Learning phonotactics from linguistic informants

Interactions between Formal and Computational Linguistics (IFLG) Seminar

February 14th, 2024

Canaan Breiss
cbreiss@usc.edu // cbreiss.com
USC University of Southern California

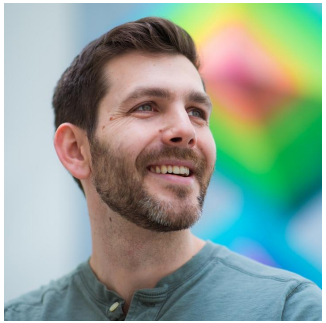
Alexis Ross
alexisro@mit.edu // alexisjihyeross.github.io
Massachusetts Institute of Technology

Acknowledgments

Team:

Amani Maina-Kilaas
(MIT)

Roger Levy
(MIT)



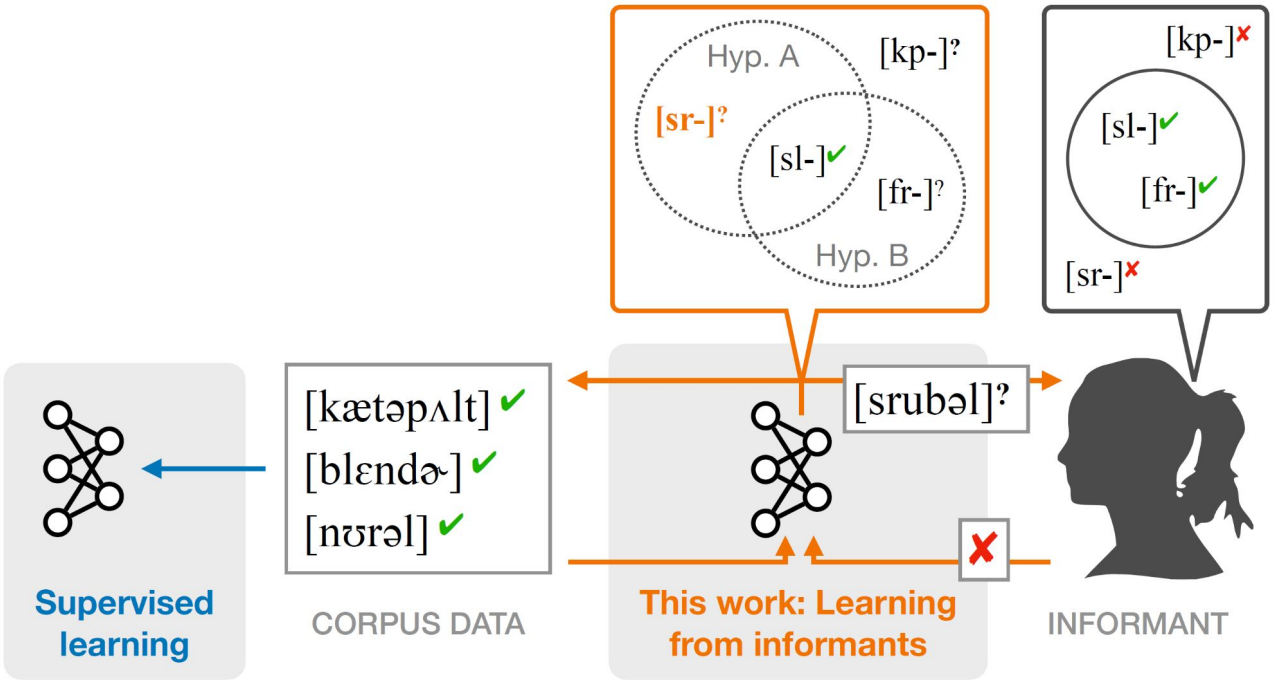
Jacob Andreas
(MIT)

Funding:

NSF

MIT-IBM
Watson
AI Lab

Problem setting



Phonotactics: constraints on sound sequences

Existing English words	Possible English words
back	
nick	*blick?* --> ✅
black	*bnick?* --> ❌
...	...

Cf. Chomsky & Halle (1968), Bailey & Hahn (2001), Hayes & Wilson (2008), ...

Iterative querying procedure

Algorithm 1: Iterative Query Procedure

Input: policy π, total timesteps T

$(\underline{x}, \underline{y}) \leftarrow [\,]$; $t \leftarrow 0$;

while $t < T$ **do**

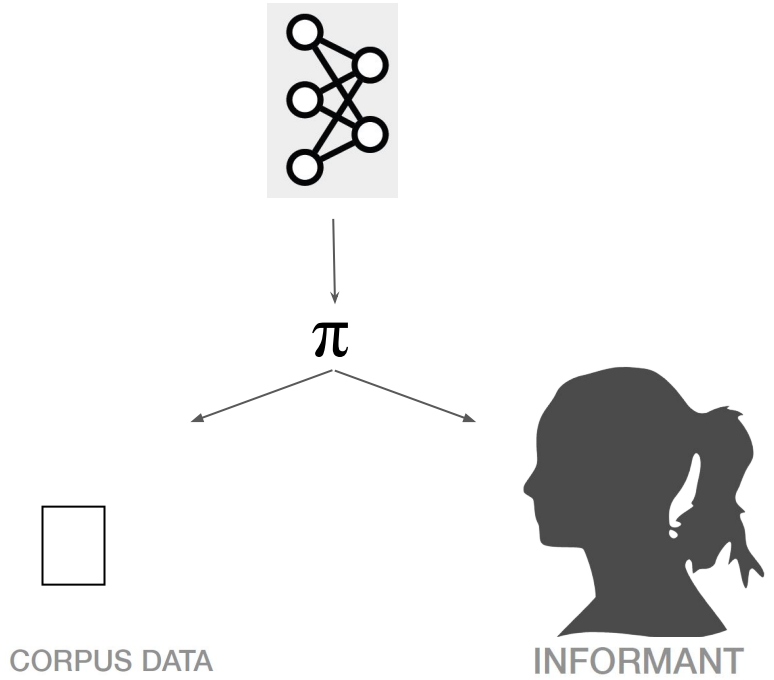
  $x_t \leftarrow \pi(x \mid \underline{x}, \underline{y})$;

  $y_t \leftarrow \text{informant}(x_t)$;

  append (x_t, y_t) to $(\underline{x}, \underline{y})$;

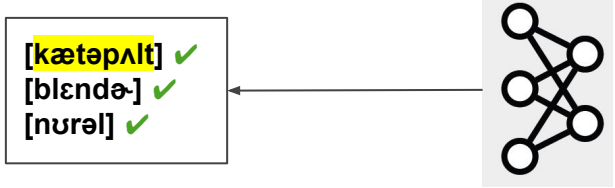
  $t \leftarrow t + 1$;

end



Policies: Basic

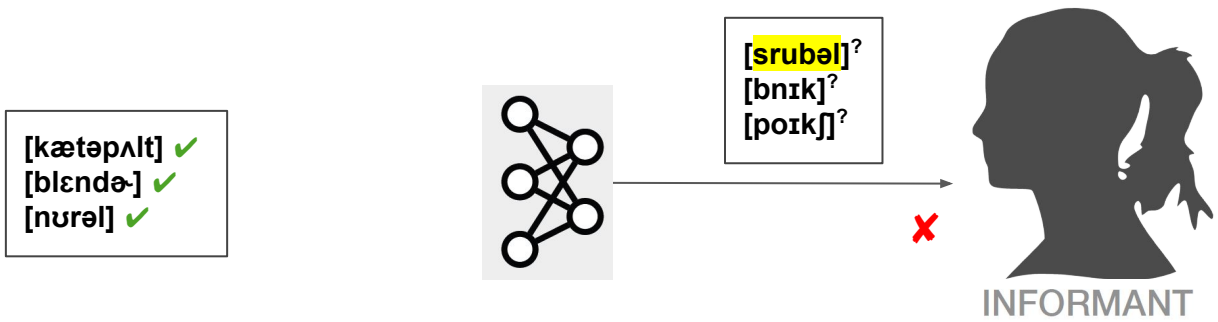
Train



Policies: Basic

Train

Uniform

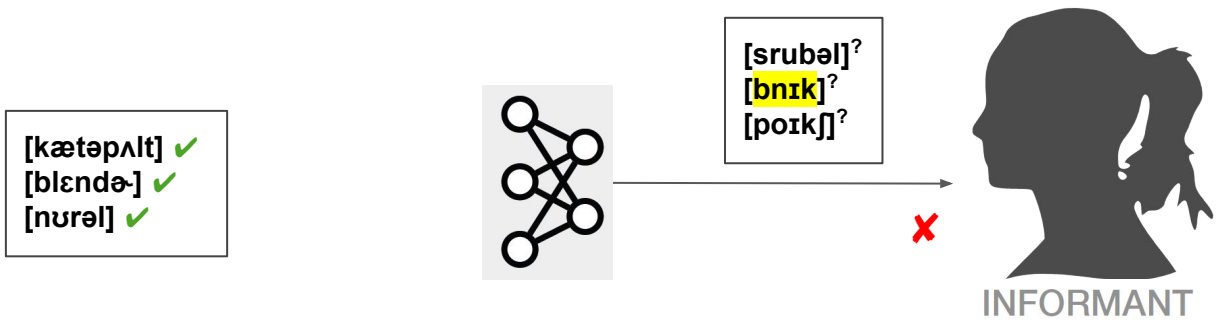


Policies: Basic

Train

Uniform, **Label Entropy**

$$x^* = \arg\max_{x \in \mathcal{X}} \mathcal{H}(y \mid x, \underline{x}, \underline{y}).$$



Policies: Basic

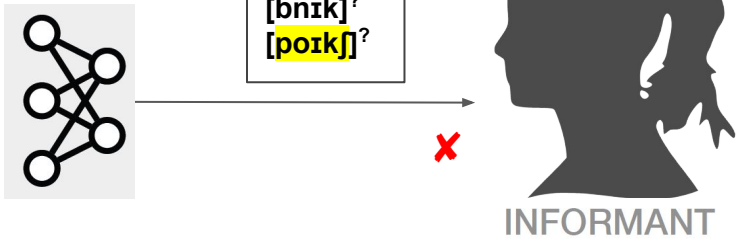
Train

Uniform, Label Entropy, **Expected Info Gain**

$$V_{\mathrm{IG}}(x, y; \underline{x}, \underline{y}) = \mathcal{H}(\boldsymbol{\theta} \mid \underline{x}, \underline{y}) - \mathcal{H}(\boldsymbol{\theta} \mid x, y, \underline{x}, \underline{y}).$$

$$x^* = \arg\max_{x \in \mathcal{X}} V_{\mathrm{IG}}(x, y = 1; \underline{x}, \underline{y}) \cdot p(y = 1 \mid x, \underline{x}, \underline{y}) + V_{\mathrm{IG}}(x, y = 0; \underline{x}, \underline{y}) \cdot p(y = 0 \mid x, \underline{x}, \underline{y}),$$

[kætəpʌlt] ✔
[blɛndə˞] ✔
[nʊrəl] ✔



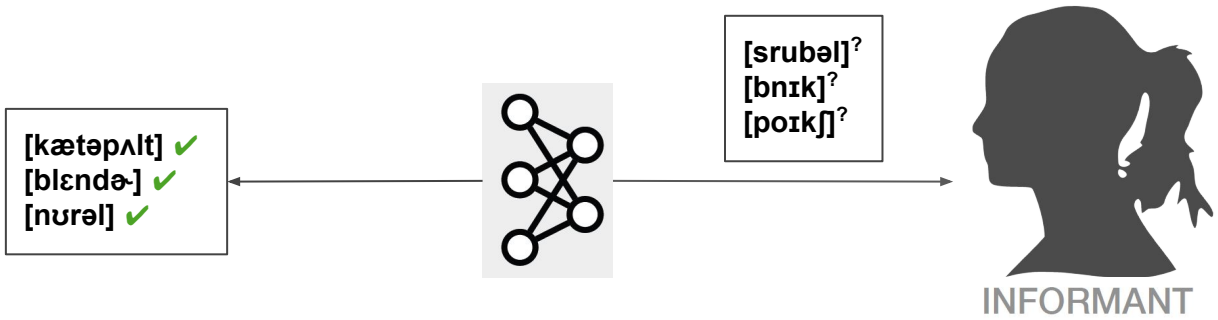
Policies: Hybrid

Train

Uniform, Label Entropy, **Expected Info Gain**

Choosing Basic Policies:

History: empirical information gains
Model: expected information gains
Hybrid: history (train) *and* model (EIG)



Data: Vowel harmony

	[high]	[low]	[ATR]
i	+	−	+
ɪ	+	−	−
e	−	−	+
ɛ	−	−	−
a	−	+	0

ATR = Advanced Tongue Root

Individual vowels:

/i/ --> [+high, -low, +ATR]

Classes of vowels:

[+high] --> /i, ɪ/

[+ATR] --> /i, e/

[+low] --> /a/

...

ATR Harmony and Procedurally-Generated Languages

*$V_{[\alpha ATR]}C_0V_{[-\alpha ATR]}$= no adjacent ATR-mismatches

[tekɪ], [kɪpɛ] --> ✅ [tekɪ], [kipɛ] --> ❌

[katipi], [pɛkɪka] --> ✅

[pɪpake], [pɛkati] --> ✅

Train ~ $\Sigma^*(*V_{[\alpha ATR]}C_0V_{[-\alpha ATR]})$

Test ~ $\Sigma^*(*V_{[\alpha ATR]}C_0V_{[-\alpha ATR]})$

ATR Harmony uses 16 parameters

$Language_i$ ~ prior(hyperparameters)

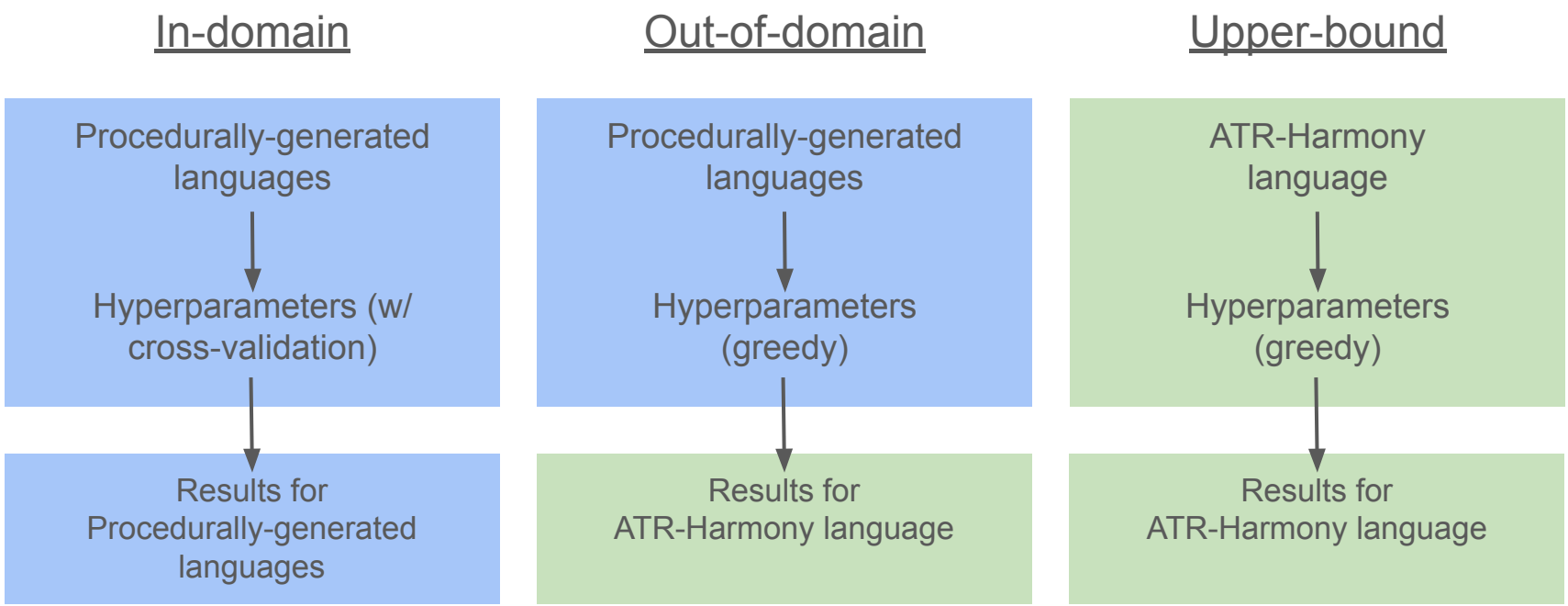
$Train_i$ ~ $\Sigma^*(Language_i)$

$Test_i$ ~ $\Sigma^*(Language_i)$

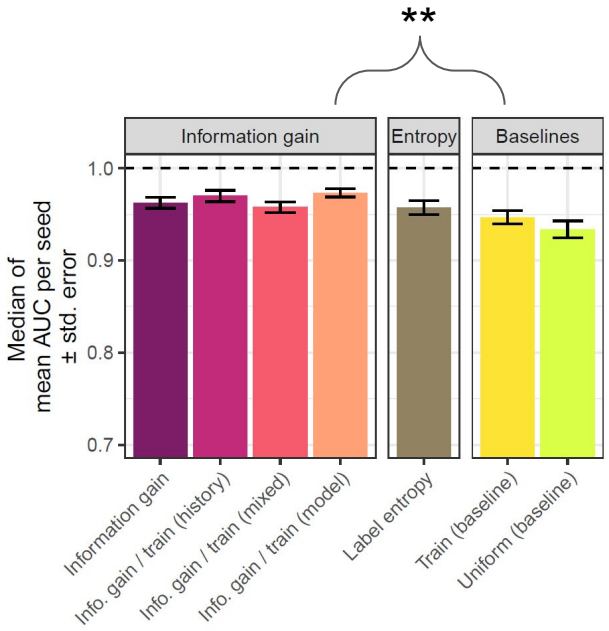
i = 1...10

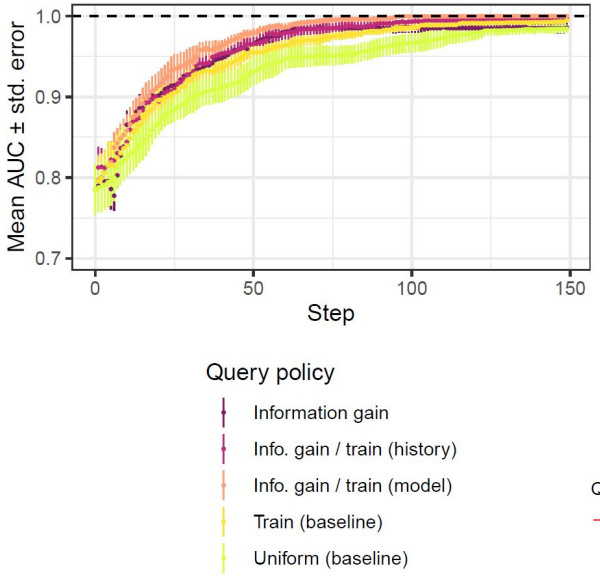
- 16 parameters sampled per language
- Distribution of word lengths matched

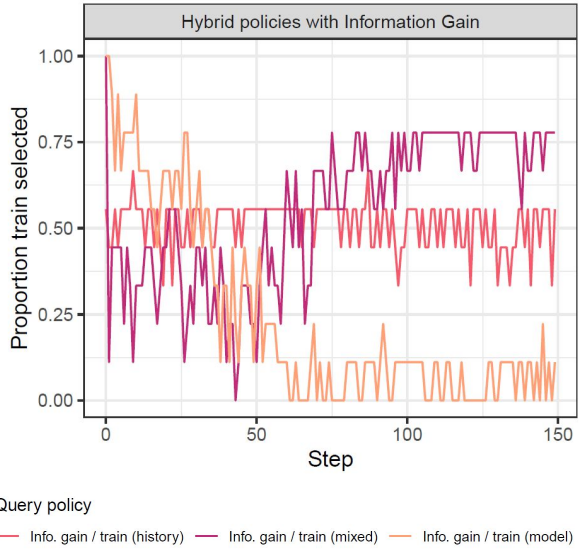
Coping with hyperparameters: three analyses



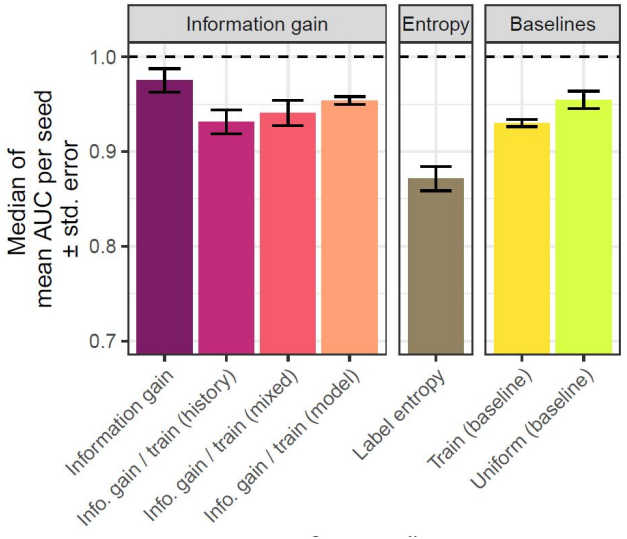
In-domain Analysis

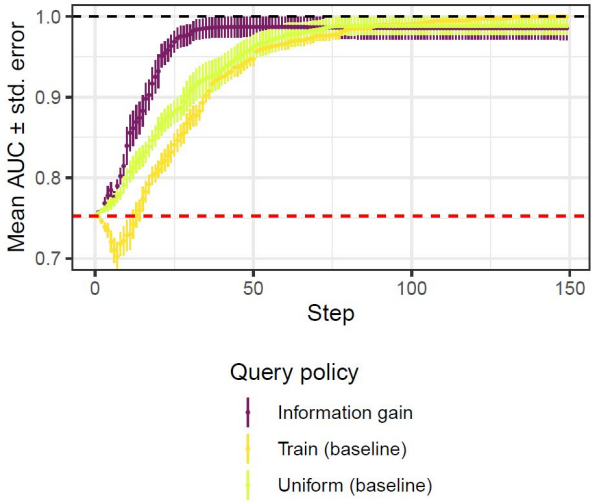


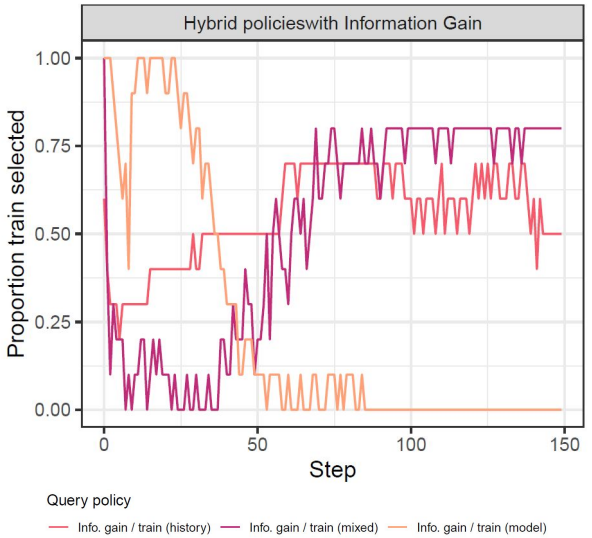




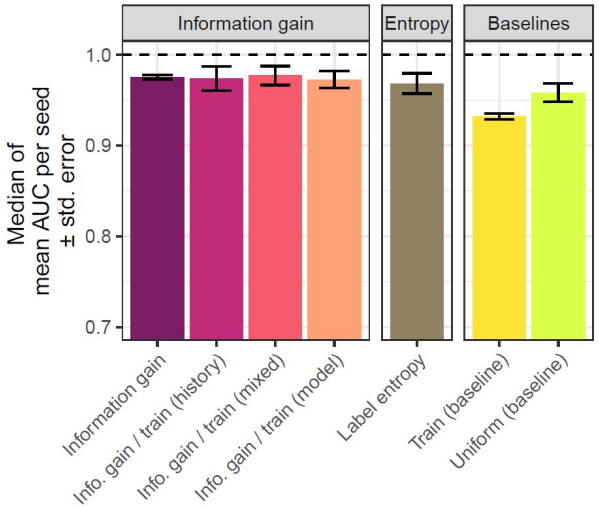
Out-of-domain Analysis

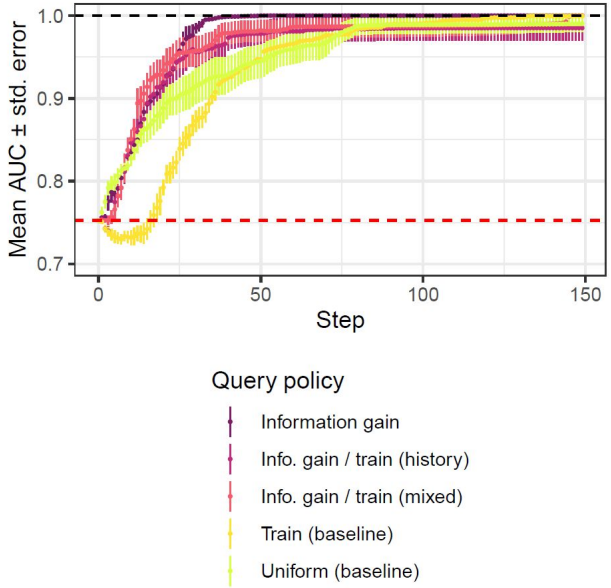


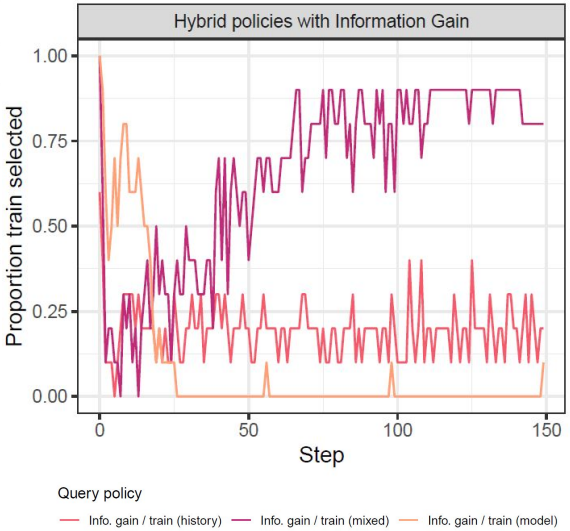




Upper-bound Analysis



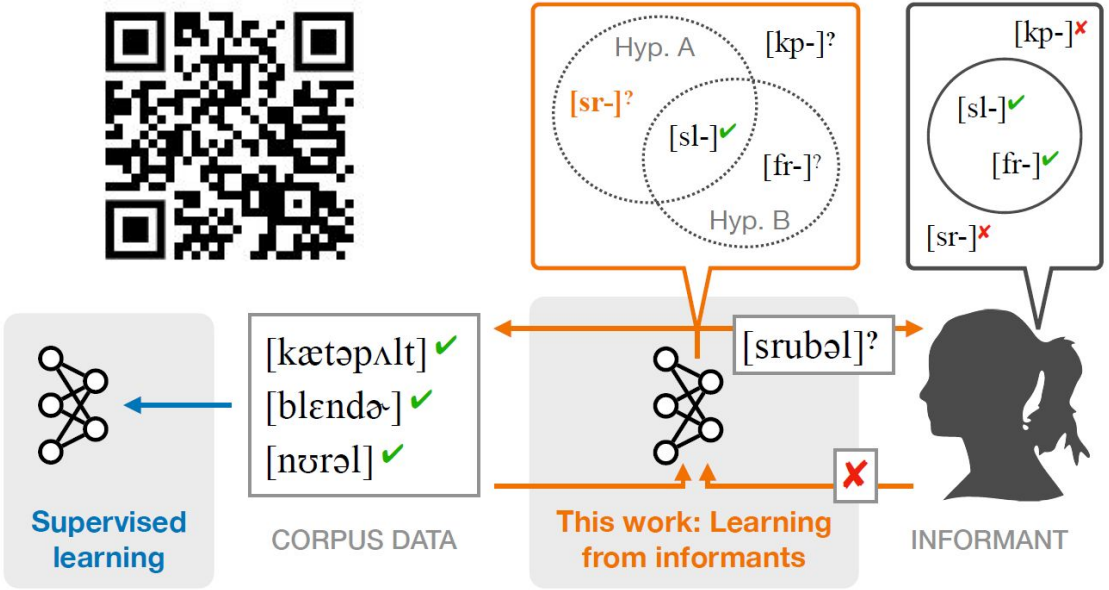


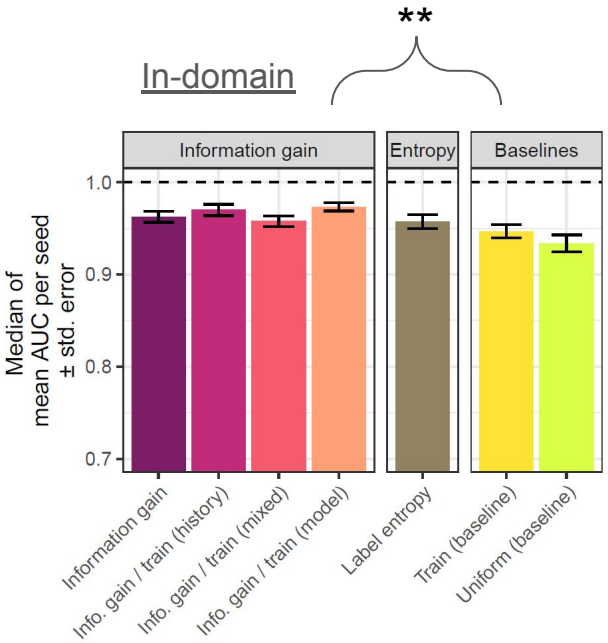


Summary

Contact: cbreiss@usc.edu; alexisro@mit.edu

Paper:





Questions?

Hybrid Policies

History

$$S^{\mathbf{EMP}}(\pi; \underline{x}, \underline{y}) = \frac{\sum_{i \in I_\pi} V(x_i, y_i; \underline{x}_{<i})}{|I_\pi|},$$

$$\text{where } I_\pi = \{i \mid x_i \text{ was selected by } \pi, i < t\}.$$

$$\pi^* = \arg\max_{\pi \in [\hat{\pi}, \pi_{\text{train}}]} S^{\text{EMP}}(\pi; \underline{x}, \underline{y}).$$

Model

$$S^{\mathbf{EXP(y)}}(\hat{\pi}; \underline{x}, \underline{y}) = \mathop{\mathbb{E}}_{y \in [0,1]} V(x^*, y; \underline{x}, \underline{y}), x^* \sim \hat{\pi}.$$

$$S^{\mathbf{EXP(x)}}(\pi_{\text{train}}; \underline{x}, \underline{y}) = \mathop{\mathbb{E}}_{x \in L} V(x, y = 1; \underline{x}, \underline{y}).$$

$$\pi^* = \arg\max_{\pi \in [\hat{\pi}, \pi_{\text{train}}]} S(\pi; \underline{x}, \underline{y}).$$

Hybrid

$$S(\hat{\pi}; \underline{x}, \underline{y}) = S^{\text{EXP}(y)}(\hat{\pi}; \underline{x}, \underline{y}),$$

$$S(\pi_{\text{train}}; \underline{x}, \underline{y}) = S^{\text{EMP}}(\pi_{\text{train}}; \underline{x}, \underline{y}),$$

$$\pi^* = \arg\max_{\pi \in [\hat{\pi}, \pi_{\text{train}}]} S(\pi; \underline{x}, \underline{y}).$$