



Error Detection and Alternation Subsets

Bruce Tesar

Linguistics Dept. / Center for Cognitive Science

Rutgers University, New Brunswick

NECPhon 3, MIT. October 24, 2009.

Outline

- Error Detection with Output-Driven Maps
- Alternation Subsets
- Estimating Restrictiveness using Output-Driven Maps

Error-Driven Learning

- Determine if the learner's current hypothesis (ranking + lexicon) correctly generates an observed word.
 - If not, attempt to modify the hypothesis.
- If the learner's current hypothesis consists of fully set underlying forms, this is straight-forward.
 - Construct the input from the underlying forms.
 - Map the input to the surface realization using the ranking.

Learning with Unset Features

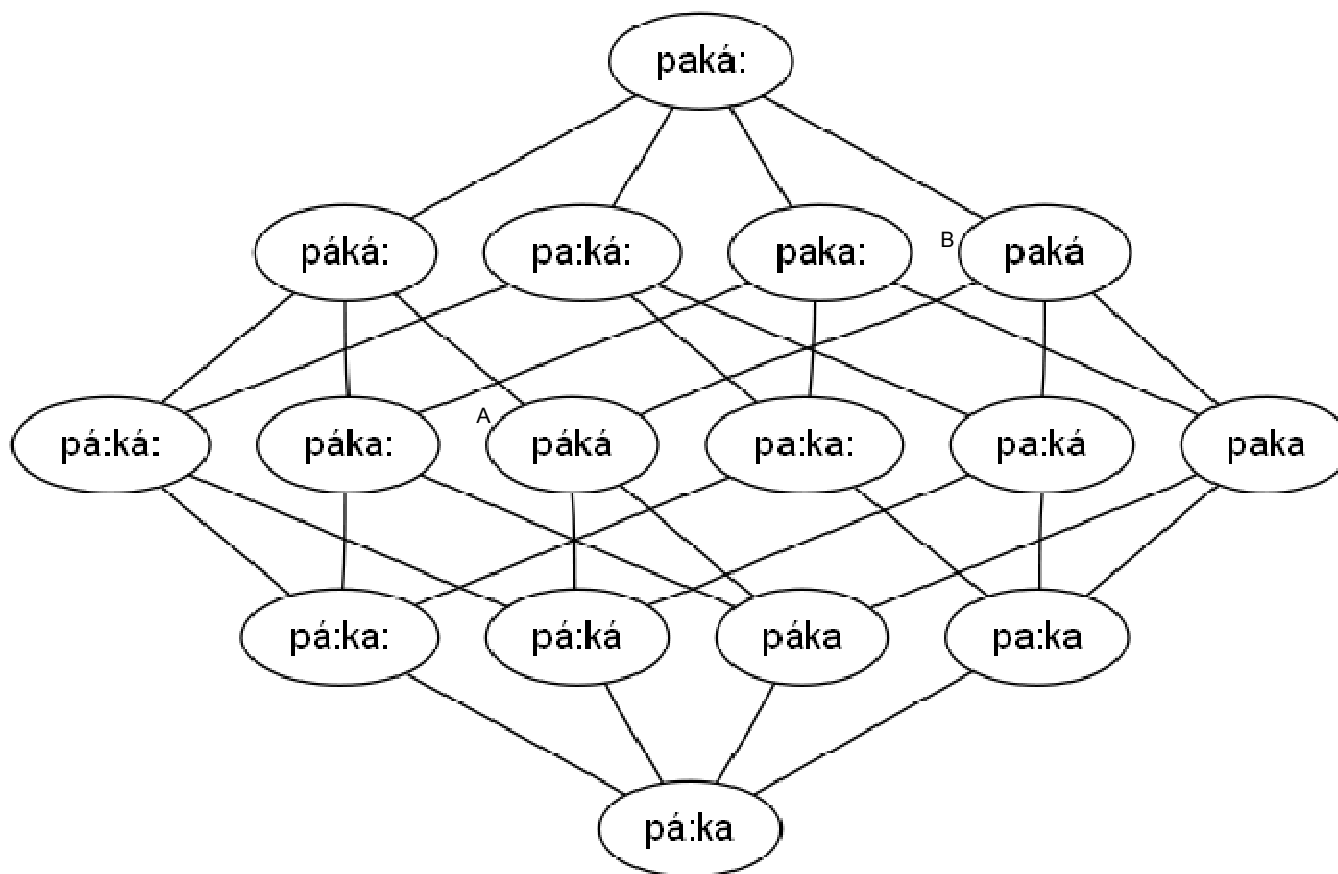
- Set features only when necessary (Tesar 2006, Merchant 2008).
 - Underlying forms have unset features during learning.
 - Some features may never be set.
- How do you determine if a word is generated by a hypothesis? Three possibilities:
 - Define an interpretation of inputs with unset features.
 - Separately try all possible combinations of values for unset features.
 - Exploit output-driven maps.

Output-Driven Maps

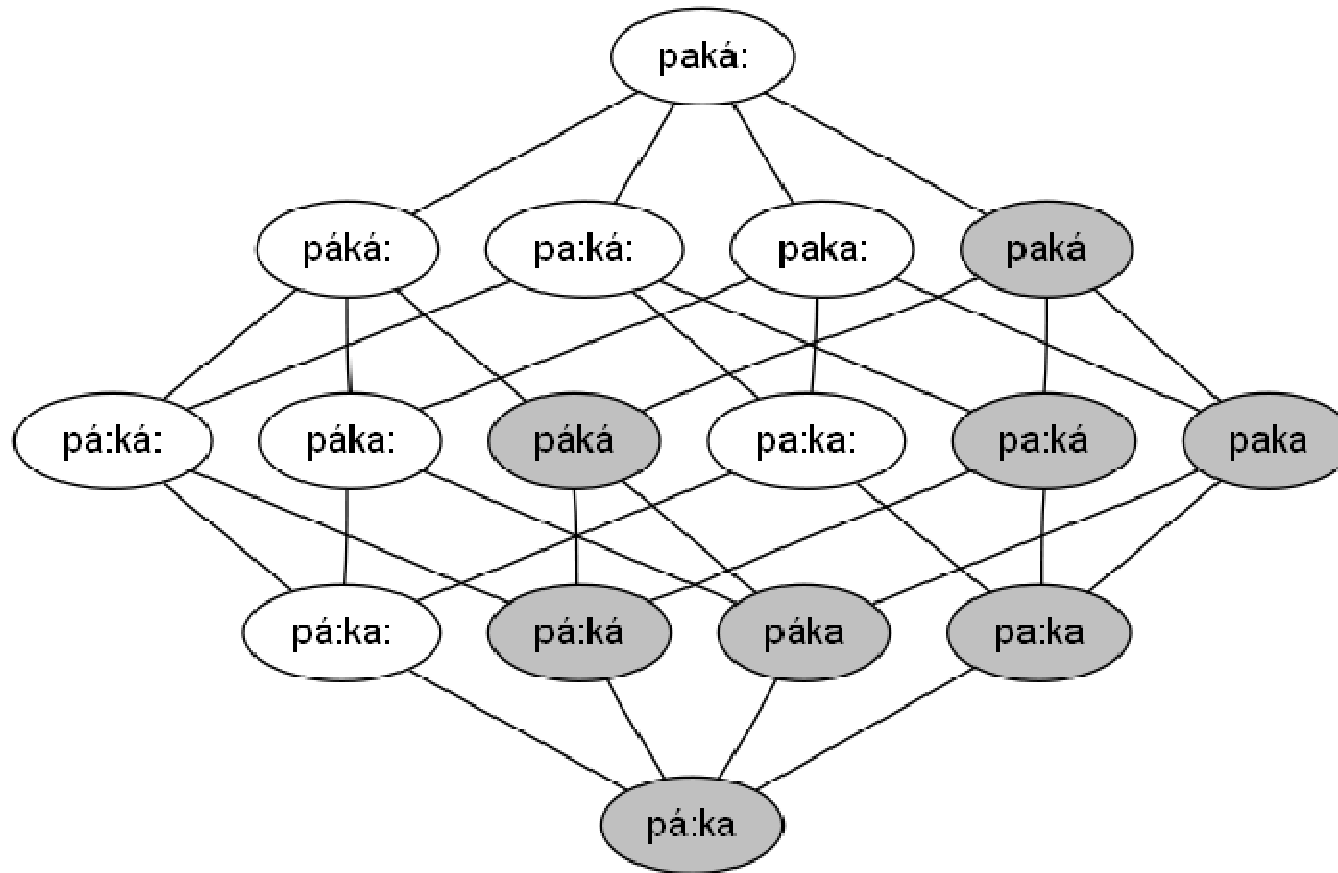
(Tesar 2008)

- A map is output-driven if:
 - for every grammatical candidate $A \rightarrow X$ of the map:
 - if candidate $B \rightarrow X$ (same output) has greater similarity than $A \rightarrow X$,
 - then $B \rightarrow X$ is also grammatical.
- Simplified:
 - for every grammatical candidate $A \rightarrow X$ of the map:
 - if input B is more similar to X than A is,
 - then B also maps to X .

Relative Similarity (up = greater similarity)



Input space with suffix already set to +long



The Least Similar Input

- The unset features of an input define the space of possible actual inputs for a word.
- If all inputs in the space currently map to the correct output, there is nothing more to be learned from that word (at that time).
- ODM: if the least similar input maps correctly, then all of them do.
 - Only one input need be tested.
- ODM: if it doesn't map correctly, test each unset feature (Tesar to appear).
 - Minimal disparity inputs, one per unset feature.

Input Subspace Evaluation

- For error detection purposes, we can equate a lexical subspace (sublattice) with its least similar member.
 - The bottom of the sublattice.
 - If the current ranking maps it correctly, it maps every member of the subspace correctly.
- This can be exploited elsewhere in learning as well.
 - Contending with alternation subsets.

A System for Illustration

- Words: root + suffix
 - Both roots and suffixes are monosyllabic.
- Each vowel has two features:
 - Vowel length: long (+) or short (–)
 - Main stress: stressed (+) or unstressed (–)
- Example surface words:
 - *páka pá:ka paká páka: pa:ká: pa:ká*
 - Each word has two morphemes
 - Each word has exactly one main stress in the output.

The Constraints

- Six Constraints

MainLeft	main stress on the initial syllable
MainRight	main stress on the final syllable
*V:	no long vowels
WSP	long vowels are stressed
FaithStress	correspondents have equal stress value
FaithLength	correspondents have equal length value

(McCarthy & Prince 1993, 1995; Prince 1990; Rosenthal 1994)

L8 (subset)

r1=/pa/	r2=/pa:/	r3=/pá/	r4=/pá:/	
<i>pa</i> ká	<i>pá</i> :ka	<i>pa</i> ká	<i>pá</i> :ka	s1=/-ka/
<i>pa</i> ká:	<i>pa</i> ká:	<i>pa</i> ká:	<i>pa</i> ká:	s2=/-ka:/
<i>pa</i> ká	<i>pá</i> :ka	<i>pa</i> ká	<i>pá</i> :ka	s3=/-ká/
<i>pa</i> ká:	<i>pa</i> ká:	<i>pa</i> ká:	<i>pa</i> ká:	s4=/-ká:/

Ranking: WSP \gg FL \gg {*V:, MR} \gg ML \gg FS

Stress attracted to length (default final), long vowels shorten in unstressed position.

Neutralized: r1/r3, r2/r4, s1/s3, s2/s4

Phonotactic Inventory: *pa*ká *pa*ká: *pá*:ka

L8 compressed

r1=/pa/	r2=/pa:/	
<i>paká</i>	<i>pá:ka</i>	s1=/-ka/
<i>paká:</i>	<i>paká:</i>	s2=/-ka:/

Ranking: WSP \gg FL \gg $\{^*V:, MR\}$ \gg ML \gg FS

r1 / ? - / r2 / ? + / s1 / ? - / s2 / ? + /

Phonotactic Inventory: *paká* *paká:* *pá:ka*

Learning L8: Phonotactics and Single-Form

Word	Input	win ~ lose	WSP	FL	*V:	MR	ML	FS
r2-s2	/pa:-ká:/	paká: ~ pa:ká:	W	L	W			
r1-s2	/pa-ká:/	paká: ~ paká		W	L			
r2-s1	/pá:-ka/	pá:ka ~ paká		W	L	L	W	W
r1-s1	/pa-kál/	paká ~ páka				W	L	W

r1 / ? **?** / r2 / ? + / s1 / ? - / s2 / ? + /

Unable to set the length feature for r1.

(contrast pairs won't help)

L7 (superset)

<i>/pa/</i>	<i>/pa:/</i>	<i>/pá/</i>	<i>/pá:/</i>	
<i>paká</i>	<i>pá:ka</i>	<i>páka</i>	<i>pá:ka</i>	<i>/-ka/</i>
<i>paká:</i>	<i>paká:</i>	<i>páka</i>	<i>pá:ka</i>	<i>/-ka:/</i>
<i>paká</i>	<i>paká</i>	<i>paká</i>	<i>pá:ka</i>	<i>/-ká/</i>
<i>paká:</i>	<i>paká:</i>	<i>paká:</i>	<i>paká:</i>	<i>/-ká:/</i>

Ranking: WSP \gg FS \gg FL \gg $\{^*V:, MR\}$ \gg ML

Lexical stress (default final), long vowels shorten in unstressed position.

Neutralized: <none>

Phonotactic Inventory: *paká paká: pá:ka páka*

L8 a Phonotactic Subset of L7

L8 Phonotactic Inventory: *pa**ká* *pa**ká:* *pá:**ka*

L7 Phonotactic Inventory: *pa**ká* *pa**ká:* *pá:**ka* *páka*

L8: WSP \gg FL \gg $\{^*V:, MR\}$ \gg ML \gg FS

L7: WSP \gg FS \gg FL \gg $\{^*V:, MR\}$ \gg ML

What about alternations?

L8 an Alternation Subset of L7

<i>/pa/</i>	<i>/pa:/</i>	<i>/pá/</i>	<i>/pá:/</i>	
<i>paká</i>	<i>pá:ka</i>	<i>páka</i>	<i>pá:ka</i>	<i>/-ka/</i>
<i>paká:</i>	<i>paká:</i>	<i>páka</i>	<i>pá:ka</i>	<i>/-ka:/</i>
<i>paká</i>	<i>paká</i>	<i>paká</i>	<i>pá:ka</i>	<i>/-ká/</i>
<i>paká:</i>	<i>paká:</i>	<i>paká:</i>	<i>paká:</i>	<i>/-ká:/</i>

Same alternation pattern as L8.

Root length contrast replaced with a stress contrast.

L8 roots: */pa/* */pa:/* (both roots stress irrelevant)

L7 roots: */pa:/* */pá:/* (first root length irrelevant)

Alternation Subsets

- The choice cannot be expressed solely via a more restrictive ranking; the underlying forms differ.
 - L8: r1 has uf space / ?stress, –long /
 - L7: r1 has uf space / –stress, ?long /
- The learner needs to set at least one feature for r1.
- Each value is consistent with some ranking.
- The learner needs to choose the underlying form consistent with the most restrictive ranking (L8).
 - Needs to set r1 to –long.

r1's Length Feature

- L8 requires r1 to be –long.
- L7 allows r1 to be +long, consistent with the given data.
- Inconsistency detection won't set r1's length feature.
- Restrictiveness considerations clearly favor L8.
 - How can the learner reach this conclusion?

Restrictiveness in the Lexicon

- In more restrictive languages, the same outputs are mapped onto by more inputs.
- Jarosz (2006) used this in phonotactic learning.
 - Maximize # inputs mapped to each observed output.
 - Equivalent to maximum likelihood with a uniform prior dist.
- Recast in feature setting terms:
 - Select the hypothesis in which fewer features are set.
 - This means a larger equivalence class of underlying forms.
- Here we use it to learn non-phonotactic underlying forms.

Which Form Needs Fixing?

- Recall our (partially) learned lexicon for L8.

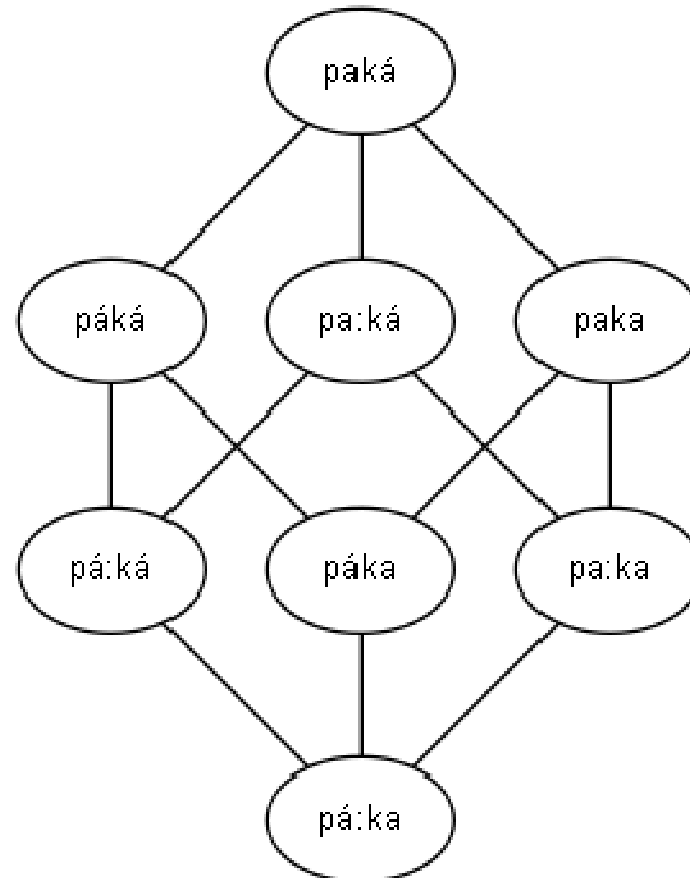
r1 / ? ? / r2 / ? + / s1 / ? - / s2 / ? + /

- Every morpheme has at least one unset feature.
 - Which word needs to have additional features set?
- Error Detection.
 - An error indicates more learning is required.
 - For each word, test the input with the most disparities.

Error Detected on r1s1

- An error is detected on word r1s1.
 - */pá:ka/* incorrectly maps to [pá:ka], instead of [paká].
- The other three words have no detected error.
 - the greatest disparity input maps to the correct output.
- Error detection focuses the learner on r1s1.

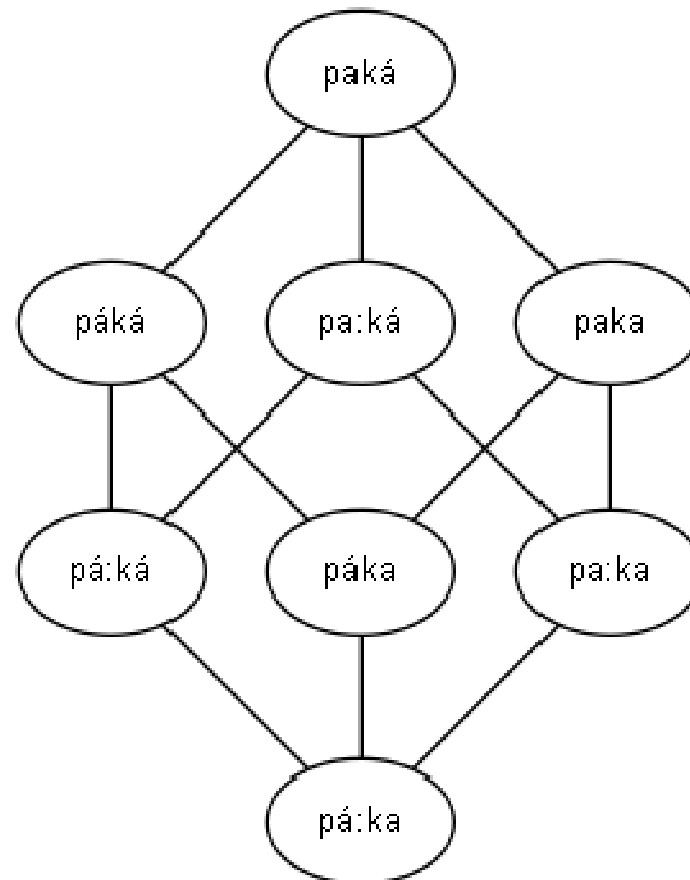
Effective Lexical Space for r1s1



Setting Features for r1s1

- Three unset features: $r1 / ? ? /$ $s1 / ? - /$
- Setting r1 to $-long$ works with the current hierarchy.
 - /páka/ → [paká] with the current ranking.
 - By ODM, four inputs for [paká]: /paka/ /páka/ /paká/ /páká/
- r1 $+long$ (unset) only works if r1 is set to $-stress$, s1 to $+stress$.
 - /pa:ká/ → [paká] with the current ranking.
 - By ODM, two inputs for [paká]: /pa:ká/ /paká/
- Restrictiveness favors setting r1 to $-long$.
 - assigns a larger portion of the input space to the observed [paká].

Larger Share of the Input Space



Summary

- Alternation subsets cannot be solved via ranking biases alone when differing analyses have conflicting underlying forms.
- An entire input sublattice can be evaluated using the bottom element, under output-driven maps.
 - Error detection for UF feature setting can be done by evaluating only one input form, the least similar one available.
 - The consistency of different partial lexical hypotheses can be evaluated via the bottom of the associated input sublattice.
- The relative restrictiveness implications of competing underlying forms may be estimated by comparing the sizes of the associated input sublattices.

References

- Jarosz, Gaja. 2006. Rich Lexicons and Restrictive Grammars - Maximum Likelihood Learning in Optimality Theory. PhD. dissertation, The Johns Hopkins University, Baltimore, MD. ROA-884.
- McCarthy, John J., and Prince, Alan. 1993. Generalized alignment. In *Yearbook of Morphology*, ed. by Geert Booij and Jaap Van Marle, 79-154. Dordrecht: Kluwer.
- McCarthy, John J., and Prince, Alan. 1995. Faithfulness and Reduplicative Identity. In *University of Massachusetts Occasional Papers 18: Papers in Optimality Theory*, ed. by Jill Beckman, Laura Walsh Dickey, and Suzanne Urbanczyk, 249-384. Amherst, MA: GLSA, University of Massachusetts.
- Merchant, Nazarré. 2008. Discovering underlying forms: Contrast pairs and ranking. PhD. dissertation, Rutgers University, New Brunswick. ROA-964.
- Merchant, Nazarré, and Tesar, Bruce. 2008. Learning underlying forms by searching restricted lexical subspaces. *CLS 41 (2005), vol. II: The Panels*, 33-47. ROA-811.

Prince, Alan. 1990. Quantitative consequences of rhythmic organization. *CLS 26 vol. II: Papers from the Parasession on the Syllable in Phonetics and Phonology*, 355-398.

Prince, Alan, and Tesar, Bruce. 2004. Learning phonotactic distributions. In *Constraints in Phonological Acquisition*, eds. René Kager, Joe Pater and Wim Zonneveld, 245-291. Cambridge: Cambridge University Press.

Rosenthal, Sam. 1994. Vowel/glide alternation in a theory of constraint interaction. PhD. dissertation, University of Massachusetts, Amherst.

Tesar, Bruce. 2006. Learning from paradigmatic information. *NELS 36*, 619-638. GLSA. ROA-795.

Tesar, Bruce. 2008. Output-driven maps. Ms., Linguistics Dept., Rutgers University. ROA-956.

Tesar, Bruce. to appear. Learning Phonological Grammars for Output-Driven Maps. *NELS 39*. ROA-1013.