmingibna: ‘he rolls someone’ *unchanged*

ndΛ-gtigmingibna: ‘I roll someone’ *novel form*

(earlier 1 sg. form: **ngutgumnΛgbina:**)

- Comparable changes happened **throughout the vocabulary.**

Where does the “crazy” prefix [ndʌ-] come from?

- **Recutting.** The [n] is part of the old prefix, and the [dʌ] comes from misapprehension of morpheme boundaries in the old alternations.
- **Historical derivation**

ʌgo:dʒɪn	nɪ-ʌgo:dʒɪn	‘hang, I hang’
—	nɪdʌgo:dʒɪn	resolve hiatus with [d]
(ʌgó:)(dʒín)	(nɪdʌ)(gó:)(dʒín)	iambic stress
(əgó:)(dʒín)	(nədʌ)(gó:)(dʒín)	vowel reduction
go:dʒɪn	ndʌgo:dʒɪn	syncope

- Justifying the recutting:

$$\begin{array}{c|c}
 & \text{g o: dʒ I n} \\
 \text{n d } \Lambda & \text{g o: dʒ I n}
 \end{array}$$

- So [ndΛ-] is a prefix!
- Similar prefixes arose from other recut stem material, like [ndɪ-].
- These prefix allomorphs now compete with one another, with a non-etymological distribution, and much free variation.

Upshot

- The phonetic drift of Vowel Reduction into full deletion induced a catastrophe:
 - massive stem reshaping
 - novel prefix allomorph system.
- Bowers: dating of the sources suggests that the changes occurred — in a still-vibrant language — *the moment that reduction became crossed the line to deletion.*

And it wasn't just Odawa

- Bowers: Old Russian, Old Irish, likewise had alternating stress, reduction developing into syncope.
- They likewise restructured radically, as soon as syncope had thoroughly kicked in.

What do these cases mean?

- *Human children are hopeless at acquiring phonology?*
- This seems unlikely to me — plenty of interesting phonology can be stable.
- It makes sense to try to **localize** the acquisition problem.

Bowers's conjecture

- The data pattern that the restructuring Odawa children encountered, unusually, requires genuine **serial derivation** for its analysis.
- You must first assign stress, to know where to “syncopate.” After syncope, the alternating count that governed stress is no longer present.

/ni-gʊtɪgʊmɪnʌgɪbɪnɑː/

(ni-gú)(tɪgú)(mɪnʌ)(gɪbí)(náː)

∅ ∅ ∅ ∅

[ngʊtɪgʊmɪnʌgɪbɪnɑː]

UR

Stress Assignment

Syncope

Surface representation

- Maybe phonology isn't serial?

The controversy over serialism in phonology

- Classical Optimality Theory (Prince and Smolensky 1993, McCarthy and Prince 1995) radically introduced:
 - **single-step derivation**
 - **parallel evaluation** of many candidates.
- This turned out to be **surprisingly viable**, with well-motivated strategies to cover phenomena that people had thought required serialism. See below.
- It's only rather unusual cases — like Odawa — that require **faithfulness to intermediate representations** — hence serialism.

Serial versions of Optimality Theory

- John McCarthy and colleagues have recently proposed — and ably defended — *serial* versions of OT (Candidate Chain Theory, Harmonic Serialism)
 - Sample references: McCarthy (2007, 2008, 2010)
 - Candidates are not single representations but (roughly) **sequences** of representations.
 - These theories work extremely well (like rule-based phonology) for constructing phonological analyses that **mimic the historical origins of synchronic patterns**.
- Stress-syncope interactions form one of the best arguments for serial Optimality Theory — McCarthy (2008).

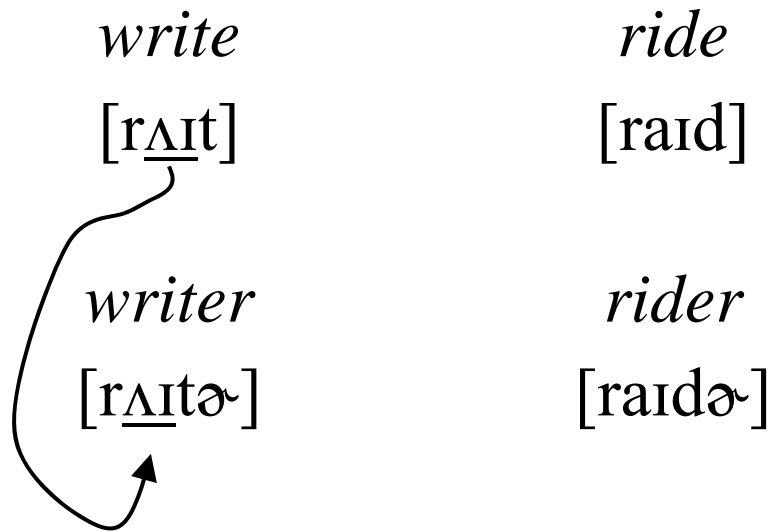
- But if the crucial cases involve breakdown of acquisition, the shoe is on the other foot — serialism may be too powerful!

Example of eliminating derivations I: Non-serial account of counterbleeding in *writer*

- Possible historical origin, sequence of sound changes:

<i>write</i>	<i>writer</i>	<i>rider</i>	
/raɪt/	/raɪt-ə/	/raɪdə/	proto-American English
Λɪ	Λɪ	—	<i>Raising</i>
			aɪ → Λɪ / ___ [− voice]
—	ɾ	ɾ	<i>Tapping</i>
			t,d → ɾ / V ___ Ǟ
[rΛɪt]	[rΛɪɾə]	[rΛɪɾə]	contemporary forms

- This is often nonserially analyzed with Faithfulness (“Output-Output”) to **other forms in the paradigm** — *writer* gets [ΛI] by inheritance from *write*, not derivationally.



Example of eliminating derivations II: Non-serial account of counterfeeding in Western Basque (Hualde 1991)

/aa/	/ea/	Underlying representation
—	i	/e/ Raising before vowels
e	—	/a/ Raising before vowels
[ea]	[ia]	Surface representation

- This is commonly analyzed as **distantial Faithfulness**:
/a/ → *[i] is “too long a phonetic journey” and violates an undominated Faithfulness constraint (Kirchner 1996 et seq.)

Upshot of the Odawa discussion

- Reconsideration of the Kiparskyan research program suggests a possible resolution to the serialism debate.
- Most of the evidence for serialism received sensible reanalyses *before* serial versions of OT appeared on the scene.
- Serialism is perhaps dispensable — *if* the cases for which serialism is absolutely necessary are those that language learners eschew, preferring to restructure.
- If this works out, it is strong vindication for the Kiparskian approach, which tells us not to take data patterns necessarily at their historical face value.

II. Renewing the Kiparskyan paradigm in contemporary research

How to explain why children sometimes acquire phonology imperfectly?

- Stupidity
- Bias

Are kids just dumb when it comes to learning phonology?

- This is an uncharitable reading of Hooper (1976), a work that took very seriously the Kiparskyan criterion of predicting language breakdown.
- I think recent research refutes this view: kids are actually **virtuosi**.
 - In many ways, they outperform phonologists in apprehending the data pattern of a language.

Kids notice amazing amounts of detail

- ... insofar as we can determine from how they take wug-tests when they reach adulthood.
- See e.g. Ernestus and Baayen (2003), Hayes, Zuraw, Siptár and Londe (2009), Gouskova and Becker (2013)
- Said detail often is quite arbitrary.
- Example:
 - All verbs in English that **end in a voiceless fricative** are regular.
 - Albright and Hayes's (2003) wug test shows that speakers **particularly prefer regular pasts** for wug stems of this type.

Kids match lexical frequencies with striking precision

- Again the support comes from wug-test data on adults.
- Example:
 - Vowel height in stems has quantitative effects on Hungarian vowel harmony: lower front vowels trigger harmony in more stems than higher.
 - This quantitative pattern gets noticed and replicated in wug-test studies (Hayes and Londe 2006; Hayes et al. 2009)

Summing up: kids not dumb

- Both the ability to notice detail and frequency-matching make people perform very well on wug tests.
- I think the average published phonology of a language is much smaller than what native speakers actually know.

Theory II: kids bring *biases* to phonological acquisition

- I.e. they expect certain patterns *a priori* and are skeptical about other patterns.
- This could be taken to be a Kiparskyan idea; e.g. his suggestion of learning bias for particular rule orderings.
- In modern guise: work such as Wilson (2006), Moreton (2008) has made bias a leading idea in contemporary theorizing.

An important bias treated in the Kiparskian program

- **Paradigm uniformity**

“Allomorphy tends to be minimized in a paradigm.”
(Kiparsky 1982, 65)

- Comes in two flavors:

- morphemes should not alternate *at all*
- morphemes should alternate in *phonetically non-salient ways*

Kiparsky's Swiss German example (1982:19-20)

- Conservative Northeastern dialects of Swiss German:

- /o/ has the allophone [ɔ] before nonlateral coronals

[hɔrn], [rɔss], [xrɔttə], [bɔdə], [pɔft] vs.

[grob], [ops], [ofə], [xoxxə], [rokx], [bogə]

[foll], [gold]

- But the older process of **Umlaut**, triggered in plurals, derives [ö] from underlying /o/:

- singular [bogə], plural [bögə]

- singular [bɔdə], plural [bödə]

- Innovating dialects of Swiss German
 - The Umlauted version of /o/ before nonlateral coronals is now [ö] — low front rounded
 - singular [bɔgə], plural [böɡə] (same)
 - singular [bɔdə], plural [bödə]
 - *Not* due to lowering of [ö]! [ö] when not derived from /o/ did not lower: [plötsli], [fröss]
- What triggered the new [ö]? Gradient phonetic paradigm uniformity: restoring Umlaut as an alternation of backness only, not height and backness.

III. The paradigm uniformity bias in contemporary theory

Kiparsky in the 1970's is ambivalent about paradigm uniformity

- Considerable data support it ...
- But nothing in the rule-based framework of the time could **accommodate it as part of formal analysis.**

How things are different now

- **Constraint-based grammars** let us incorporate paradigm uniformity as an actual ingredient of analysis, rather than a functional principle lurking around the periphery.
- You need several specific ingredients to do this.

Element I: Output-output correspondence constraints

- Source: Benua (1997) and much later work
- These penalize lack of faithfulness between a candidate and the **base form of the paradigm** in which it occurs.

Element II: the *P-map*

- Source: Steriade (2001, 2008)
- A data structure thought to be compiled by children during acquisition, encoding the **perceptual distance between all pairs of potentially-alternating segments.**
- Requisite to enforcing phonetic paradigm uniformity
- A tiny P-map from White (2013); obtained by maxent modeling of a confusion matrix (Wang and Bilger 1973)

	t	d	ð
t	0	1.98	3.57
d	1.98	0	0.02
ð	3.57	0.02	0

Element III: *MAP constraints

- Source: Zuraw (2007, 2013)
 - These generalize output-to-output IDENT() constraints, but may penalize larger, multi-feature distances.
- *MAP(t-d): “Assess a violation when a candidate has a [d] where its morphological base has [t]”
(same as IDENT (voice))
- *MAP(t-ǰ): “Assess a violation when a candidate has a [ǰ] where its morphological base has [t]”

Element IV: ranking bias (Zuraw)

$x \dots y \dots z$

- If x is phonetically farther from z than y is (on the P-map), language learners expect:

$*\text{MAP}(x-z) \gg * \text{MAP}(y-z)$

Element V: Learning algorithms

- Various constraint based frameworks let us model language acquisition with **algorithms that rank the constraints**. (Tesar and Smolensky 2000, Boersma and Hayes 2001)
- More accurate and capable algorithms are available in the **Harmonic Grammar** framework (Legendre et al. 1990, Smolensky and Legendre 2006, Pater 2009, Jesney 2010, Potts et al. 2010, Jesney and Tessier 2011)
 - closely related to OT
 - Constraints not ranked but are assigned **weights** (real numbers reflecting their strength)

Framework and algorithm to be used here

- Framework: the **maxent** flavor of Harmonic Grammar (Goldwater and Johnson 2003)
- Learning algorithm: the **Conjugate Gradient** algorithm (Press et al. 1992)
- Software: **Maxent Grammar Tool** (Wilson and George 2009)
 - www.linguistics.ucla.edu/people/hayes/MaxentGrammarTool

Element VI: mathematical implementation of learning bias

- Origin: Wilson (2006)
- In Maxent Harmonic Grammar, we can specify a **prior** weight (μ) for each constraint, letting it serve as the value that will emerge from learning unless the data override it.

Summing up the above and applying it to the Kiparskyan research paradigm

- We assemble the elements above:
 - OO-correspondence constraints, taking the form of *MAP, with preferred weightings deriving from the P-map.
- We assemble data similar to what the innovating generation of children faced.
- If all goes well, we can “postdict” the innovating change with our learning algorithm, starting from principled assumptions.

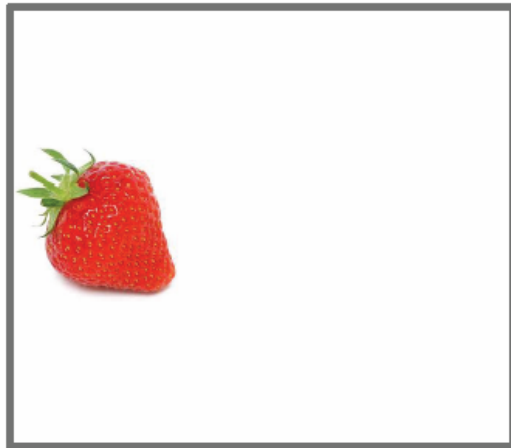
IV. A simulation study of paradigm uniformity bias using experimental data

Work of James White and collaborators

- White (2013, 2014), Hayes and White (forthcoming), White and Sundara (2014).
- Goal is to study acquisition difficulty using the tools just given, with data from an **artificial language experiment**.

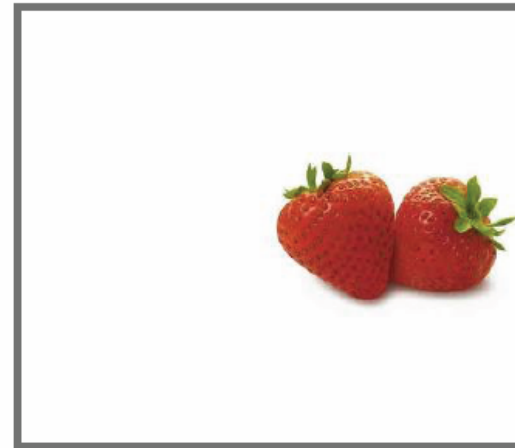
White's experiments: informal overview

- Subjects learned to produce plurals, trained on singular-plural pairs.
- Sample training stimulus:



[luman]!

...



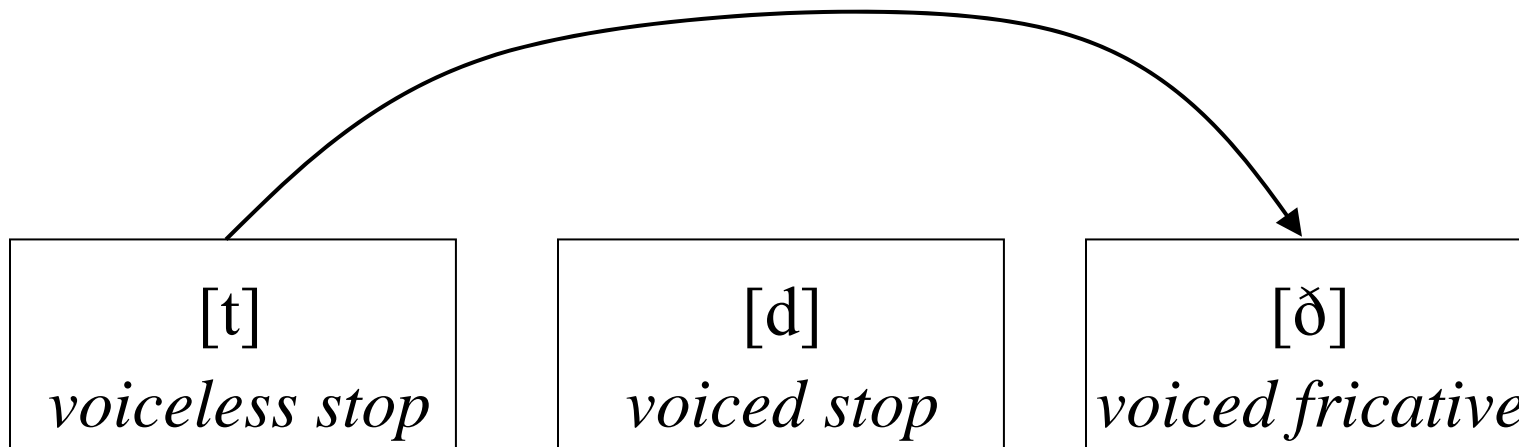
[lumani]!

Schematic examples of the words employed

<i>Sing.</i>	<i>Plural</i>	<i>classification</i>
[luman]	[lumani]	bland ordinary form (suffixation only)
[gimal]	[gimali]	bland ordinary form (suffixation only)
[ʃar it]	[ʃari ð i]	intervocalic spirantization/voicing of /t/
[mas id]	[mas i di]	nothing happens to intervocalic /d/

“Saltation”

- Hayes and White (forthcoming) call alternations like [t] ~ [ð] **saltatory**, since [t] “leaps over” invariant [d] to arrive at [ð].



Saltation violates Zuraw's learning bias

- E.g. it's hard *not* to alternate [d] with [ð] when you are already alternating [t] with [ð].
- Reason: the greater distance [t] ~ [ð] alternation is penalized by a constraint with a preferred-higher weight.
- If the Zurovian learning bias is true, saltation should be **hard to learn**.

White's experiments confirm this

- Not so hard to learn a [t] - [ð] alternation
- But when you do, [d] gets **carried along**, becoming [ð] as well: *[masiði] for correct [masidi].
- This happens even when the learning data includes ample instances of non-alternating [d].
- An followup study with infants (White and Sundara 2014) indicates that is true for them too.

Analyzed as a “marked” OT grammar

a. /t/ becomes [ð] intervocalically

/ata/	*MAP (d, ð)	*V[–voice]V	*V[–cont]V	*MAP (t, ð)	*MAP (t, d)
☞ aða				*	
*ada			*!		
*ata		*!	*		*

b. /d/ is stable

/ada/	*MAP (d, ð)	*V[–voice]V	*V[–cont]V	*MAP (t, ð)	*MAP (t, d)
☞ ada			*		
*aða	*!				

The grammar is “marked” because it violates the P-map principle

- Transitivity argument:

$*MAP(d, \check{d}) \gg *V[-\text{continuant}]V$

$*V[-\text{continuant}]V \gg *MAP(t, \check{d})$

- Therefore, a non-P-map-compliant ranking:

$*MAP(d, \check{d}) \gg *MAP(t, \check{d})$

Modeling the subjects' behavior in Maxent Harmonic Grammar

- White's procedure:
 - μ values (preferred constraint weights) for $*MAP(t, d)$, $*MAP(d, \check{d})$, $*MAP(t, \check{d})$ are the values from White's experimentally-derived P-map, given above.
 - Feed the maxent learning software the same data that the experimental participants got.

How the weights changed in the course of learning

<i>Constraint</i>	μ	<i>Weight post-learning</i>
*V[−voice]V	0	2.04
*V[−contin]V	0	.48
*MAP(t, d)	1.98	2.74
*MAP(t, ð)	3.57	1.04
*MAP(d, ð)	0.02	1.51

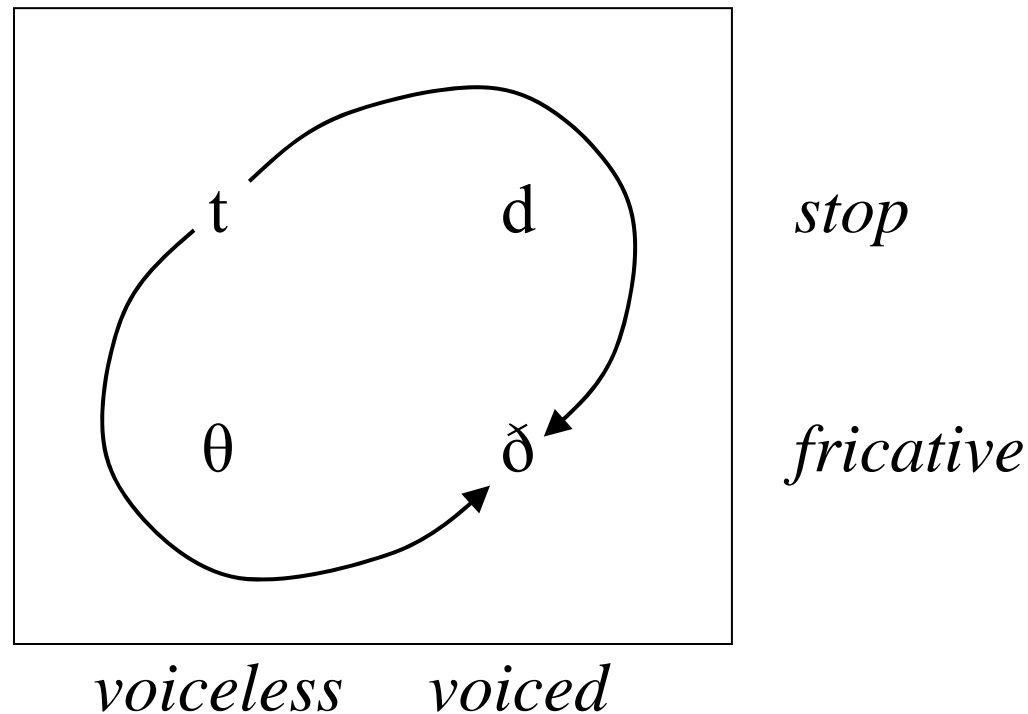
- MAP(t, ð) and MAP(d, ð) **swap places** — but **not enough** to match the data fully; bias holds them back.

Testing the learned grammar

- Test the trained grammar with the same test items that the experimental participants got.
- Results: **same mistakes** that the experimental participants made
 - [masid] ~ *[masiði] preferred over correct [masid] ~ [masidi]

The fricative stimuli

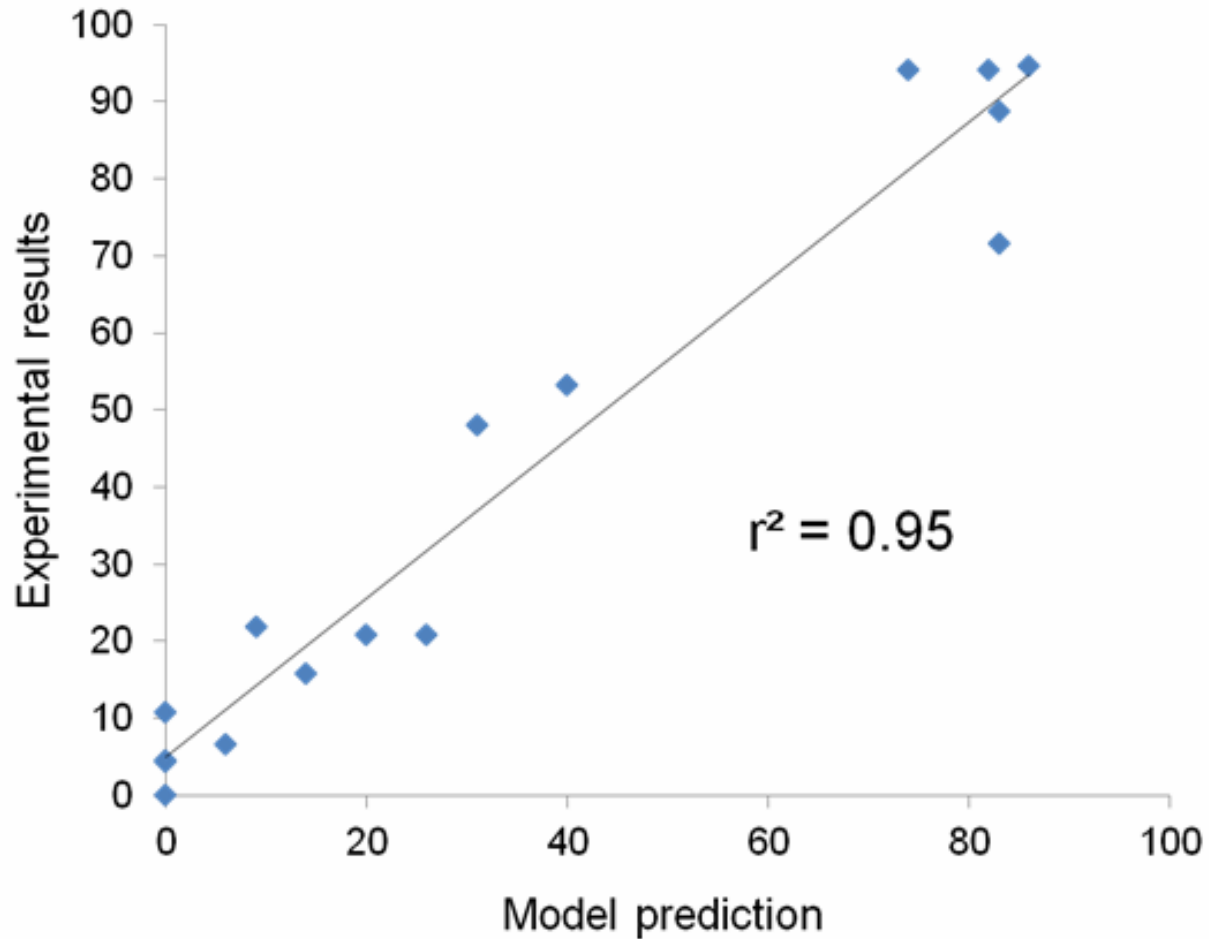
- The experiment also included **fricative stems** like [puriθ] ~ [puriθi] — here, [t] saltates over [θ] to get to [ð].
- The two saltations in the experiment are compared here:



The fricative stimuli: experimental outcome and modeling result

- Again, subjects often err, producing *[purið̥i] for [puriθ̥i].
- But not *as often* as with *[masið̥i] for [masidi].
- Why? Perceptual data show that [θ] - [ð̥] are further apart than [d] - [ð̥] — making *[purið̥i] a bigger “journey” and reducing its appeal as a candidate.
- White’s model predicts the difference accurately.

White's model fit (all experiments, all predictions vs. observed)



White's experiments are modeled on a real-life example

- Various dialects of **Sardinian** actually instantiate the scenarios White tested.

The historical evolution of Sardinian dialects through phonetic change and restructuring

- Scenario here is from Hayes and White (forthcoming), following Bolognesi (1993), Ladd and Scobbie (2003).

Stage I: creation of saltation

'30' 'the 30' 'house' 'the house'

[trinta]	[s:u trinta]	[dɔmu]	[s:a dɔmu]	ur-forms
—	d	—	ð	chain-shift lenition I
—	ð	—	∅	chain-shift lenition II
—	—	—	d	analogical restoration
[trinta]	[s:u ðrinta]	[dɔmu]	[s:a dɔmu]	attested forms

- Leveling the extreme [d] ~ ∅ alternation created the **Sestu** dialect (Bolognesi 1993)
- Alternation was “extreme” because stem-initial (Beckman 1997, 1998), neutralizing /b,d,g/ to null.

The historical evolution of Sardinian dialects, stage II: repair of saltation

'30' 'the 30' 'house' 'the house'

[trinta][s:u ðrinta][dɔmu][s:a dɔmu]	as above
— — — ð	saltation repair
[trinta][s:u ðrinta][dɔmu][s:a ðɔmu]	observed forms

- The very same error made by White's subjects created the pattern of the **Logudorese** dialect (Ladd and Scobbie 2003)

Future research?

- We're now in a position to try to implement the Kiparskian program in full computational explicitness.
- Historical reconstruction creates knowledge of data patterns faced by the children of yore.
- Bias-based learning simulations should, if the theory is right, be able to model the large-scale changes that took place.

Summing up

- The Kiparskyan orientation — that studying what systems children *fail* to learn can be as important as studying the systems they *do* learn — remains as relevant today as in the 1970's.
- It suggests a possible basis for skepticism about the need for serial frameworks in phonology.
- The original research program can be strengthened with contemporary formal models and research methods:
 - Constraint based grammars
 - Learning simulations
 - Experiments
 - Explicit theories of learning bias

Thank you

A downloadable copy of these slides, with references included, is available at:

[www.linguistics.ucla.edu/people/hayes/papers/
HayesLSAPlenaryTalkSlidesJan8_2015.pdf](http://www.linguistics.ucla.edu/people/hayes/papers/HayesLSAPlenaryTalkSlidesJan8_2015.pdf)

Thanks to Dustin Bowers and other members of the UCLA Phonology Seminar for their help in preparing this talk.

References

- Albright, Adam and Bruce Hayes (2003). Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* **90**. 119-161
- Beckman, Jill (1997). Positional faithfulness, positional neutralization, and Shona vowel harmony. *Phonology* **14**. 1–46.
- Beckman, Jill N. (1998). *Positional faithfulness*. PhD dissertation, University of Massachusetts, Amherst.
- Benua, Laura (1997). *Tranderivational identity: phonological relations between words*. Ph.D. dissertation, University of Massachusetts, Amherst.
- Bloomfield, Leonard. 1957. *Eastern Ojibwa: Grammatical sketch, texts and word list*. Ann Arbor: University of Michigan Press.
- Boersma, Paul and Bruce Hayes (2001). Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* **32**. 45-86.

- Bolognesi, Roberto (1998). *The phonology of Campidanian Sardinian: a unitary account of a self-organizing structure*. The Hague: Holland Institute of Generative Linguistics.
- Ernestus, Miriam and Harald Baayen (2003). Predicting the unpredictable: Interpreting neutralized segments in Dutch, *Language* 79, 5-38.
- Goldwater, Sharon & Mark Johnson. 2003. Learning OT Constraint Rankings Using a Maximum Entropy Model. In Jennifer Spenader, Anders Eriksson & Östen Dahl (eds.), *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, 111–120. Stockholm: Stockholm University.
- Gouskova, Maria and Michael Becker (2013) Nonce words show that Russian yer alternations are governed by the grammar. *NLLT* 31: 735–765.
- Hayes, Bruce (1995). *Metrical stress theory: principles and case studies*. Chicago: University of Chicago Press.

- Hayes, Bruce and Zsuzsa Cziráky Londe (2006). Stochastic phonological knowledge: the case of Hungarian vowel harmony. *Phonology* **23**: 59-104.
- Hayes, Bruce and James White (forthcoming) Saltation and the P-map. Provisionally accepted in *Phonology*.
- Hayes, Bruce, Kie Zuraw, Péter Siptár, and Zsuzsa Londe (2009). Natural and unnatural constraints in Hungarian vowel harmony. *Language* **85**: 822-863.
- Hooper, J. B. (1976). *An introduction to natural generative phonology*. New York: Academic Press.
- Hualde, José I. (1991) *Basque Phonology*. London: Routledge.
- Jesney, Karen (2010). Licensing in multiple contexts: an argument for Harmonic Grammar. *CLS* **45**. 287-301.
- Jesney, Karen and Anne-Michelle Tessier (2011). Biases in Harmonic Grammar: the road to restrictive learning. *NLLT* **29**. 251-290.

- Keating, Patricia A. (1985) Universal phonetics and the organization of grammars. *Phonetic Linguistics*, ed. V. Fromkin, Academic Press, 115-132.
- Kiparsky, Paul. 1970. Historical linguistics. In *New horizons in linguistics*, ed. J Lyons, 302-316. Penguin Books.
- Kiparsky, Paul (1982) *Explanation in phonology*. Dordrecht: Foris.
- Kirchner, Robert (1996) Synchronic chain shifts in Optimality Theory. *Linguistic Inquiry* 27341-350.
- Labov, William (1994). *Principles of linguistic change. Volume 1: Internal factors*. Oxford: Blackwell.
- Ladd, D. Robert & James M. Scobbie (2003). External sandhi as gestural overlap? Counter-evidence from Sardinian. In John Local *et al.* (eds), *Papers in Laboratory Phonology VI*. Cambridge: Cambridge University Press. 164–182.
- Legendre, Géraldine, Yoshiro Miyata and Paul Smolensky. 1990. *Harmonic Grammar: A formal multi-level connectionist theory*

of linguistic well-formedness: Theoretical foundations. Report CU-CS-465-90. Computer Science Department, University of Colorado at Boulder. (Online: www.cs.colorado.edu/departments/publications/reports/docs/CU-CS-465-90.pdf.)

McCarthy, John. 2007. *Hidden generalizations: Phonological opacity in Optimality Theory*. Equinox Publishing.

McCarthy, John. 2008. The serial interaction of stress and syncope. *Natural Language and Linguistic Theory* 26: 499–546.

McCarthy, John. 2010. An introduction to Harmonic Serialism. *Language and Linguistics Compass* 4 (10):1001–1018.

Mohanan, K. P. (1986) *The Theory of Lexical Phonology*. Dordrecht: Reidel.

Moreton, Elliott (2008). Analytic bias and phonological typology. *Phonology* 25.83–127.

Potts, Christopher, Joe Pater, Karen Jesney, Rajesh Bhatt & Michael Becker (2010). Harmonic Grammar with linear

- programming: From linear systems to linguistic typology.
Phonology 27, 77-117.
- Prince, Alan & Paul Smolensky (1993). *Optimality theory: constraint interaction in generative grammar*. Ms, Rutgers University & University of Colorado, Boulder. Published 2004, Malden, Mass. & Oxford: Blackwell.
- McCarthy, John & Alan Prince (1995). Faithfulness and reduplicative identity. In Jill Beckman, Suzanne Urbanczyk & Laura W. Dickey (eds.) *University of Massachusetts occasional papers in linguistics 18: Papers in Optimality Theory*. 249–384.
- Moreton, Elliott (2008). Analytic bias and phonological typology.
Phonology **25**. 83–127.
- Pater, Joe (2009). Weighted constraints in generative linguistics.
Cognitive Science **33**. 999-1035.
- Press, William H., Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. 1992. *Numerical recipes in C: the art of scientific computing*. Cambridge: Cambridge University Press.

- Rhodes, Richard. 1976. The morphosyntax of the Central Ojibwa verb. PhD diss, University of Michigan.
- Rhodes, Richard. 1985a. Eastern Ojibwa-Chippewa-Ottawa dictionary. Mouton.
- Rhodes, Richard (1985b). Lexicography and Ojibwa Vowel Deletion. *Canadian Journal of Linguistics* 30: 453–471.
- Steriade, Donca (2000). Paradigm uniformity and the phonetics-phonology boundary. In M. B. Broe & J. B. Pierrehumbert (Eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon* (pp. 313–334). Cambridge: Cambridge University Press.
- Steriade, Donca (2008). The phonology of perceptibility effects: the P-map and its consequences for constraint organization. In S. Inkelas & K. Hanson (Eds.), *The Nature of the Word: Studies in Honor of Paul Kiparsky* (pp. 151–180). Cambridge: MIT Press.

- Smolensky, Paul and Géraldine Legendre (2006). *The harmonic mind: From neural computation to optimality-theoretic grammar*. Cambridge: MIT Press.
- Tesar, Bruce, and Paul Smolensky. 2000. *Learnability in optimality theory*. Cambridge, MA: MIT Press.
- Wang, Marilyn D. and Robert Bilger (1973) Consonant confusions in noise. *JASA* 54:1248-66.
- White, James (2013). *Bias in phonological learning: evidence from saltation*. PhD dissertation, UCLA.
- White, James (2014) Evidence for a learning bias against saltatory phonological alternations. *Cognition*.
- White, James and Megha Sundara (2014) Biased generalization of newly learned phonological alternations by 12-month-old infants. *Cognition* 133:85-90.
- Wilson, Colin. 2006. Learning phonology with substantive bias: an experimental and computational investigation of velar palatalization. *Cognitive Science* 30:945–982.

- Wilson, Colin and Ben George. 2009. Maxent Grammar Tool.
<http://www.linguistics.ucla.edu/people/hayes/MaxentGrammarTool/>.
- Zuraw, Kie. 2007. The role of phonetic knowledge in phonological patterning: Corpus and survey evidence from Tagalog.
Language 83.277-316
- Zuraw, Kie (2013). *MAP constraints. Ms, UCLA.
www.linguistics.ucla.edu/people/zuraw/dnldpprs/star_map.pdf.