Learning Covert URs via Disparity Minimization

Jonathan C. Paramore

jcparamo@ucsc.edu

Society for Computation in Linguistics (SCiL)
July 19th, 2025

What is Abstractness?

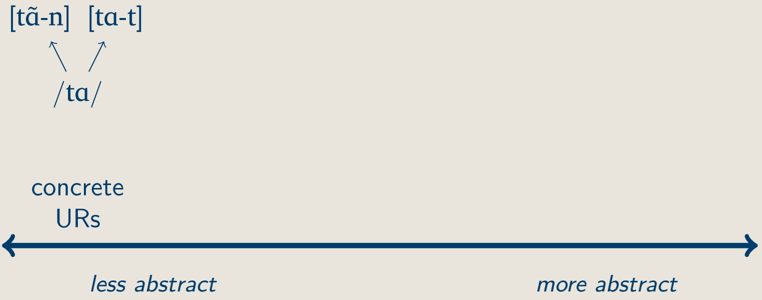
Abstractness in phonology: Any stored representation not directly heard or experienced.

- Abstraction levels (Kenstowicz & Kisseberth, 1977; Wang & Hayes, 2025)

What is Abstractness?

Abstractness in phonology: Any stored representation not directly heard or experienced.

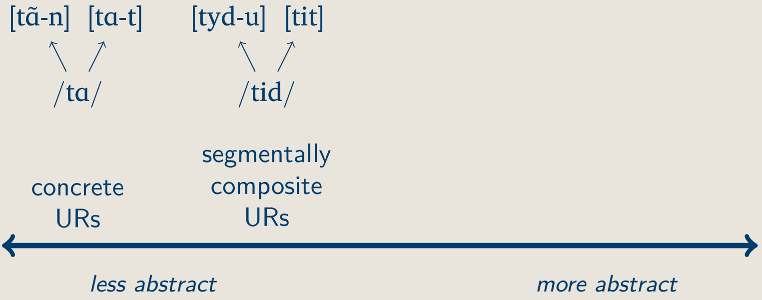
- Abstraction levels (Kenstowicz & Kisseberth, 1977; Wang & Hayes, 2025)



What is Abstractness?

Abstractness in phonology: Any stored representation not directly heard or experienced.

- Abstraction levels (Kenstowicz & Kisseberth, 1977; Wang & Hayes, 2025)



What is Abstractness?

Abstractness in phonology: Any stored representation not directly heard or experienced.

- Abstraction levels (Kenstowicz & Kisseberth, 1977; Wang & Hayes, 2025)

[tɑ̃-n] [tɑ-t] ← /tɑ/ — concrete URs

[tyd-u] [tit] ← /tid/ — segmentally composite URs

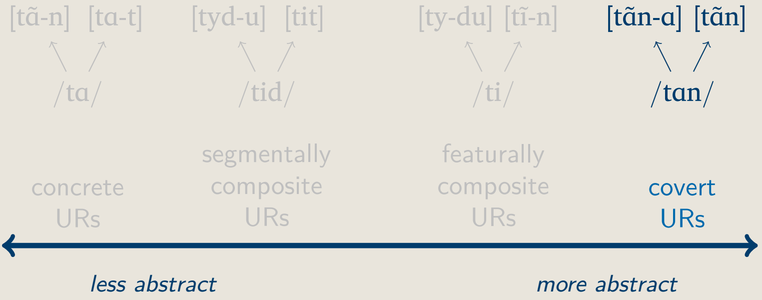
[ty-du] [tĩ-n] ← /ti/ — featurally composite URs

less abstract ⟷ *more abstract*

What is Abstractness?

Abstractness in phonology: Any stored representation not directly heard or experienced.

- Abstraction levels (Kenstowicz & Kisseberth, 1977; Wang & Hayes, 2025)



What is Abstractness?

Abstractness in phonology: Any stored representation not directly heard or experienced.

- Abstraction levels (Kenstowicz & Kisseberth, 1977; Wang & Hayes, 2025)

[tã-n] [tɑ-t]	[tyd-u] [tit]	[ty-du] [tĩ-n]	[tãn-ɑ] [tãn]
/tɑ/	/tid/	/ti/	/tɑn/
concrete URs	segmentally composite URs	featurally composite URs	covert URs

less abstract ⟷ *more abstract*

Motivation for Covert URs: evidence from Punjabi

- Vowel nasality is contrastive: [sɔɔ] '100' vs. [sɔ̃ɔ̃] 'sleep'
- Nasality Processes (Paramore, 2023)
  - ***VN** - Vowels are categorically nasal before nasal codas.
  - **SPRD-L$_{[+nas]}$** - nasality spreads from contrastive nasal vowels to glides and vowels.
- Nonalternating pre-N vowels are covertly oral in Punjabi.
  - Phonetically identical to contrastive nasal vowels.
  - they don't trigger harmony.

i.	/sɑɑ-ʋɑ̃ɑ̃/	→	[sɑ̃ɑ̃ʋ̃ɑ̃ɑ̃]	'breath-PL'
ii.	/tɑɑʋɑɑɳ/	→	[tɑɑʋɑ̃ɑ̃ɳ]	'penalty'

Table 1: *Nasal Harmony in Punjabi.*

Issues in Learning Highly Abstract URs

1) Need a mechanism that minimizes abstraction (O'Hara, 2017).
   - All else equal, learner should prefer /ɑ/ → [ɑ̃] over /æ/ → [ɑ̃]

2) Infeasible to search the entire space of potential URs (Tesar, 2014; Wang & Hayes, 2025).
   - The space of potential URs for [ɑ̃] is potentially infinite: /ɑ/, /æ/, /ɑɑ/, /ɑɑɑ/, etc.

This Paper

I directly address the problem of preferring minimally abstract URs

- Previous approaches do not generate a preference for minimal abstraction for non-alternating pre-n vowels in Punjabi.
- Propose and implement an update to MaxLex (O'Hara, 2017):
  - Add a Disparity Bias to the objective function.
  - Enforces minimal abstraction by penalizing disparities between representations in UR → SR mappings.

I don't address the computational cost of the increased search space size.

MaxLex

MaxEnt Learner of Grammars (Hayes & Wilson, 2008) composed of two learning stages

- Phonotactic stage: focus on acquiring phonotactic patterns
  - UR→SR mappings are not considered
  - Must learn constraint weights that maximize the likelihood of observing the surface forms
- Morphologically aware stage: learner aware of morphological relatedness and assigns a probability distribution to each morpheme's set of potential URs (Jarosz, 2006)
  - Must learn optimal constraint weights
  - *and* UR probabilities

Morphologically Aware Objective Function

- The MaxLex algorithm minimizes an objective function made up of two components.

Morphologically Aware Objective Function

- The MaxLex algorithm minimizes an objective function made up of two components.

$$\mathcal{O}_{Lex}(\mathbf{w}, \boldsymbol{\pi}) = \underbrace{-\ln\left[\prod_{i=1}^{n} \left(\mathbb{P}[O_i \mid (\mathbf{w}, \boldsymbol{\pi})]\right)\right]}_{\text{Negative Log Likelihood}} + \underbrace{\sum_{w_i \in \mathbf{w}} \frac{(w_i - c_i)^2}{\sigma_i^2}}_{\text{L2 Gaussian Prior}}$$

- Minimizing the NLL maximizes the likelihood of observing the surface forms provided to the learner.

Morphologically Aware Objective Function

- The MaxLex algorithm minimizes an objective function made up of two components.

$$\mathcal{O}_{Lex}(\mathbf{w}, \boldsymbol{\pi}) = \underbrace{-\ln\left[\prod_{i=1}^{n}\left(\mathbb{P}[O_i \mid (\mathbf{w}, \boldsymbol{\pi})]\right)\right]}_{\text{Negative Log Likelihood}} + \underbrace{\sum_{w_i \in \mathbf{w}} \frac{(w_i - c_i)^2}{\sigma_i^2}}_{\text{L2 Gaussian Prior}}$$

- Prior increases restrictivity with preference for $M > F$.
  - $c_i = 100$ for markedness, $c_i = 0$ for faithfulness

Morphologically Aware Objective Function

$$\mathcal{O}_{Lex}(\mathbf{w}, \boldsymbol{\pi}) = \underbrace{-\ln\left[\prod_{i=1}^{n}\left(\mathbb{P}[O_i \mid (\mathbf{w}, \boldsymbol{\pi})]\right)\right]}_{\text{Negative Log Likelihood}} + \underbrace{\sum_{w_i \in \mathbf{w}} \frac{(w_i - c_i)^2}{\sigma_i^2}}_{\text{L2 Gaussian Prior}}$$

- Constraints only deviate from their ideal weights (c_i) if doing so sufficiently decreases the NLL.
- **Byproduct**: *emergence* of minimal UR abstraction:
  - Faithfulness constraints active in phonotactic patterning are favored over inactive constraints to model alternations.

MaxLex and Punjabi Pre-N Vowels

- The Prior does not minimize abstraction for non-alternating pre-N vowels in Punjabi.
- Punjabi forms fed to MaxLex

i.	[sɑɑ]	'breath'	ii.	[sɑ̃ɑ̃-ʋ̃ɑ̃ɑ̃]	'breaths'
iii.	[ʋʃɑɑ]	'morning'	iv.	[ʋʃɑ̃ɑ̃-ʋ̃ɑ̃ɑ̃]	'mornings'
v.	[gɑ̃ɑ̃]	'cow'	vi.	[gɑ̃ɑ̃-ʋ̃ɑ̃ɑ̃]	'cows'
vii.	[tʃʰɑ̃ɑ̃]	'shade'	viii.	[tʃʰɑ̃ɑ̃-ʋ̃ɑ̃ɑ̃]	'shades'
ix.	[tɑɑʋɑ̃ɑ̃ɳ]	'penalty'	x.	[prəʋɑ̃ɑ̃ɳ]	'accepted'

Table 2: *Punjabi surface forms fed to MaxLex*

Constraints used in Model

i. SPRD-L[+nas] (Walker, 2003, p. 47): Nasal Harmony
ii. *NASOBS (Walker, 2003, p. 51): Penalize nasal obstruents.
iii. *NASG (Walker, 2003, p. 51): Penalize nasal glides.
iv. ID[nas]: Penalize changes in nasality.
v. IDFIN[nas]: Penalize nasality changes on the word-final segment.
vi. *VN: Penalize oral vowels before nasal codas.
vii. **ID[nas]/___V** (Hauser & Hughto, 2020): 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*.
viii. **ID[rd]**: Penalize changes in rounding.
ix. ***LOWRD**: Penalize low round vowels.

Constraint weights after phonotactic learning

Constraint	Type	initial w	final w
ID[nas]	faith.	50.00	51.37
IDFIN[nas]	faith.	50.00	44.83
SPRD-L[+nas]	mark.	50.00	92.83
*NASOBS	mark.	50.00	100.00
*NASG	mark.	50.00	99.48
ID[nas]/_V	context.faith.	50.00	100.00
*VN	mark.	50.00	100.00
ID[rd]	faith.	50.00	0.00
*LOWRD	mark.	50.00	100.00

Crucial Phonotactic Pattern

- Low Round vowels never surface
  - ID[rd] is minimized.
  - *LOWRD is maximized.

/sɒɒ/	*LOWRD 100.00	ID[rd] 0.00	$\mathcal{H}$	$\tilde{\mathcal{P}}$
a. ☞ sɑɑ		-1	0	1.0
b. sɒɒ	-1		-100	$4e^{-44}$

Morphological stage

- Learner becomes morphologically aware.
- Attempts to learn UR probability distributions for each morpheme.
- Optimizes constraint weights and UR distributions to maximize the likelihood of observing the data.
- For each morpheme with a pre-n vowel, 3 potential URs were provided. e.g., for [taaʋããɳ] 'penalty':
  - Concrete: /taaʋããɳ/
  - Covert: /taaʋaaɳ/
  - Highly Abstract: /taaʋɒɒɳ/

Learned Weights and $\mathbb{P}(UR)$ under MaxLex

- The minimally abstract covert UR is not preferred.
  - Faithfulness constraints do not drive the surface realizations.
  - No difference in the constraint weights for the two covert URs.

Constraint	Type	initial w	final w
ID[nas]	faith.	51.37	0.07
IDFIN[nas]	faith.	44.83	100.00
SPRD-L	mark.	92.83	5.42
*NASOBS	mark.	100.00	100.00
*NASG	mark.	99.48	0.02
ID[nas]/_V	cont.faith.	100.00	100.00
*VN	mark.	100.00	100.00
ID[rd]	faith.	0.00	0.00
*LOWRD	mark.	100.00	100.00

UR	$\mathbb{P}$	*Type*
/taaʋããɳ/	0.0	Concrete
/taaʋaaɳ/	0.5	Covert
/taaʋɒɒɳ/	0.5	Highly abstract

Learning via Disparity Minimization

- **Disparity Bias**

$$D(IO_j) = \sum_{i=1}^{k_j} \Bigg[\underbrace{\mathbf{1}_{\{s^I_{ij} \oplus s^O_{ij} = \varnothing\}}}_{\text{Insertion/Deletion}} + \underbrace{\sum_{f \in F} \mathbf{1}_{\{s^I_{fij} \neq s^O_{fij}\}}}_{\text{Feature Changes}} \Bigg]^2$$

- Sums and squares disparities between corresponding input-output segments for the *j*-th morpheme.
- **Result**: URs containing segments with more disparities are dispreferred by the learner

Updated Objective Function

$$\mathcal{O}_{Lex}(\mathbf{w}, \boldsymbol{\pi}) = \mathbf{NLL} + \mathbf{Prior} + \sum_{\mathbf{j=1}}^{\mathbf{n}} \mathsf{D}(\mathsf{IO}_{\mathbf{j}})$$

Weights and $\mathbb{P}(UR)$ with Disparity Bias

Constraints	Type	initial w	final w
ID[nas]	faith.	51.37	0.00
IDFIN[nas]	faith.	44.83	100.00
SPRD-L	mark.	92.83	4.61
*NASOBS	mark.	100.00	100.00
*NASG	mark.	99.48	0.00
ID[nas]/_V	cont.faith.	100.00	100.00
*VN	mark.	100.00	100.00
ID[rd]	faith.	0.00	0.00
*LOWRD	mark.	100.00	100.00

UR	$\mathbb{P}$	*Type*
/taaʋããɳ/	$2e^{-15}$	Concrete
/taaʋaaɳ/	1.00	Covert
/taaʋɒɒɳ/	$9e^{-15}$	Highly abstract

Traversing the Search Space

- The Disparity Bias prefers minimally abstract URs over increasingly abstract alternatives.
- But, it does not solve the computational cost of the massive search space caused by permitting covert URs.
- **Two Potential Solutions**:
  - Constrain the size of the search space (Wang & Hayes, 2025).
  - Organize the space for efficient exploration (Tesar, 2014, 2016).

Traversing the Search Space

- The Disparity Bias prefers minimally abstract URs over increasingly abstract alternatives.
- But, it does not solve the computational cost of the massive search space caused by permitting covert URs.
- **Two Potential Solutions**:
  - Constrain the size of the search space (Wang & Hayes, 2025).
  - Organize the space for efficient exploration (Tesar, 2014, 2016).

Conclusion and Future Directions

Proposed a solution for minimizing abstraction when covert URs are permitted.

- Implemented a **Disparity Bias** that generates a preference for minimal UR abstraction.

Several outstanding issues require future research.

- Are there case in which the Disparity Bias and MaxLex L2 prior conflict? If so, how is learning resolved?
- The Updated learner should be tested on other languages with varying degrees of abstraction (c.f. Wang and Hayes (2025)).
- A framework for efficiently searching the space of potential URs must be developed.

References I

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

Hayes, Bruce, & Wilson, Colin. (2008). A maxiumum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, *39*(3), 379–440.

Jarosz, Gaja. (2006). *Rich lexicons and restrictive grammars - maximum likelihood learning in Optimality Theory* (Doctoral dissertation). Johns Hopkins University.

Kenstowicz, Michael, & Kisseberth, Charles. (1977). *Topics in phonological theory*. Academic Press.

O'Hara, Charlie. (2017). How abstract is more abstract? learning abstract underlying representations. *Phonology*, *34*, 325–345.

Paramore, Jonathan Charles. Covert URs: evidence from Pakistani Punjabi (talk). In: In *Formal approaches to South Asian languages (FASAL) 14*. 2023.

Tesar, Bruce. (2014). *Output-Driven Phonology*. Cambridge: Cambridge University Press.

Tesar, Bruce. (2016). Phonological learning with output-driven maps. *Language Acquisition*, *24*(2), 148–167.

Walker, Rachel. (2003). Reinterpreting transparency in nasal harmony. In Jeroen van de Weijer, Vincent J. van Heuven, & Harry van der Hulst (Eds.), *The phonological spectrum, part I: Segmental structure* (pp. 37–72). Amsterdam: John Benjamins.

Wang, Yang, & Hayes, Bruce. (2025). Learning phonological underlying representations: The role of abstractness. *Linguistic Inquiry*.

Organizing the Search Space

- The Disparity Bias provides implicit structure to the search space (Tesar, 2014, 2016).
- If $UR_A \rightarrow SR_1$ has *n* disparities & is ungrammatical, any $UR_i \rightarrow SR_1$ with a **superset** of *n* disparities is ungrammatical.
  - Eliminate all UR_i once UR_A is ruled out.

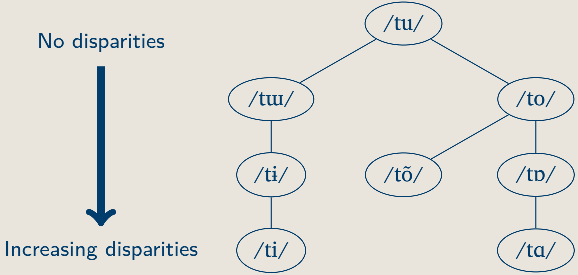


Figure 1: Lattice for the output form [tu]

Organizing the Search Space

- The Disparity Bias provides implicit structure to the search space (Tesar, 2014, 2016).
- If $UR_A \rightarrow SR_1$ has *n* disparities & is ungrammatical, any $UR_i \rightarrow SR_1$ with a **superset** of *n* disparities is ungrammatical.
  - Eliminate all UR_i once UR_A is ruled out.

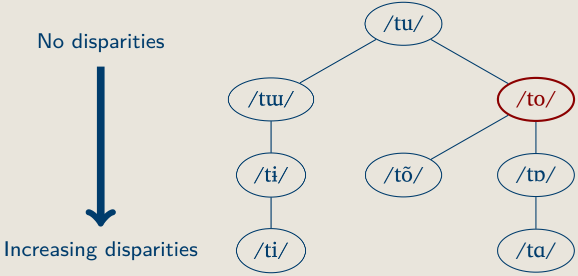


Figure 1: Lattice for the output form [tu]

Organizing the Search Space

- The Disparity Bias provides implicit structure to the search space (Tesar, 2014, 2016).
- If $\mathrm{UR_A}\rightarrow\mathrm{SR_1}$ has *n* disparities & is ungrammatical, any $\mathrm{UR_i}\rightarrow\mathrm{SR_1}$ with a **superset** of *n* disparities is ungrammatical.
  - Eliminate all $\mathrm{UR_i}$ once $\mathrm{UR_A}$ is ruled out.

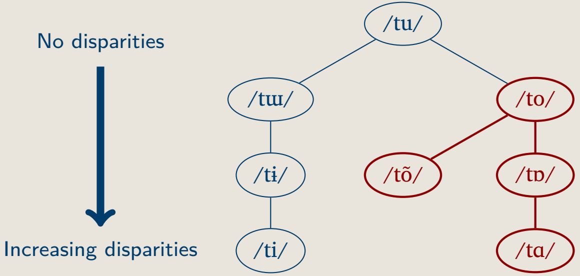


Figure 1: Lattice for the output form [tu]

Limitation of Output-Driven Phonology

- Current system only rules out non-grammatical UR chains.
- No mechanism to stop searching once a sufficiently good UR is found.
- **Upshot**: The search space is still too large to examine exhaustively.

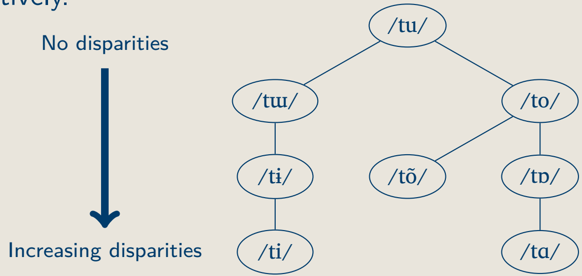


Figure 2: Lattice for the output form [tu]

Limitation of Output-Driven Phonology

- Current system only rules out non-grammatical UR chains.
- No mechanism to stop searching once a sufficiently good UR is found.
- **Upshot**: The search space is still too large to examine exhaustively.

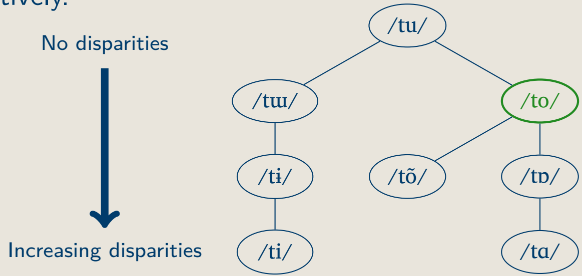


Figure 2: Lattice for the output form [tu]

Limitation of Output-Driven Phonology

- Current system only rules out non-grammatical UR chains.
- No mechanism to stop searching once a sufficiently good UR is found.
- **Upshot**: The search space is still too large to examine exhaustively.

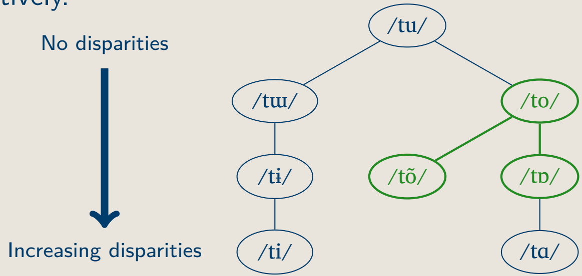


Figure 2: Lattice for the output form [tu]

Limitation of Output-Driven Phonology

- Current system only rules out non-grammatical UR chains.
- No mechanism to stop searching once a sufficiently good UR is found.
- **Upshot**: The search space is still too large to examine exhaustively.

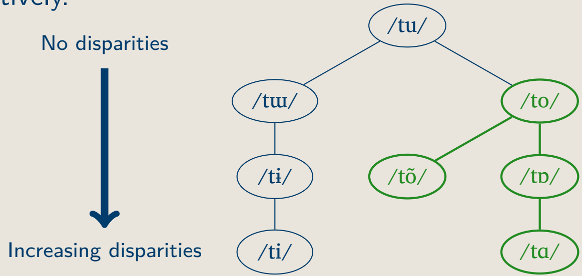


Figure 2: Lattice for the output form [tu]

Potential UR Learning Process

1. Serial search explores URs in batches, in increasing order of disparities.
2. Minimum Likelihood Threshold ensures search continues until the data is sufficiently likely.
3. Marginal Improvement Threshold stops the search once adding disparities fail to yield a substantial gain in likelihood.
4. Disparity Bias encodes a preference for the minimally abstract UR.

- These components have the potential to ensure that:
  - The entire space does not need to traversed.
  - The learner arrives at a UR that sufficiently explains the data.
  - Abstraction is introduced **only to the extent necessary**.