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)

[tɑ̃-n] [tɑ-t]	[tyd-u] [tit]	[ty-du] [tĩ-n]	[tɑ̃n-ə] [tɑ̃n]
/tɑ/	/tid/	/ti/	/tɑn/
concrete URs	segmentally composite URs	featurally composite URs	covert URs

less abstract ⟷ *more abstract*

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

Vowel Nasality in Punjabi

- Vowel nasality is contrastive (Bashir and Conners, 2019)

  [sɑ̃ː] ‘I was’ vs. [sɑː] ‘breath’

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

  [sɑ̃ːɳ] ‘grindstone’ vs. *[sɑːɳ]

- 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)

/sɑː-ʋɑ̃ː/	→	[sɑ̃ːʋ̃ɑ̃ː]	'breath-PL'
/mətʃiː-jɑ̃ː/	→	[mətʃĩːj̃ɑ̃ː]	'fish-PL'

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

/ɑːʋɑːm/	→	[ɑːʋɑ̃ːm]	'public'
/siːjɑːɳ/	→	[siːjɑ̃ːɳ]	'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 **H**ypothesis **S**pace **E**xpansion 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.	**[sɑː]**	**'breath'**	ii.	**[sɑ̃ː-ʋ̃ɑ̃ː]**	**'breath-PL'**
iii.	[gɑ̃ː]	'cow'	iv.	[gɑ̃ː-ʋ̃ɑ̃ː]	'cow-PL'
v.	**[tɑːʋɑ̃ːɳ]**	**'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	[sɑː]	[sɑ̃ː]	[tɑːʋɑ̃ːɳ]
0-disparity URs	sɑː	sɑ̃ː	tɑːʋɑ̃ːɳ

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'	/sɑː/	0.5
	/sɑ̃ː/	0.5
'penalty'	/tɑːʋɑ̃ːɳ/	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’	/sɑː/	0.5	**0.5**
	/sɑ̃ː/	0.5	**0.5**
‘penalty’	/tɑːʋɑ̃ːɳ/	1.0	1.0

- No satisfactory solution exists in the 0-disparity hypothesis space
  - Learner can't jointly model harmony in [sɑ̃ːʋ̃ɑ̃ː] and non-harmony in [tɑːʋɑ̃ːɳ]
  - Distributes probability across competing URs for ‘breath’

Step III: Likelihood Threshold Evaluation

SR	Likelihood
[sɑː]	0.50
[sɑ̃ːʋ̃ɑ̃ː]	0.48
[tɑːʋɑ̃ːɳ]	1.00

Likelihood Threshold of $> .95$ not met!

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

Morpheme	‘breath’		‘penalty’		
SRs	[saː]	[sãː]		[taːʋãːɳ]	
0-disparity URs	saː	sãː		taːʋãːɳ	
1-disparity URs			taːʋ̃ãːɳ	tãːʋãːɳ	taːʋaːɳ

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'	/saː/	0.5	**1.0**
	/sãː/	0.5	0.0
'penalty'	/taːʋãːɳ/	0.25	0.0
	/taːʋaːɳ/	0.25	**1.0**
	/taːʋ̃ãːɳ/	0.25	0.0
	/tãːʋãːɳ/	0.25	0.0

- 1-disparity UR space resolves the learning problem
  - Successfully models harmony in [sãːʋ̃ãː]
  - Successfully models non-harmony in [taːʋãːɳ]

Step III: Likelihood Evaluation

SR	Likelihood
[sɑː]	1.00
[sãːʊ̃ãː]	0.96
[tɑːʋãːɳ]	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 Tractibility**
   - 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!

References I

Bashir, E. and Conners, T. J. (2019). *A descriptive grammar of Hindko, Panjabi, and Saraiki*. Mouton-CASL Grammar Series. De Gruyter Mouton.

Bhatia, T. K. (1993). *Punjabi*. Descriptive Grammars. Routledge.

Hauser, I. and Hughto, C. (2020). Analyzing opacity with contextual faithfulness constraints. *Glossa: a journal of general linguistics*, 5(1):1–33.

Jarosz, G. (2006a). *Rich lexicons and restrictive grammars - Maximum likelihood learning in Optimality Theory*. PhD thesis, Johns Hopkins University.

Jarosz, G. (2006b). Richness of the base and probabilistic unsupervised learning in Optimality Theory. In *Proceedings of the Eighth Meeting of the ACL Special Interest Group on Computational Phonology*, pages 50–59.

Jarosz, G. (2009). Restrictiveness and phonological grammar and lexicon learning. In Elliot, M., Kirby, J., Sawada, O., Staraki, E., and Yoon, S., editors, *Proceedings of the 43rd Annual Meeting of the Chicago Linguistics Society*, volume 43, pages 125–134.

Jarosz, G. (2010). Implicational markedness and frequency in constraint-based computational models of phonological learning. *Journal of Child Language. Special Issue on Computational Models of Child Language Learning*, 37(3):565–606.

Jarosz, G. (2015). Expectation driven learning of phonology. University of Massachusetts manuscript.

Jarosz, G. (2019). Computational modeling of phonological learning. *Annual Review of Linguistics*, 5:67–90.

Kenstowicz, M. and Kisseberth, C. (1977). *Topics in Phonological Theory*. Academic Press.

Kiparsky, P. (1973). *Abstractness, opacity, and global rules*, pages 57–86. Tokyo: TEC.

Paramore, J. C. and Bennett, R. T. (2025). Covert URs: evidence from nasalization in Pakistani Punjabi. Manuscript, UC Santa Cruz.

References II

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69(1):99–118.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, 63(2):129–138.

Walker, R. (2003). *Reinterpreting transparency in nasal harmony*, pages 37–72. Amsterdam: John Benjamins.

Wang, Y. and Hayes, B. (2025). Learning phonological underlying representations: the role of abstractness. *Linguistic Inquiry*, pages 1–39.

Zahid, S. and Hussain, S. (2012). An acoustic study of vowel nasalization in Punjabi. In *Proceedings of the Conference on Language & Technology*, volume 61.

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*.

Appendix: Constraint Definitions II

vii. IdFin[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* in the final syllable of a prosodic word.

viii. Id[nas]/__V
Let *A* be a segment that occurs before an oral vowel, V, in the input. Assign one violation if the output correspondent of *A* does not have the same specifications for [nas] as *A*.

ix. Id[rd]
For every segment, *A*, assign a violation if the output value for the [rd] feature dominated by *A* does not match the input value for the [rd] feature dominated by *A*.

x. *LowRd
Assign a violation for every vowel that dominates a [rd] feature and a [low] feature simultaneously.