Adam Albright, UCSC
Bruce Hayes, UCLA[1]

LSA Atlanta
Jan. 4, 2003

Learning Nonlocal Environments

1. Nonlocal Environments

are cases in which the trigger of a phonological process can occur an extended distance from the target, as for example in long-distance harmony.

2. Navajo Sibilant Harmony (Sapir and Hoijer 1967)

- Example: the s-perfective prefix /sì-/ is realized as
  - [šì-] if the **first segment** of the stem is a [–anterior] sibilant ([č, č’, $č^{h}$, š, ž])
    /sì-**č**ìd/ → [**š**ì-**č**ìd] ‘he is stooping over’
  - Either [šì-] or [sì-] if **somewhere later in the stem** is a [–anterior] sibilant
    /**s**ì-té:**ž/** → [**š**ì-té:**ž**], [**s**ì-té:**ž**] ‘they two are lying’
  - [sì-] **otherwise**
    /[**s**ì-tĩ□]/ → [**s**ì-tĩ□] ‘he is lying’

3. The Learnability Problem for Nonlocal Processes

- Local environments: the number of logically possible environments for an affix allomorph is roughly proportional to the number of natural classes in the language.
- Nonlocal environments: the number of logically possible environments for an affix allomorph rises exponentially with the length of the longest string being considered.

4. Approaches to Solving the Nonlocal Learnability Problem

- UG responses: constrain the class of possible phonological environments *a priori*, so that the search space is smaller.
  - Approach 1: nonlocal processes are actually local
    - Autosegmentalism (Goldsmith 1976): All nonlocal operations are local on a tier; the set of tiers thus constrains the set of “nonlocal” environments.
    - Articulatory approach (Stampe 1979, Gafos 1999, Ní Chiosáin and Padgett 2001, and others): All nonlocal processes are local when one considers the articulatory gestures involved
  - Approach 2: nonlocal processes are nonlocal, but formally constrained
    - Relevancy condition (Jensen 1973, Odden 1994): only certain material can intervene between the target and the trigger

---

[1] Authors’ email addresses: albright@ling.ucsc.edu, bhayes@humnet.ucla.edu.

- Limited set of innate constraints (many Optimality theorists, e.g., Tesar and Smolensky 2000): learner pre-equipped with finite set of constraints like $*\begin{bmatrix}+\text{sibilant}\\ \alpha\text{anterior}\end{bmatrix}\ldots\begin{bmatrix}+\text{sibilant}\\ -\alpha\text{anterior}\end{bmatrix}$.

- UG-agnostic responses:
  - ➢ Assume that at least some processes really are non-local, and that the correct constraints are not handed to the learner by UG, but must be discovered by an inductive learning mechanism.
  - ➢ If an inductive approach succeeds, the evidence that any constraint it learns is in UG becomes weaker; if it fails without the help of assisting principles of UG, then the evidence for those principles is strengthened (Gildea and Jurafsky 1996).

5. Our General Approach to Finding Environments

as developed for the study of local environments (Albright, Andrade and Hayes 2001; Albright, in press; Albright and Hayes 2002):

I. Generalize bottom-up from the lexicon to find candidate environments.
II. Use an evaluation metric to decide which environments to keep.

- This approach is used here as well, but a more sophisticated scheme is needed when environments can be nonlocal.
- The rest of the talk addresses I, then II, applied to the problem of discovering sibilant harmony.

6. Sample Language of Application: "Pseudo-Navajo"

- A language with the pattern of sibilant harmony described in (2)
- Based on real Navajo: **whole words** from Young, Morgan and Midgette (1992), to which we prefixed either [sì-] or [šì-], following the principles of sibilant harmony.
- Ultimately, we would like to do real Navajo, but for now the lack of a morphologically parsed electronic dictionary limits us to pseudo-Navajo.

FINDING NONLOCAL ENVIRONMENTS

7. Basic Approach

- We wish to explain the difference between [šì-] stems and [sì-] stems by looking for something that all [šì-] stems (or [sì-] stems) have in common.

8. Input to the Learner

- A set of pairs:

[tà̃š], [šìtà̃š]	[gàn], [sìgàn]	[čʰò:jìn], [šìčʰò:jìn]
[tĩ[illegible], [sìtĩ[illegible]	[sí:ʔ], [sìsí:ʔ]	[bà:ʔ], [sìbà:ʔ]
[č’ìɬ], [šìč’ìɬ]	[kéšgà̃:], [šìkéšgà̃:]	etc.
[t͡ɬé:ž], [šìt͡ɬé:ž]	[k’àz], [sìk’àz]	

9. Parse Into Morphemes; Group Forms by Change

I. Prefix [sì-]	
a. [tĩɸ]	[sì-tĩɸ]
b. [bàːʔ]	[sì-bàːʔ]
c. [gàn]	[sì-gàn]
d. [síːʔ]	[sì-síːʔ]
e. [k'àz]	[sì-k'àz]

II. Prefix [šì-]	
a. [tã̀š]	[šì-tã̀š]
b. [t͡ɬéːž]	[šì-t͡ɬéːž]
c. [kéšgã̀ː]	[šì-kéšgã̀ː]
d. [čʰòːjìn]	[šì-čʰòːjìn]
e. [č'ìɬ]	[šì-č'ìɬ]

10. For Each Change, Try to Find an Environment

- Building up incrementally:
  - Step 1: Treat each learning pair as a rule, with a word-specific environment:

    a. ∅ → šì / [___ tã̀š]
    b. ∅ → šì / [___ t͡ɬéːž]
    c. ∅ → šì / [___ kéšgã̀ː] etc.

  - Step 2: Compare pairs of rules which both take [šì-] (or both take [sì-]), and extract what their environments have in common, to form a generalized rule.
  - Step 3: Iterate the process, so that ever more general rules get discovered.

11. Comparing Pairs of Rules

Starting with two word-specific rules:

∅ → šì / [___ tã̀š]
∅ → šì / [___ t͡ɬéːž]

we collapse them together (details below), using features:

	∅	→	šì	/	[	___	t	ã̀	š	]
+	∅	→	šì	/	[	___	t͡ɬ	éː	ž	]
=	∅	→	šì	/	[	___	$\begin{bmatrix}\text{–sonorant}\\ \text{–continuant}\\ \text{–spread gl.}\\ \text{+anterior}\end{bmatrix}$	$\begin{bmatrix}\text{+syllabic}\\ \text{–high}\\ \text{–round}\end{bmatrix}$	$\begin{bmatrix}\text{–sonorant}\\ \text{+continuant}\\ \text{–anterior}\\ \text{+strident}\end{bmatrix}$	]

- This particular rule looks unpromising—but with further generalization, the same process arrives quickly at the right answer (below).
- But what should be collapsed with what? [tã̀š], [t͡ɬéːž] seems obvious, but what of (say) [čʰòːjìn], [č'ìɬ]?
- To find the crucial triggering elements, we use **similarity-based alignment**.

12. Similarity-Based Alignment

- Here is an intuitively good alignment:

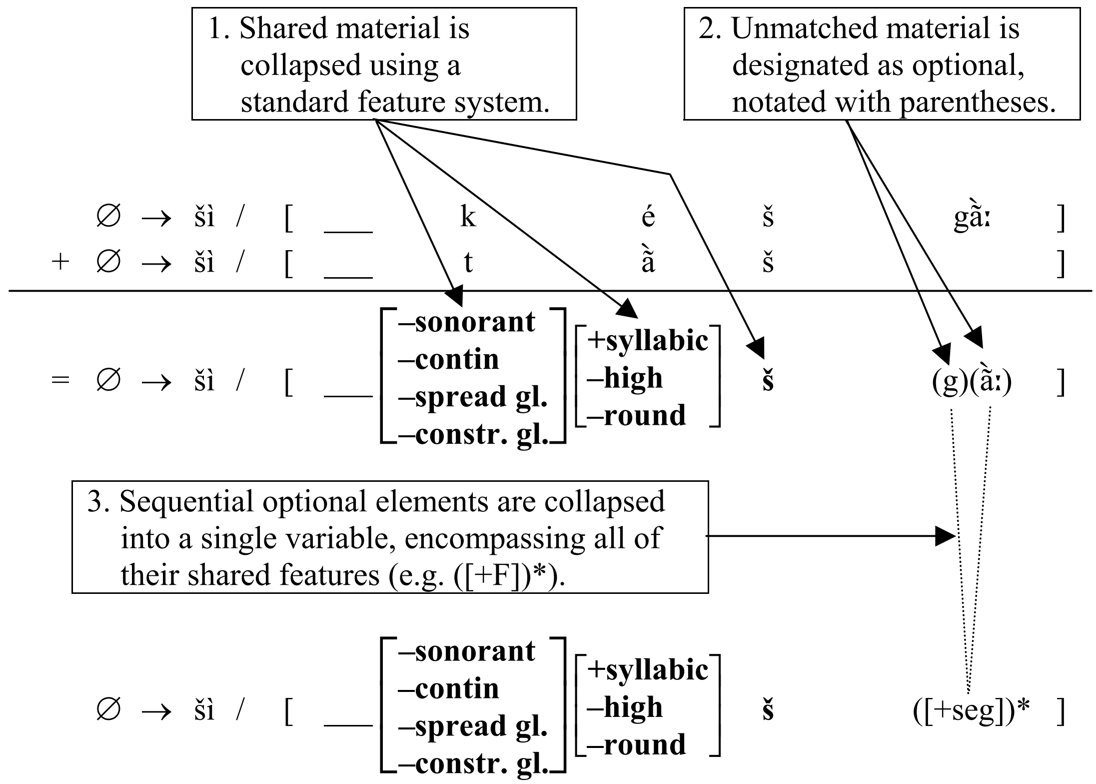
$\check{c}^h$	òː	j	ì	n
|			|	|
č’			ì	ɬ

- Good alignments have two properties:
  - They match phonetically-similar segments.
  - They avoid leaving too many segments unpaired.

- We use existing methods to find the optimal alignment:
  - The theory of phonetic similarity from Frisch, Broe and Pierrehumbert (1997) to match the most similar segments with each other.
  - A cost-minimizing search of all possible alignments (minimum string edit distance; Kruskal 1983)

13. Rule Generalization By Collapsing Aligned Pairs of Forms

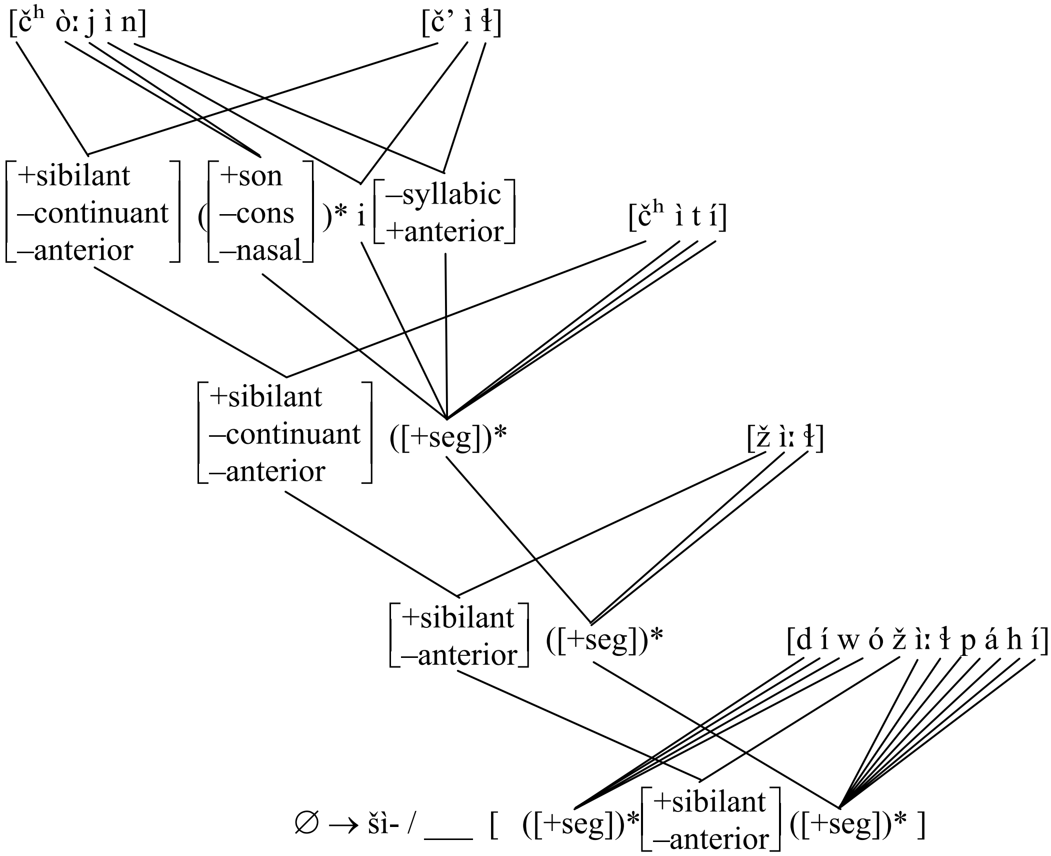
- Align the forms optimally as described above, and collapse.
- Three rules of generalization:



- When collapsing across two segments that have no features in common, we use the SPE notation [+seg].
- Iterate by generalizing with the other words in the training data.
- Periodically trim back the hypothesis set, keeping only those rules that perform best.[2]
- Learning terminates when no new "keeper" rules are found.

[2] Specifically: (a) for each word in the training set, keep the most reliable rule (in the sense of Albright and Hayes 2002) that derives it; (b) for each change, keep the rule that derives more forms than any other.

14. Finding the Environment for Nonlocal Sibilant Harmony By Iterative Generalization



VERIFYING THE APPROACH: A SIMULATION

15. Training Set

- 200 whole Navajo words, taken at random from Young, Morgan, and Midgette (1992). [3]
- Prefixes were attached following the rules of (2):
  - [šì-] if the "stem" began with a nonanterior sibilant (24 stems).
  - Two copies of the stem, one with [šì-], one with [sì-], if the stem contained but did not begin with a nonanterior sibilant (34 stems, 2 copies each)
  - [sì-] otherwise (142 stems)

[3] We repeated the learning process on nine other sets of 200 forms, obtaining similar results each time.

16. The Correct Environments are Learned (among others)

$\varnothing \rightarrow$ [šì-] / [___ $\begin{bmatrix}\text{+sibilant}\\ \text{–anterior}\end{bmatrix}$ ([+seg])*]	Environment for obligatory local harmony.
$\varnothing \rightarrow$ [šì-] / [___ ([+seg])* $\begin{bmatrix}\text{+sibilant}\\ \text{–anterior}\end{bmatrix}$ ([+seg])*]	Environment for optional harmony, when the stem includes a nonanterior sibilant that is not initial
$\varnothing \rightarrow$ [sì-] / [___ ([+seg])*]	Context free environment, takes [sì-] by default—below, we show how to limit this case to instances where the [šì-] environments just given are not met.
+ 88 others, discussed below.	

SELECTING THE CORRECT RULES FROM THE LEARNED SET

17. Some Potentially Harmful Hypotheses

- Among the 88 other generalizations, many hold exceptionlessly true of the learning data, entirely by accident.
- For example, consider the following environment for [sì-]:

  / [___ ([–nasal])* $\begin{bmatrix}\text{–cont}\\ \text{–syllabic}\\ \text{+anterior}\end{bmatrix}$ ($\begin{bmatrix}\text{–nasal}\\ \text{–high}\end{bmatrix}$)* $\begin{bmatrix}\text{–sonorant}\\ \text{–round}\end{bmatrix}$ ($\begin{bmatrix}\text{+sonorant}\\ \text{–round}\end{bmatrix}$)*]

  This works for 40/200 forms; there are no counterexamples in the training data.
- This environment would have catastrophic effects if taken seriously. For example, if the speaker later encountered the form /šátàt/ (not in the training set), *[sì-šátàt] would be derived.

  [sì-] / ___ ([–nasal])* $\begin{bmatrix}\text{–cont}\\ \text{–syllabic}\\ \text{+anterior}\end{bmatrix}$ ($\begin{bmatrix}\text{–nasal}\\ \text{–high}\end{bmatrix}$)* $\begin{bmatrix}\text{–sonorant}\\ \text{–round}\end{bmatrix}$ ($\begin{bmatrix}\text{+sonorant}\\ \text{–round}\end{bmatrix}$)*

  š á   t   à   t   ∅

- In order to discard such accidentally true hypotheses, we need an **evaluation metric** (Chomsky and Halle 1968).

18. Evaluating Environments by Constraint Ranking

- Various scholars (e.g. Boersma 1998, Russell 1999, Burzio 2002) propose to treat morphological mappings as Optimality-theoretic constraints.

- Rules are trivially restated as constraints; e.g.

$$\varnothing \rightarrow [\check{s}\grave{\imath}\text{-}]\ /\ [\ \_\_\_ \begin{bmatrix}+\text{sibilant}\\ -\text{anterior}\end{bmatrix} ([+\text{seg}])^*\ ]_{[+\text{s-perfective}]}$$

is restated as:

$$\text{“U{\small SE} } [\check{s}\grave{\imath}\text{-}]\ /\ [\ \_\_\_ \begin{bmatrix}+\text{sibilant}\\ -\text{anterior}\end{bmatrix} ([+\text{seg}])^*\ ] \text{ to form the s-perfective”}$$

- This constraint is violated by forms that begin with a $\begin{bmatrix}+\text{sibilant}\\ -\text{anterior}\end{bmatrix}$ segment, but use something other than [šì-] to form the s-perfective. For example:

Morphological Base	Candidates that obey USE [šì-] / ___ $\begin{bmatrix}+\text{sib}\\ -\text{ant}\end{bmatrix}$	Candidates that violate USE [šì-] / ___ $\begin{bmatrix}+\text{sib}\\ -\text{ant}\end{bmatrix}$
[šáp]	[šì-šáp]	*[sì-šáp], *[mù-šáp], etc.
[táp]	all	none

19. Using the Gradual Learning Algorithm as an Evaluation Metric

Goal: rank bad constraints like (17) so low that they never affect the outcome.

Method: Provide the constraints with initial ranking values, and submit to the Gradual Learning Algorithm (GLA; Boersma 1997, Boersma and Hayes 2001) to establish the correct ranking.

20. Initial Rankings Based On Generality (Boersma 1998)

- Junk constraints are true, but non-general.
  - ➢ Thus, they start low—and they stay there. The GLA is error-driven, and the errors that would promote the junk constraints are already averted by more general constraints.
- Good constraints with specific contexts, like "USE [šì-] / ___ $\begin{bmatrix}+\text{sib}\\ -\text{ant}\end{bmatrix}$", are also nongeneral—but appropriately so.
  - ➢ They start low, but they are needed to avert errors like *[sì-šáp], so they are promoted by the GLA to the top of the grammar.

21. A Numerical Characterization of Generality

$$= \frac{\text{number of forms that a constraint applies to}}{\text{total number of forms exhibiting the change that the constraint requires}}$$

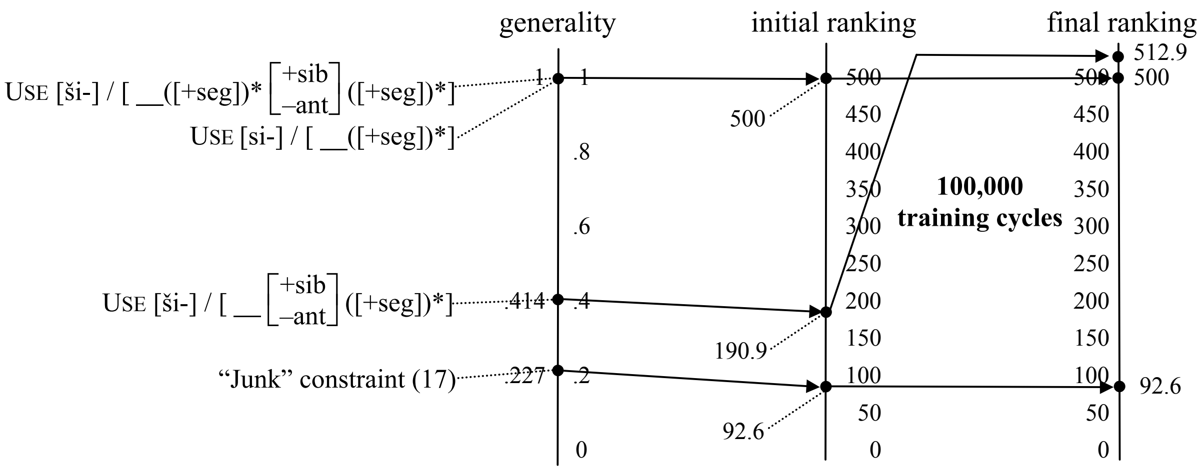
22. Calculating Generality in the 200-Word Navajo Simulation

Constraint	(a) Relevant forms	(b) Forms with this change	(c)Generality = (a)/(b)
USE [šì-] / [___ $\begin{bmatrix}+\text{sibilant}\\-\text{anterior}\end{bmatrix}$ ([+seg])*]	24	58 [šì-] forms	.414
USE [šì-] / [___ ([+seg])* $\begin{bmatrix}+\text{sibilant}\\-\text{anterior}\end{bmatrix}$ ([+seg])*]	58		1
USE [sì-] / [___ ([+seg])*]	176	176 [sì-] forms	1
Constraint (17) (“junk” constraint)	40		.227

23. Making Sure Generality Will Make a Difference

- Rescale generality so that the original generality range is converted to a very large range (0-500) on the GLA ranking scale.

24. How Ranking Proceeded



- The valid contextual constraint climbed to the top of the grammar—it was essential in stamping out learning errors like *[sì-žéː].
- The junk constraint never rose, because it was never needed to explain anything.

25. Final Ranking Obtained (Hasse Diagram)

USE [ši-] / [___ $\begin{bmatrix}\text{+sib}\\ \text{–ant}\end{bmatrix}$ ([+seg])*]
512.9

Undominated local harmony

↓

USE [sì-] / [___ ([+seg])*] ⟷ USE [ši-] / [___ ([+seg])* $\begin{bmatrix}\text{+sib}\\ \text{–ant}\end{bmatrix}$ ([+seg])*]
500.0 500.0

Free variation for non-local harmony

↓

USE [sì-] / [___ ([–nasal])* $\begin{bmatrix}\text{–cont}\\ \text{–syllabic}\\ \text{+anterior}\end{bmatrix}$ ($\begin{bmatrix}\text{–nasal}\\ \text{–high}\end{bmatrix}$)* $\begin{bmatrix}\text{–sonorant}\\ \text{–round}\end{bmatrix}$ ($\begin{bmatrix}\text{+sonorant}\\ \text{–round}\end{bmatrix}$)*]
92.6

Potentially harmful constraints like (17) safely outranked

26. End Result

- A grammar of inductively learned constraints, ranked stochastically in a way that correctly derives the pattern of Navajo sibilant harmony seen in (2)

DISCUSSION

27. Summary of the Model

- *Find* candidate environments by working upward from the training data.
  - Tools: similarity-based alignment, iterative generalization, selective retention of best hypotheses
- *Evaluate* environments by
  - recasting them as constraints
  - ranking the constraints with the Gradual Learning Algorithm
  - using initial ranking values based on generality

28. Why the Model is of Interest

- It learns nontrivial constraints inductively, isolating the correct environments from a very large search space; opens up the possibility that not all such constraints need be assumed to be innate
- It provides an inductive baseline (see below) for further exploration in phonological learning.

29. The Issue of UG

- Although this model does not incorporate an innate constraint for sibilant harmony, it *does* incorporate many hypothesized principles of Universal Grammar, among them:
    - feature system
    - generalization based on natural classes
    - theory of phonetic similarity
    - GLA
    - preference for generality

30. What is Needed to Make this Model Scale Up to Harder Cases?

- Examples:
    - processes that distinguish opaque from transparent intervening segments
    - processes that count interveners: in Hungarian vowel harmony, two neutral vowels are more opaque than one; and three more opaque than two
- Our model may well require considerable help from further principles of UG to handle these. Some likely candidates:
    - tiers (Goldsmith 1976) or gestures (Gafos 1999; Ní Chiosáin and Padgett 2001)
    - principles specifying possible interveners (Jensen 1973, Odden 1994)
    - any means of singling out from the string just the vowels (Vergnaud and Halle 1979, Archangeli and Pulleyblank 1987, Clements 1991)
- As noted above, when proposed principles of UG make learning possible in cases where a more impoverished inductive system fails, they are empirically supported.

31. Explaining the Typology of Nonlocal Processes

- Nothing in our simulation depended on the affix allomorphs including pairs of [+anterior] and [–anterior] sibilants. If the affix allomorphs had been [ka-] (before nonanterior sibilants) and [ga-] (elsewhere), the model would have behaved identically.
- Research indicates that long-distance segmental processes are overwhelmingly (though not exclusively) assimilations and dissimilations (Jensen 1973, Hansson 2001, Rose and Walker 2001).
- Can a learning model help explain the typology of phonological processes?
    - One possibility is that our model should be equipped with *learning biases*, which will explain why some processes are learned more easily or reliably than others (cf. Wilson, in progress), and thus indirectly account for phonological typology.

32. Conclusion: The Role of an Inductive-Baseline Model

- The model offers the possibility of simplifying the theory of UG by learning constraints previously hypothesized to be innate.
- The model makes it possible to test proposed principles of UG, by determining whether they are essential to learning or to explaining phonological typology.

References

Albright, Adam (in press) Phonological subregularities in inflectional classes: Evidence from Italian. To appear in *Language*. http://www.linguistics.ucla.edu/people/grads/aalbrigh/

Albright, Adam and Bruce Hayes (2002) Modeling English Past Tense Intuitions with Minimal Generalization. In Maxwell, Michael (ed) Proceedings of the Sixth Meeting of the ACL Special Interest Group in Computational Phonology. Philadelphia, July 2002. ACL.

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. http://www.linguistics.ucla.edu/people/hayes/Segenvspandiph/

Archangeli, Diana and Douglas Pulleyblank (1987) Maximal and minimal rules: effects of tier scansion. *NELS* 17, pp. 16-35.

Boersma, Paul (1997) "How we learn variation, optionality, and probability," *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam* 21:43–58. http://www.fon.hum.uva.nl/paul/

Boersma, Paul (1998) "Typology and acquisition in functional and arbitrary phonology," http://www.fon.hum.uva.nl/paul/. To appear in René Kager, Joe Pater, and Wim Zonneveld, eds., *Fixing Priorities: Constraints in Phonological Acquisition.* Cambridge: Cambridge University Press.

Boersma, Paul and Bruce Hayes (2001) "Empirical Tests of the Gradual Learning Algorithm," *Linguistic Inquiry* 32: 45-86. [http://www.linguistics.ucla.edu/people/hayes/GLA/]

Burzio, Luigi (2002) Missing players: Phonology and the past tense debate. *Lingua*, 112, 157-199.

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

Frisch, S., Broe, M., & Pierrehumbert, J. (1997). *Similarity and phonotactics in Arabic*. http://roa.rutgers.edu/files/223-1097/roa-223-frisch-2.pdf

Gafos, Adamantios I. (1999) *The Articulatory Basis of Locality in Phonology*. New York: Garland.

Gildea, Daniel and Daniel Jurafsky (1996) Learning bias and phonological-rule induction. *Computational Linguistics*, 22, 497–530.

Goldsmith, John (1976) *Autosegmental Phonology*. Bloomington : Indiana University Linguistics Club.

Hansson, Gunnar (2001) *Theoretical and typological issues in consonant harmony*. Ph.D. dissertation, University of California, Berkeley.

Jensen, John T. (1973) A constraint on variables in phonology. *Language* 50:675-686.

Kruskal, Joseph B. (1983) An overview of sequence comparison. In D. Sankoff and J. B. Kruskal, eds., *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison*. Reading, MA: Addison-Wesley.

Odden, David (1994) Adjacency parameters in phonology. *Language* 70: 289-330

Padgett, Jaye, and Máire Ní Chiosáin (2001) Markedness, segment realization, and locality in spreading. In Linda Lombardi, ed., *Constraints and representations: segmental phonology in Optimality Theory*. Cambridge: Cambridge University Press, 118-156.

Rose, Sharon and Rachel Walker (2001) "A Typology of Consonant Agreement as Correspondence," ROA 458-0801.

Russell, Kevin (1999) MOT: Sketch of an OT approach to morphology. Rutgers Optimality Archive 352. [http://roa.rutgers.edu/view.php3?roa=352]

Sapir, Edward, and Harry Hoijer (1967) *The Phonology and Morphology of the Navajo Language*. Berkeley and Los Angeles: University of California Press.

Stampe, David (1979) *A Dissertation on Natural Phonology*. Revised Ph.D. dissertation, University of Chicago. Distributed by the Indiana University Linguistics Club, Bloomington.

Tesar, Bruce & Paul Smolensky (2000) *Learnability in Optimality Theory*. Cambridge, MA: MIT Press.

Vergnaud, Jean-Roger and Morris Halle (1979) Metrical structures in phonology. Ms., MIT.

Young, Robert W., William Morgan Sr., and Sally Midgette (1992) *Analytical Lexicon of Navajo*. Albuquerque: University of New Mexico Press.