



Subregular Induction of Underlying Representations and a Phonological Grammar

Wenyue Hua, Adam Jardine, and Huteng Dai

{wenyue.hua,adam.jardine,huteng.dai}@rutgers.edu

Department of Linguistics, Rutgers University

Goals and Background

- Project goal: the **simultaneous inference** of URs and a grammar from SRs in a morphological paradigm.
(Albright, 2002; Tesar, 2014)
- The **Input Strictly Local** (ISL) functions provide a structure that can solve this problem.
(Chandlee and Heinz, 2018)

Primary result

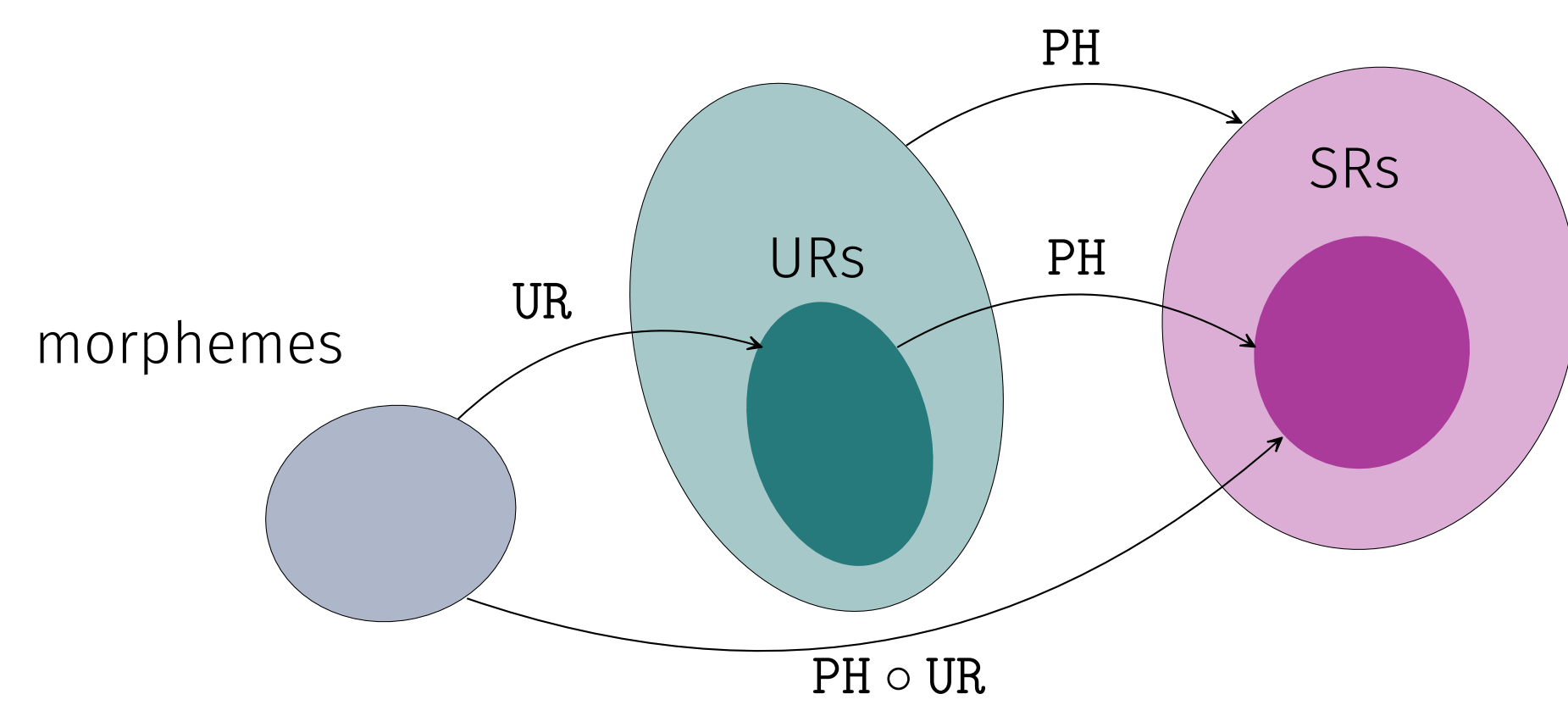
- The learner induces UR and phonological grammar from a range of ISL₂ functions (ISL function for $k = 2$), including progressive and regressive assimilation, deletion, epenthesis, and opacity.

Learning Problem

- M : finite set of morphemes $\{\text{CAT, DOG, ..., PL}\}$
 Σ : finite set of segments $\{a, b, \beta, ..., z\}$
- UR function: maps one morpheme to one UR;
 $\text{UR} : M^* \rightarrow \Sigma^*$
- PH function: maps URs to SRs;

$$\text{PH} : \Sigma^* \rightarrow \Sigma^*$$

$\text{UR}(\text{CAT}) = \text{kæt}$	$\text{PH}(\text{kæt}) = \text{kæt}$
$\text{UR}(\text{PL}) = \text{z}$	$\text{PH}(\text{dɔgz}) = \text{dɔgz}$
$\text{UR}(\text{CAT-PL}) = \text{kætz}$	$\text{PH}(\text{kætz}) = \text{kæts}$
...	...
$\text{PH}(\text{bnɪkz}) = \text{bnɪks}$	



- Given a finite sample of $\text{PH} \circ \text{UR}$, how do we identify PH and UR?

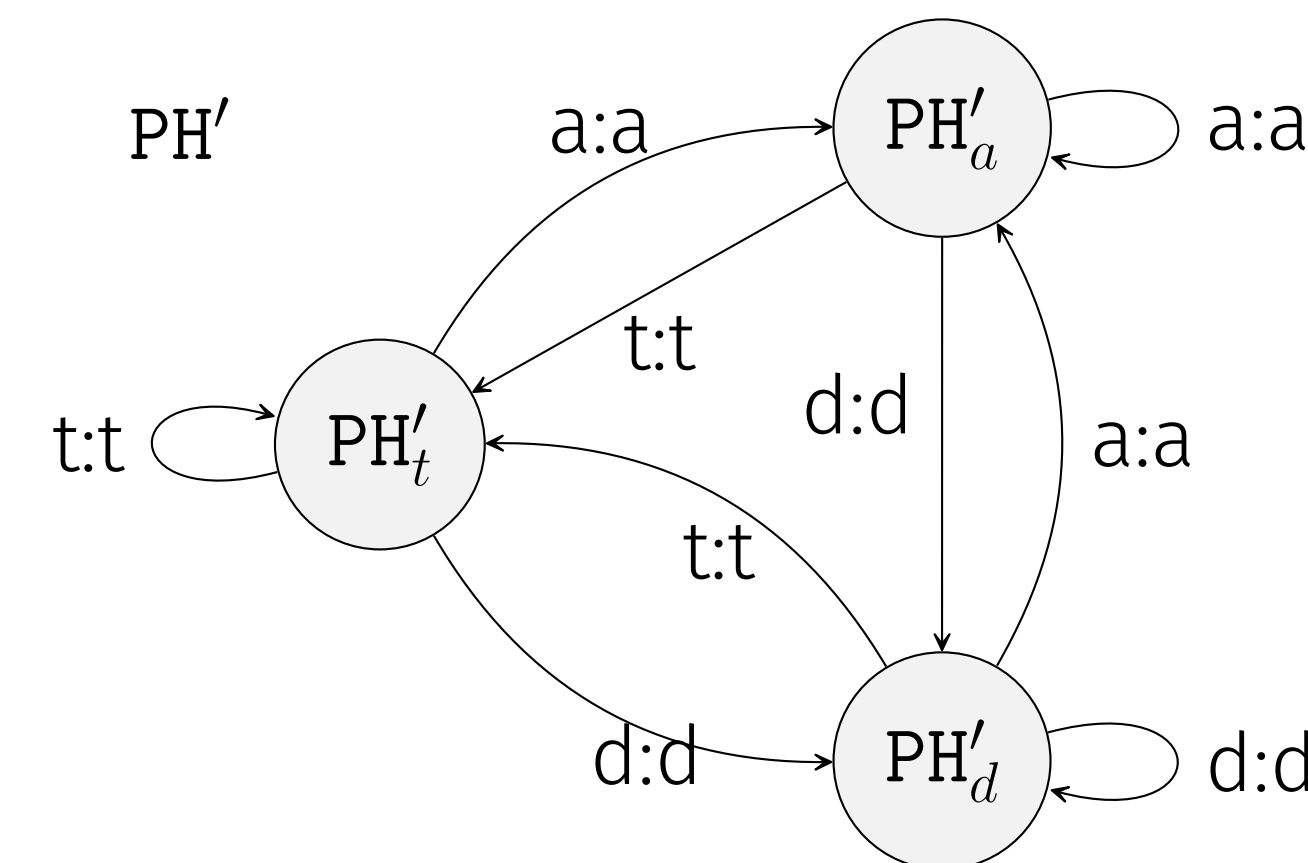
$(\text{CAT-PL, kæt s}), (\text{DOG-PL, dɔgz}), \dots, (\text{BOOK-PL, bʊks})$

Initialization

- Running example $D \subset \text{PH} \circ \text{UR}(M^*)$

Sample of $\text{PH} \circ \text{UR}$ (PROG. ASSIMILATION)					
w	$\text{PH}(\text{UR}(w))$	w	$\text{PH}(\text{UR}(w))$	w	$\text{PH}(\text{UR}(w))$
$r_1 s_1$	tatta	$r_2 s_1$	tadda	$r_3 s_1$	ata
$r_1 s_2$	tatda	$r_2 s_2$	tadda	$r_3 s_2$	ada
$r_1 s_3$	tata	$r_2 s_3$	tada	$r_3 s_3$	aa

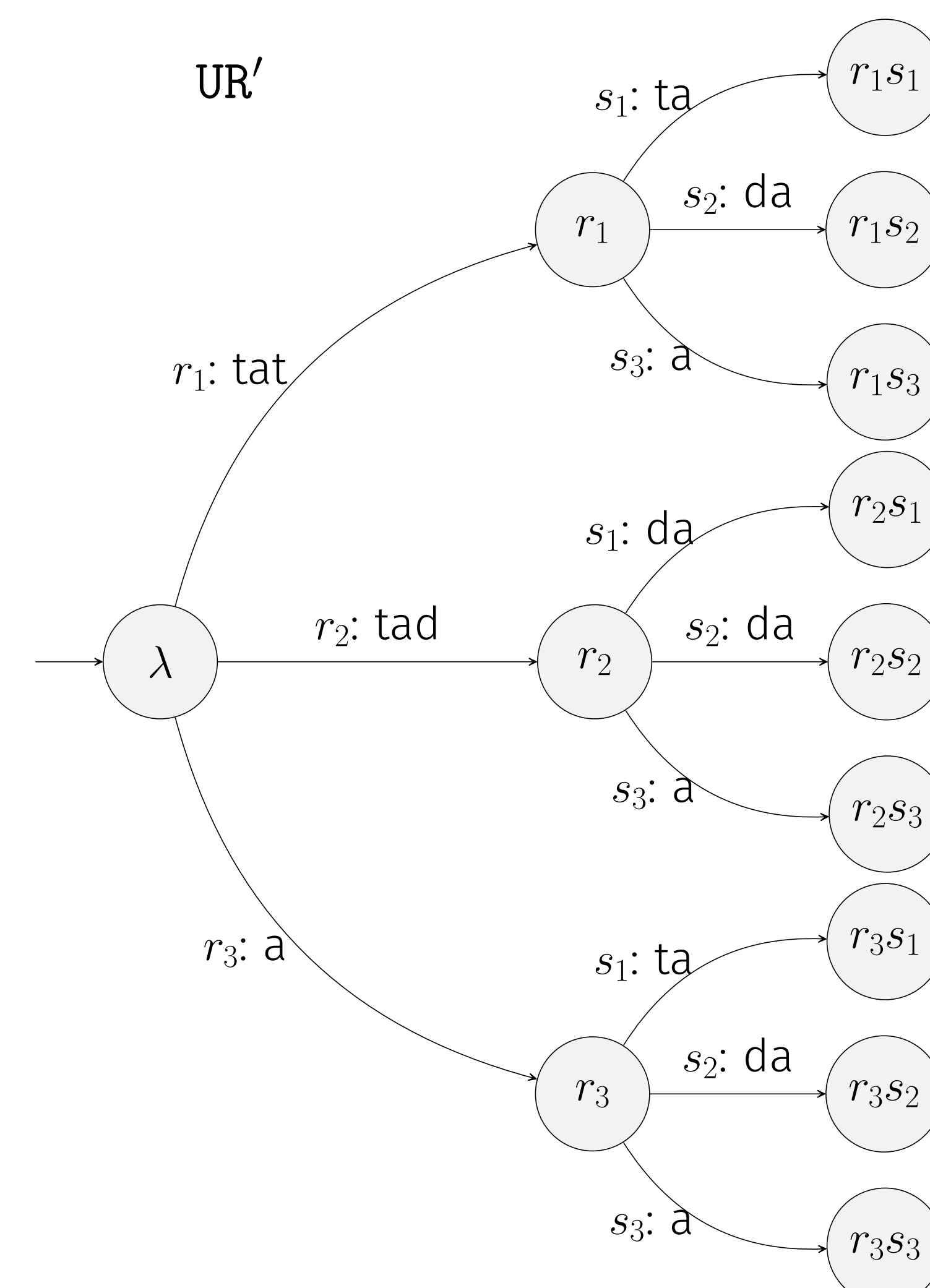
- Initialize PH' to the identity function
 $\text{PH}'(\text{tatta}) = \text{tatta}$, $\text{PH}'(\text{tatda}) = \text{tatda}$, etc.



- Initialize UR' to a **prefix tree transducer** representing D : segmentation based on **longest common prefix (lcp)**.

$$\text{lcp}(\{\text{tatta}, \text{tatda}, \text{tata}\}) = \text{tat}$$

$$\text{lcp}(\{\text{tadda}, \text{tada}\}) = \text{tad}$$



Inconsistency detection

If a morpheme is mapped to multiple SRs, the learner detects this inconsistency.

$$r_1: \text{tat} \quad r_2: \text{tad} \quad r_3: \text{a}$$

$$s_1: \text{ta, da} \quad s_2: \text{da} \quad s_3: \text{a}$$

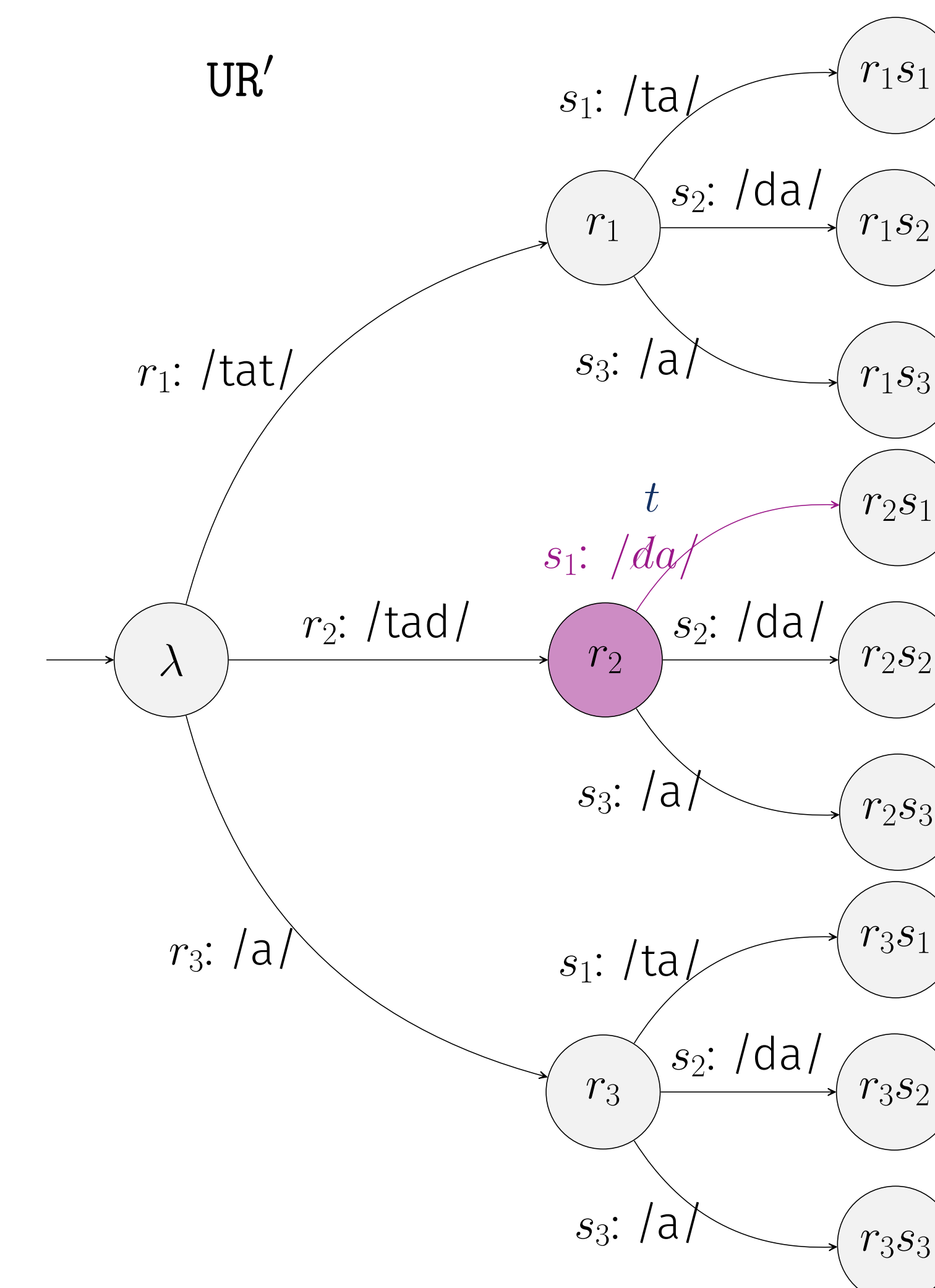
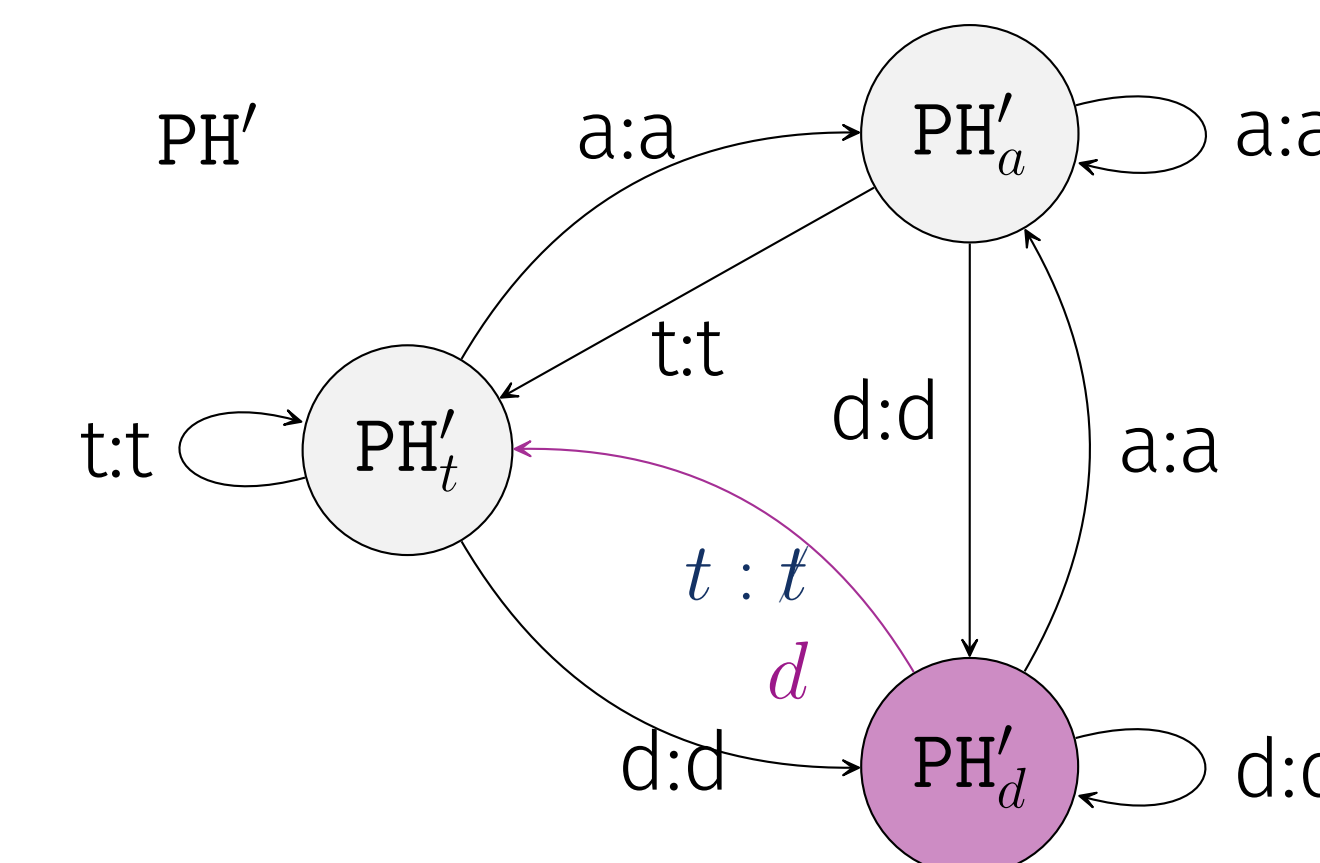
Environment collection

- ISL provides environment information
- t is the most **informative** form $\rightarrow \text{UR}$.

ws_1	env.	s_1
$r_1 s_1$	tat	ta
$r_3 s_1$	a	ta
$r_2 s_1$	tad	da

Modification

Change UR' , making the opposite change in PH'



Take-home message

From an **abstract** and **principled** perspective, learning is possible given the basic principles:

- a restrictive, structured hypothesis space
- complementarily distributed allomorphs
- a surface-driven set of URs
- One morpheme \rightarrow one UR

Future work

- Long-distance processes can be captured by different classes of subsequential functions with a similar structure;
- One example: output strictly-local class also has a restricted state structure;
(Chandlee et al., 2015)

- Abstract URs may be learnable when input alphabet is larger than output alphabet (and thus allows larger categories).

Selected References

Albright, Adam C (2002). *The identification of bases in morphological paradigms*. PhD thesis, University of California, Los Angeles.

Chandlee, Jane, Eyraud, Rémi, and Heinz, Jeffrey (2015). Output strictly local functions. In Kornai, Andras and Kuhlmann, Marco, editors, *Proceedings of the 14th Meeting on the Mathematics of Language (MoL 14)*, pages 52–63, Chicago, IL.

Chandlee, Jane and Heinz, Jeffrey (2018). Strict locality and phonological maps. *Linguistic Inquiry*, 49(1):23–60.

Tesar, Bruce (2014). *Output-driven phonology: Theory and learning*. Cambridge University Press.

Acknowledgements: We thank attendees of the Rutgers MathLing group, NECPhon 2019, and in particular Jeff Heinz, Charles Reiss, Bruce Tesar, Adam McCollum, and Colin Wilson for their insightful comments.



Take a picture to download the poster