Abductive Inference of Grammatical Constraints

Jon Rawski

SAN JOSÉ STATE UNIVERSITY

Jane Chandlee
(Haverford)

Remi Eyraud
(Aix-Marseille)

Jeff Heinz
(Stony Brook)

Adam Jardine
(Rutgers)

Overall Summary

Road Map

- Is statistical induction necessary for phonotactic learning?
- No. I will show a series of nonstatistical feature-based phonotactic algorithms
- I show they are provably correct and compare them on corpora to MaxEnt learners

Make Takeaways

- the space of possible constraints possesses significant structure (a partial order) that learners can easily exploit
- the learner akways faces multiple pairwise incomparable grammars which are surface-true
- particular constraint selection is due to the particular abductive (not inductive) principles

Phonotactics (Halle 1978)

Speakers' knowledge of possible and impossible sequences

ptak, **thole**, hlad, **plast**, sram, mgla, vlas, **flitch**, dnom, rtut

Infants show early sensitivity to phonotactic patterns (5-9mo)

- (Friederici & Wessels, 1993; Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993,Jusczyk, Luce, & Charles-Luce, 1994, Sundara & Breiss 2020)

Advantages:

- Phonotactic structures are robustly observable
- Easy access to robust corpora
- Inductive phonotactic learning algorithms for comparisons

Finite Models

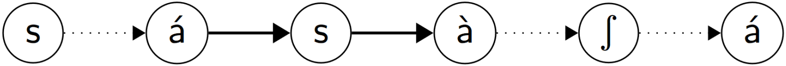
'model' is synonymous with 'structure.'

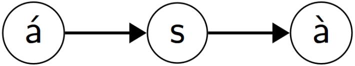
- A model of a linguistic form is a representation of it.
- A (Relational) Model contains two kinds of elements.
  - A domain: a finite set of elements.
  - Relations over domain elements.
- Every structure has a model.
- Different structures have different models.

Factors

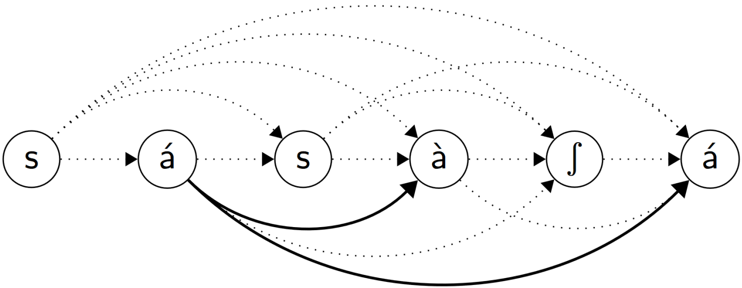
- A connected substructure A of some model B
- Properties that hold of the connected structure A also hold within B.

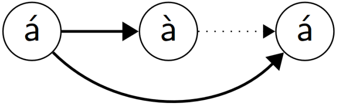
Finite Linguistic Models (Lambert, Rawski, Heinz 2021)



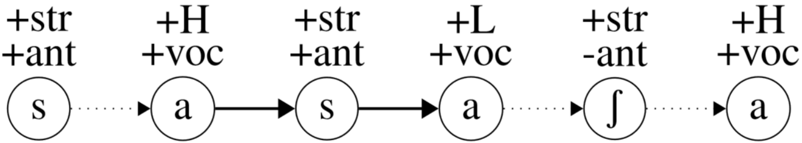


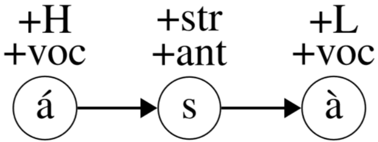
Finite Linguistic Models (Lambert, Rawski, Heinz 2021)



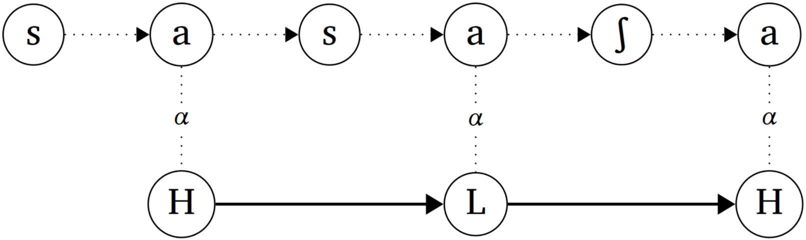


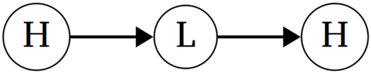
Finite Linguistic Models (Lambert, Rawski, Heinz 2021)



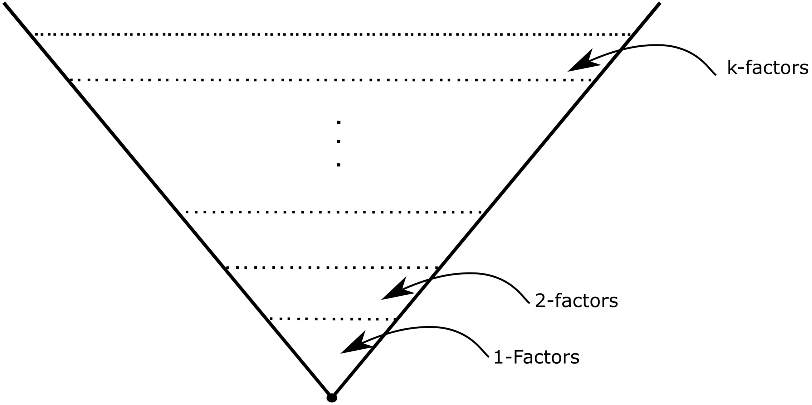


Finite Linguistic Models (Lambert, Rawski, Heinz 2021)





Grammar Space Forms a Partial Order



The Challenge of Factors with Features

With n features, we can potentially distinguish 2^n symbols in a conventional model.

The Selection Problem: Hayes & Wilson 2008

"one still faces a formidable difficulty: the fact that an enormous number of distributional generalizations are consistent with any given set of surface forms."

Wilson & Gallagher 2018

"What about... a nonstatistical model ... that learns by memorizing feature sequences ... ? The immediate problem confronting such a model is that any given segment sequence has multiple different featural representations."

Example

Imagine the sequence *nt* is not present in a corpus. There are many possible equivalent forbidden factors:

*nt
*[+nasal][+coronal]
*[+consonant][+coronal,-continuant]
*[+sonorant][-sonorant]
....

How can a learner decide which of these constraints is responsible for the absence of *nt*?

For example, the attested dorsal-tier trigram [oqa] could be represented

- with very general classes (e.g., [+syll][-syll][+syll] = VCV),
- with maximally specific classes (i.e., [+syll, -high, -low, +back][-cont, -son, +dorsal, -high, -cg][+syll, -high, +low] = [oqa]),
- or at intermediate levels of granularity (e.g., [+ syll, -high, -low][- cont, -son, +dorsal, -high][+ syll, -high, +low] = EQA).

Hayes & Wilson 2008

> *To solve the selection problem, we assume that UG determines the feature inventory and the format of constraints, yielding a search space that is quite large and hence compatible with the inductive baseline approach. Nevertheless, in our experience it is effectively searchable, provided the right search heuristics are used*

MaxEnt models: define a feature system and constraint format, input a corpus, and use statstical heuristics to select weighted factors via hill-climbing

Wilson & Gallagher again

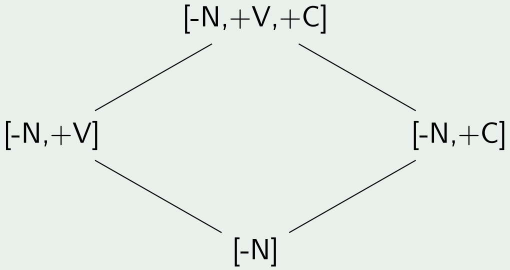
"If a hypothetical [model] judged a substring to be legal as long as it satisfied any attested featural description, it would tolerate (among other structures) every VCV trigram and thus massively overgeneralize. If the model instead required all feature representations of a substring to be attested, it would be equivalent to [memorizing segmental trigrams] . . . Lacking a method for deciding which representations are relevant for assessing well-formedness—precisely the role played by statistics [here]—learning ...is doomed.

Grammatical Entailments

Factor Ideals

If $\mathbf{s}$ is a factor of $\mathbf{t}$ for grammar G and G generates $\mathbf{t}$, then G generates $\mathbf{s}$.

Example

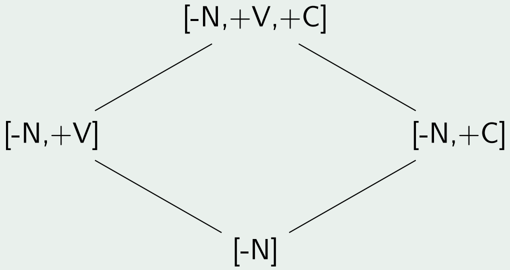


Grammatical Entailments

Factor Ideals

If $\mathbf{s}$ is a factor of $\mathbf{t}$ for grammar G and G generates $\mathbf{t}$,
then G generates $\mathbf{s}$.

Example

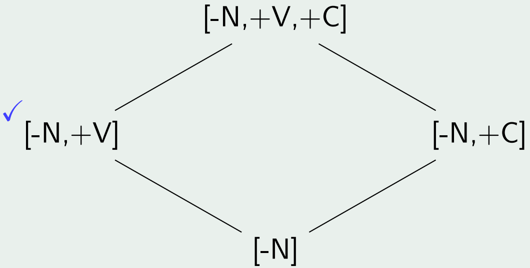


Grammatical Entailments

Factor Ideals

If $\mathbf{s}$ is a factor of $\mathbf{t}$ for grammar G and G generates $\mathbf{t}$, then G generates $\mathbf{s}$.

Example

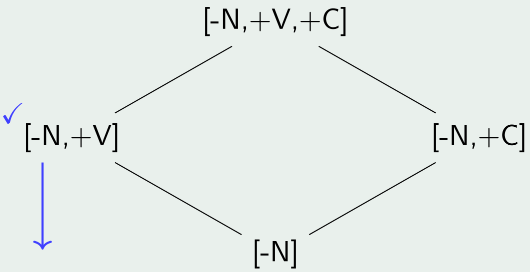


Grammatical Entailments

Factor Ideals

If $\mathbf{s}$ is a factor of $\mathbf{t}$ for grammar G and G generates $\mathbf{t}$, then G generates $\mathbf{s}$.

Example

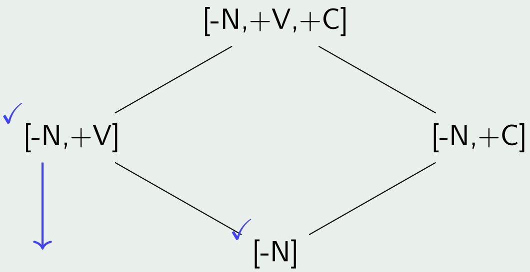


Grammatical Entailments

Factor Ideals

If $\mathbf{s}$ is a factor of $\mathbf{t}$ for grammar G and G generates $\mathbf{t}$, then G generates $\mathbf{s}$.

Example

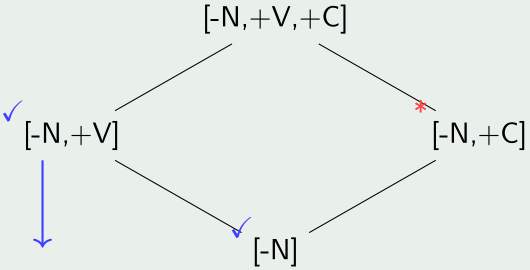


Grammatical Entailments

Factor Ideals

If **s** is a factor of **t** for grammar G and G generates **t**, then G generates **s**.

Example

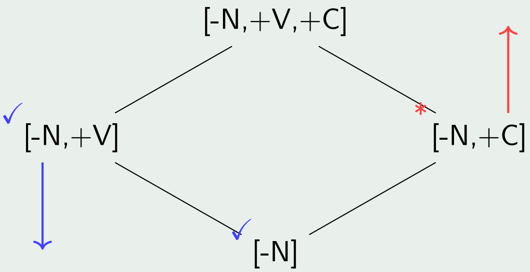


Grammatical Entailments

Factor Ideals

If **s** is a factor of **t** for grammar G and G generates **t**, then G generates **s**.

Example

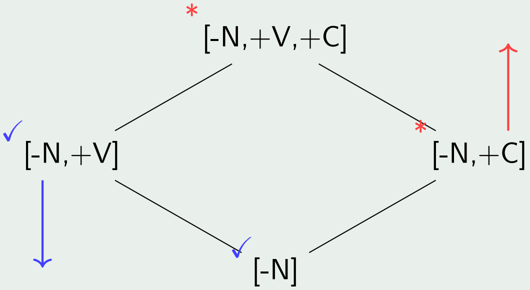


Grammatical Entailments

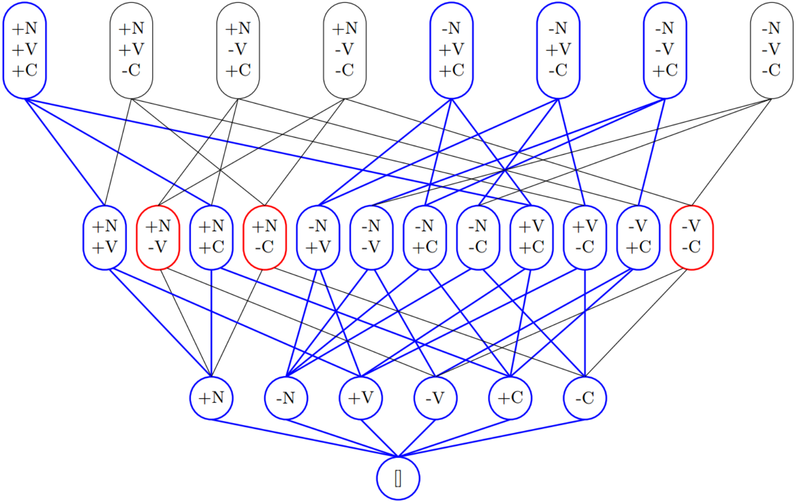
Factor Ideals

If $\mathbf{s}$ is a factor of $\mathbf{t}$ for grammar G and G generates $\mathbf{t}$, then G generates $\mathbf{s}$.

Example

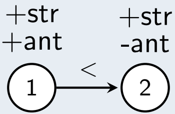


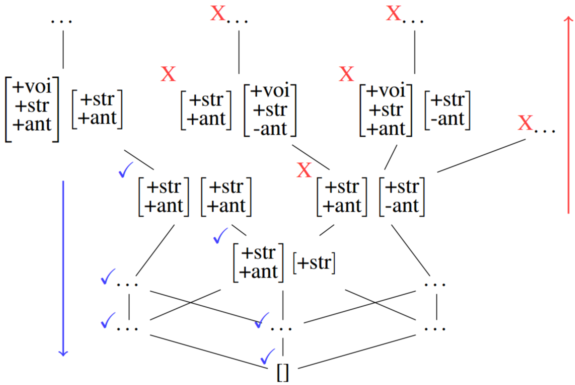
Example with Singular Segments



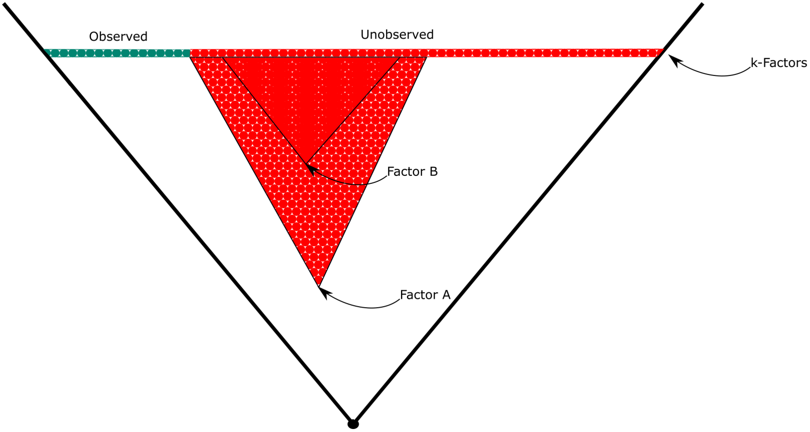
Example: Samala Long-Distance Sibilant Harmony

Banned Structure





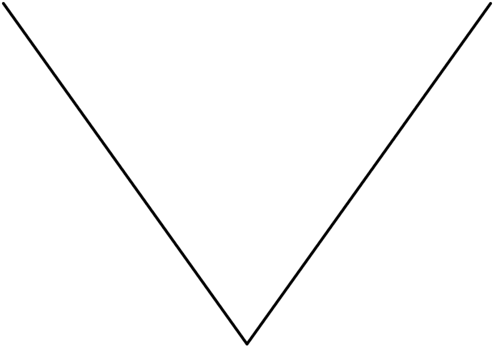
The Learning Problem



Chandlee et al 2019

Bottom-Up Factor Inference Algorithm (BUFIA)

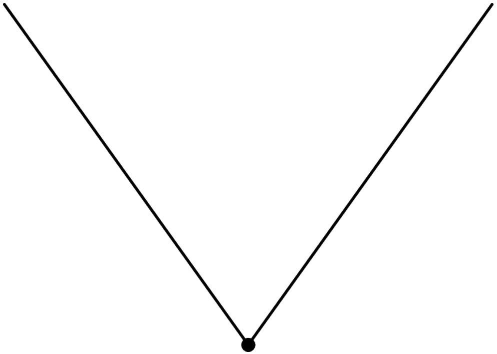
- Prunes Factor space according to ordering relation
- Provably identifies most general correct constraints for data



Chandlee et al 2019

Bottom-Up Factor Inference Algorithm (BUFIA)

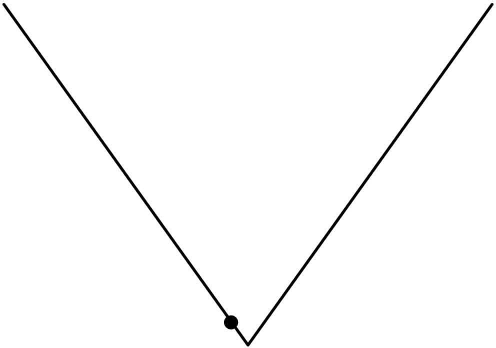
- Prunes Factor space according to ordering relation
- Provably identifies most general correct constraints for data



Chandlee et al 2019

Bottom-Up Factor Inference Algorithm (BUFIA)

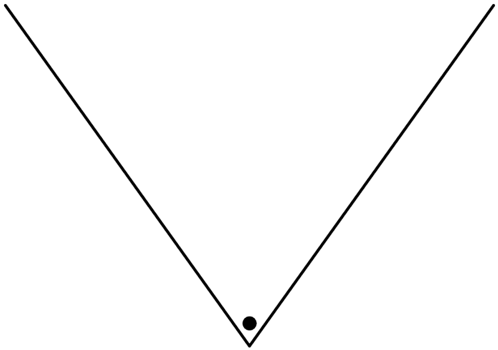
- Prunes Factor space according to ordering relation
- Provably identifies most general correct constraints for data



Chandlee et al 2019

Bottom-Up Factor Inference Algorithm (BUFIA)

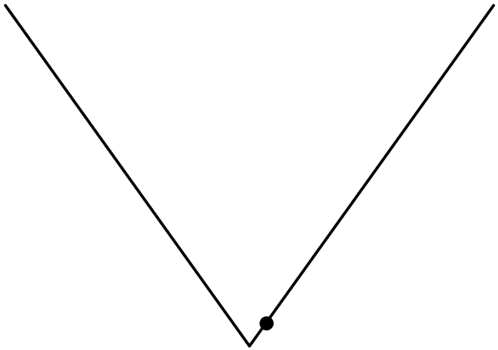
- Prunes Factor space according to ordering relation
- Provably identifies most general correct constraints for data



Chandlee et al 2019

Bottom-Up Factor Inference Algorithm (BUFIA)

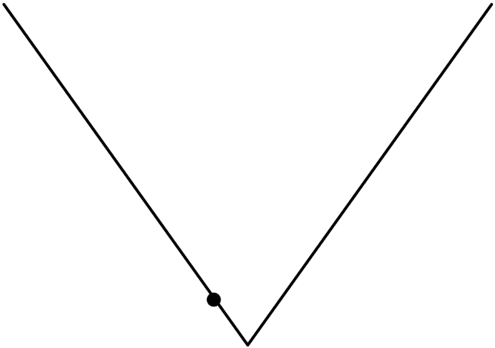
- Prunes Factor space according to ordering relation
- Provably identifies most general correct constraints for data



Chandlee et al 2019

Bottom-Up Factor Inference Algorithm (BUFIA)

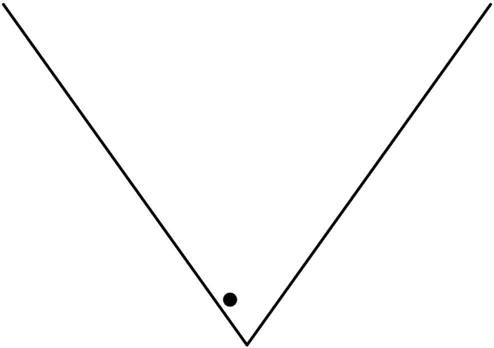
- Prunes Factor space according to ordering relation
- Provably identifies most general correct constraints for data



Chandlee et al 2019

Bottom-Up Factor Inference Algorithm (BUFIA)

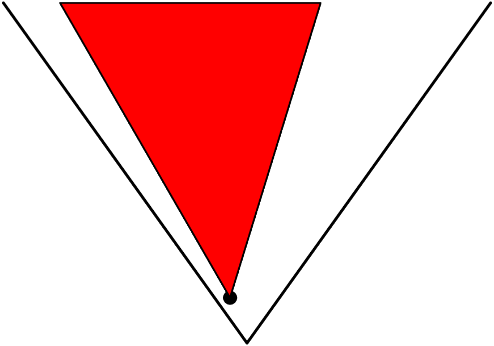
- Prunes Factor space according to ordering relation
- Provably identifies most general correct constraints for data



Chandlee et al 2019

Bottom-Up Factor Inference Algorithm (BUFIA)

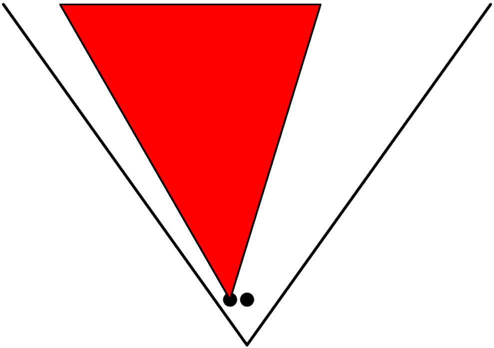
- Prunes Factor space according to ordering relation
- Provably identifies most general correct constraints for data



Chandlee et al 2019

Bottom-Up Factor Inference Algorithm (BUFIA)

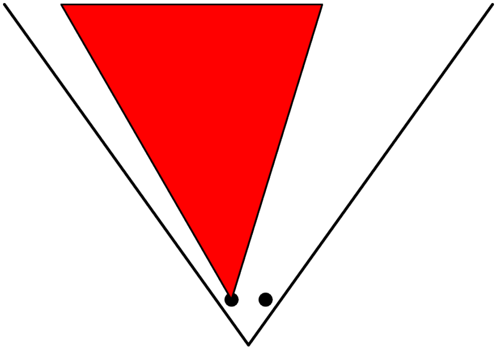
- Prunes Factor space according to ordering relation
- Provably identifies most general correct constraints for data



Chandlee et al 2019

Bottom-Up Factor Inference Algorithm (BUFIA)

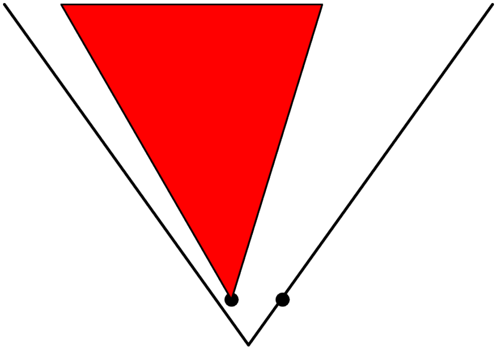
- Prunes Factor space according to ordering relation
- Provably identifies most general correct constraints for data



Chandlee et al 2019

Bottom-Up Factor Inference Algorithm (BUFIA)

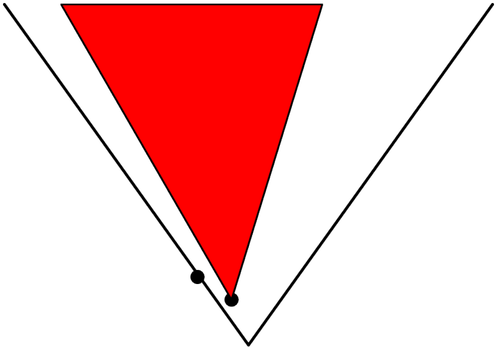
- Prunes Factor space according to ordering relation
- Provably identifies most general correct constraints for data



Chandlee et al 2019

Bottom-Up Factor Inference Algorithm (BUFIA)

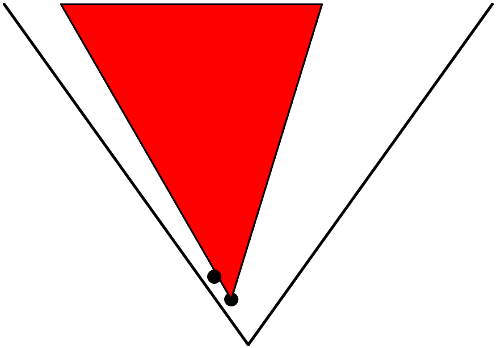
- Prunes Factor space according to ordering relation
- Provably identifies most general correct constraints for data



Chandlee et al 2019

Bottom-Up Factor Inference Algorithm (BUFIA)

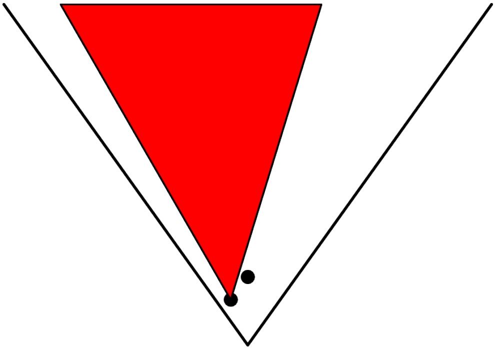
- Prunes Factor space according to ordering relation
- Provably identifies most general correct constraints for data



Chandlee et al 2019

Bottom-Up Factor Inference Algorithm (BUFIA)

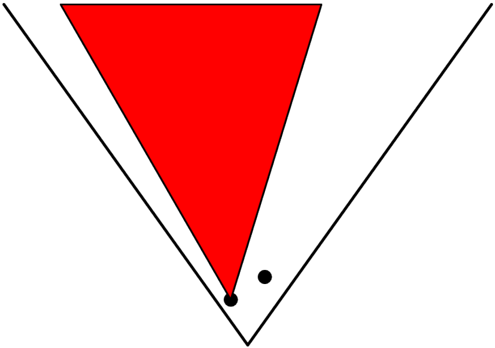
- Prunes Factor space according to ordering relation
- Provably identifies most general correct constraints for data



Chandlee et al 2019

Bottom-Up Factor Inference Algorithm (BUFIA)

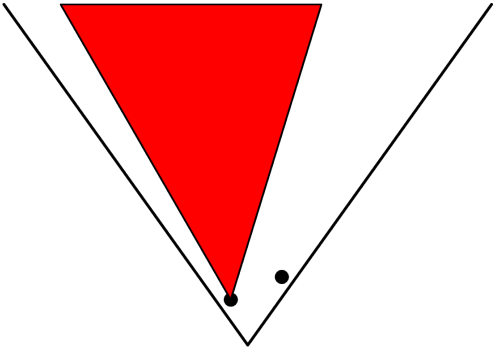
- Prunes Factor space according to ordering relation
- Provably identifies most general correct constraints for data



Chandlee et al 2019

Bottom-Up Factor Inference Algorithm (BUFIA)

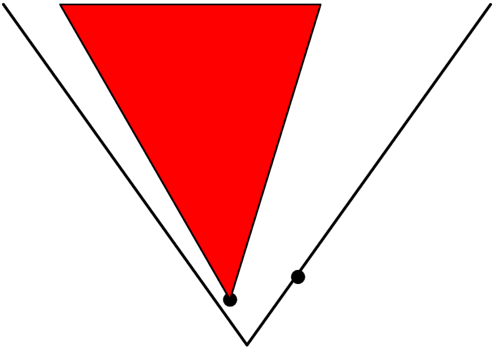
- Prunes Factor space according to ordering relation
- Provably identifies most general correct constraints for data



Chandlee et al 2019

Bottom-Up Factor Inference Algorithm (BUFIA)

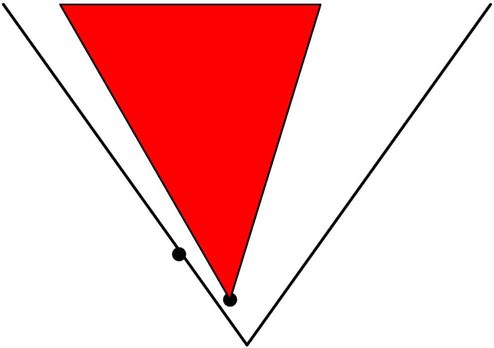
- Prunes Factor space according to ordering relation
- Provably identifies most general correct constraints for data



Chandlee et al 2019

Bottom-Up Factor Inference Algorithm (BUFIA)

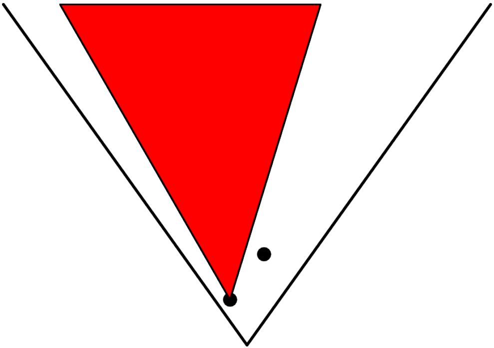
- Prunes Factor space according to ordering relation
- Provably identifies most general correct constraints for data



Chandlee et al 2019

Bottom-Up Factor Inference Algorithm (BUFIA)

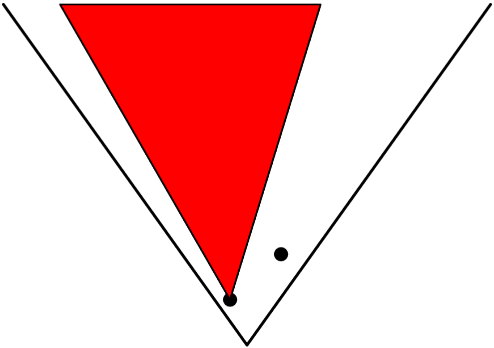
- Prunes Factor space according to ordering relation
- Provably identifies most general correct constraints for data



Chandlee et al 2019

Bottom-Up Factor Inference Algorithm (BUFIA)

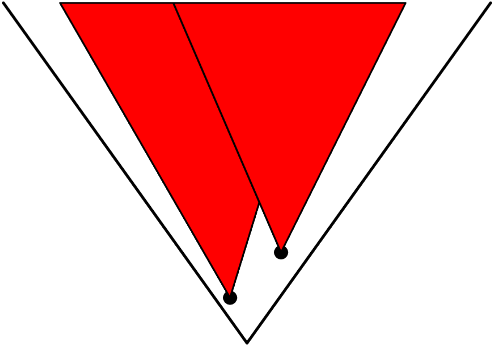
- Prunes Factor space according to ordering relation
- Provably identifies most general correct constraints for data



Chandlee et al 2019

Bottom-Up Factor Inference Algorithm (BUFIA)

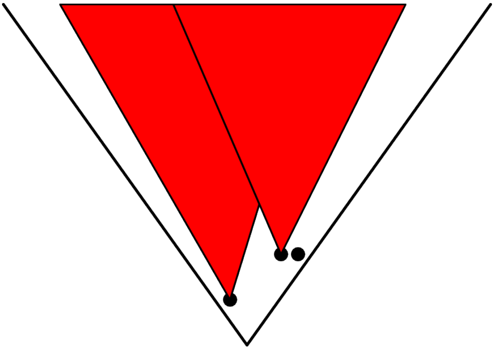
- Prunes Factor space according to ordering relation
- Provably identifies most general correct constraints for data



Chandlee et al 2019

Bottom-Up Factor Inference Algorithm (BUFIA)

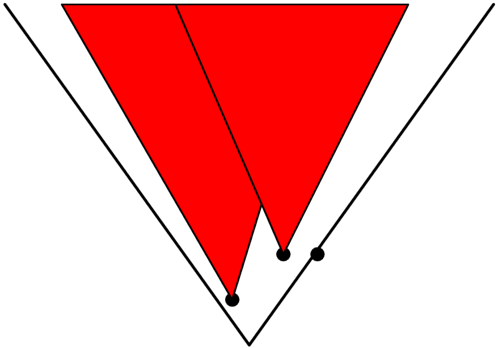
- Prunes Factor space according to ordering relation
- Provably identifies most general correct constraints for data



Chandlee et al 2019

Bottom-Up Factor Inference Algorithm (BUFIA)

- Prunes Factor space according to ordering relation
- Provably identifies most general correct constraints for data



Learning Guarantees

Theorem

Given a finite positive data sample, BUFIA finds a factor-based grammar G such that:

1. *G is consistent, i.e. it covers the data:*
   - $D \subseteq L(G)$
2. *$L(G)$ is the smallest language in $\mathscr{L}$ which covers the data*
   - *for all $L \in \mathscr{L}$ where $D \subseteq L$, $L(G) \subseteq L$*
3. *each factor size $\leq k$*
4. *G includes structures S that are restrictions of structures S' included in other grammars G' that also satisfy (1,2,3)*
   - *for all $S' \in G'$, there exists $S \in G$ such that $S \sqsubseteq S'$.*

Toy Example: BUFIA vs UCLA MaxEnt Learner

	i	u	e	o	a
high	+	+	−	−	−
back	−	+	−	+	−
low	−	−	−	−	+

Input Data: $\{ii, aa\}$

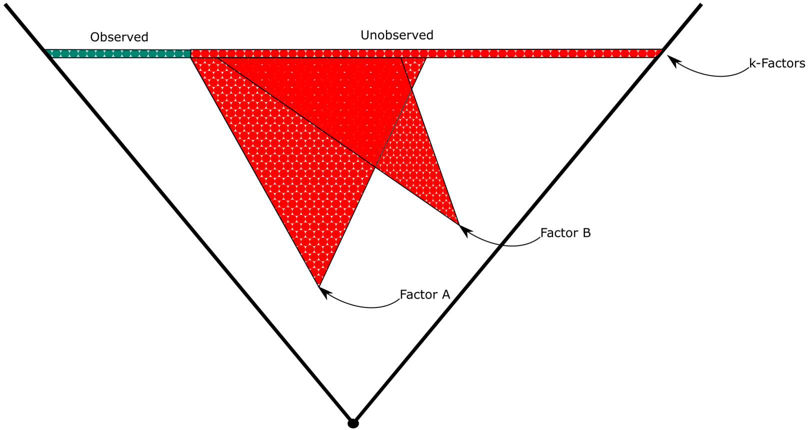
BUFIA	UCLA Learner	MaxEnt Score
[+ back]	[+ back]	6.186
[-low, -high]	[-low, -high]	2.162
[-low][-high]	[-low][-high]	5.766
[-high] [-low]	[-high] [-low]	5.766
[+ low][+ high]		
[+high][−high]		
[−high][+high]		
[+low][+high]		
[+low][−low]		
[−low][−low]		

Induction vs Abduction

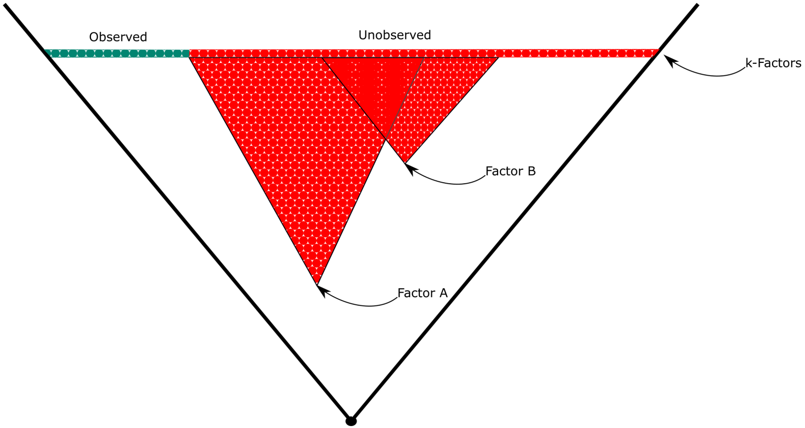
> *"In a very real sense, stating the problem is half the solution!" – Haig 1982*

Induction: inference via data-driven model-selection
Abduction: inference via intersecting constraint-driven hypotheses restrictions
Problems as Constraints on possible solutions
"inference to the best (loveliest) explanation"

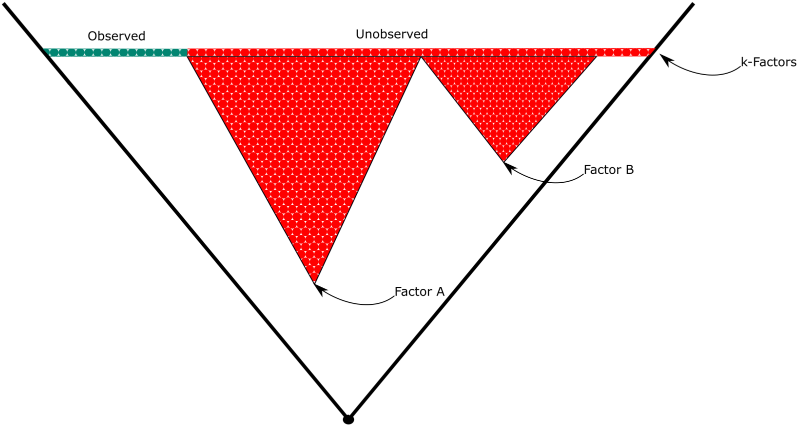
Induction to Abduction



Induction to Abduction



Induction to Abduction



Adding Abductive Constraints

BUFIA + Principle 1	BUFIA + Principle 2
[+ back]	[+ back]
[-low, -high]	[-back][−back, -high, -low]
[+high][−high]	[+low][-back, +high]
[−high][+high]	[−back, -high, -low][+low]
	[−back, -high, -low][−back, +high]

- Why are the second two constraints different from MaxEnt?
- Answer: features proposed alphabetically

Abductive Inference of Phonotactic Constraints

- BUFIA $\rightarrow$ BUFIA + Principle 1
- Forbidden factors up to size 2

Experiment 1: Hayes and Wilson's CMU initial cluster data

- 469 $\rightarrow$ 32 (feature-based successor model)

Experiment 2: Gallagher's Quechua roots (p.c.)

- 1913 $\rightarrow$ 89 (feature-based successor model)
- 320 $\rightarrow$ 22 (feature-based precedence model)
- Example P1 factors: [+syll][+syll], [+low][+low]

Factor Explosion (Hayes and Wilson 2008)

As we add segments and features, the amount of possible hypotheses grows larger. How much larger?

Table 2
Number of possible constraints for various values of $|C|$ and n

		$\lvert C \rvert$			
		30	**100**	**200**	**400**
	1	30	100	200	400
	2	900	10,000	40,000	160,000
n	**3**	27,000	1,000,000	8 million	64 million
	4	810,000	100 million	1.6 billion	26 billion
	5	24 million	10 billion	320 billion	10 trillion

$|C|$ is the number of natural classes and
n is the length of the constraint.

Many Ways to Extend This Work

- Run BUFIA with different abductive principles on different corpora (van Rooij & Baggo 2021).
- Compare BUFIA outputs to learner judgments (Durvasula & Liter, Durvasula 2020 AMP)
- Explore grammar learning with BUFIA with other kinds of representations (autosegmental, syllable structure)
- Combine BUFIA with a syntactic parser for learning constraints on syntactic trees.

Conclusion

- Structure Is everywhere
- Representations structure grammars, which structure learning
- Statistics and Structure play off Each other

Thanks!

References I

Applegate, R.B. 1972. *Ineseno chumash grammar*. Doctoral Dissertation, University of California,Berkeley.