

Learning Latent Representations with Progressive Hypothesis Space Expansion

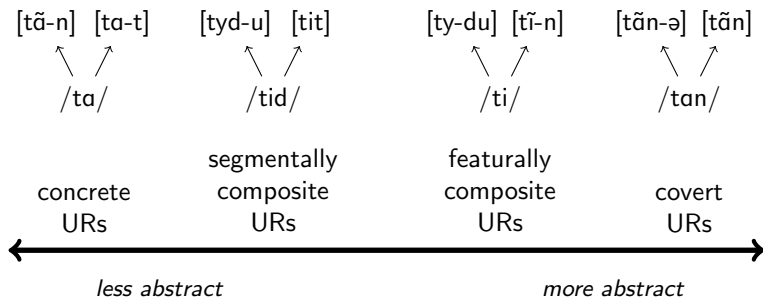
Jonathan C. Paramore
jcparamo@ucsc.edu

University of California, Santa Cruz
July 3rd, 2026



Abstraction in Phonology

- Debate in phonology: **How abstract can Underlying Representations (URs) be?** (Kenstowicz and Kisseberth, 1977; Kiparsky, 1973)



- Covert URs are analytically necessary (e.g., Punjabi vowel nasality)

Vowel Nasality in Punjabi

- ▶ Vowel nasality is contrastive (Bashir and Connors, 2019)

[sã:] ‘I was’ vs. [sa:] ‘breath’

- ▶ Contrast neutralized before nasal consonants (Zahid and Hussain, 2012)

[sã:ŋ] ‘grindstone’ vs. *[sa:ŋ]

- ▶ pre-N vowels are phonetically identical to contrastive nasal vowels in terms of nasality (Paramore and Bennett, 2025)

Nasal Harmony in Punjabi

- ▶ pre-N vowels and contrastive nasal vowels behave differently in regard to nasal harmony

- ▶ Contrastive nasal vowels trigger nasal harmony (Bhatia, 1993)

/sa:-u[̃]ɑ:/ → [s[̃]ɑ:u[̃]ɑ:] ‘breath-PL’

/mætʃi:-j[̃]ɑ:/ → [mætʃi[̃]j[̃]ɑ:] ‘fish-PL’

- ▶ pre-N vowels **do not** trigger nasal harmony (Paramore and Bennett, 2025)

/ɑ:va:m/ → [ɑ:va:m] ‘public’

/si:ja:ŋ/ → [si:ja:ŋ] ‘recognition’

- ▶ Simplest analysis treats pre-N vowels as **covert**: they are underlyingly oral but always surface as nasal
 - only underlying nasal vowels trigger harmony

Abstraction & Learnability

- ▶ Phonological patterns like Punjabi vowel nasality motivate highly abstract covert URs
- ▶ But are Covert URs **learnable**?

Abstraction raises two general learnability problems (Jarosz, 2019):

- 1. Learners should prefer less abstract URs whenever possible.**
 - How is this preference enforced?
- 2. Permitting covert URs makes the hypothesis space computationally intractable.**
 - Computationally infeasible to search an infinite space for the optimal UR

Covert URs should not be permitted because they result in learning failures (Wang and Hayes, 2025)

UR Progressive Hypothesis Space Expansion Learner

UR PHaSE Learner solves both learnability issues by searching through the hypothesis space progressively

- ▶ Considers URs one disparity level at a time
- ▶ Stops as soon as it finds a satisfactory explanation of the data

(Simon, 1955, 1956)

UR PHaSE Learner

Four iterative learning steps

Step I: phonotactic learning

Step II: UR Expectation Maximization learning

- Determines optimal grammar
- Using a finite set of UR candidates

Step III: Likelihood Threshold Evaluation

- Evaluates whether observed data is sufficiently likely under the current grammar
- Learning successful if likelihood of SRs $> .95$

Step IV: UR Candidate Expansion

- By “1 level” of abstraction

Case Study: Punjabi Vowel Nasality

Input to the Learner

- ▶ Small Punjabi lexicon (30 forms) that demonstrates the nasality patterns, e.g.,:

i.	[sa:]	‘breath’	ii.	[sã:-ũã:]	‘breath-PL’
iii.	[gã:]	‘cow’	iv.	[gã:-ũã:]	‘cow-PL’
v.	[ta:ũa:ŋ]	‘penalty’			

- ▶ 10 constraints (see appendix for definitions)
 - 9 standard faithfulness/markedness constraints
 - 1 contextual faithfulness constraint (Hauser and Hughto, 2020) to handle opaque application of nasal harmony

Step I: Phonotactic Learning

Goal: Calibrate constraint weights to acquire phonotactic patterns

Constraint	Type	initial w	final w
SPRD-L[NAS]	mark.	50.00	0.13
*NASOBS	mark.	50.00	100.00
*NASG	mark.	50.00	0.47
*NASV	mark.	50.00	0.00
*VN	mark.	50.00	100.00
*LOWRD	mark.	50.00	100.00

Table 1: Markedness constraint weights during phonotactic learning

Step II: EM Learning with 0-disparity URs

Assumption of the learner: Introduce UR→SR disparities (abstraction) only when learning fails at lower disparity levels.

- ▶ 0-Disparity URs for each morpheme

Morpheme	'breath'		'penalty'
SRs	[sa:]	[sã:]	[ta:vã:n]
0-disparity URs	sa:	sã:	ta:vã:n

Step II: EM Learning with 0-disparity URs

- Use EM learning to search for the combination of URs and weights that maximize data likelihood (Jarosz, 2006a,b, 2009, 2010, 2015)

Morpheme	UR candidates	Prior $\mathbb{P}$
'breath'	/sa:/	0.5
	/sã:/	0.5
'penalty'	/ta:vã:n/	1.0

E-step: Update probability assigned to each candidate UR.



M-step: Update constraint weights to best explain the data under posterior UR probabilities

Step II: EM Learning with 0-disparity URs

Constraint	initial w	final w
SPRD-L[NAS]	0.13	62.92
*NASOBS	100.00	100.00
*NASG	0.47	60.24
*NASV	0.00	23.16
*VN	100.00	100.00
ID[NAS]	0.00	83.44
IDFin[NAS]	0.00	89.46
ID[NAS]/_V	0.00	100.00
ID[RD]	0.00	6.29
*LOWRD	100.00	100.00

Morpheme	UR candidates	Prior $\mathbb{P}$	Posterior $\mathbb{P}$
'breath'	/sa:/	0.5	0.5
	/sã:/	0.5	0.5
'penalty'	/ta:vã:ŋ/	1.0	1.0

- No satisfactory solution exists in the 0-disparity hypothesis space
 - Learner can't jointly model harmony in [sã:ũã:] and non-harmony in [ta:vã:ŋ]
 - Distributes probability across competing URs for 'breath'

Step IV: UR Candidate Expansion

- ▶ Adjusting grammar fails to produce successful learning
 - Learner considers more abstract URs
- ▶ Expand the UR candidate space of each morpheme by introducing 1-disparity URs
- ▶ Expansion does not include all features
 - Identify markedness constraints at end of EM learning with intermediate weights, violated by 1+ SR
 - Only features referenced by those constraints are eligible to be flipped to create new URs

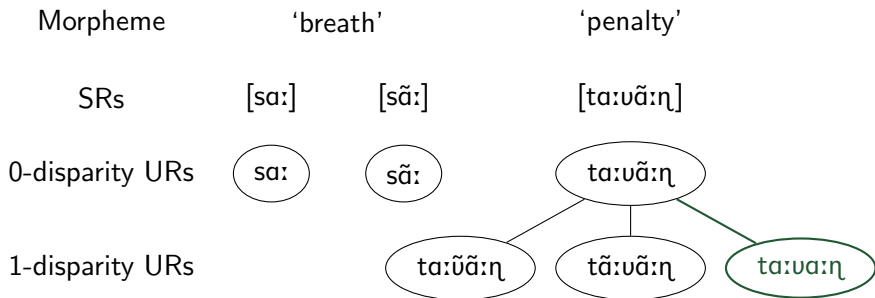
Step IV: UR Candidate Expansion

- Only the [nas] feature is eligible to be flipped

Constraint	final w
SPRD-L[NAS]	62.92
*NASOBS	100.00
*NASG	60.24
*NASV	23.16
*VN	100.00
*LOWRD	100.00

Step IV: UR Candidate Expansion

- Expand UR space: create 1-disparity candidates by flipping [NAS] on eligible segments



Step II: EM Learning with 1-disparity URs

Constraint	initial w	final w
SPRD-L[NAS]	0.13	48.45
*NASOBS	100.00	100.00
*NASG	0.47	48.45
*NASV	0.00	0.00
*VN	100.00	100.00
ID[NAS]	0.00	0.00
IDFin[NAS]	0.00	100.00
ID[NAS]/_V	0.00	100.00
ID[RD]	0.00	6.28
*LOWRD	100.00	100.00

Morpheme	UR candidates	Prior $\mathbb{P}$	Posterior $\mathbb{P}$
'breath'	/sa:/	0.5	1.0
	/sã:/	0.5	0.0
'penalty'	/ta:vã:ŋ/	0.25	0.0
	/ta:va:ŋ/	0.25	1.0
	/ta:ũã:ŋ/	0.25	0.0
	/tã:vã:ŋ/	0.25	0.0

- ▶ 1-disparity UR space resolves the learning problem
 - Successfully models harmony in [sã:ũã:]
 - Successfully models non-harmony in [ta:vã:ŋ]

Step III: Likelihood Evaluation

SR	Likelihood
[sa:]	1.00
[sã:ũã:]	0.96
[ta:vã:ŋ]	1.00

Likelihood Threshold satisfied!

A Solution to Learnability Problems

1. Preference for minimally abstract representations

- Abstraction is introduced only to the extent necessary by nature of the serial expansion of UR candidate space

2. Search Space Tractability

- An unbounded search space can be searched efficiently if:
 - (i) the space is structured appropriately
 - (ii) search for a satisfactory solution rather than the optimal one

Thank you!

Appendix: Constraint Definitions I

- i. **SPRD-L[NAS]** (cf. Walker, 2003, p.47)
For every occurrence of a [+NAS] feature in a prosodic word, if that [+NAS] feature is dominated by some segment, assign a violation for every segment to the left of that segment that does not dominate the [+NAS] feature.
- ii. ***NASOBS** (Walker, 2003, p.51)
Assign a violation for every obstruent that dominates a [+NAS] feature.
- iii. ***NASG** (Walker, 2003, p.51))
Assign a violation for every glide that dominates a [+NAS] feature.
- iv. ***NASV** (Walker, 2003, p.51))
Assign a violation for every vowel that dominates a [+NAS] feature.
- v. ***VN**
Assign a violation for every vowel that dominates a [-NAS] feature when directly preceding a nasal consonant.
- vi. **ID[NAS]**
For every segment, *A*, assign a violation if the output value for the [NAS] feature dominated by *A* does not match the input value for the [NAS] feature dominated by *A*.

