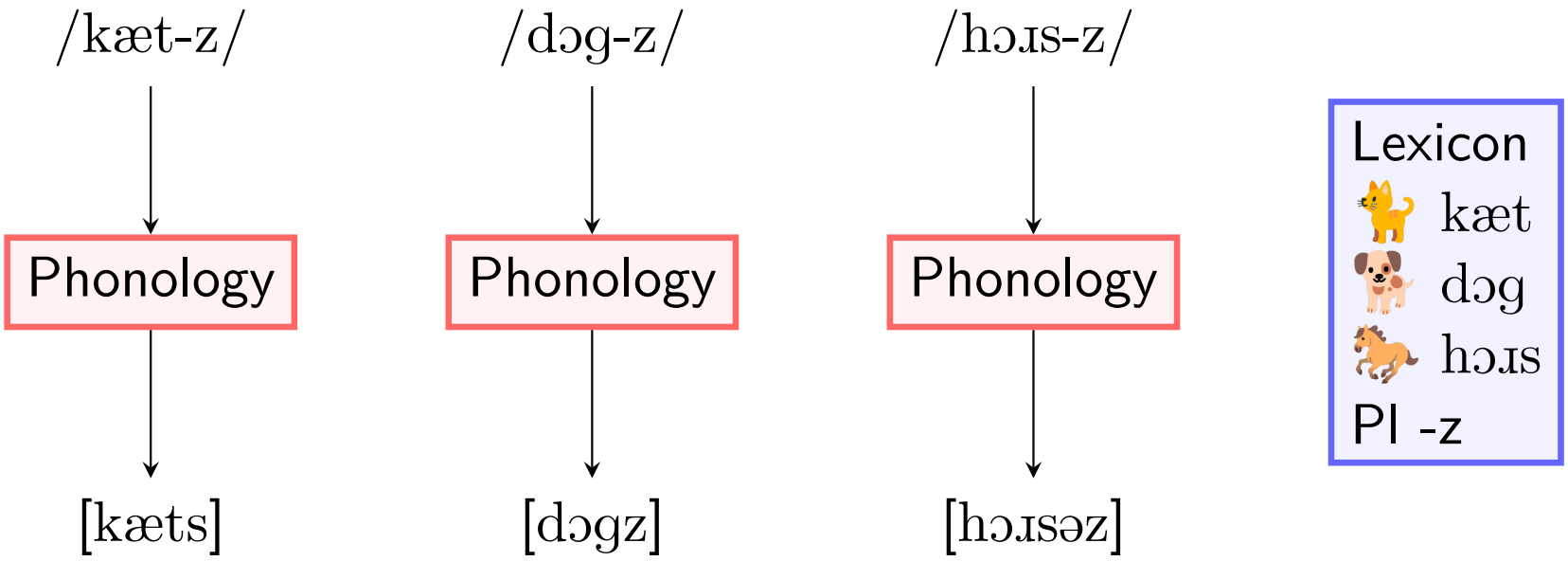
Learning underlying representations via function decomposition

Jane Chandlee

Haverford College

PLC 48
17 March 2024

Underlying representations: a hidden structure problem



Previous work on learning URs

OT	Maximum Likelihood	Jarosz (2006a,b, 2009, 2015)
	GLA w/ UR constraints	Apoussidou (2007)
	Constraint Demotion w/ contrast pairs	Merchant (2008)
	CD w/ output-driven maps	Tesar (2014)
HG	MaxEnt	O'Hara (2017); Wang and Hayes (2022)
	MaxEnt w/ UR constraints	Pater et al. (2012); Nelson (2019)
SPE	Minimum Description Length	Rasin et al. (2020, 2021)

- Structured hypothesis space of grammars
- Rich conception of UG (e.g., innate CON)

Structural inference

- The invariant formal properties of phonological mappings also structure the hypothesis space in ways that can be leveraged by a learner.
- The phonological mapping is constant across theories of the grammar:
  - {(kætz, kæts), (dɔgz, dɔgz), (hɔɹsz, hɔɹsəz), ...}

Preliminaries

- **Morphology**, $\mathcal{M}$: orders sets of meanings
  - $\mathcal{M}(\{\text{PL}, \text{🐈}\}) = \text{🐈}\text{-PL}$
- **Lexicon**, $\mathcal{L}$: supplies the URs
  - $\mathcal{L}(\text{🐈}\text{-PL}) = /\text{kæt-z}/$
- **Phonology**, $\mathcal{P}$: maps URs to SRs
  - $\mathcal{P}(\text{kætz}) = [\text{kæts}]$

$$\mathcal{P} \circ \mathcal{L} \circ \mathcal{M}\ (\{\text{🐈}, \text{PL}\}) = [\text{kæts}]$$

(Adapted from Hua and Jardine (2021); see also Cotterell et al. (2015))

The learning problem

Given examples of $\mathcal{P} \circ \mathcal{L} \circ \mathcal{M}$, identify $\mathcal{M}$, $\mathcal{L}$, and $\mathcal{P}$.

- For today, though, the learning of $\mathcal{M}$ is being factored out and the learner is provided with explicit morphological knowledge.
- Learning approach: function decomposition (Hua et al., 2021)

$$\mathcal{P} \circ \mathcal{L}(\text{🐈Pl}) = [\text{kæts}]$$

$$\mathcal{P}(\mathcal{L}(\text{🐈Pl})) = \mathcal{P}(/\text{kætz}/) = [\text{kæts}]$$

Function decomposition

- In the general case, function decomposition is an unsolvable problem.

$$f(x) = x + 1$$
$$g(y) = y - 1$$
$$f \circ g(1) = 1, f \circ g(2) = 2, f \circ g(3) = 3...$$

- However, the learner capitalizes on the assumption that the phonological function is a **local function** (Berstel, 1982; Vaysse, 1986; Lind and Marcus, 1995; Chandlee, 2014).
- In addition, the learning paradigm is *semi-strong learning* (Hua and Jardine, 2021), rather than exact identification (Gold, 1967).

Malay alternations

Johor Malay (Austronesian; Peninsular Malaysia; Onn, 1976) /pəŋ-/ (nominalizer) and /məŋ-/ (active voice) both have five variants.

- Before obstruents: nasal place assimilation

/pəŋ-boroŋ/	[pəmboroŋ]	'wholesaler'
/pəŋ-jahit/	[pəɲjahit]	'tailor'
/pəŋ-daki/	[pəndaki]	'climber'
/pəŋ-gali/	[pəŋgali]	'digger'
/pəŋ-arah/	[pəŋarah]	'director'

- Before sonorant consonants: nasal deletion

/pəŋ-layan/	[pəlayan]	'waitress'
/pəŋ-ɲaɲi/	[pəɲaɲi]	'singer'
/pəŋ-malu/	[pəmalu]	'shame'
/pəŋ-rayu/	[pərayu]	'appeal'

Additional processes

- Final devoicing and regressive devoicing

/jawab/	[jawap]	'to answer'
/pəŋ-jawab-an/	[pəɲjawaban]	'the answering'
/məŋ-jawab-kan/	[məɲjawapkan]	'to cause to answer for'

- Velar stops become glottal in coda position

/masak/	[masaʔ]	'to cook'
/pəŋ-masak-an/	[pəmasakan]	'the cooking'
/məŋ-masak-kan/	[məmasaʔkan]	'to cause to cook for'

Additional processes

- /r/ deletes in coda position

/kisar/	[kisa]	‘revolve’
/kisar-an/	[kisaran]	‘revolution’
/kisar-kan/	[kisakan]	‘to cause to revolve for’

- Word-final /a/ reduces to schwa.

/bawa/	[bawə]	‘to carry’
/bawa-kan/	[bawakan]	‘to cause to carry for’

Data

A	/ikat/	‘tie’	B	/asut/	‘instigate’
D	/gali/	‘dig’	E	/bayar/	‘pay’
F	/bawa/	‘carry’	I	/dakap/	‘embrace’
L	/ladaŋ/	‘farm’	M	/naik/	‘ascend’
N	/jawab/	‘answer’	P	/cərca/	‘revile’
S	/rompak/	‘rob’	T	/ŋaŋa/	‘open’
U	/main/	‘play’	V	/ɲaɲi/	‘sing’

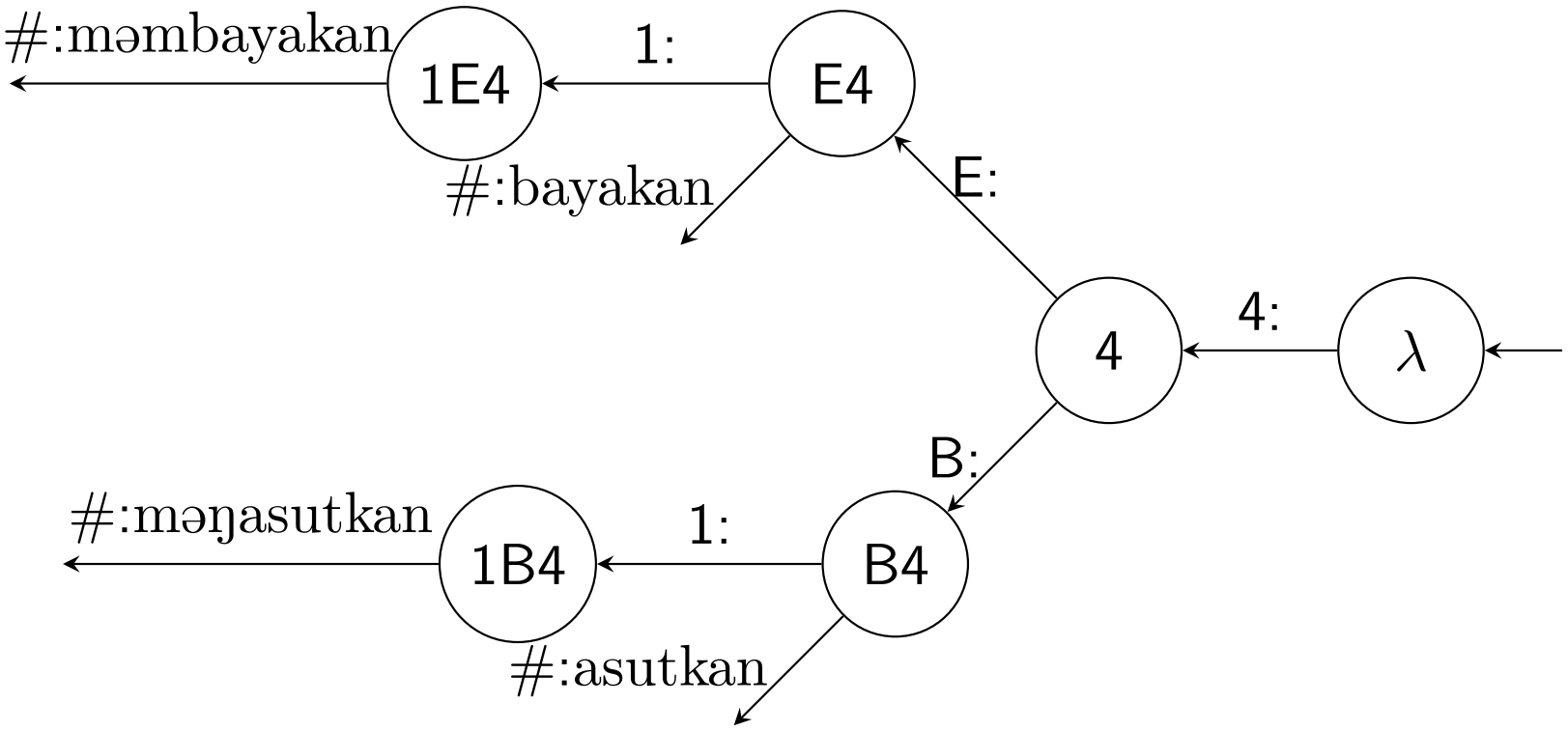
1	/məŋ-/	active
2	/pəŋ-/	nominalizer
3	/-i/	causative
4	/-kan/	causative benefactive
5	/-an/	nominalizer

Data

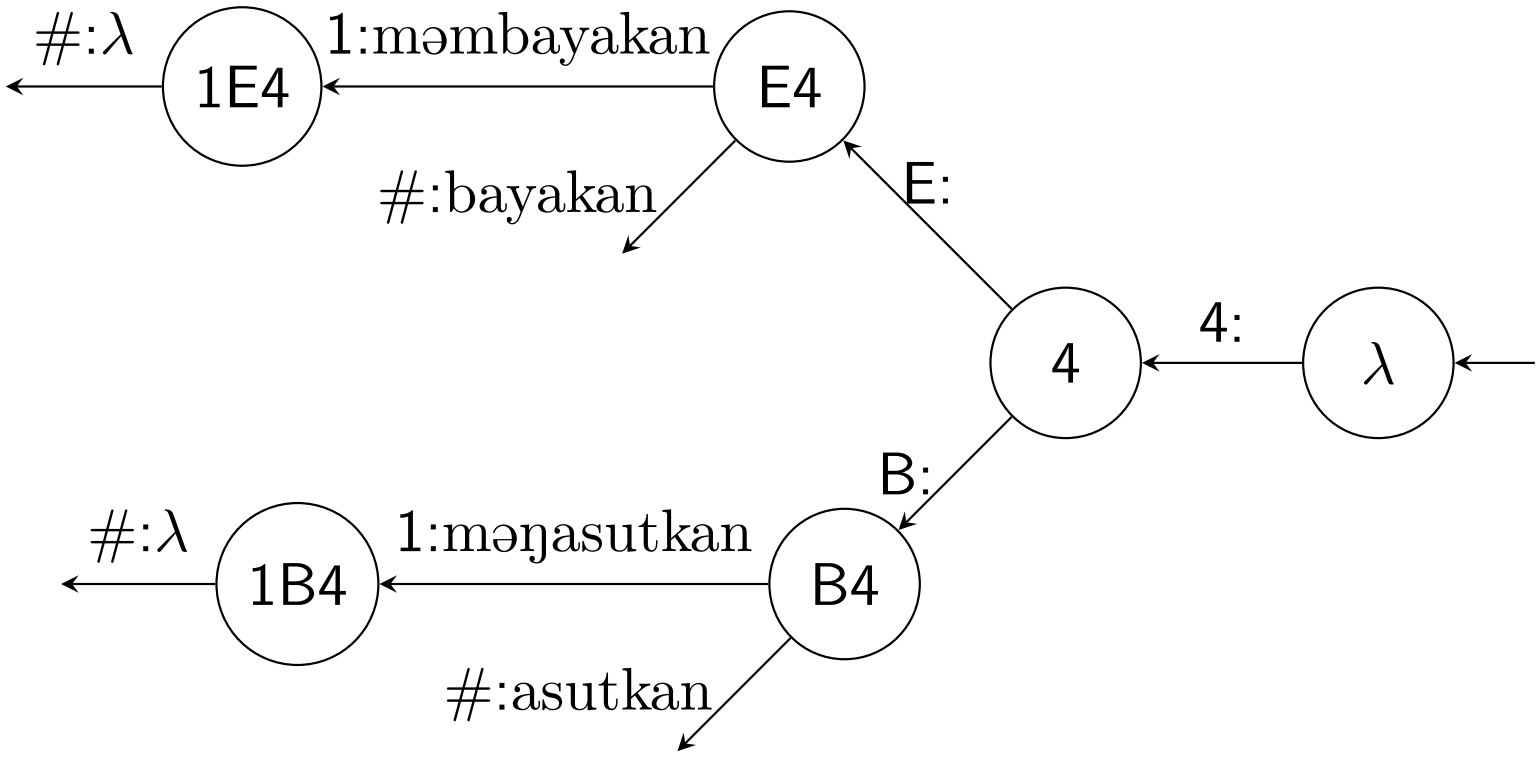
/bayar/	
baya	‘to pay’
məmbaya	‘to pay (active)’
bayari	‘to cause to pay’
bayakan	‘to cause to pay for’
məmbayari	‘to cause to pay (active)’
məmbayakan	‘to cause to pay for (active)’
pəmbaya	‘payment’

- {(E, baya), (1E, məmbaya), (E3, bayari), (E4, bayakan), (1E3, məmbayari), (1E4, məmbayakan), (2E, pəmbaya), ...}

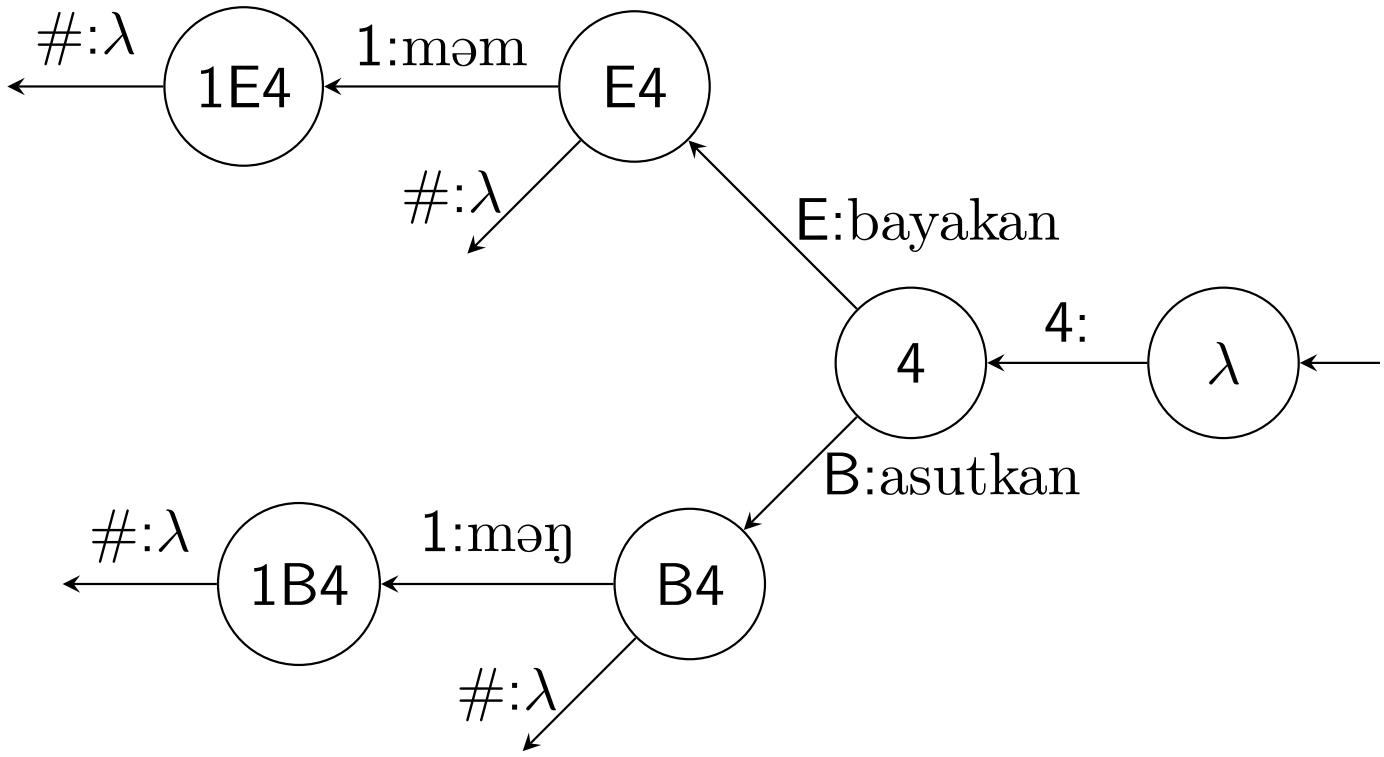
Morpheme segmentation: onward prefix tree transducer



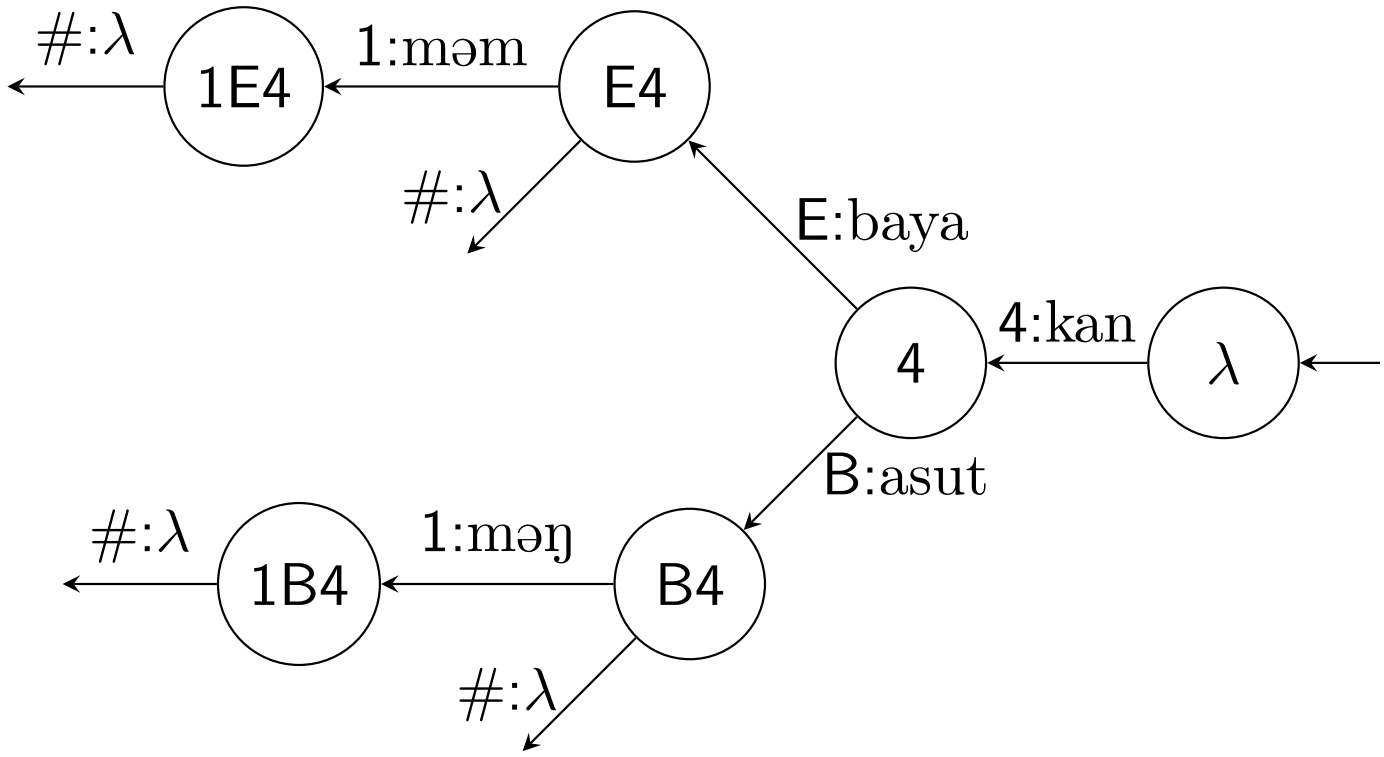
Morpheme segmentation: onward prefix tree transducer



Morpheme segmentation: onward prefix tree transducer



Morpheme segmentation: onward prefix tree transducer



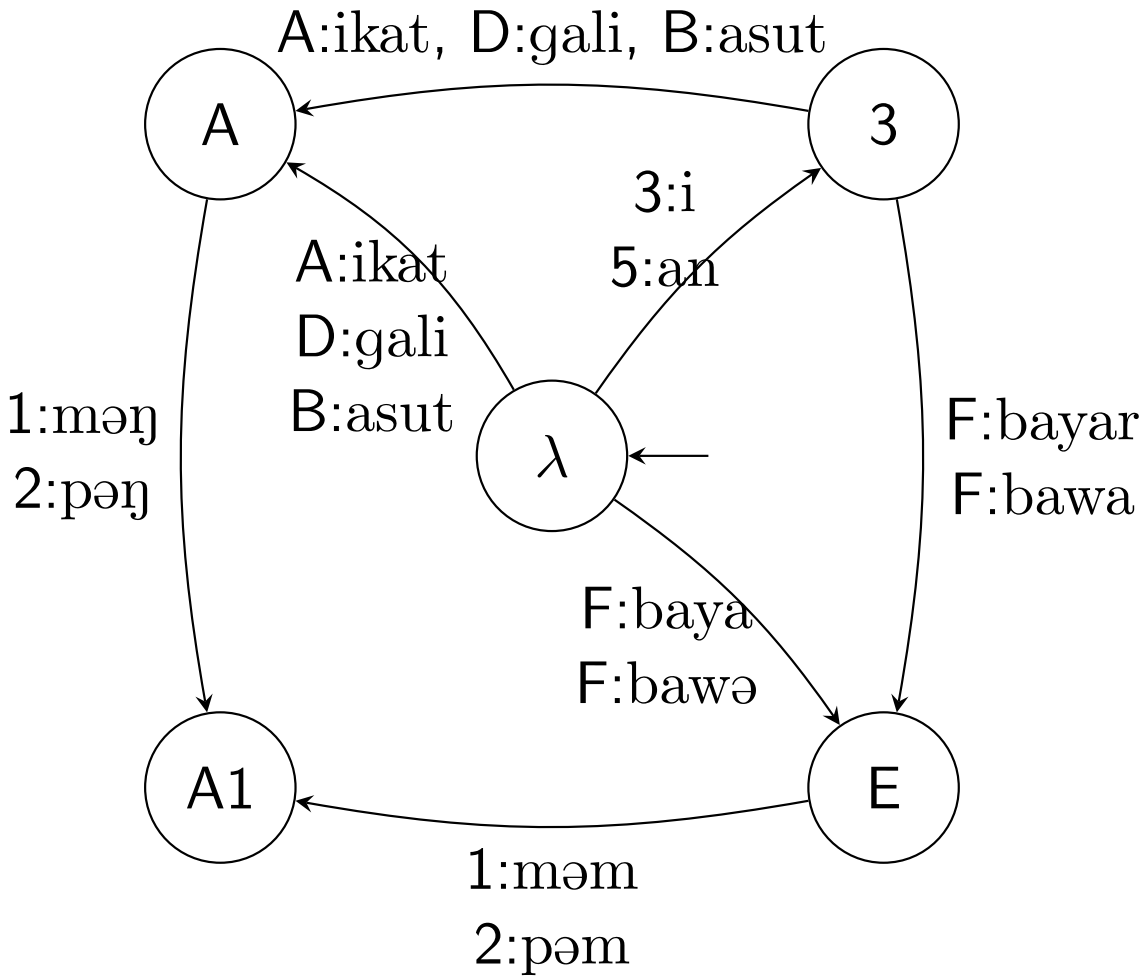
State merging

- Same method used by Oncina et al. (1993) for learning subsequential (i.e., deterministic) functions (Onward Subsequential Transducer Inference Algorithm).
- States are merged provided they have the same sets of **tails**:

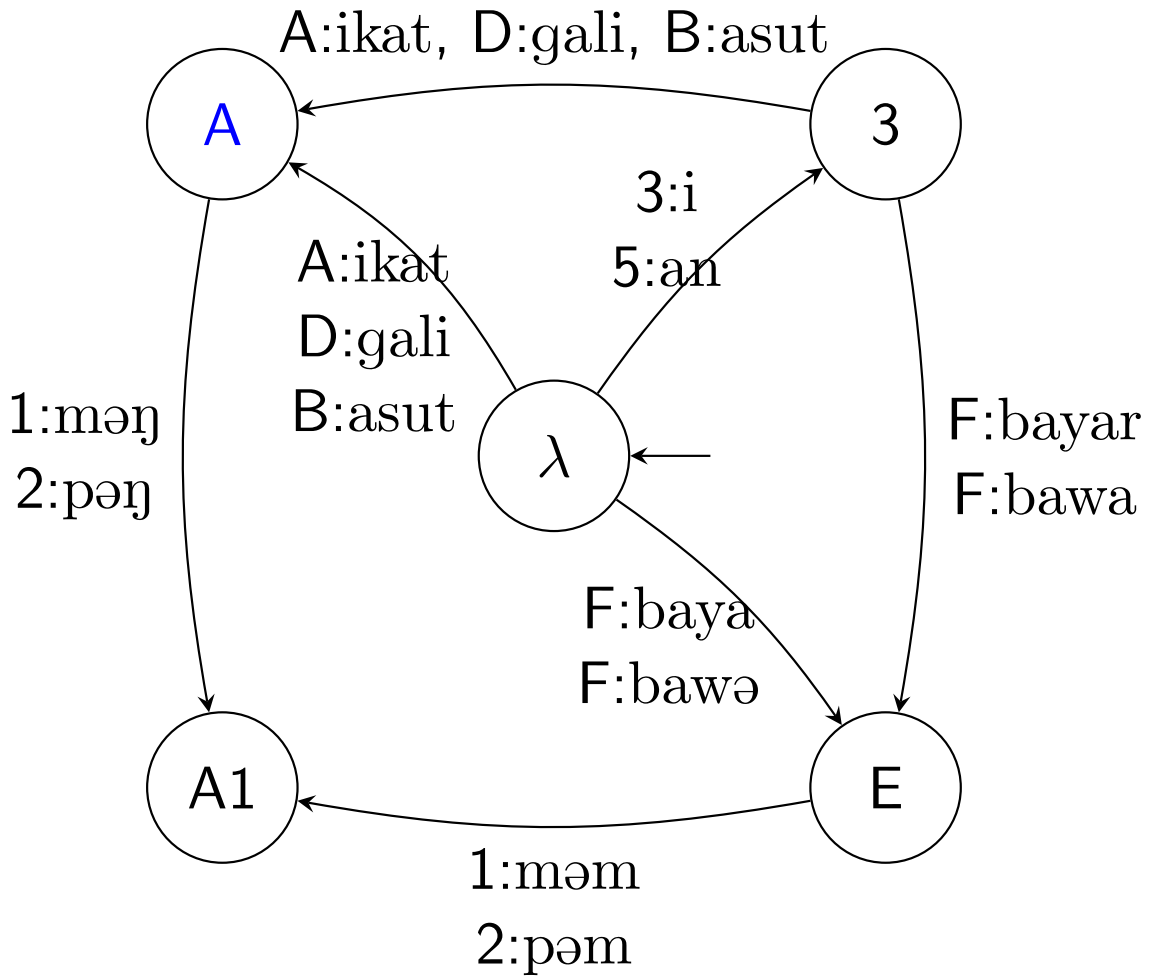
$$\texttt{tails}_f(x) = \{(y, v) | f(xy) = uv \wedge u = \texttt{lcp}(f(x\Sigma^*))$$

- Any merge that creates irreparable non-determinism is rejected.

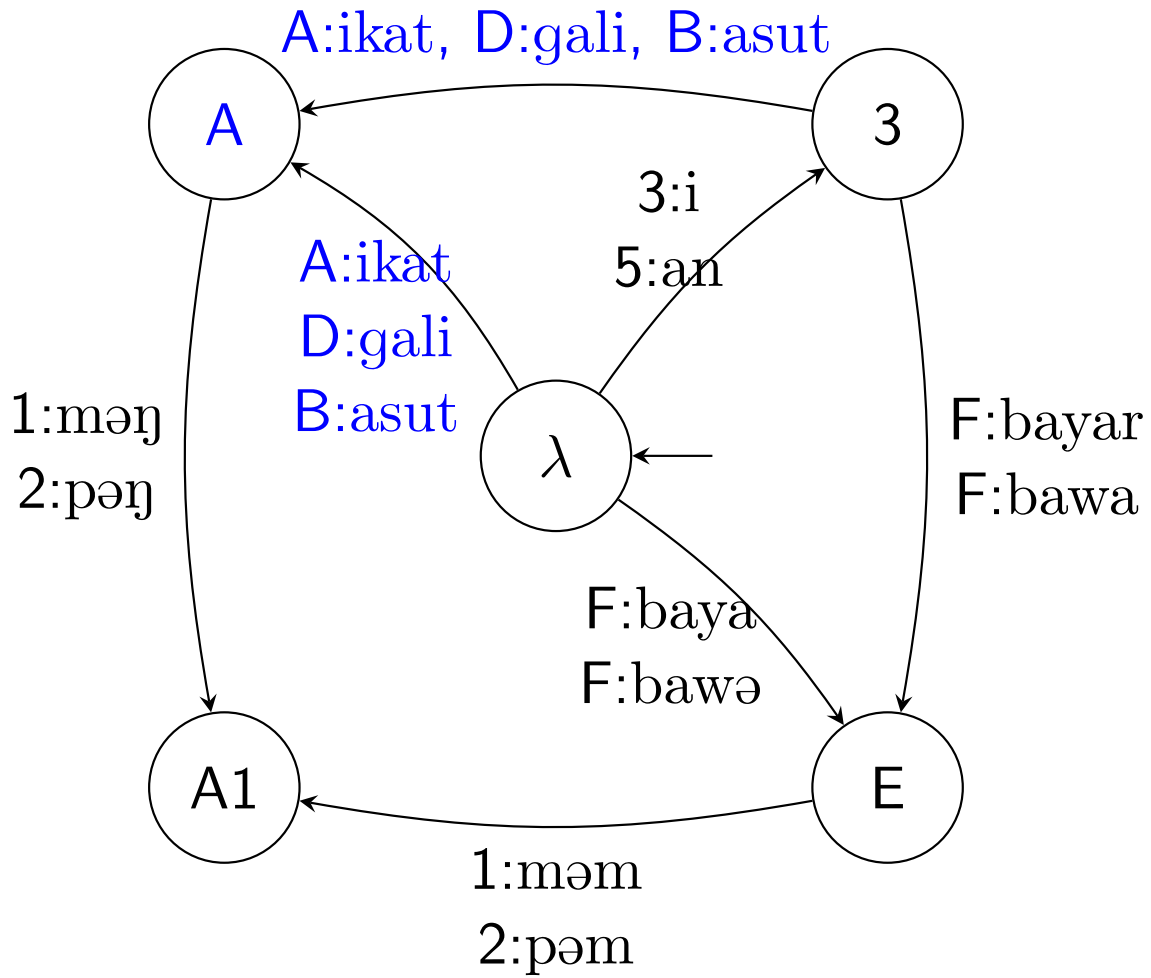
After state merging (partial result)



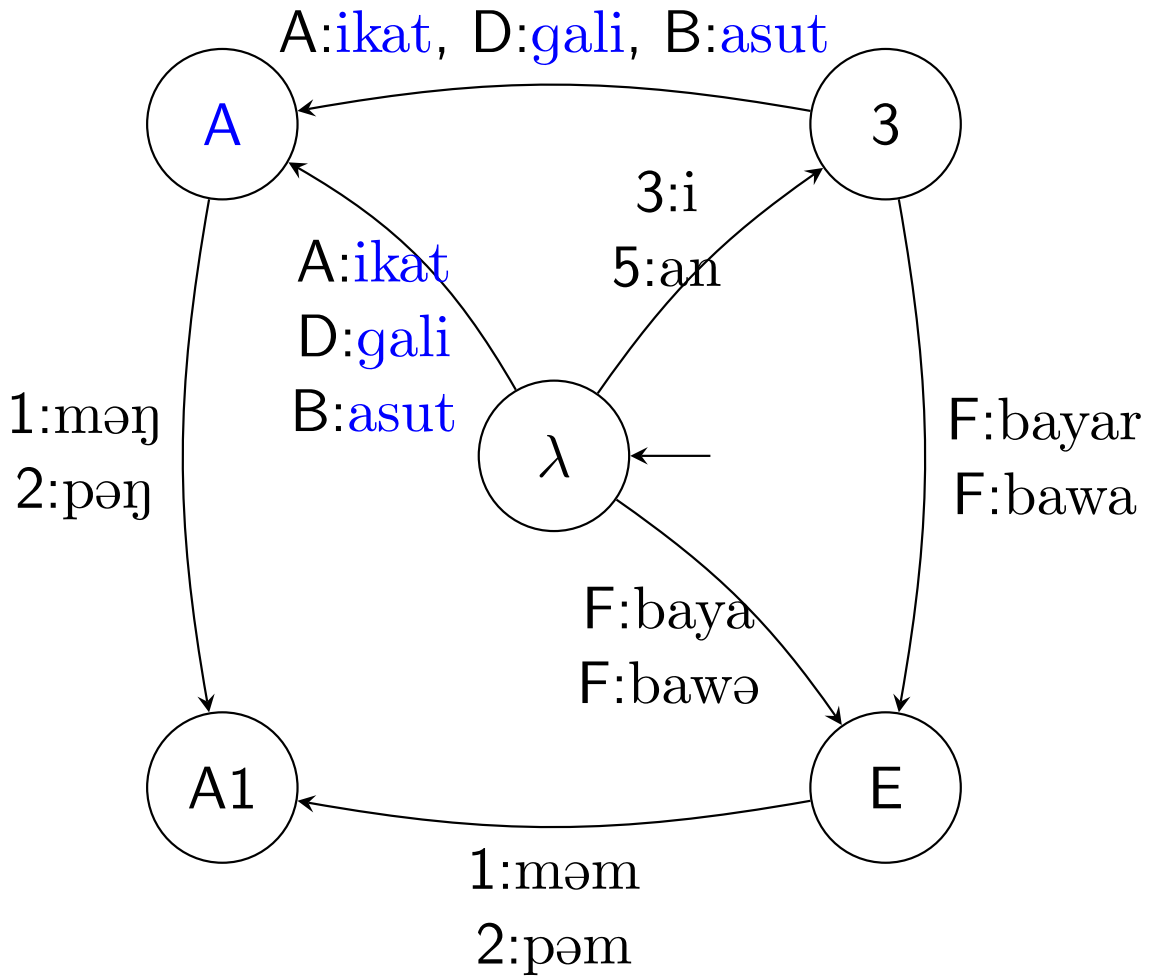
Locality assumption: collect 1-suffixes



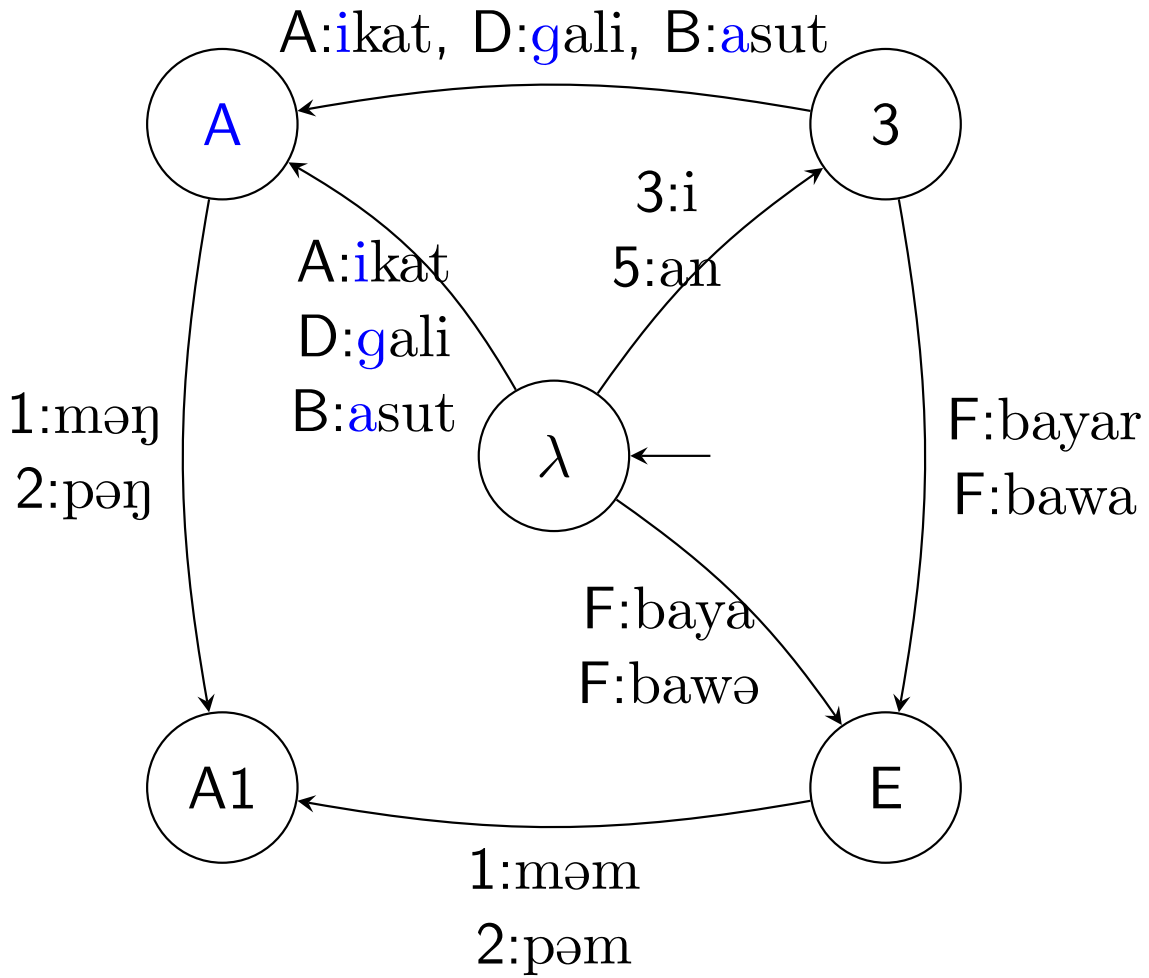
Locality assumption: collect 1-suffixes



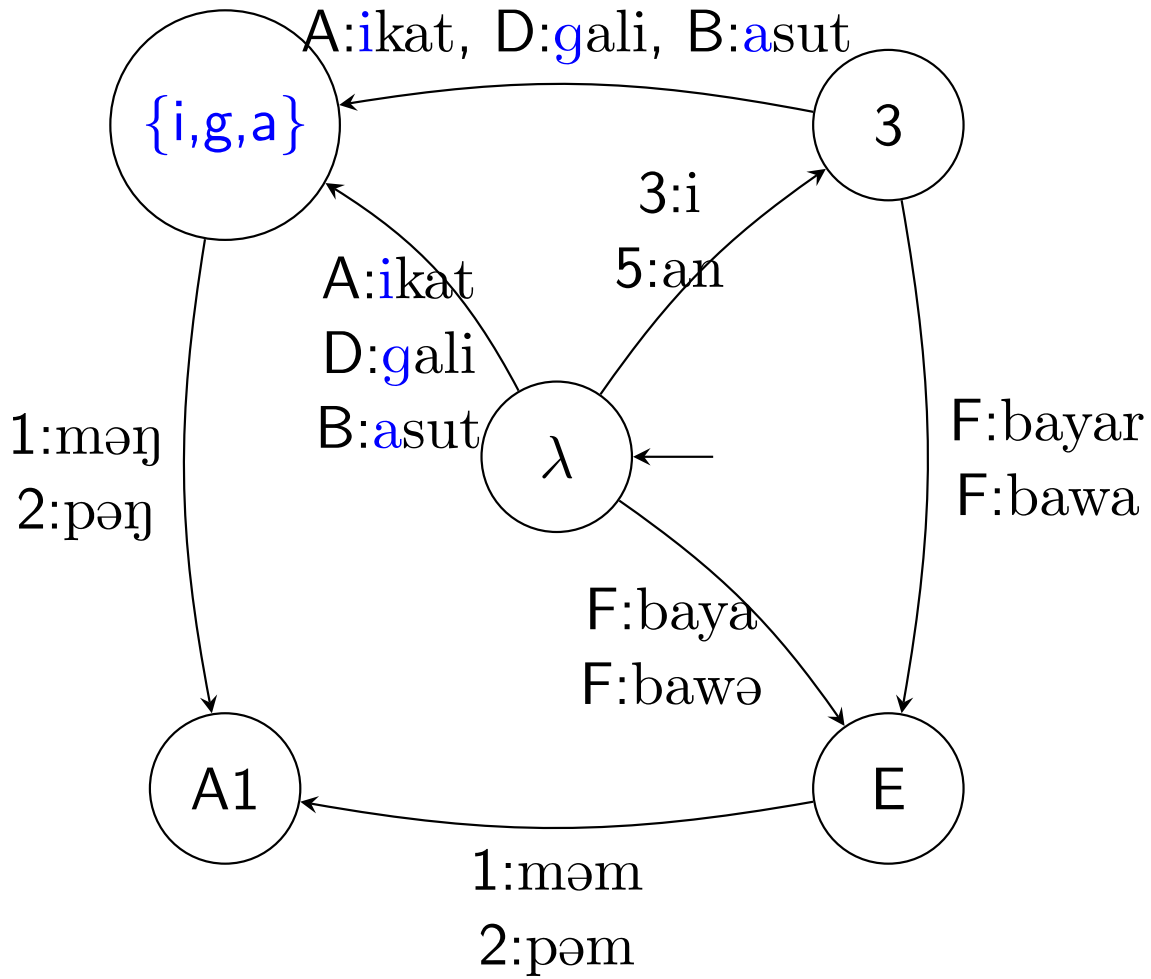
Locality assumption: collect 1-suffixes



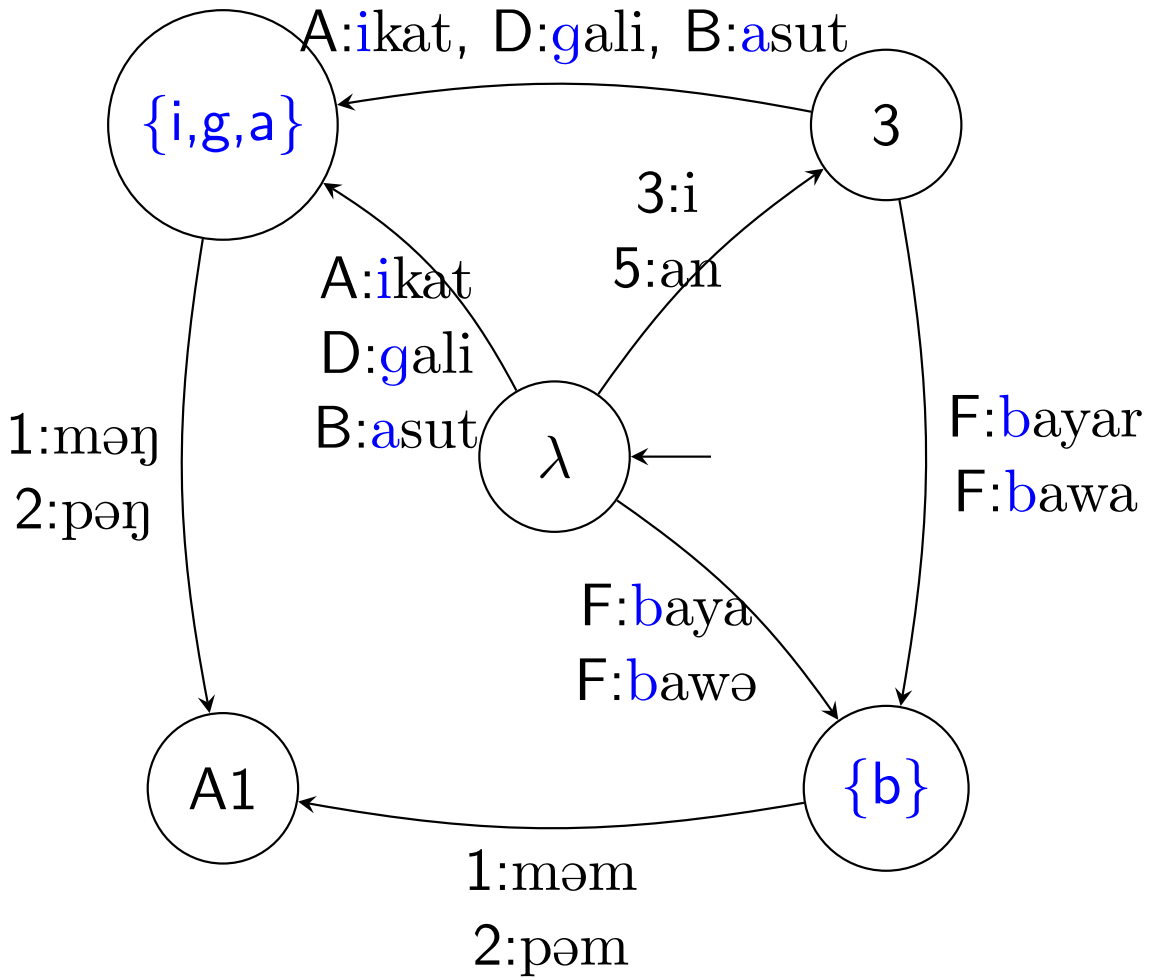
Locality assumption: collect 1-suffixes



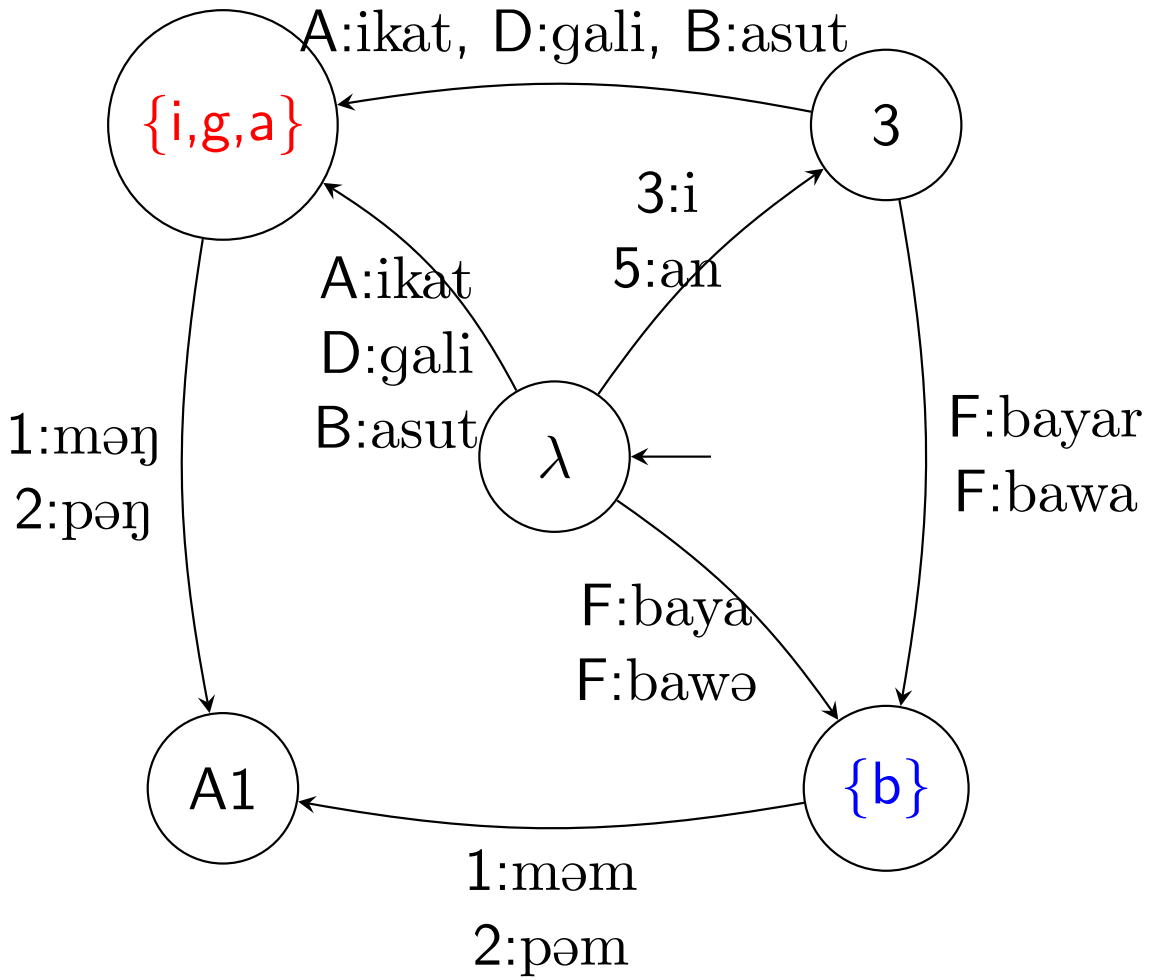
Locality assumption: collect 1-suffixes



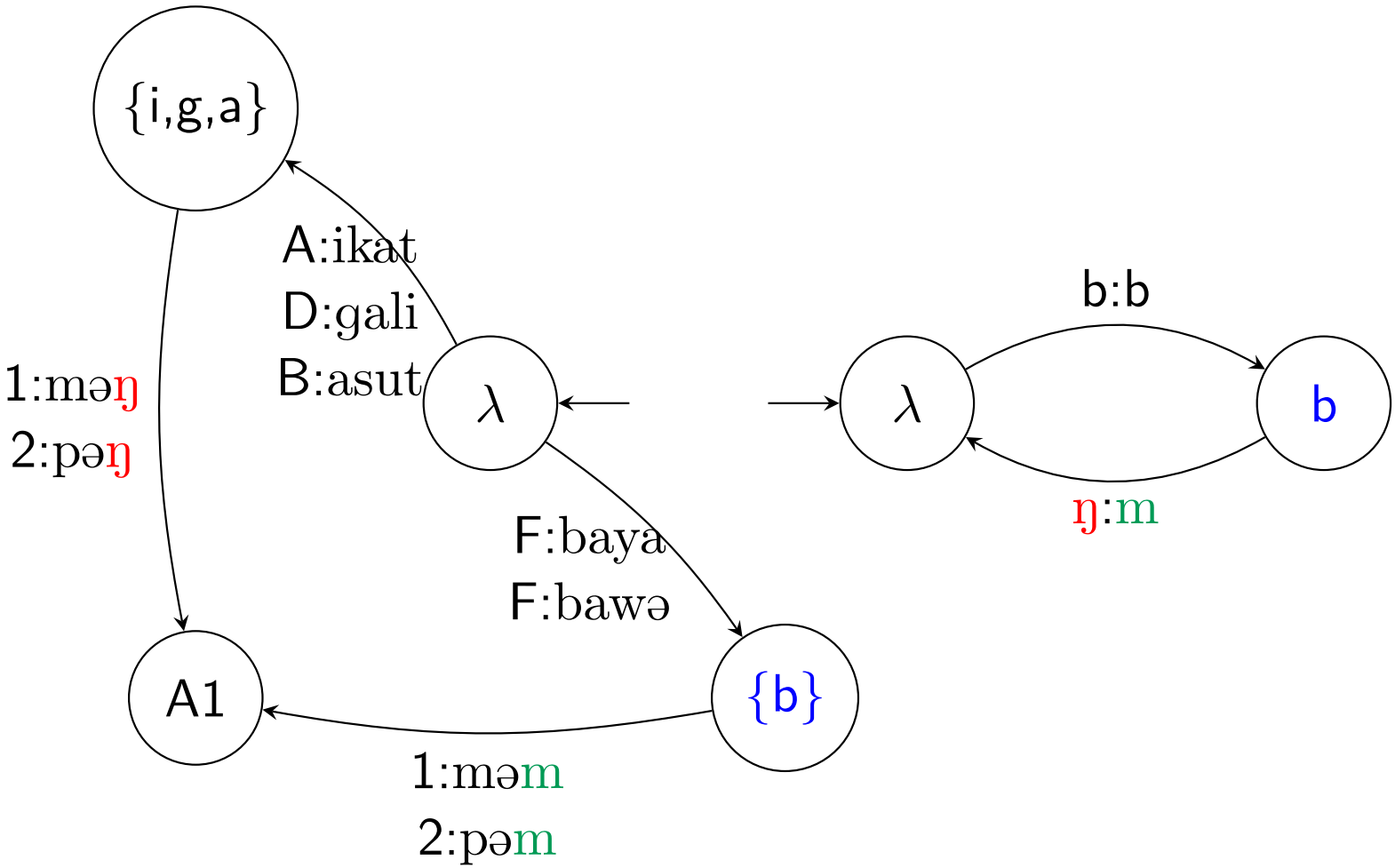
Locality assumption: collect 1-suffixes



UR selection: natural class bias



Constructing phonology function



Malay test case results

- Original lexicon successfully reconstructed, and the resulting phonology function enacts multiple processes simultaneously:

  $\eta \rightarrow [\alpha\text{place}] \; / \; \_\_ \; [-\text{son}, \alpha\text{place}]$

  $\eta \rightarrow \emptyset \; / \; \_\_ \; [+\text{son}, -\text{syl}]$

  $[-\text{son}] \rightarrow [-\text{voice}] \; / \; \_\_ \; \{[-\text{voice}], \#\}$

  $\{\text{k}, \text{g}\} \rightarrow \text{ʔ} \; / \; \_\_ \; \{\text{C}, \#\}$

  $\text{r} \rightarrow \emptyset \; / \; \_\_ \; \{\text{C}, \#\}$

  $\text{a} \rightarrow \text{ə} \; / \; \_\_ \; \#$

Scaling up

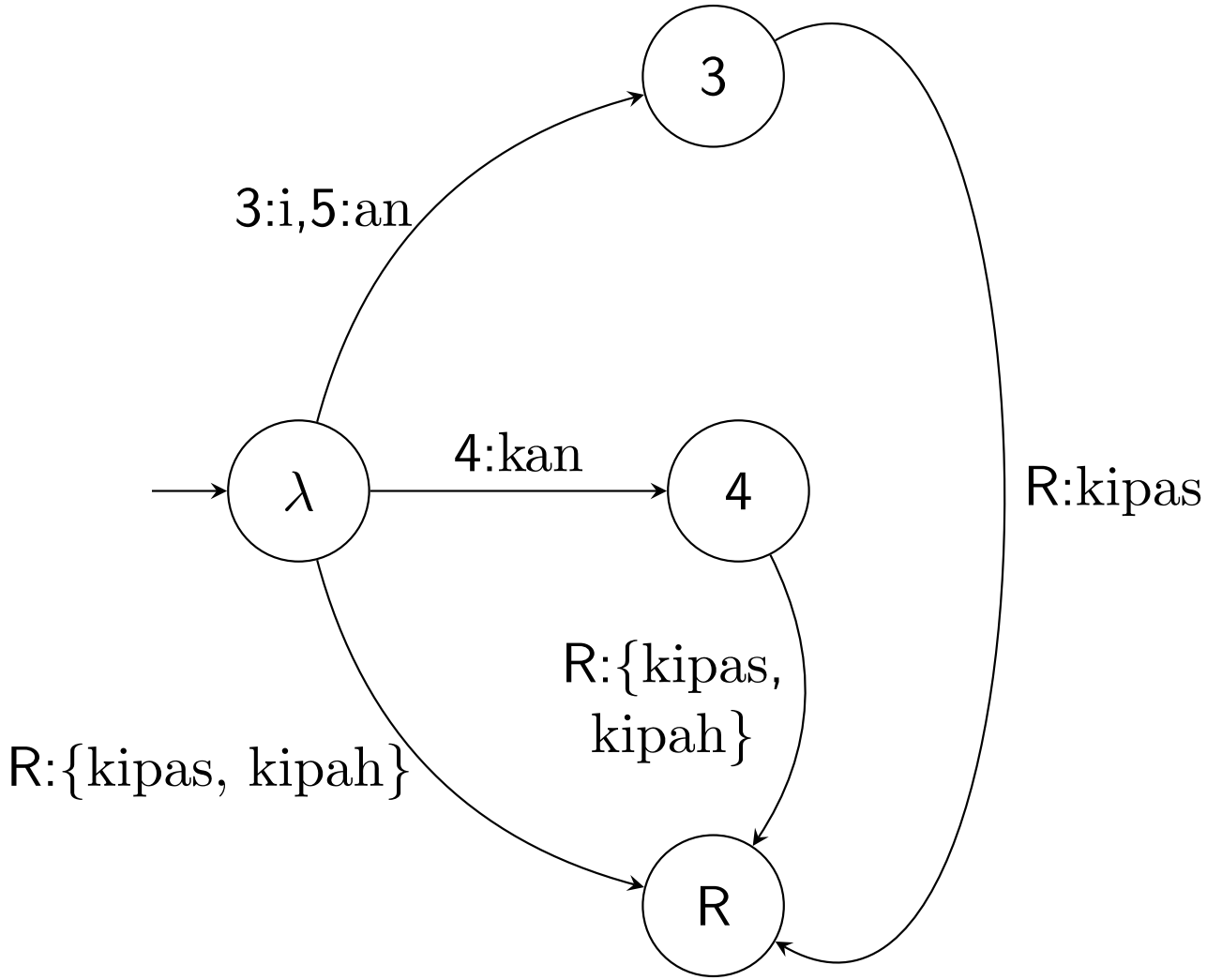
- Directionality (right-to-left, left-to-right) and locality window (2-local, 3-local, etc.) are parameters of the learner, but the presence of non-determinism in the output of state merging can be used as a cue that they are set incorrectly.

Semi-determinism

- Data that reflects optionality and variation can be addressed using *semi-determinism* (Beros and de la Higuera, 2016; Heinz, 2020)
- /s/ optionally becomes [h] in coda position

/kipas/	[kipas] ∼ [kipah]	'fan'
/kipas-kan/	[kipaskan] ∼ [kipahkan]	'to cause to fan for'

Semi-determinism

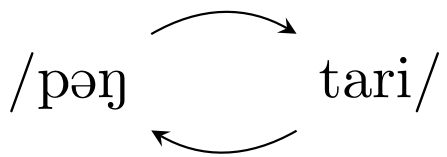


Advantages

- **Modular** approach that can be used with alternative:
  - formalizations of locality (e.g., tier-based, autosegmental)
  - theories of possible URs
  - UR selection heuristics
- **Interpretable**, bottom-up approach:
  - We know exactly what can be learned under what conditions (i.e., data content and parameter settings).

Fusion allomorph

- /pəŋ-tari/ → [pənari], ‘dancer’
- 2-local, but bidirectional



References I

Apoussidou, D. (2007). *The Learnability of Metrical Phonology*. PhD thesis, University of Amsterdam.

Beros, A. A. and de la Higuera, C. (2016). A canonical semi-deterministic transducer. *Fundam. Informaticae*, 146(4):431–459.

Berstel, J. (1982). Fonctions rationnelles et addition, actes de l'ecole de printemps de théorie des langages. *Laboratoire d'Informatique de Paris*, pages 177–183.

Chandlee, J. (2014). *Strictly Local Phonological Processes*. PhD thesis, University of Delaware.

Cotterell, R., Peng, N., and Eisner, J. (2015). Modeling word forms using latent underlying morphs and phonology. *Transactions of the Association for Computational Linguistics*, 3:433–447.

Gold, E. (1967). Language identification in the limit. *Information and Control*, (10):447–474.

Heinz, J. (2020). Deterministic analyses of optional processes. Talk given at University of Leipzig.

Hua, W., Dai, H., and Jardine, A. (2021). Learning input strictly local functions from their composition. In Reisinger, D. K. E. and Huijsmans, M., editors, *Proceedings of the 37th West Coast Conference on Formal Linguistics*, pages 143–151. Cascadilla Proceedings Project.

Hua, W. and Jardine, A. (2021). Learning input strictly local functions from their composition. In Chandlee, J., Eyraud, R., Heinz, J., Jardine, A., and van Zaanen, M., editors, *Proceedings of the Fifteenth International Conference on Grammatical Inference*, volume 153 of *Proceedings of Machine Learning Research*, pages 47–65. PMLR.

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 in Computational Phonology*. Association for Computational Linguistics.

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*, pages 125–134. Chicago Linguistics Society.

Jarosz, G. (2015). Expectation driven learning of phonology. Ms., UMass Amherst.

References II

Lind, D. and Marcus, B. (1995). *Symbolic Dynamics and Coding*. Cambridge UP.

Merchant, N. N. (2008). *Discovering Underlying Forms: Contrast Pairs and Ranking*. PhD thesis, Rutgers.

Nelson, M. (2019). Segmentation and UR acquisition with UR constraints. In *Proceedings of the Society for Computation in Linguistics*, volume 2. Article 8.

Oncina, J., García, J., and Vidal, E. (1993). Learning subsequential transducers for pattern recognition interpretation tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (15:5):448–457.

Oncina, J. and Varò, M. A. (1996). Using domain information during the learning of a subsequential transducer. *Lecture Notes in Computer Science - Lecture Notes in Artificial Intelligence*, pages 313–325.

Onn, F. M. (1976). *Aspects of Malay Phonology and Morphology*. PhD thesis, University of Illinois at Urbana-Champaign.

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

Pater, J., Jesney, K., Staubs, R., and Smith, B. (2012). Learning probabilities over underlying representations. In *Proceedings of the 12th Meeting of the Special Interest Group on Computational Morphology and Phonology (SIGMORPHON)*, pages 62–71. ACL.

Rasin, E., Berger, I., Lan, N., Shefi, I., and Katzir, R. (2021). Approaching explanatory adequacy in phonology using minimum description length. *Journal of Language Modelling*, 9(1):17–66.

Rasin, E., Shefi, I., and Katzir, R. (2020). A unified approach to several learning challenges in phonology. In Asatryan, M., Song, Y., and Whitmal, A., editors, *Proceedings of NELS 50*, volume 3, pages 73–87. GLSA.

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

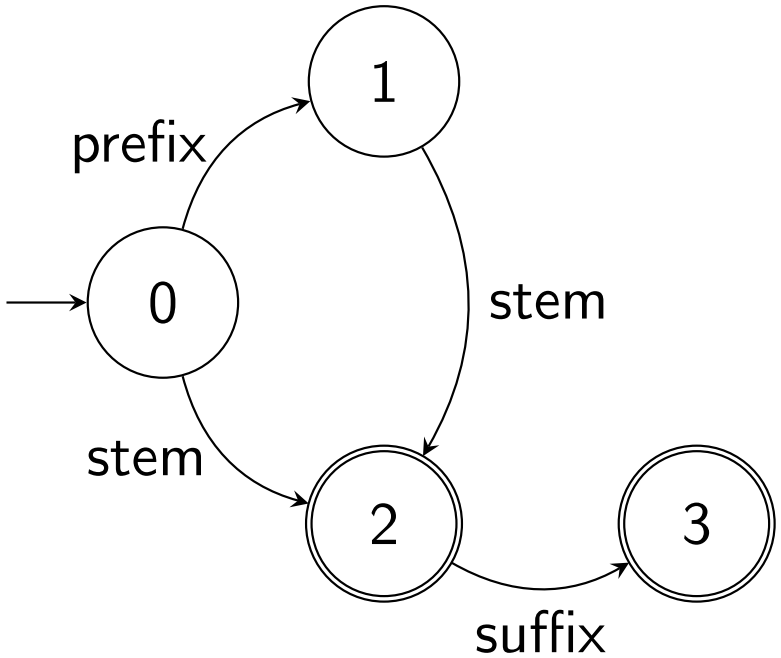
Vaysse, O. (1986). Addition molle et fonctions p-locales. *Semigroup Forum*, 34:157–175.

Wang, Y. and Hayes, B. (2022). Learning underlying representations: An approach guided by the Kenstowicz-Kisseberth UR hierarchy. Unpublished Ms., UCLA, and poster at AMP 2022.

State merging using domain information

- Modified version of Oncina and Varò (1996) adds a condition: merged states must have the same tails *and* reach the same state in an FSA representation of the domain of the function (in this case the output language of $\mathcal{M}$).

Domain FSA



Epenthesis

- Glide epenthesis if first vowel is high:
  /bantu-an/ [bantuwan] 'aid, relief'
  /tari-an/ [tariyan] 'dance'
- Otherwise glottal epenthesis:
  /məŋ-gula-i/ [məŋgulaʔi] 'to cause to sweeten'
  /pəŋ-buka-an/ [pəmbukaʔan] 'opening'
- At prefix boundary, always glottal:
  /di-ambil/ [diʔambel] 'to take (passive)'
- Nonderived environment blocking:
  /main/ [main] 'to play'
  /naik/ [naik] 'to ascend'

Epenthesis

- Adirectional and still 2-local, but requires a more detailed UR decision procedure

/cərca/, 'to revile'

SRs:	[cərcə]	[cərcaʔ]	[cərca]
Contexts:	__#	__V	__C

Hypothesis testing

/cərcə/	/cərcaʔ/	/cərca/
ə → a / __ C	aʔ → ə / __ #	a → ə / __ #
ə → aʔ / __ V	ʔ → ∅ / __ C	∅ → ʔ / V __ V
*[carcakan]	ʔ isn't phonemic	