# LI 511: Computational Models of Sound Change

James Kirby and Morgan Sonderegger

18 July 2013

# Sound change

- Typology of sound changes
  - conditioned vs. unconditioned
  - phonemic vs. nonphonemic (allophonic)
  - regular vs. sporadic

- Our models and the typology

# Conditioned vs. unconditioned

- *Unconditioned*: context-free
  - Spanish *l*$^{y}$ > Latin Amer Spanish *j*

- *Conditioned:* context-sensitive
  - Latin *vita* > Spanish *vida*, but
  - Latin *tempus* > Spanish *tiempo*, **diempo*

# Phonemic vs. nonphonemic

- In essence: shift w/in existing system of contrasts, or creation/deletion of phonemes?
- Can also be conditioned or unconditioned
- Nonphonemic (allophonic, phonetic)
  - English *u* > *ʉ*, V > Ṽ / __ [+nasal]
- Can our models do these?

# Phonemic change

- Campbell's typology:
  - Merger (A, B > B or A, B > C)
  - Split (A > B, C)
- Differs somewhat from other accounts
- One man's splits, etc.

# Unconditioned merger

- Total loss of contrast between A and B in all environments (*cot* ~ *caught*)

| | *PIE* | *Greek* | *Latin* | *Gothic* | *OHG* | |
|---|---|---|---|---|---|---|
| *o | *oktṓ(u)- | oktṓ | octo | axtau | ahto | 'eight' |
| *ə | *pətēr | patḗr | pater | fadar | fater | 'father' |
| *a | *agro- | agrós | ager | akrs | ackar | 'field' |

- Rare: most mergers are conditioned

# Conditioned merger
## (a.k.a. primary split)

- some *allophones* of A merge with B

- Ex.: final 'devoicing' in German, Dutch:
  - *Rad, Rat* > [ʁɑt]

- Ex.: Latin rhotacism
  - *rural* (< *rūs-al*) but *rustic* (< *rūs-ticus*)

# Split (a.k.a. secondary split)

- Sounds involved in splits don't actually change

- Ex.: Slavic palatalization
  - *krov$^j$ĭ 'blood' vs. *krovŭ 'shelter'
  - /ĭ ŭ/ > Ø, so allophones [v$^j$ v] > /v$^j$ v/

- Follows 'merger with Ø' (deletion)

# Phonemic split

TABLE 2.2: Historical derivation of 'mouse', 'mice', 'foot', 'feet'

| | *mouse* | *mice* | *foot* | *feet* |
|---|---|---|---|---|
| Stage 1 (no changes) | /mu:s/ | /mu:s-i/ | /fo:t/ | /fo:t-i/ |
| | [mu:s] | [mu:s-i] | [fo:t] | [fo:t-i] |
| Umlaut | /mu:s/ | /mu:s-i/ | /fo:t/ | /fo:t-i/ |
| | [mu:s] | [my:s-i] | [fo:t] | [fø̄:t-i] |
| Loss of *-i* (= split after merger) | /mu:s/ | /my:s/ | /fo:t/ | /fø̄:t/ |
| | [mu:s] | [my:s] | [fo:t] | [fø̄:t] |
| Unrounding | /mu:s/ | /mi:s/ | /fo:t/ | /fe:t/ |
| | [mu:s] | [mi:s] | [fo:t] | [fe:t] |
| Great Vowel Shift | /maus/ | /mais/ | /fu:t/ | /fi:t/ |

# Chain shifts

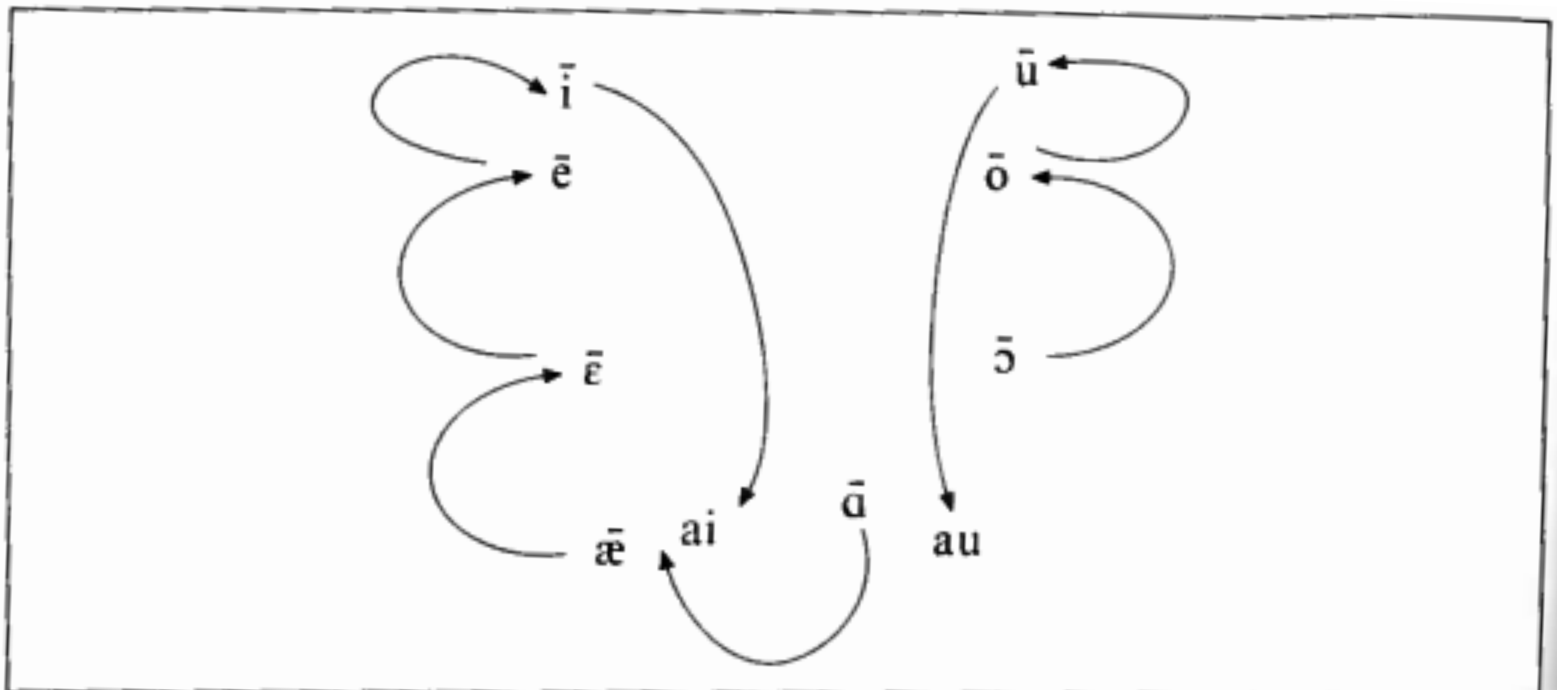



FIGURE 2.1: The Great Vowel Shift in English

# Attic Greek chain shifts

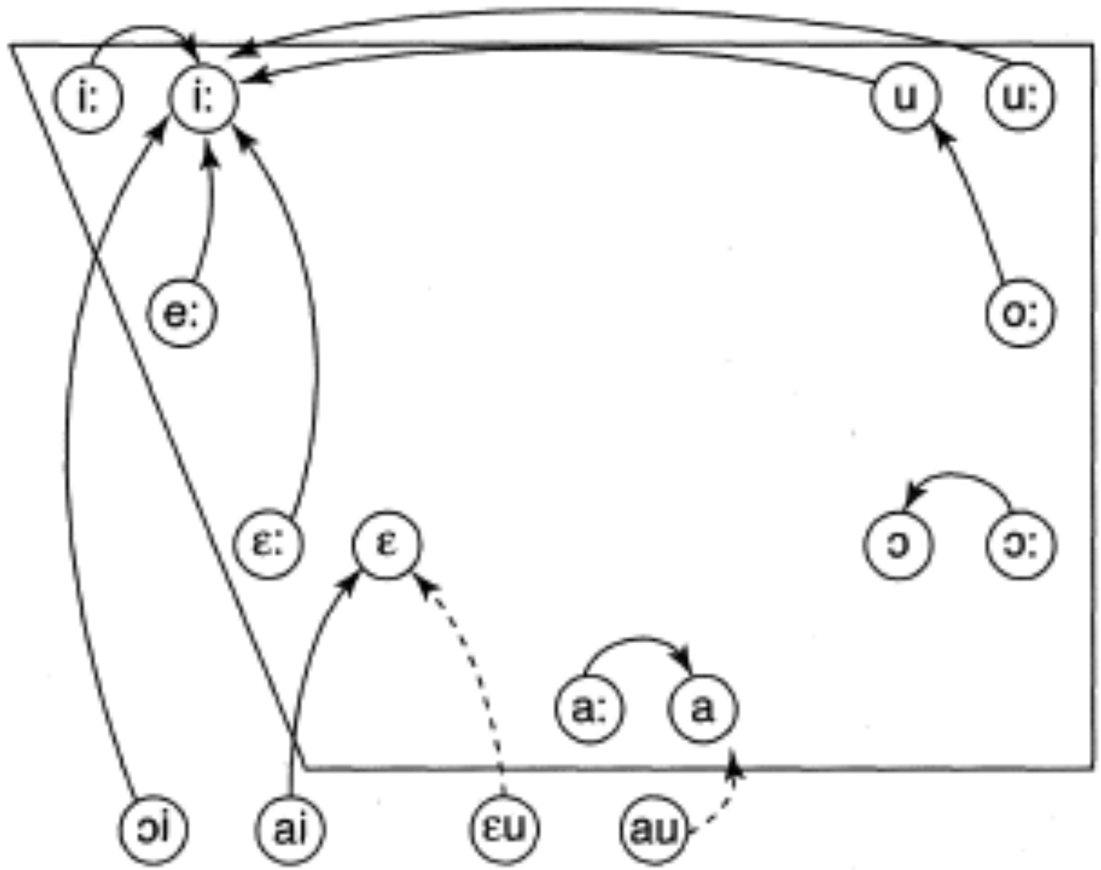



Fig. 4.4 The history of the Greek vowel system

# Mandarin: order effects

- The emergence of palatals:
  - *EM* = Early Mandarin, *Zhongyuan Yinyun* around 1324 AD;
  - Man. = Mandarin, Standard Chinese, Putonghua.

| | | | | | | |
|---|---|---|---|---|---|---|
| EM | si | xi | ts$^{h}$i | k$^{h}$i | tsi | ki |
| | ↘ | ↓ | ↘ | ↓ | ↘ | ↓ |
| Man. | | ɕi | | tɕ$^{h}$i | | tɕi |

# Mandarin: order effects

- The emergence of palatals:
  - Velars palatalize first, then dentals

| | | | | | | |
|---|---|---|---|---|---|---|
| EM | si | xi | ts$^h$i | k$^h$i | tsi | ki |
| | si | ɕi | ts$^h$i | tɕ$^h$i | tsi | tɕi |
| Man. | | ɕi | | tɕ$^h$i | | tɕi |

# Axioms

- 'Mergers are irreversible'
- 'Splits follow mergers'
- Do our models obey these axioms?

# Beddor (2009)

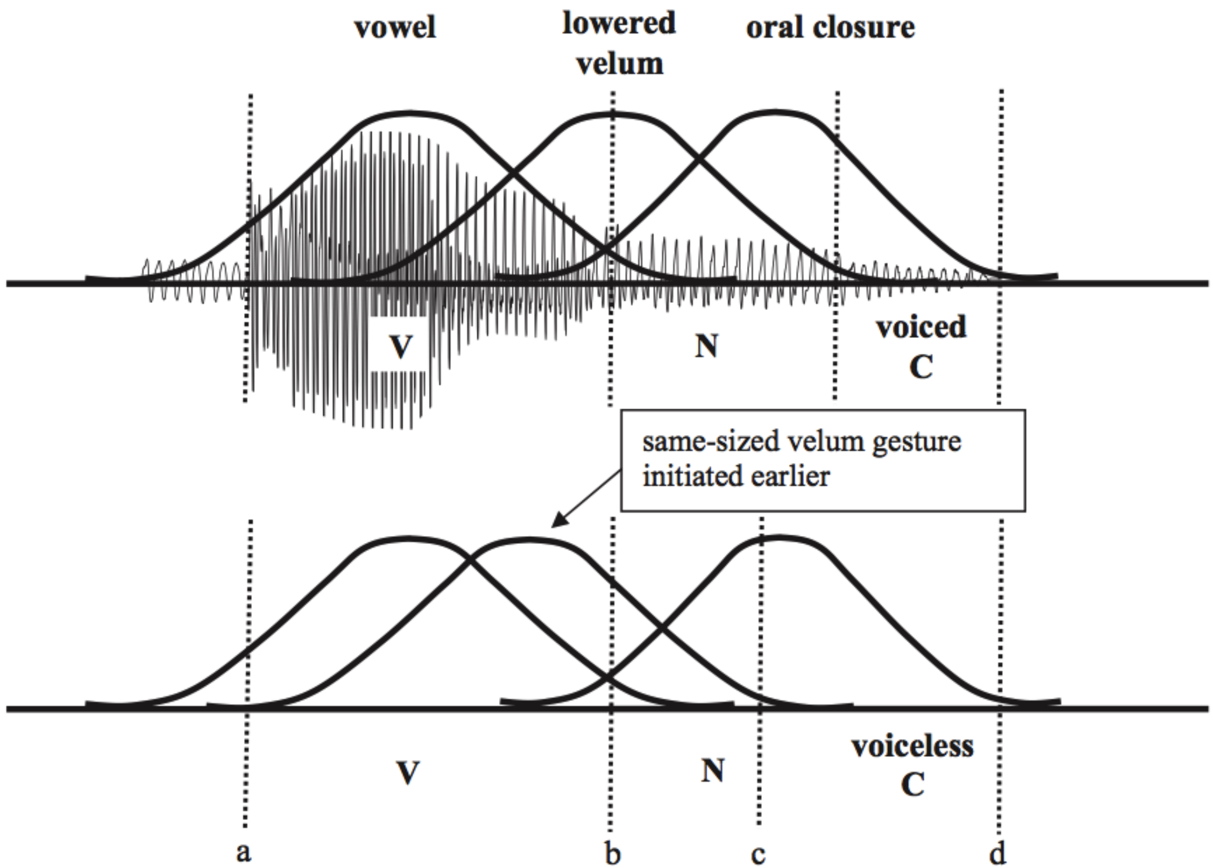



FIGURE 1. Schematic representation of the consequences for vowel nasalization, the nasal consonant, and the postnasal oral constriction if the velum gesture is initiated earlier in voiceless (bottom) than in voiced (top) contexts. Dashed lines indicate acoustic segmentation.

# Beddor (2009)

- Don't 'factor out' coarticulatory effects, but selectively weight dimensions
- Maybe solves some Ohalaian problems?
- Relation to Kirby's/mixture models?

# Types of sound change

- Assimilation/dissimilation
- Deletion (syncope, apocope, aphaeresis)
- Epenthesis (prothesis, anaptyxis, excrescence)
- Compensatory lengthening
- Metathesis
- ...

# Computational historical linguistics

- Growing area since ≈ mid-90s
- For a classic HL task:
  - Adapt computational, mathematical methods
  - Often from analogous problems in computational biology or linguistics
- Advantages, disadvantages vs traditional methods

# One example: Phylogeny

- HL task: Given cognate sets in modern languages, reconstruct tree
- Methods: Algs from computational biology
  - Given modern genetic info, reconstruct tree
  - Not intermediate stages
- Why it works:
  - Assumptions about evolution ("normal" sound change, vertical vs. horizontal transmission…) very similar to genetics

- Problems: Unclear how to incorporate syntactic change, extensive contact, non-simple sound changes

- Payoff:
  - Leverage (way) more data
  - Deeper reconstruction; time estimates (maybe)
  - Uncertainty estimates

- Applied to Austronesian, Indo-European, Pama-Nyungan, …
  (Bowern & Atkinson, 2012; Dunn et al., 2002; Gray & Atkinson, 2003; Warnow et al., 1995 *passim*…)

Review (old): Nichols & Warnow (2008), *Language & Linguistics Compass*

# Other examples

- Determining cognate sets
  (e.g. Ellison, 2007; Hall & Klein, 2010, 2011)

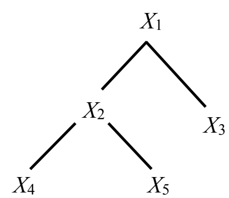
- Reconstructing sound changes, proto-forms
  (Oakes, 2000; Kondrak, 2002)

# Bouchard-Côté et al. (2013): High level

## Input: modern words

| | 'fish' | 'fear' |
|---|---|---|
| Hawaiian | iʔa | makaʔu |
| Samoan | iʔa | mataʔu |
| Tongan | ika | |

## Probabilistic model of change




$$X_2 | X_1 \sim \mathrm{Mut}(X_1; \eta)$$

$$X_3 | X_1 \sim \mathrm{Mut}(X_1; \eta)$$

$$\vdots$$

$$X_i | X_{\mathrm{par}(i)} \sim \mathrm{Mut}(X_{\mathrm{par}(i)}; \eta)$$

## Reconstruction

| | 'fish' |
|---|---|
| POc | *ika |

or

| | 'fish' |
|---|---|
| POc | *iʔa |

⋮

## Cognates

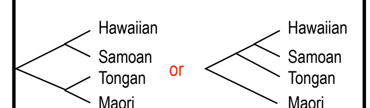
| | 'fish' (1) |
|---|---|
| Hawaiian | iʔa |
| Samoan | iʔa |
| Tongan | ika |
| Marshallese | yapil |

or

| | 'fish' (1) | 'fish' (2) |
|---|---|---|
| Hawaiian | iʔa | |
| Samoan | iʔa | |
| Tongan | ika | |
| Marshallese | | yapil |

..

## Trees



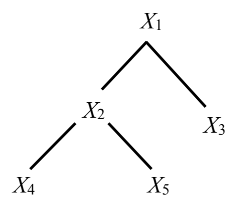

Thanks for slides: Alexandre Bouchard-Côté

# Bouchard-Côté et al (2013): High level

## Input: modern words

| | 'fish' | 'fear' |
|---|---|---|
| Hawaiian | iʔa | makaʔu |
| Samoan | iʔa | mataʔu |
| Tongan | ika | |

## Probabilistic model of change




$$X_2 | X_1 \sim \mathrm{Mut}(X_1; \eta)$$

$$X_3 | X_1 \sim \mathrm{Mut}(X_1; \eta)$$

$$\vdots$$

$$X_i | X_{\mathrm{par}(i)} \sim \mathrm{Mut}(X_{\mathrm{par}(i)}; \eta)$$

### Reconstruction

| | 'fish' |
|---|---|
| POc | *ika |

or

| | 'fish' |
|---|---|
| POc | *iʔa |

⋮

### Cognates

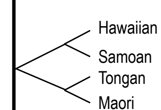
| | 'fish' (1) |
|---|---|
| Hawaiian | iʔa |
| Samoan | iʔa |
| Tongan | ika |
| Marshallese | yapil |

or

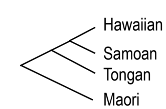
| | 'fish' (1) | 'fish' (2) |
|---|---|---|
| Hawaiian | iʔa | |
| Samoan | iʔa | |
| Tongan | ika | |
| Marshallese | | yapil |

..

### Trees




or




..

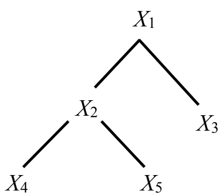
The dream: all three jointly

# Bouchard-Côté et al (2013): High level

Input: modern words

| | 'fish' | 'fear' |
|---|---|---|
| Hawaiian | iʔa | makaʔu |
| Samoan | iʔa | mataʔu |
| Tongan | ika | |

Probabilistic model of change




$$X_2 | X_1 \sim \mathrm{Mut}(X_1; \eta)$$
$$X_3 | X_1 \sim \mathrm{Mut}(X_1; \eta)$$
$$\vdots$$
$$X_i | X_{\mathrm{par}(i)} \sim \mathrm{Mut}(X_{\mathrm{par}(i)}; \eta)$$

Reconstruction

| | 'fish' |
|---|---|
| POc | *ika |

or

| | 'fish' |
|---|---|
| POc | *iʔa |

⋮

Cognates

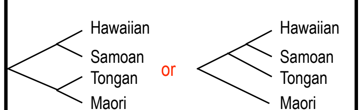
| | 'fish' (1) |
|---|---|
| Hawaiian | iʔa |
| Samoan | iʔa |
| Tongan | ika |
| Marshallese | yapil |

or

| | 'fish' (1) | 'fish' (2) |
|---|---|---|
| Hawaiian | iʔa | |
| Samoan | iʔa | |
| Tongan | ika | |
| Marshallese | | yapil |

...

Trees




The paper: trees given,
better performance if cognates given

# Bouchard-Côté et al (2013): High level

- Explicit, probabilistic model of sound change, word innovation at each tree edge

- Output: Posterior distributions over:
  - Sound changes for each tree edge
  - Proto-forms, intermediate forms

# Bouchard-Côté et al (2013): High level

- Explicit, probabilistic model of sound change, word innovation at <u>each</u> tree edge

- Output: Posterior distributions* over:
  - Sound changes for each tree edge
  - Proto-forms, intermediate forms

- Major advance vs. previous work

*: given what?

1. Model of sound change

2. Algorithm

3. Validation

4. Visualizing results

5. Application: Functional load

# Model of sound change

- Single-char substitution, insertions, deletions
  - Can be sensitive to left context
- Any $x > y$, $\varnothing > x$, $x > \varnothing$ can happen
  - At each edge, with some edge-dependent probability
  - Within a cognate set
- Same probability model for all cognate sets

# Model of sound change

- Single-char substitution, insertions, deletions
  - Can be sensitive to left context
- Any $x > y$, $\varnothing > x$, $x > \varnothing$ can happen
  - At each edge, with some edge-dependent probability
  - Within a cognate set
- Same probability model for all cognate sets
- What is/isn't here?

- Bouchard-Côté slides 22-38

- Note:
  - Separate evolution for each cognate set
  - Shared probabilistic model of SC across cognate sets

- Where does the regularity of SC come from here?

# Data

- Austronesian Basic Vocab Database (Greenhill et al., 2008)
  - 659 languages, 140k wordforms, 7.7k cognate sets
- Trees:
  - Manual: Ethnologue (2009)
  - Automatic: Gray et al (2009)
- Manual reconstructions (for validation)
  - Proto-Oceanic (x2): 50% of langs (Blust, 1993; Pawley, 2009)
  - Proto-Austronesian (x1) (Blust, 1999)

# Algorithm

- Bouchard-Côté  slides 49-51

# Valididation: Comparing automatic & manual reconstructions

| | 'fish' | 'fear' |
|---|---|---|
| Hawaiian | iʔa | makaʔu |
| Samoan | iʔa | mataʔu |
| Tongan | ika | |
| Proto-Oceanic | *ika | *mataku |

| | 'fish' | 'fear' |
|---|---|---|
| Hawaiian | iʔa | makaʔu |
| Samoan | iʔa | mataʔu |
| Tongan | ika | |
| Proto-Oceanic | ??? | ??? |

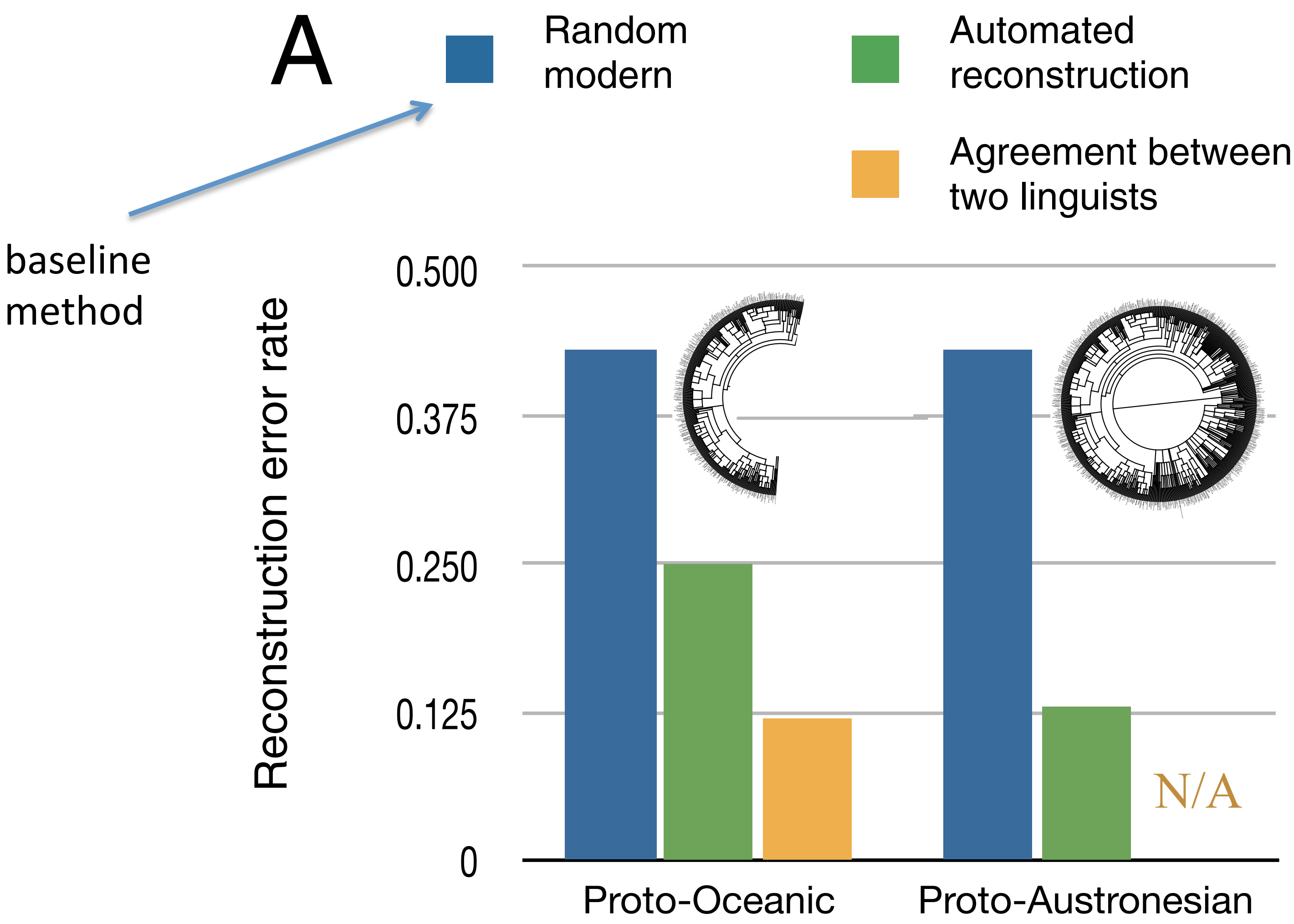
**Evaluation criterion: Edit distance**

Smallest number of substitutions, insertions and deletions needed to go from one string to the other

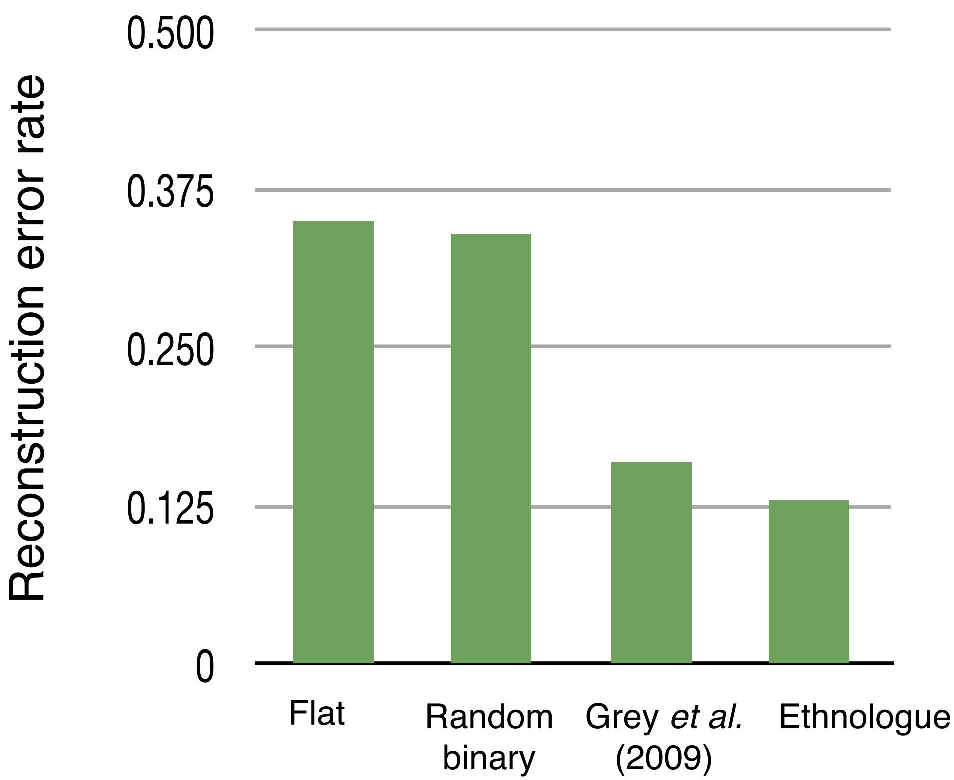
e.g.: d( /ika/ , /ga/ ) = 2

/ika/ ➔ /iga/ ➔ /ga/

# Comparison to manual reconstruction



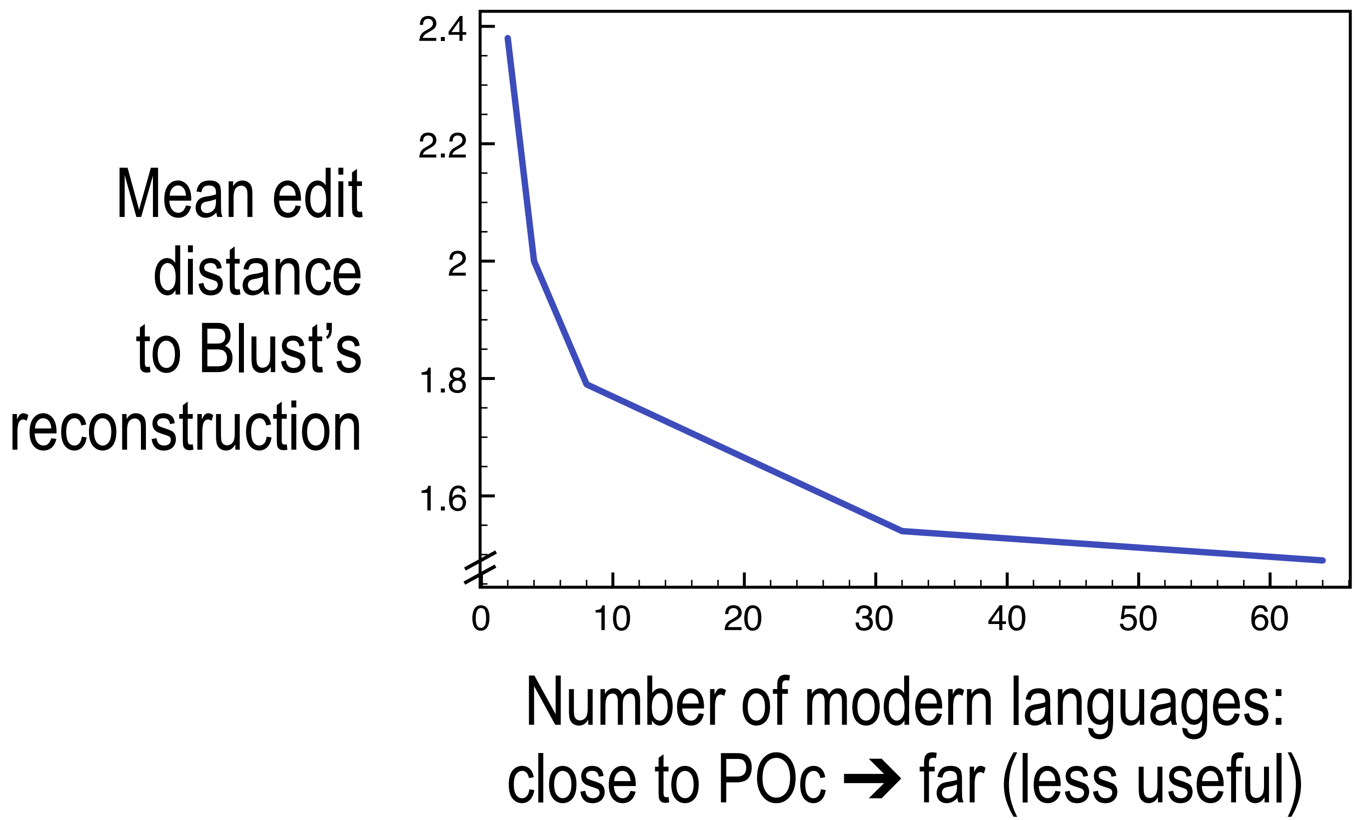
# Input to algorithm matters a lot




- tree quality helps ⇒ good trees needed

- Error rate increases without labeled cognate sets

- Amount of data matters (not obvious!)




- ⇒ need large, high-quality datasets of modern forms

# What changes and protoforms are learned?

- Bouchard-Côté slides 69-83

# Application

- Functional load: Information carried by a contrast between /x/ and /y/ in a language
  - English: p/b (high) vs. t/θ (low)

- Hypothesis (Martinet, 1955; etc.): Sound changes are less likely if they merge phonemes with high FL

- Previous work (King, 1967; Hockett, 1967; Surendran & Niyogi, 2006, etc.)
  - Mixed results
  - Small number of languages

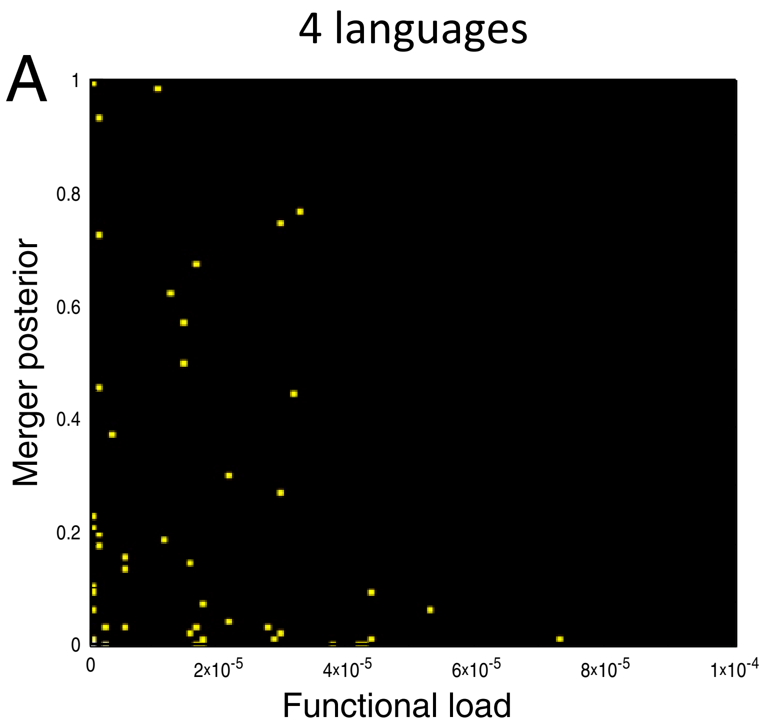
# Testing the FL hypothesis

- Can test using probabilistic model of sound change induced for Austronesian tree

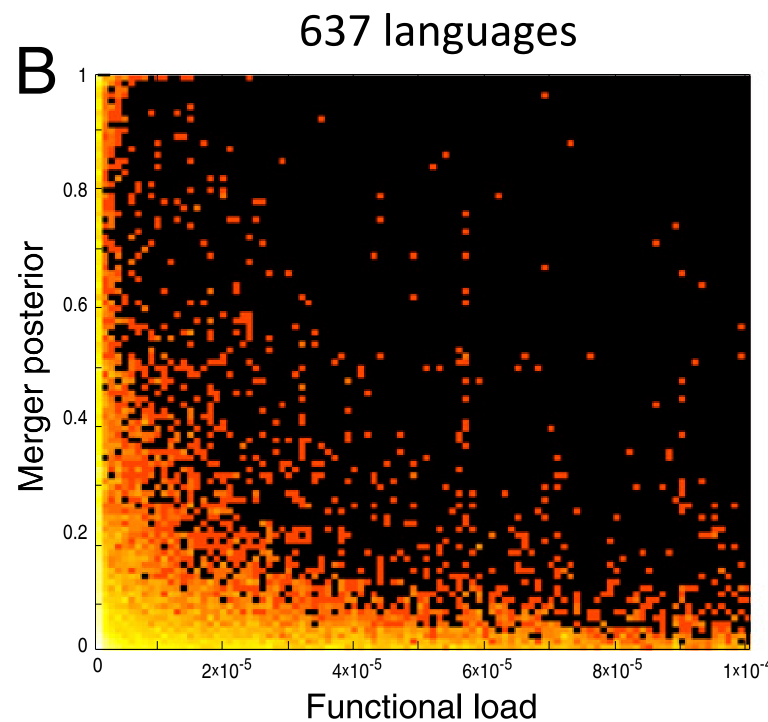
- Merger probability at language *L* :
  - Posterior over all *x* > *y* changes at *L* , relative to other *x* > *z* changes

- Functional load:
  - Posterior prediction for measure used by King (1967)

# Results

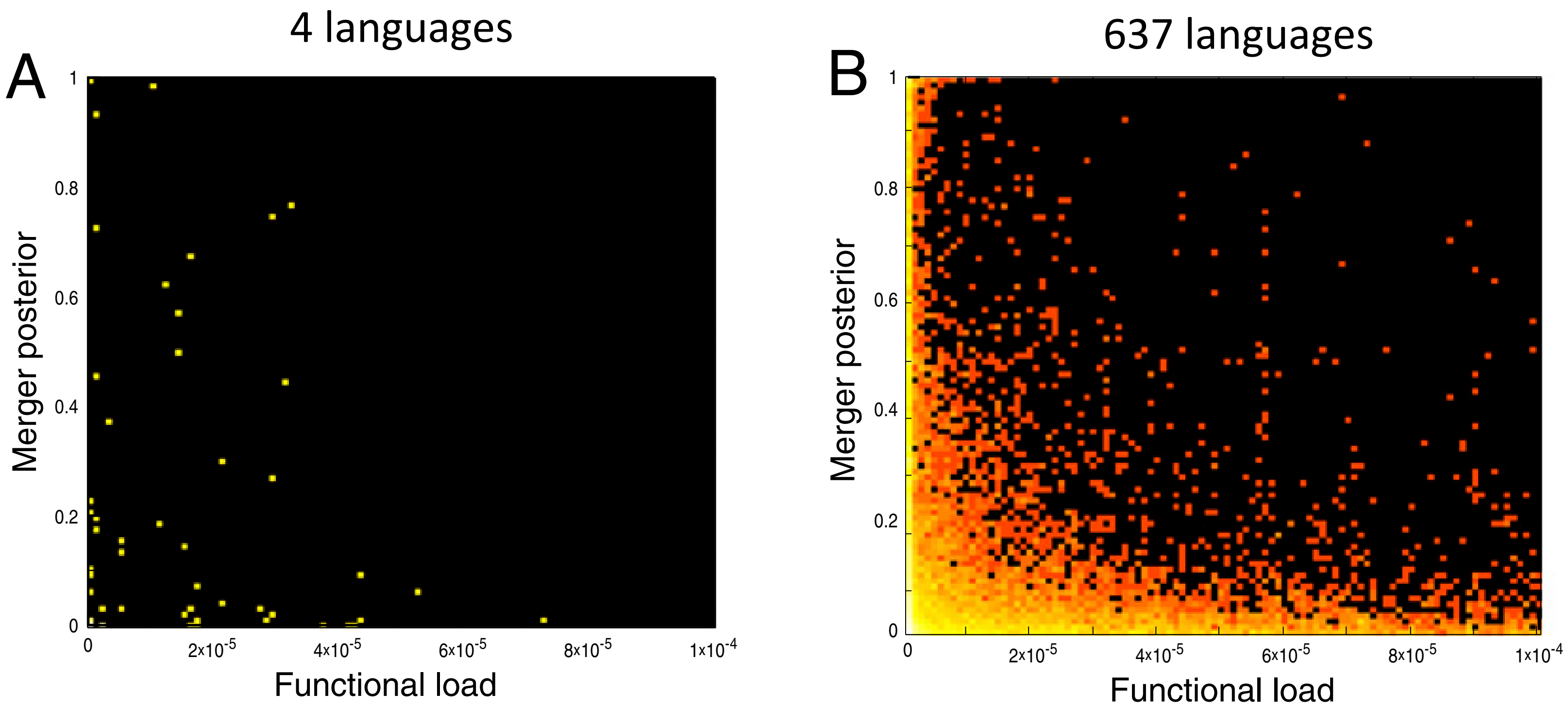


4 languages
A
1
0.8
0.6
0.4
0.2
0
Merger posterior
0
2x10-5
4x10-5
6x10-5
8x10-5
1x10-4
Functional load

637 languages
B
1
0.8
0.6
0.4
0.2
0
Merger posterior
0
2x10-5
4x10-5
6x10-5
8x10-5
1x10-4
Functional load

# Results




- High FL inhibits merger
  (c.f. Wedel et al. , 2013)

- Some issues: Definition of FL, merger, …
  - Should check finding varying these ‘parameters’

- Still: Very clear pattern, thanks to
  - Large dataset
  - Probabilistic model of SC

# What have we learned?

- Many parameters impact rates of change rather than qualitative outcomes
  - But not all (e.g., learning algorithm)
  - Not necessarily obvious a priori

- On its own, simple mistransmission (e.g., lenition bias) predicts unrealistic dynamics (at least with simple models)

# What have we learned?

- Modeling stability is non-trivial
  - Multiple stables states even harder?
  - Models considered here needed a stabilizing force (entrenchment, strong prior, …)
  - Possibly emergent (de Boer?)

- Differences between individual- and population-level models

# What have we learned?

- Insights into the general behavior of simple models (individuals, populations)

- Convergence: similar results from multiple (simple) models suggests something about the properties of the 'true' (complex) model

# Future directions

- Modeling as a two-way street
  - more complex scenarios (e.g. Greek, Mandarin)
  - more complex (real) data (e.g. Beddor)
- ...