Probabilistic Comparative Linguistic Reconstruction

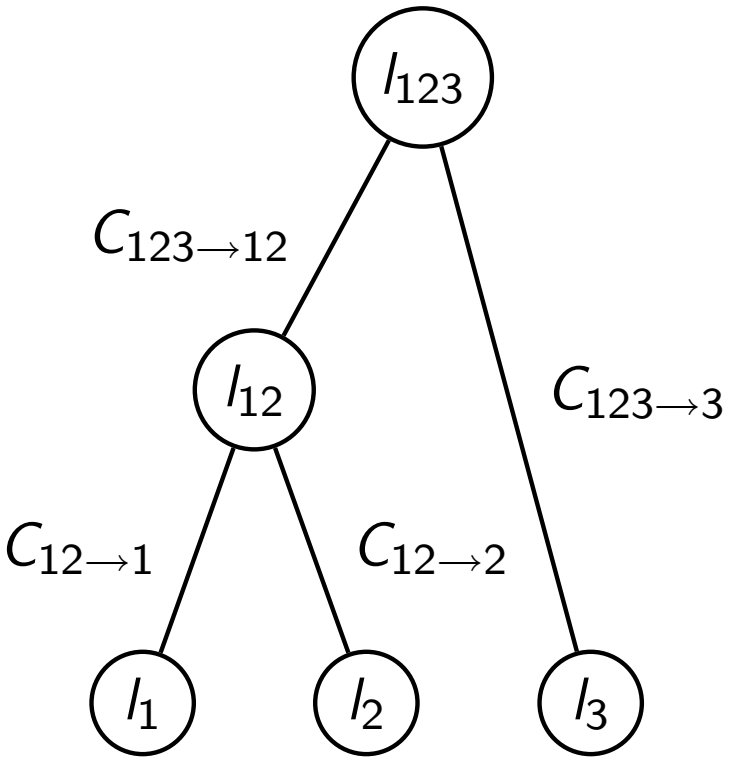
Terry Szymanski

Linguistics Student Colloquium

December 5, 2008

A Probabilistic Model of Language Change

Probabilistic Model of Language Change



Probability Distributions

Define:

$p(l)$	A probability distribution over languages
$p(s)$	A probability distribution over tree splits
$p(c)$	A probability distribution over changes

Probability Distributions

Define:

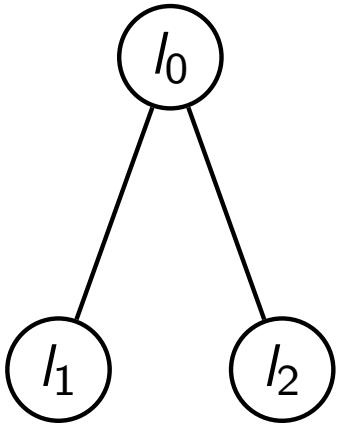
$p(l)$	A probability distribution over languages
$p(s)$	A probability distribution over tree splits
$p(c)$	A probability distribution over changes

Then the probability of a reconstruction r is:

$$p(r) = p(l_{root}) \cdot \prod_{s_i \in r}(s_i) \cdot \prod_{c_i \in r} p(c_i)$$

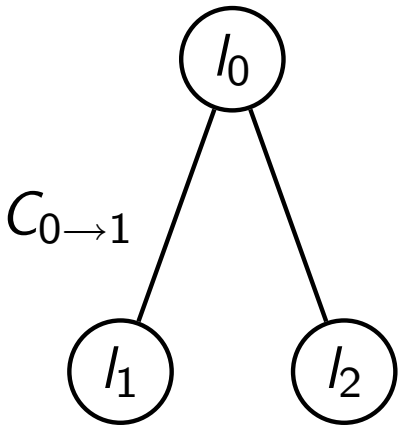
As a Generative Process

As a Generative Process



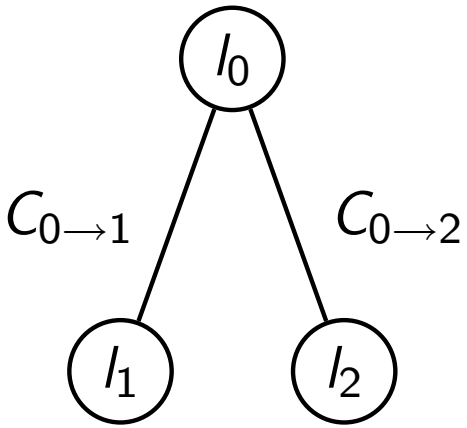
$$l_1 = l_0$$
$$l_2 = l_0$$

As a Generative Process



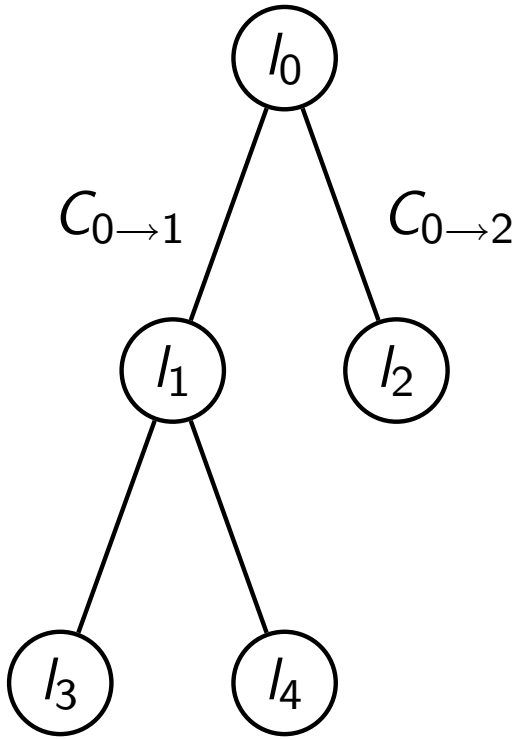
$$l_1 \neq l_0$$
$$l_2 = l_0$$

As a Generative Process



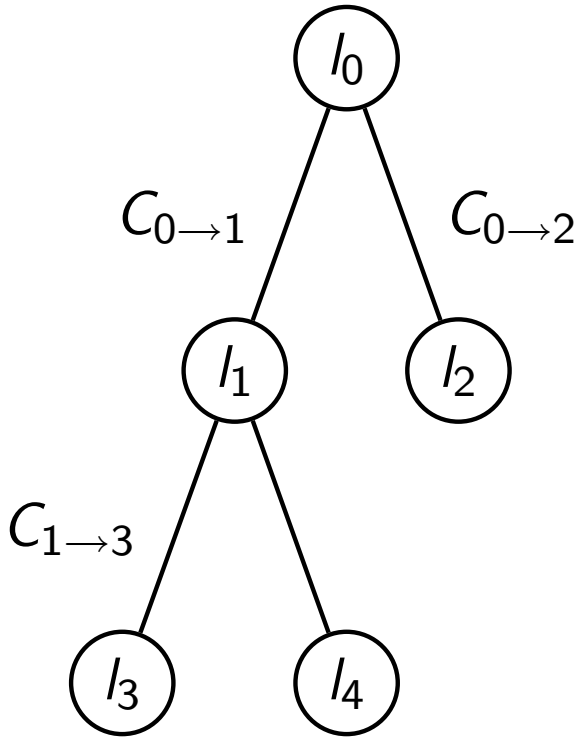
$$l_1 \neq l_0$$
$$l_2 \neq l_0$$

As a Generative Process



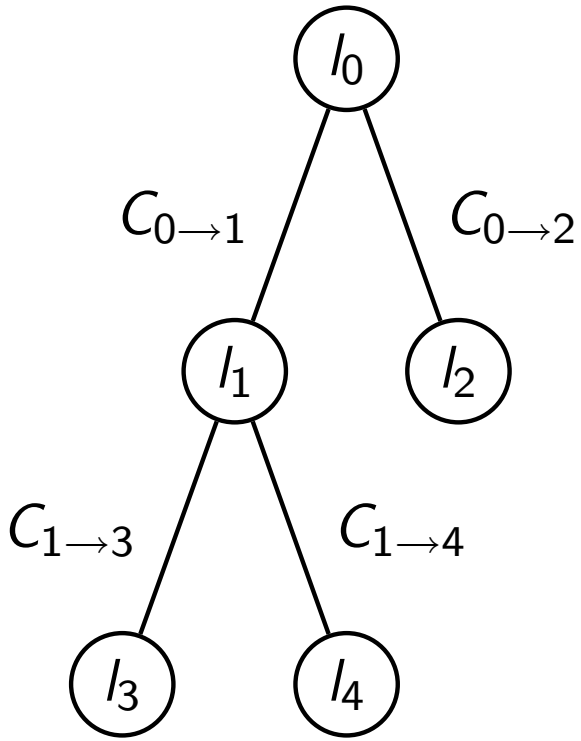
$$l_3 = l_1$$
$$l_4 = l_1$$

As a Generative Process



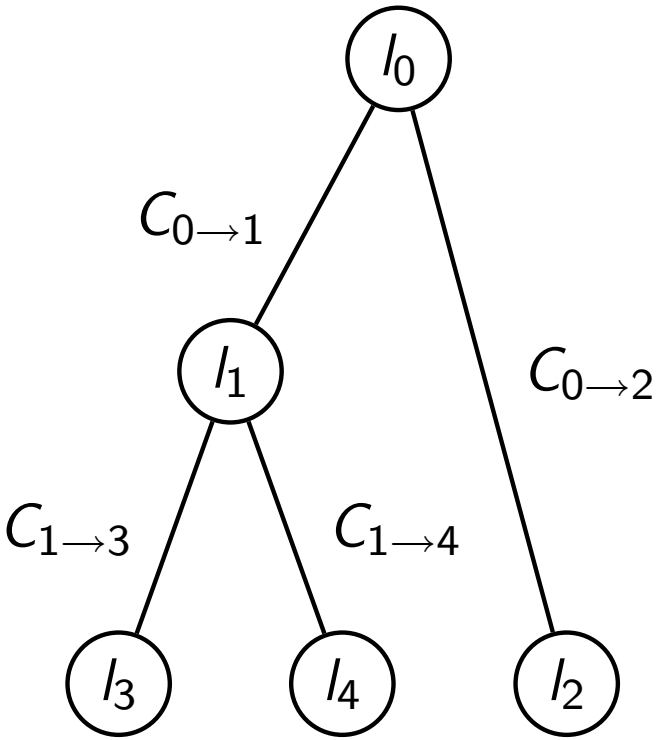
$$l_3 \neq l_1$$
$$l_4 = l_1$$

As a Generative Process

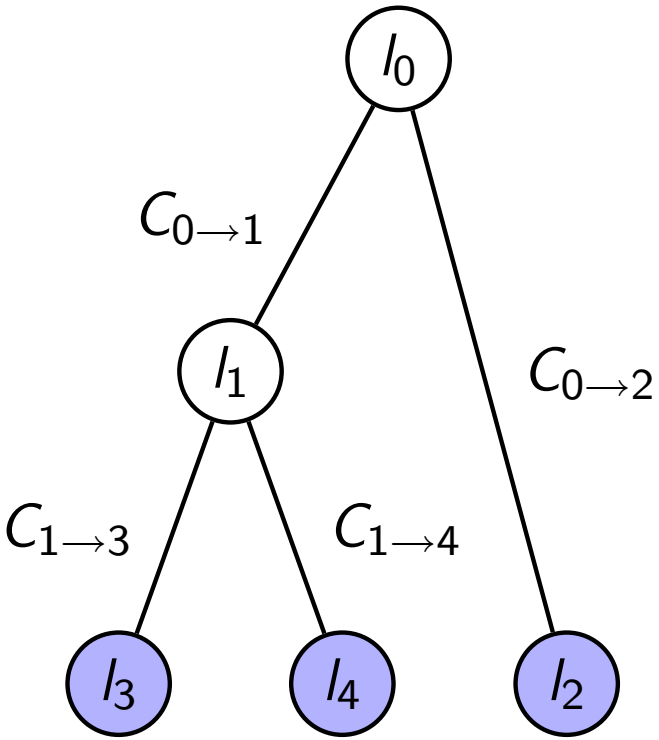


$$l_3 \neq l_1$$
$$l_4 \neq l_1$$

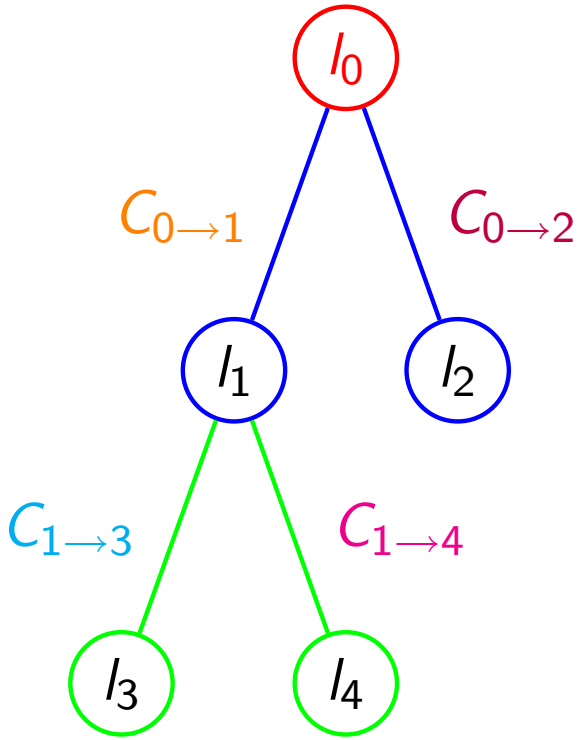
As a Generative Process



As a Generative Process



The Probability of a Language Family



$$p(r) = p(l_0) \cdot p(\curlywedge) \cdot p(C_{0\to1}) \cdot p(C_{0\to2}) \cdot p(\curlywedge) \cdot p(C_{1\to3}) \cdot p(C_{1\to4})$$

The Tree Model

- Each node is a **language** (*l*). Need to define:
  - What exactly is a "language" (for my purposes)

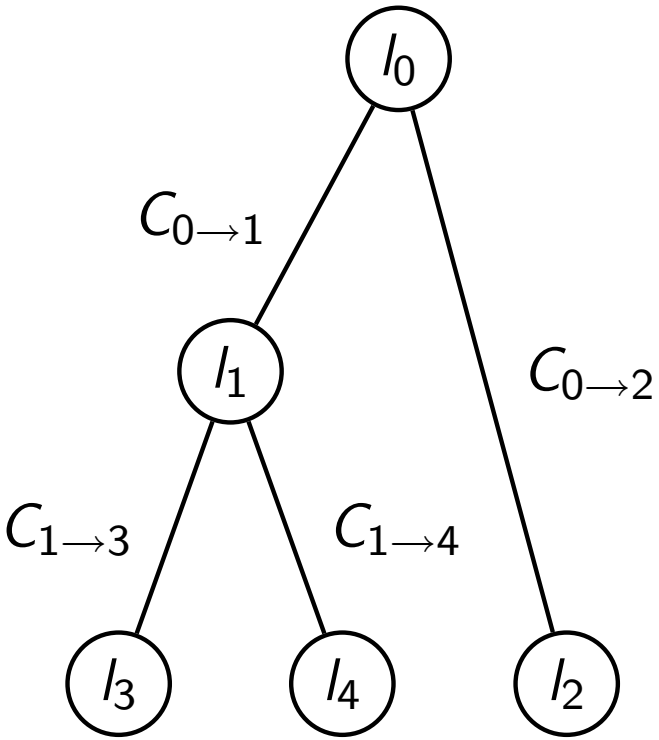
The Tree Model

- Each node is a **language** (l). Need to define:
  - What exactly is a "language" (for my purposes)
- Each edge contains **changes** (C). Need to define:
  - The set of possible changes
  - The effects of changes on a language
  - The probability distribution over changes

The Tree Model

- Each node is a **language** (l). Need to define:
  - What exactly is a "language" (for my purposes)
- Each edge contains **changes** (C). Need to define:
  - The set of possible changes
  - The effects of changes on a language
  - The probability distribution over changes
- Each edge derives its daughter from its mother.

The Tree Model



Types of Language Change

Type of Change	Representation of language	Types of changes
sound change	words	regular sound change.

Types of Language Change

Type of Change	Representation of language	Types of changes
sound change	words	regular sound change.
morphological	paradigms	leveling, analogy, etc.

Types of Language Change

Type of Change	Representation of language	Types of changes
sound change	words	regular sound change.
morphological	paradigms	leveling, analogy, etc.
syntactic	utterances	word-order, etc.

Sound Change Model

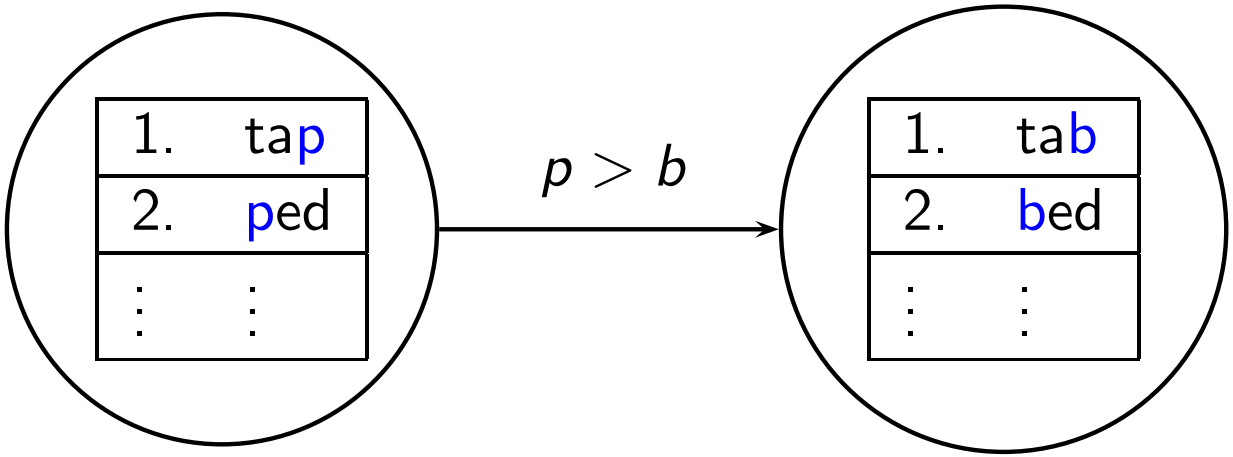
- A *language* is an indexed list of words

Sound Change Model

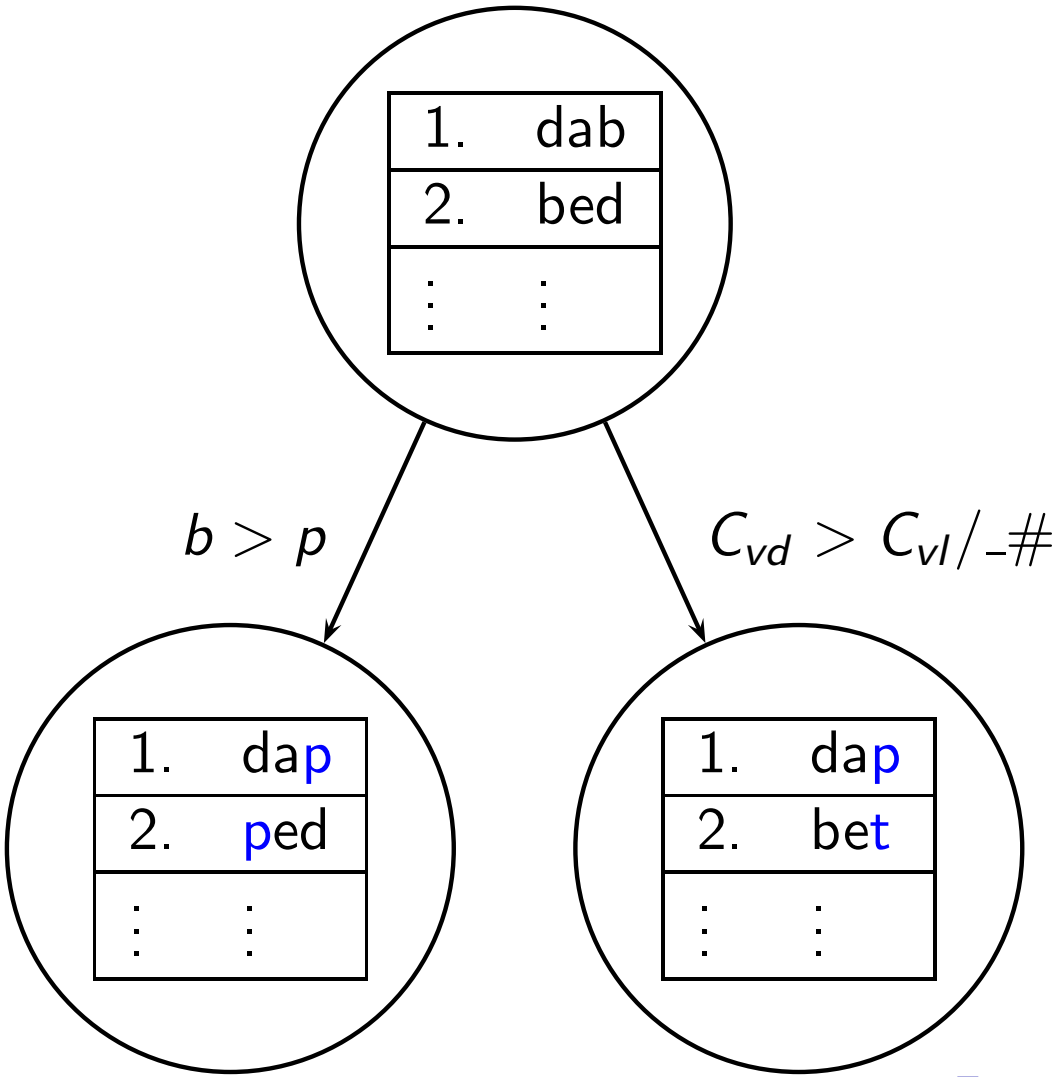
- A *language* is an indexed list of words
- A *change* is a regular sound change.

Sound Change Model

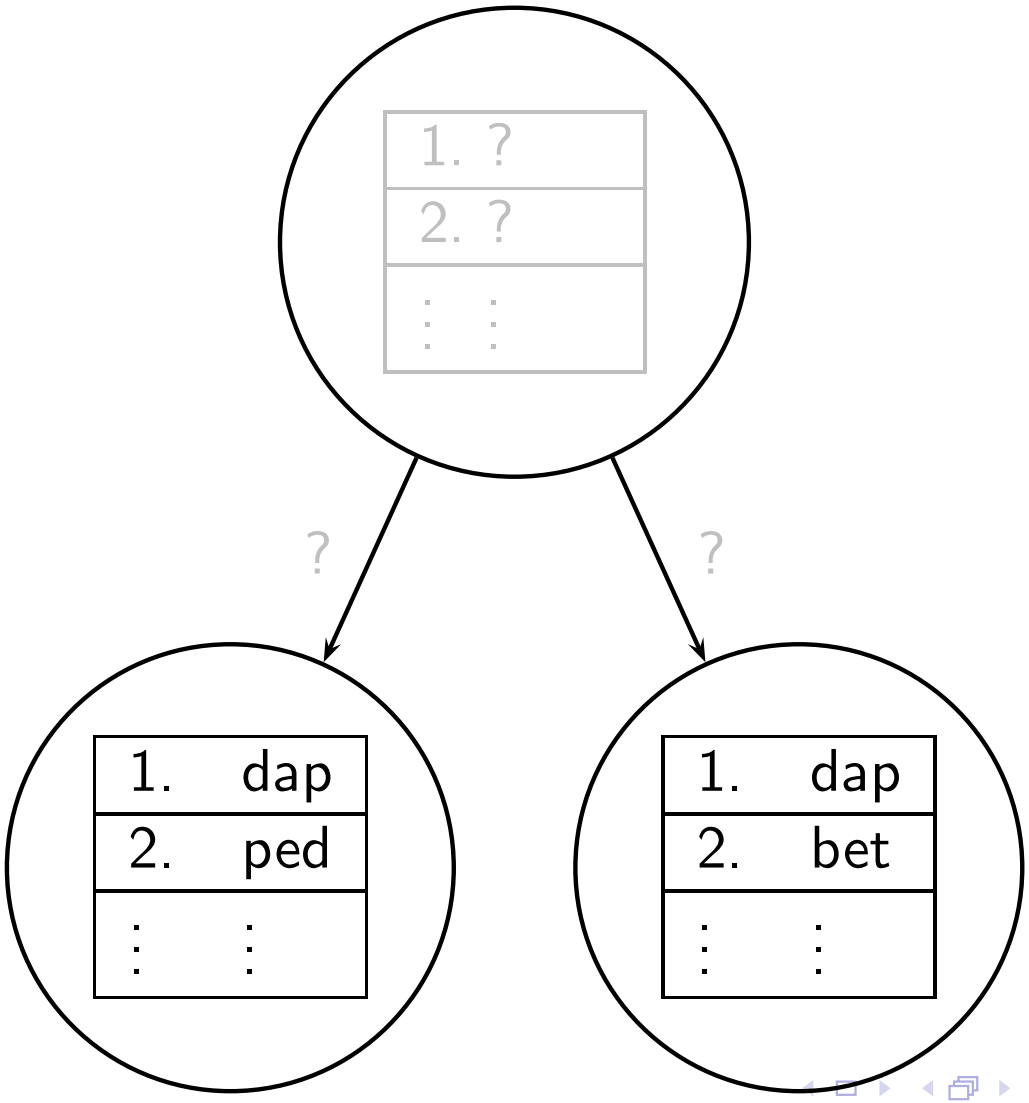
- A *language* is an indexed list of words
- A *change* is a regular sound change.



A Hypothetical Example



The Reconstruction Scenario



A Reconstruction Algorithm

A Reconstruction Algorithm

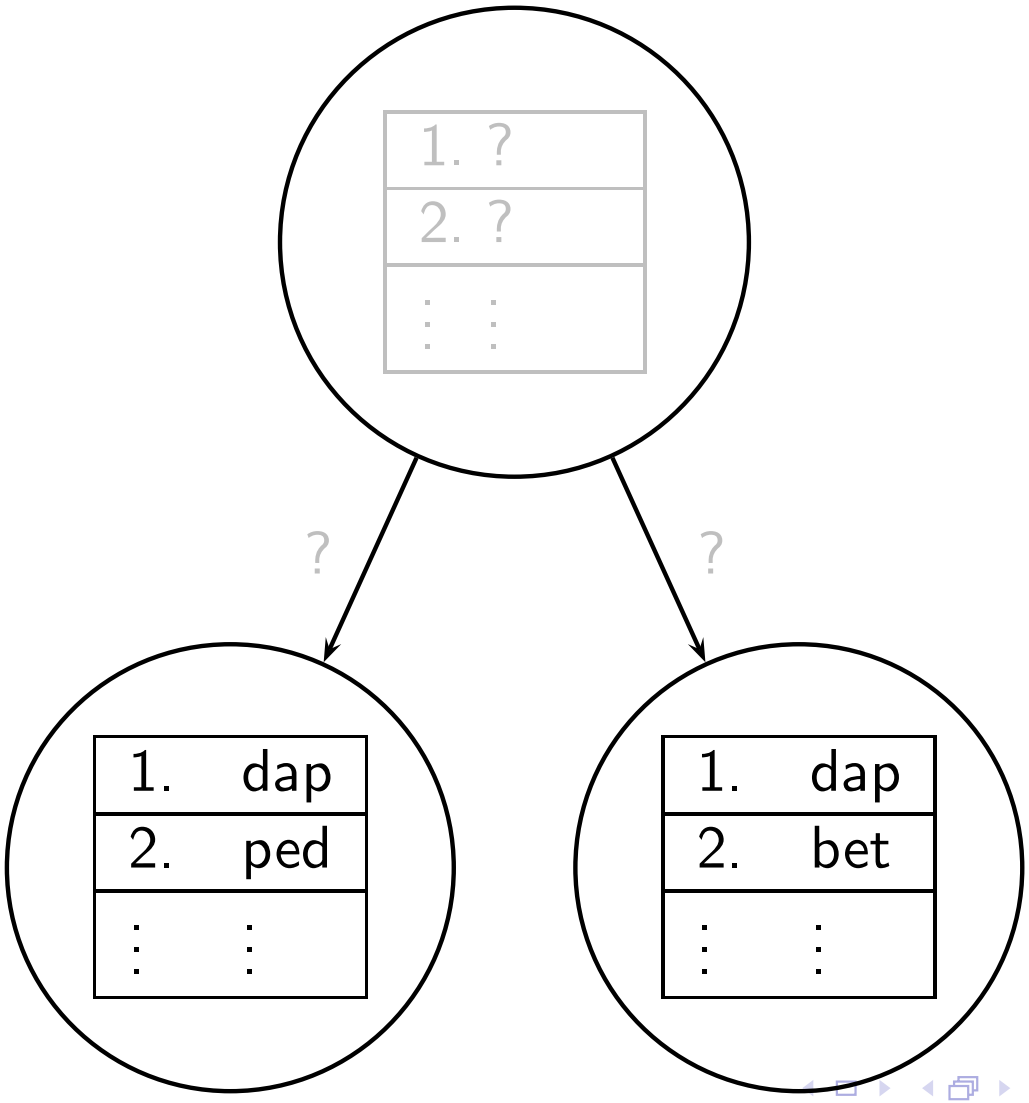
Select the optimal reconstruction from among the set of reconstructions produced by the following algorithm:

1. Align segments in cognate words to produce correspondence sets

2a. Reconstruct a proto-sound for each correspondence set

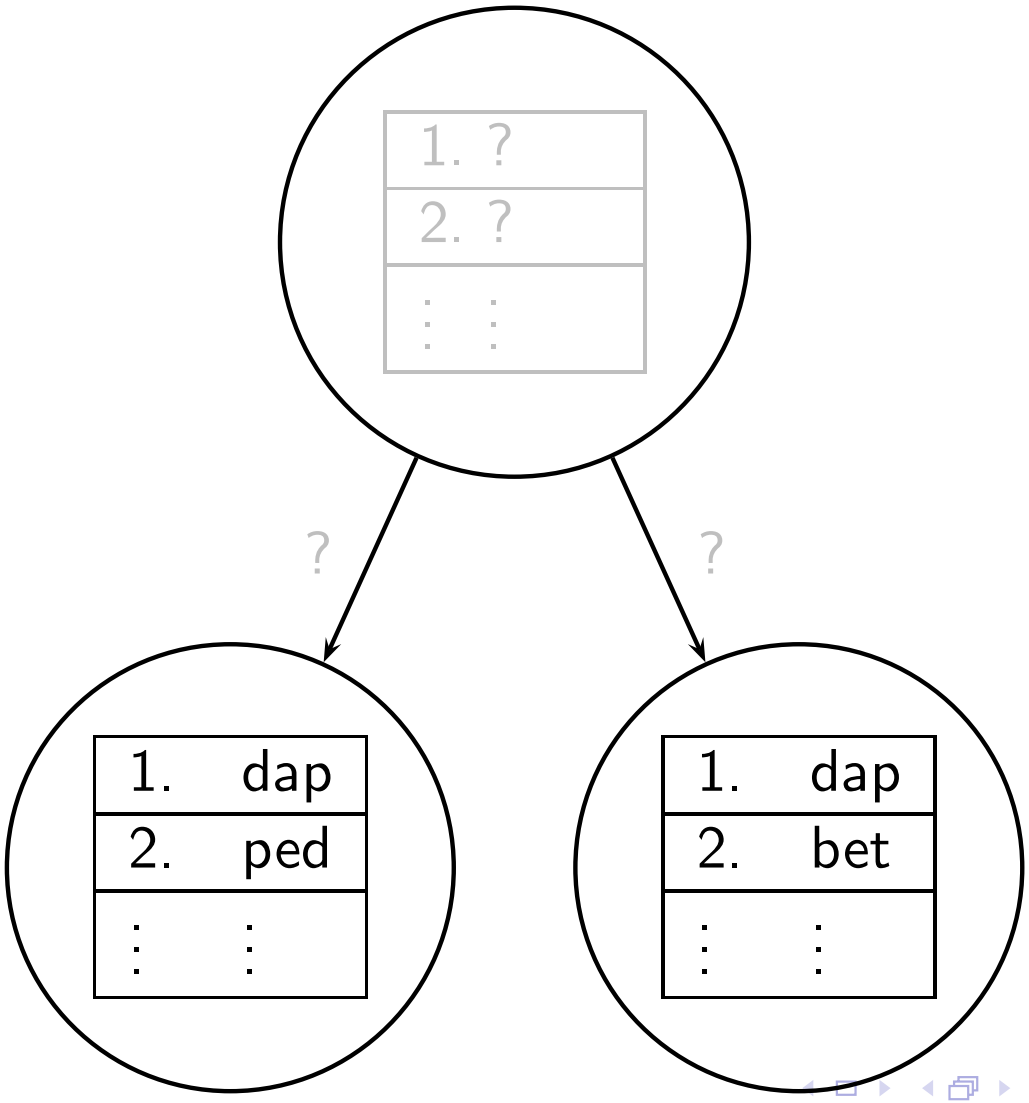
2b. Determine the necessary sound changes

The Reconstruction Scenario



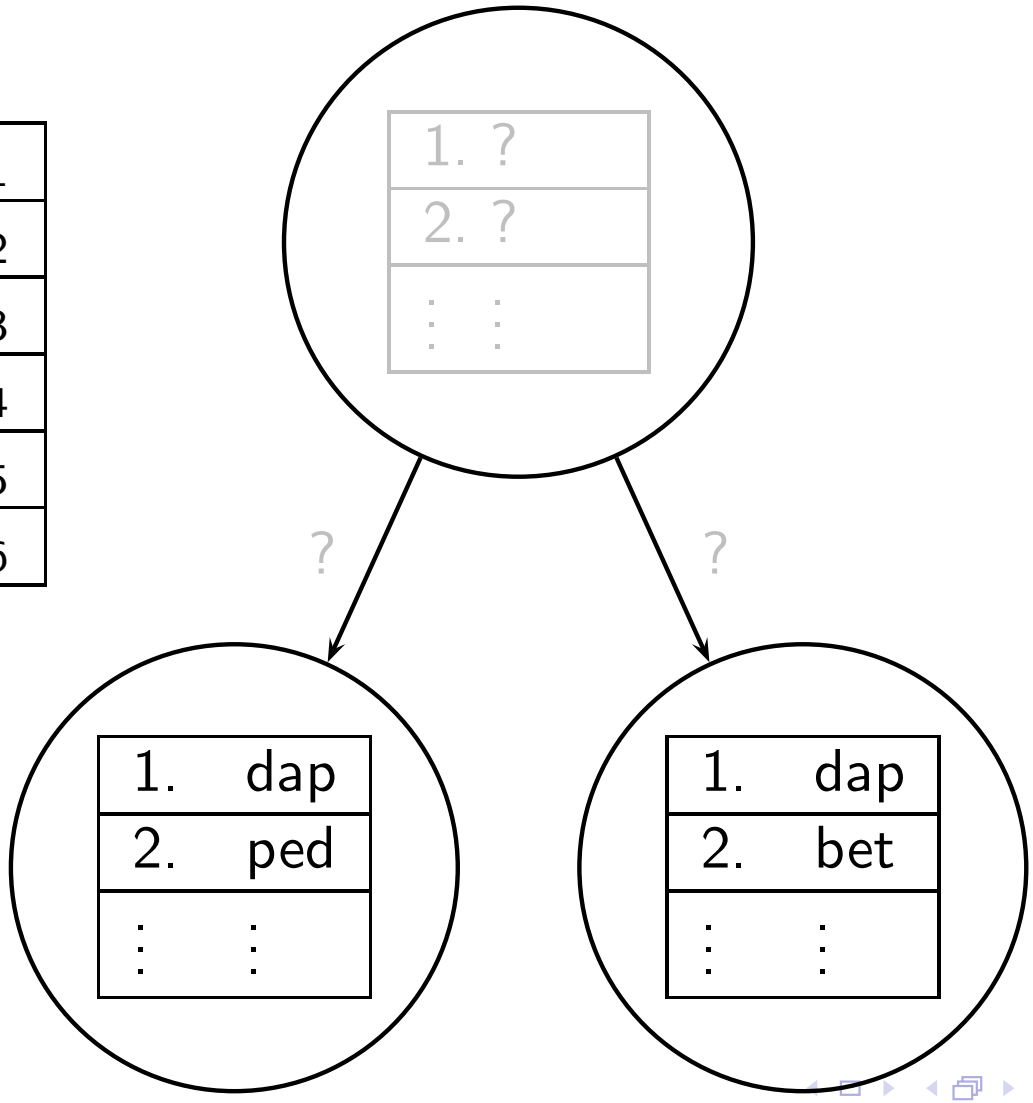
Correspondence Sets

d	:	d
a	:	a
p	:	p
p	:	b
e	:	e
d	:	t



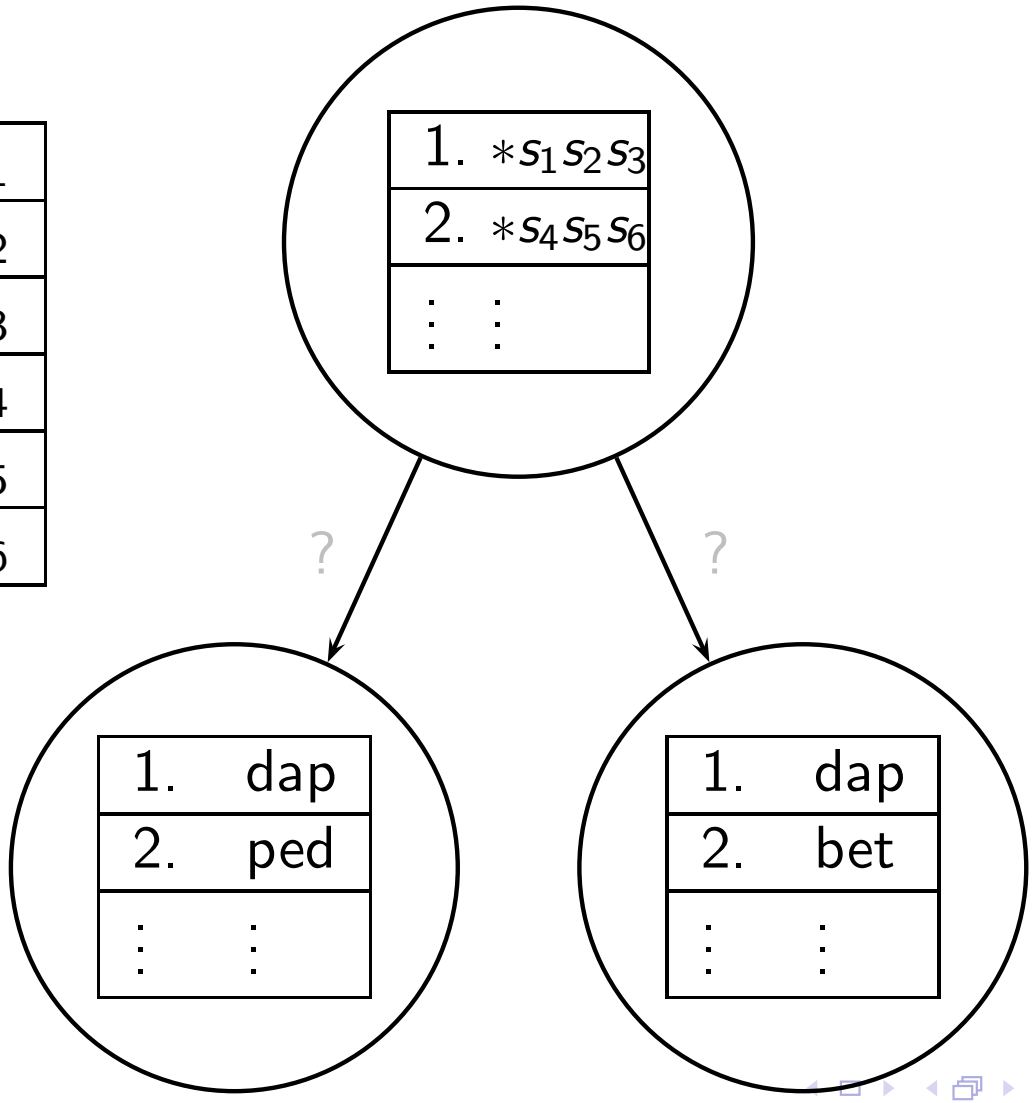
Proto Sounds

d	:	d	$*s_1$
a	:	a	$*s_2$
p	:	p	$*s_3$
p	:	b	$*s_4$
e	:	e	$*s_5$
d	:	t	$*s_6$



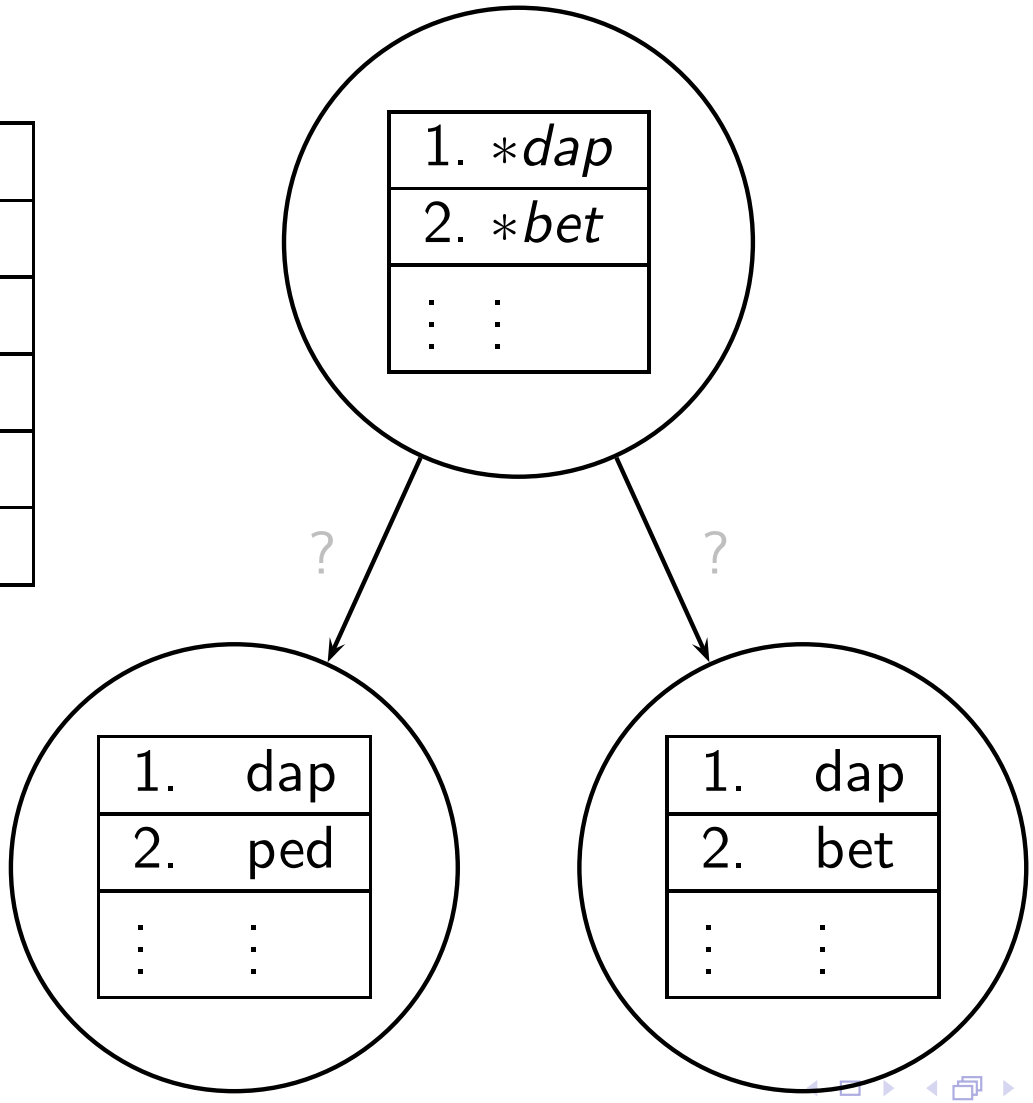
Proto Words

d	:	d	$*s_1$
a	:	a	$*s_2$
p	:	p	$*s_3$
p	:	b	$*s_4$
e	:	e	$*s_5$
d	:	t	$*s_6$



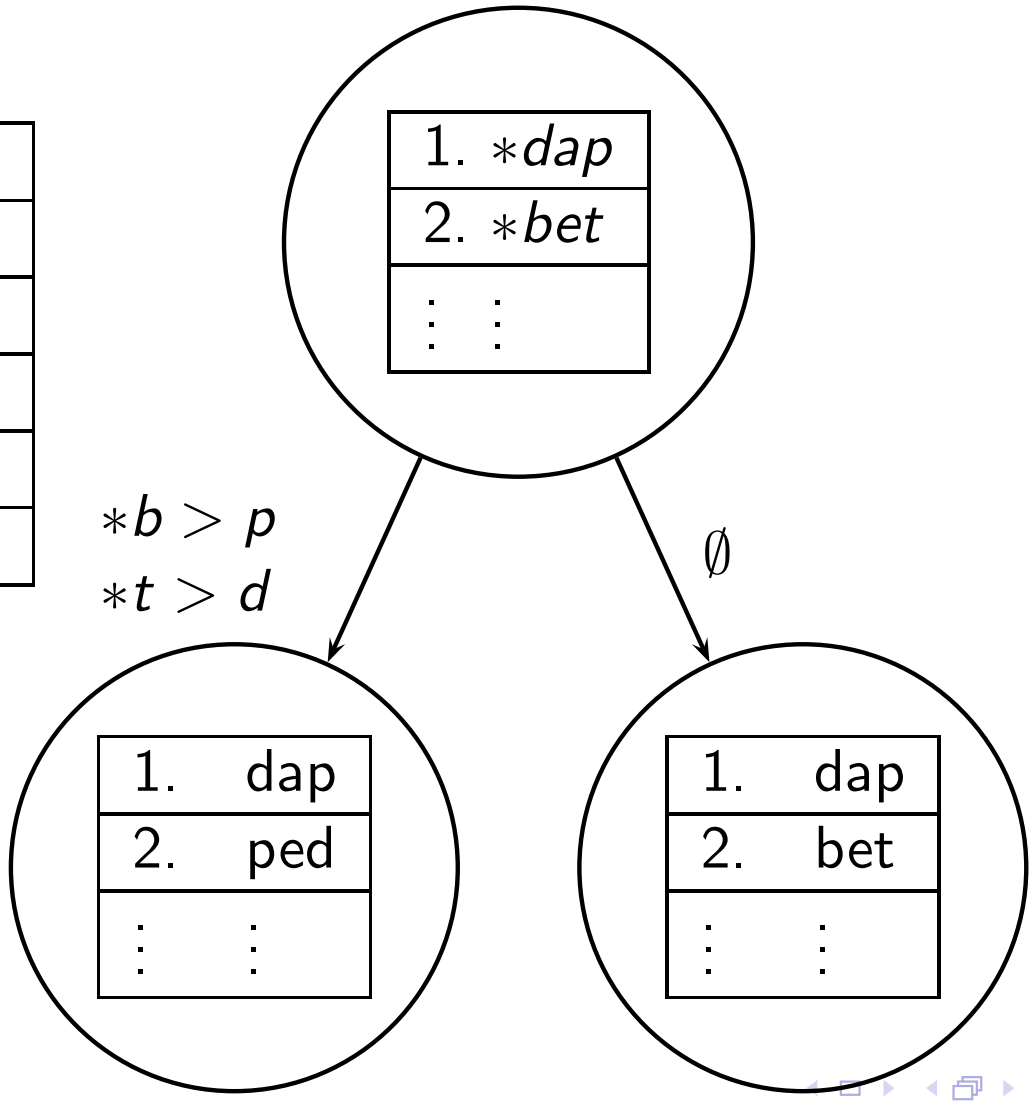
Phonetic Value

d	:	d	$*d$
a	:	a	$*a$
p	:	p	$*p$
p	:	b	$*b$
e	:	e	$*e$
d	:	t	$*t$

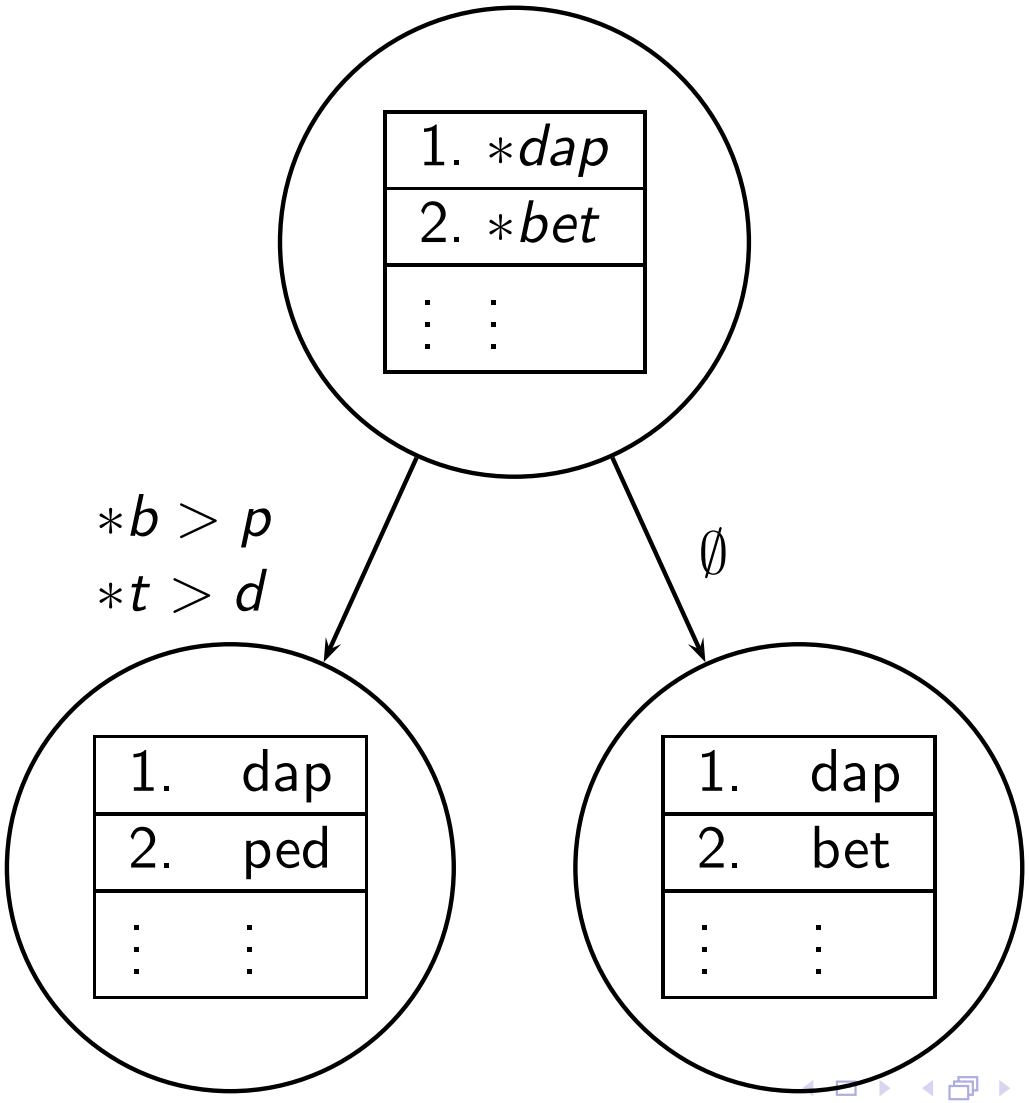


Sound Changes

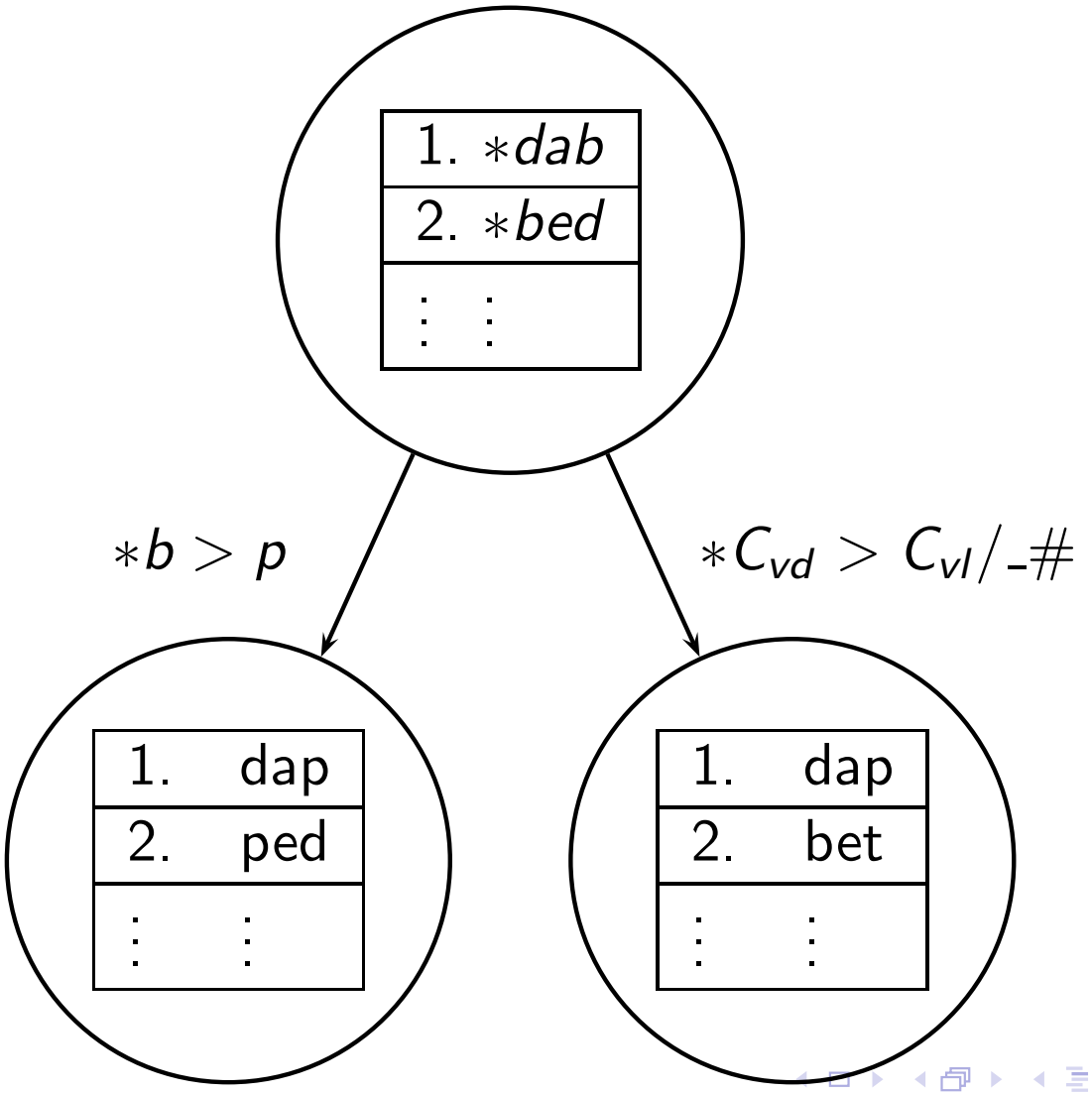
d	:	d	$*d$
a	:	a	$*a$
p	:	p	$*p$
p	:	b	$*b$
e	:	e	$*e$
d	:	t	$*t$



A Complete Reconstruction



Another Reconstruction



Evaluating Hypotheses

- Both reconstructions are *valid*

Evaluating Hypotheses

- Both reconstructions are *valid*
  - They adequately account for the data

Evaluating Hypotheses

- Both reconstructions are *valid*
  - They adequately account for the data
- But is one *better* than the other?

Evaluating Hypotheses

- Both reconstructions are *valid*
  - They adequately account for the data
- But is one *better* than the other?
  - How can the "goodness" of a reconstruction be evaluated?

The Optimal Reconstruction

- The *optimal reconstruction* is the reconstruction, from the set of all valid reconstructions, with the highest probability:

$$\hat{r} = \underset{r}{\operatorname{argmax}}\, p(r)$$

The Optimal Reconstruction

- The *optimal reconstruction* is the reconstruction, from the set of all valid reconstructions, with the highest probability:

$$\hat{r} = \underset{r}{\operatorname{argmax}}\, p(r)$$

- Maximize the likelihood of the hypothesis, given the data.

The Optimal Reconstruction

- The *optimal reconstruction* is the reconstruction, from the set of all valid reconstructions, with the highest probability:

$$\hat{r} = \underset{r}{\operatorname{argmax}}\, p(r)$$

- Maximize the likelihood of the hypothesis, given the data.
- For simplicity, ignore $p(l)$, $p(s)$ (all languages, all tree topologies equally likely)

The Optimal Reconstruction

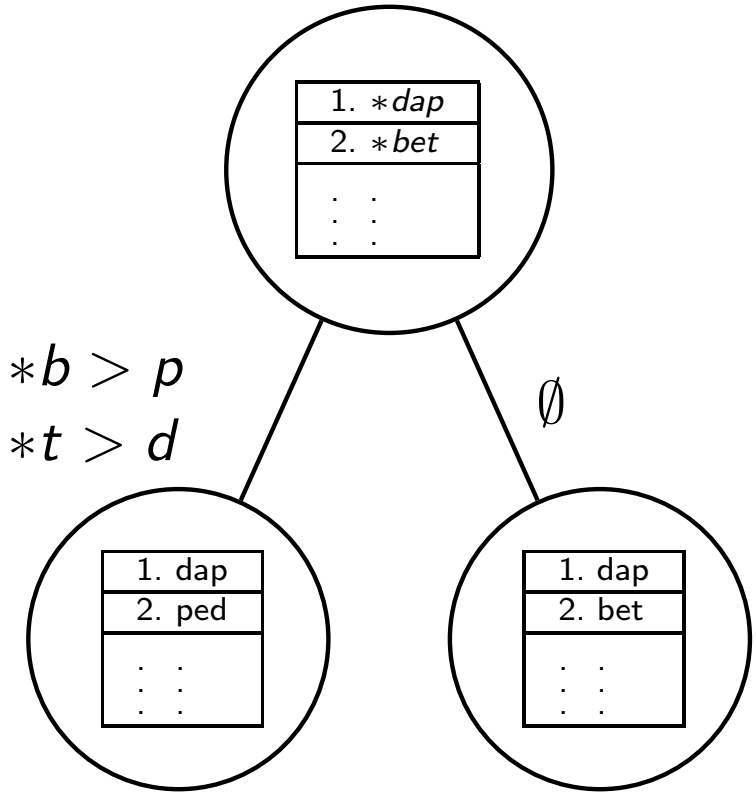
- The *optimal reconstruction* is the reconstruction, from the set of all valid reconstructions, with the highest probability:

$$\hat{r} = \underset{r}{\operatorname{argmax}}\, p(r)$$

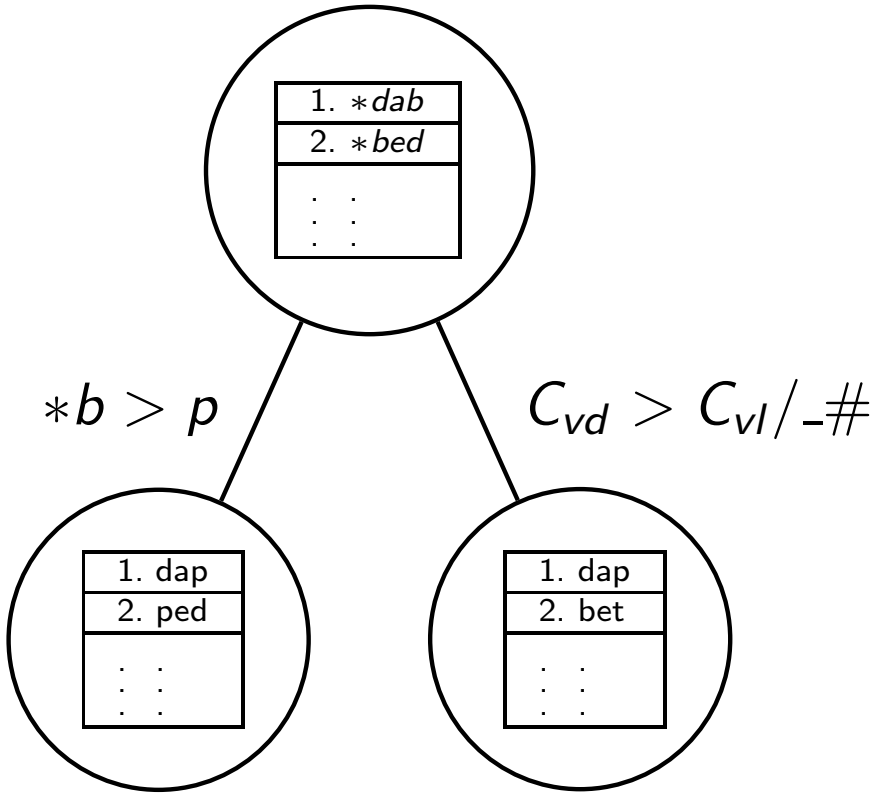
- Maximize the likelihood of the hypothesis, given the data.
- For simplicity, ignore $p(l)$, $p(s)$ (all languages, all tree topologies equally likely)
-

$$\hat{r} = \underset{r}{\operatorname{argmax}} \prod_{c_i \in r} p(c_i)$$

Evaluating Hypotheses



$$p(r_1) \propto p(*b > p) \cdot p(*t > d)$$



$$p(r_2) \propto p(*b > p) \cdot p(*C_{vd} > C_{vl}/\_\#)$$

Constraining the Search Space

- The space of valid reconstructions is very large / infinite.

Constraining the Search Space

- The space of valid reconstructions is very large / infinite.
- Limit the search space by:

Constraining the Search Space

- The space of valid reconstructions is very large / infinite.
- Limit the search space by:
  - Intelligently constraining the possible alignments

Constraining the Search Space

- The space of valid reconstructions is very large / infinite.
- Limit the search space by:
  - Intelligently constraining the possible alignments
  - Efficiently searching through proto-sound values for each correspondence set

Constraining the Search Space

- The space of valid reconstructions is very large / infinite.
- Limit the search space by:
  - Intelligently constraining the possible alignments
  - Efficiently searching through proto-sound values for each correspondence set
- Limiting the search space means that the found result is not necessarily the global optimum.

Constraining the Search Space

- The space of valid reconstructions is very large / infinite.
- Limit the search space by:
  - Intelligently constraining the possible alignments
  - Efficiently searching through proto-sound values for each correspondence set
- Limiting the search space means that the found result is not necessarily the global optimum.
- Any alternative approach, as long as it maximizes the objective function, could work as well or better than this.

Reconstruction Sub-Tasks

Within the reconstruction process, there are several non-trivial sub-tasks, including:

- Segmental word alignment
- Sound change formulation (including conditioning factors)

Segment Alignment

- Segment-to-segment alignment is a necessary step to produce correspondence sets.

Segment Alignment

- Segment-to-segment alignment is a necessary step to produce correspondence sets.
- Accurate alignment is not as easy as it might seem.

Segment Alignment

Previous work on alignment algorithms:

- Alignment based on hand-written costs based on phonetic similarity (Covington 1996, 1998)
- Alignment based on phonetic similarity costs, accounting for extra unanalyzed morphological material (Kondrak 2002)
- Probabilistic alignment of graphemes and phonemes (Knight and Graehl 1998), with many-to-many correspondences (Jiampojamarn et al. 2007)

Alignment Example

English	:	Greek
hundred	:	*hekaton*

Alignment Example

English	:	Greek
hundred	:	*hekaton*

1. A naïve alignment:

h	ə	n	d	r	ɪ	d
h	ɛ	k	a	t	o	n

Alignment Example

English	:	Greek
hundred	:	*hekaton*

2. A possible phonetically-driven alignment:

h	ə	n	d	r	ɪ	d		
h	ɛ		k		a	t	o	n

Alignment Example

English	:	Greek
hundred	:	*hekaton*

3. A more accurate alignment:

	h	ə	n	d	(rɪd)
(hɛ)	k	a		t	(on)

Probabilistic Alignment

a_1

h	ə	n	d	r	ɪ	d		
h	ɛ		k		a	t	o	n

a_2

	h	ə	n	d	(rɪd)
(hɛ)	k	a		t	(on)

$$\Pr(a_1) = \Pr(\text{h:h}) \cdot \Pr(\text{ə:}\varepsilon) \cdot \Pr(\text{n:}\emptyset) \cdot \Pr(\text{d:k}) \ldots$$

$$\Pr(a_2) = \Pr(\emptyset\text{:h}\varepsilon\text{-}) \cdot \Pr(\text{h:k}) \cdot \Pr(\text{ə:a}) \cdot \Pr(\text{n:}\emptyset) \ldots$$

Probabilistic Alignment

a_1

h	ə	n	d	r	ɪ	d		
h	ɛ		k		a	t	o	n

a_2

	h	ə	n	d	(rɪd)
(hɛ)	k	a		t	(on)

$$\Pr(a_1) = \Pr(\text{h:h}) \cdot \Pr(\text{ə:}\varepsilon) \cdot \Pr(\text{n:}\emptyset) \cdot \boxed{\Pr(\text{d:k})} \ldots$$

$$\Pr(a_2) = \Pr(\emptyset\text{:hɛ-}) \cdot \boxed{\Pr(\text{h:k})} \cdot \Pr(\text{ə:a}) \cdot \Pr(\text{n:}\emptyset) \ldots$$

Conditioning Sound Changes

- If more than one correspondence set reconstructs the same proto-sound, then the changes to that proto-sound must be conditioned somehow.

Conditioning Sound Changes

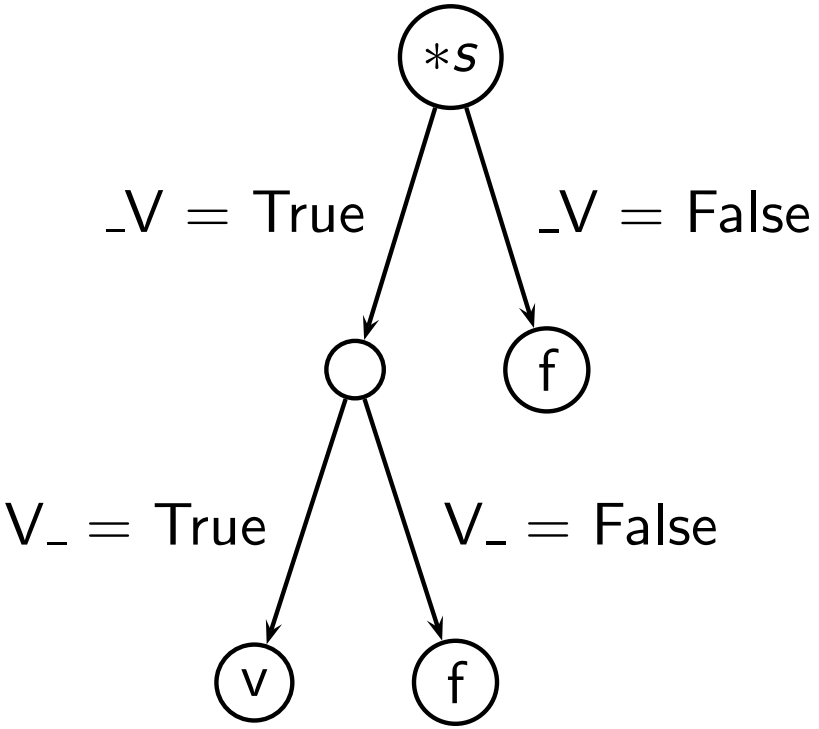
- If more than one correspondence set reconstructs the same proto-sound, then the changes to that proto-sound must be conditioned somehow.
- Treat this as a classification problem
  - Some instances of s_1 undergo c_1
  - Other instances of s_1 undergo c_2
  - What environmental features separate the two classes?

Conditioning Sound Changes

- If more than one correspondence set reconstructs the same proto-sound, then the changes to that proto-sound must be conditioned somehow.
- Treat this as a classification problem
  - Some instances of s_1 undergo c_1
  - Other instances of s_1 undergo c_2
  - What environmental features separate the two classes?
- Use decision tree classifier, and extract phonological rules from the tree.

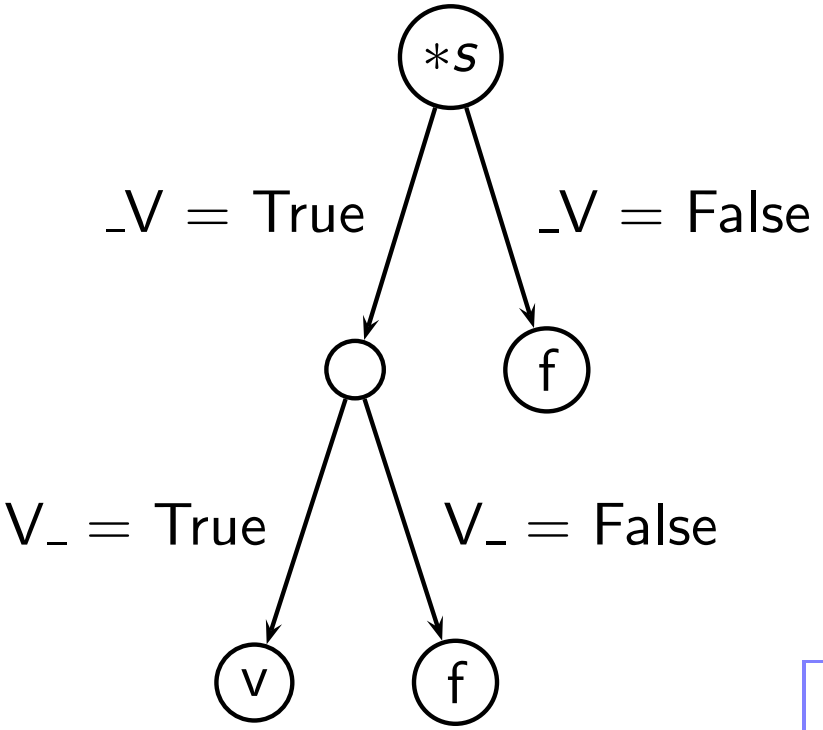
Decision Tree Example

Suppose a proto-sound $*s$ has reflexes [f] and [v] in the same language.



Decision Tree Example

Suppose a proto-sound $*s$ has reflexes [f] and [v] in the same language.



$*f > v / V\_\_V$

Simulation Experiments

Simulated Experiments

- Generate a simulated two-daughter language family:
  1. Assume a three-node topology (two daughters)
  2. Define $p(c)$, $p(l)$
  3. Sample a proto-language $l_0 \sim p(l)$
  4. Sample changes $c_1, c_2 \sim p(c)$ for each branch
  5. Derive l_1 and l_2 by applying the sampled changes

Simulated Experiments

- Generate a simulated two-daughter language family:
  1. Assume a three-node topology (two daughters)
  2. Define $p(c)$, $p(l)$
  3. Sample a proto-language $l_0 \sim p(l)$
  4. Sample changes $c_1, c_2 \sim p(c)$ for each branch
  5. Derive l_1 and l_2 by applying the sampled changes
- Apply the reconstruction algorithm to l_1 and l_2

Simulated Experiments

- Generate a simulated two-daughter language family:
  1. Assume a three-node topology (two daughters)
  2. Define $p(c)$, $p(l)$
  3. Sample a proto-language $l_0 \sim p(l)$
  4. Sample changes $c_1, c_2 \sim p(c)$ for each branch
  5. Derive l_1 and l_2 by applying the sampled changes
- Apply the reconstruction algorithm to l_1 and l_2
- Compare the reconstruction to the (known) true events

Simulated Experiments

- Generate a simulated two-daughter language family:
  1. Assume a three-node topology (two daughters)
  2. Define $p(c)$, $p(l)$
  3. Sample a proto-language $l_0 \sim p(l)$
  4. Sample changes $c_1, c_2 \sim p(c)$ for each branch
  5. Derive l_1 and l_2 by applying the sampled changes
- Apply the reconstruction algorithm to l_1 and l_2
- Compare the reconstruction to the (known) true events
- Compare the reconstruction to a baseline (reconstruct one sister as the proto-language).

Simulation Experiment Results

	Experiment No.			
	1	2	3	4
Changes to l_1	1	1	2	4
Changes to l_2	0	1	2	4
Sound Accuracy (B)	.931	.840	.683	.462
Sound Accuracy (R)	.952	.870	.737	.507
Perfect Rate (B)	.57	0	0	0
Perfect Rate (R)	.68	.31	.07	0
$Pr(C_{true})$	1.1E-2	1.1E-4	1.3E-8	2.7E-16
$Pr(C_{rec.})$	1.2E-2	3.2E-2	8.4E-5	1.2E-4

(B) = baseline; (R) = reconstruction algorithm

Simulation Experiment Results

	Experiment No.			
	1	2	3	4
Changes to l_1	1	1	2	4
Changes to l_2	0	1	2	4
Sound Accuracy (B)	.931	.840	.683	.462
Sound Accuracy (R)	.952	.870	.737	.507
Perfect Rate (B)	.57	0	0	0
Perfect Rate (R)	.68	.31	.07	0
$Pr(C_{true})$	1.1E-2	1.1E-4	1.3E-8	2.7E-16
$Pr(C_{rec.})$	1.2E-2	3.2E-2	8.4E-5	1.2E-4

(B) = baseline; (R) = reconstruction algorithm

Simulation Experiment Results

	Experiment No.			
	1	2	3	4
Changes to l_1	1	1	2	4
Changes to l_2	0	1	2	4
Sound Accuracy (B)	.931	.840	.683	.462
Sound Accuracy (R)	.952	.870	.737	.507
Perfect Rate (B)	.57	0	0	0
Perfect Rate (R)	.68	.31	.07	0
$Pr(C_{true})$	1.1E-2	1.1E-4	1.3E-8	2.7E-16
$Pr(C_{rec.})$	1.2E-2	3.2E-2	8.4E-5	1.2E-4

(B) = baseline; (R) = reconstruction algorithm

Simulation Experiment Results

	Experiment No.			
	1	2	3	4
Changes to l_1	1	1	2	4
Changes to l_2	0	1	2	4
Sound Accuracy (B)	.931	.840	.683	.462
Sound Accuracy (R)	.952	.870	.737	.507
Perfect Rate (B)	.57	0	0	0
Perfect Rate (R)	.68	.31	.07	0
$Pr(C_{true})$	1.1E-2	1.1E-4	1.3E-8	2.7E-16
$Pr(C_{rec.})$	1.2E-2	3.2E-2	8.4E-5	1.2E-4

(B) = baseline; (R) = reconstruction algorithm

Simulation Experiment Results

	Experiment No.			
	1	2	3	4
Changes to l_1	1	1	2	4
Changes to l_2	0	1	2	4
Sound Accuracy (B)	.931	.840	.683	.462
Sound Accuracy (R)	.952	.870	.737	.507
Perfect Rate (B)	.57	0	0	0
Perfect Rate (R)	.68	.31	.07	0
$Pr(C_{true})$	1.1E-2	1.1E-4	1.3E-8	2.7E-16
$Pr(C_{rec.})$	1.2E-2	3.2E-2	8.4E-5	1.2E-4

(B) = baseline; (R) = reconstruction algorithm

Questions with Answers

- Can this approach work on simulated data sets? Yes

Questions with Answers

- Can this approach work on simulated data sets? Yes
- Can this approach work on large, natural-langauge data sets? I expect so.

Questions with Answers

- Can this approach work on simulated data sets? Yes
- Can this approach work on large, natural-langauge data sets? I expect so.
- Is this a useful way of looking at language change? I think so.

Questions with Answers

- Can this approach work on simulated data sets? Yes
- Can this approach work on large, natural-langauge data sets? I expect so.
- Is this a useful way of looking at language change? I think so.
- Will this approach replace traditional comparative reconstruction? I doubt it.

Questions without Answers

- How difficult is it to reconstruct a family with just two languages?

Questions without Answers

- How difficult is it to reconstruct a family with just two languages?
- What are the ultimate limitations of this model? (assumptions of independence, restrictive view of language)

Questions without Answers

- How difficult is it to reconstruct a family with just two languages?
- What are the ultimate limitations of this model? (assumptions of independence, restrictive view of language)
- How can probabilities be assigned to real sound changes?

Questions without Answers

- How difficult is it to reconstruct a family with just two languages?
- What are the ultimate limitations of this model? (assumptions of independence, restrictive view of language)
- How can probabilities be assigned to real sound changes?
- What are the (dis)advantages of a computational model?

Thank You!

- Thank you!
- Questions?

Some References:

- Kondrak, Grzegorz, 2002. Algorithms for Language Reconstruction. Ph.D. thesis, University of Toronto.
- Bouchard, Alexandre, Percy Liang, Thomas Griffiths, and Dan Klein, 2007. A probabilistic approach to diachronic phonology. EMNLP-CoNLL'07, 887–96.
- Covington, Michael A., 1996. An algorithm to align words for historical comparison. Computational Linguistics, 22.4. 481–96.
- Covington, Michael A., 1998. Alignment of multiple languages for historical comparison. Proceedings of COLING-ACL'98, 275–80.