UCLA
Southern California Annual Meeting on Phonology 2026

Discerning segmental stability using phylogenetic models

Muhammad Rehan
University of California, Los Angeles
April 11, 2026 · rehanmuh@ucla.edu

Introduction

Southern California Annual Meeting on Phonology 2026

Sound change

What we know

- sound change is often regular
- conditioned by phonetic environments
- some changes recur across families

Sound change

What we know

- sound change is often regular
- conditioned by phonetic environments
- some changes recur across families

What we do not know

- how fast specific segments change
- how context changes those rates
- how much variation there is in change across language families

Sound change

What we know

- sound change is often regular
- conditioned by phonetic environments
- some changes recur across families

What we do not know

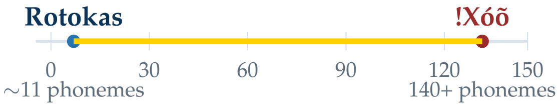
- how fast specific segments change
- how context changes those rates
- how much variation there is in change across language families

We lack a quantitative and principled model for modeling the evolution of segments over time on family trees.

Sources: Maddieson, 1984; Hock, 1991; Labov, 2020; Garrett, 2015

The scale of phonological diversity

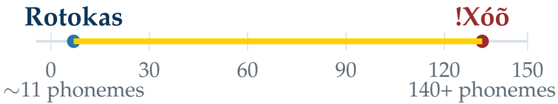
Any rate model must explain both stability and extreme variation in the make-up of phoneme inventories



- inventories vary enormously in size

The scale of phonological diversity

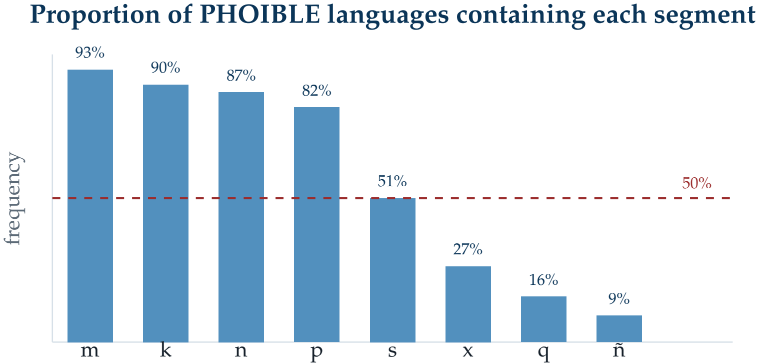
Any rate model must explain both stability and extreme variation in the make-up of phoneme inventories



- inventories vary enormously in size
- some segments are widespread; others rare and varying in stability

Sources: Maddieson, 1984; Moran and McCloy, 2019; Blasi et al., 2019

Cross-linguistic frequency as a diagnostic for stability



Cross-linguistic frequency is a proxy for diachronic stability, but static snapshots might not represent linguistic systems in equilibrium.

Sources: Moran and McCloy, 2019; Moran et al., 2021; Maddieson, 1984

Phonetic grounding

Why some changes recur across lineages

Phonetic variation → **Listener (mis)parsing** → **Phonologization**

Theoretical Consensus

Change is driven by misperception (Ohala, 1993), leading to recurrent patterns (Blevins, 2004) that are inherently phonetically biased and directional (Garrett and Johnson, 2013).

Sources: Ohala, 1981; Ohala, 1993; Blevins, 2004; Garrett and Johnson, 2013; Kirby, 2013

Phonetic grounding

Why some changes recur across lineages

Phonetic variation → Listener (mis)parsing → Phonologization

Theoretical Consensus

Change is driven by misperception (Ohala, 1993), leading to recurrent patterns (Blevins, 2004) that are inherently phonetically biased and directional (Garrett and Johnson, 2013).

Modeling implication

Rates of segmental change must be structured and not arbitrary.

Sources: Ohala, 1981; Ohala, 1993; Blevins, 2004; Garrett and Johnson, 2013; Kirby, 2013

What a probabilistic model should look like

Regularity & pooling

Same local context $\Rightarrow$ same change parameters across lexical items.

Sources: Hayes and Wilson, 2008; Hayes and Steriade, 2004; Hayes and White, 2015; Zuraw, 2010; Moore-Cantwell, 2016

What a probabilistic model should look like

Regularity & pooling

Same local context $\Rightarrow$ same change parameters across lexical items.

Gradient exceptionality

Skew, partial productivity, and near-categorical trends.

Sources: Hayes and Wilson, 2008; Hayes and Steriade, 2004; Hayes and White, 2015; Zuraw, 2010; Moore-Cantwell, 2016

What a probabilistic model should look like

Regularity & pooling

Same local context ⇒ same change parameters across lexical items.

Gradient exceptionality

Skew, partial productivity, and near-categorical trends.

Naturalness

*/p/ > [b] should look more plausible than */p/ > [r]. Asymmetries should enter the prior.

Sources: Hayes and Wilson, 2008; Hayes and Steriade, 2004; Hayes and White, 2015; Zuraw, 2010; Moore-Cantwell, 2016

What a probabilistic model should look like

Regularity & pooling

Same local context ⇒ same change parameters across lexical items.

Gradient exceptionality

Skew, partial productivity, and near-categorical trends.

Naturalness

*/p/ > [b] should look more plausible than */p/ > [r]. Asymmetries should enter the prior.

Parameter Learnability

Restricted enough that finite cognate data can identify something interpretable.

Sources: Hayes and Wilson, 2008; Hayes and Steriade, 2004; Hayes and White, 2015; Zuraw, 2010; Moore-Cantwell, 2016

Previous computational work in this domain

Model / Study	Context	Rate Var.	Naturalness	Insertions	Deletions
Bouchard-Côté et al. (2013)	✓	✓	✗	✓	✓
Hruschka et al. (2015)	✓	✗	✗	✗	✗
Moran et al. (2021)	✗	✓	✗	✗	✗
He et al. (2024)	✓	✗	✗	✓	✓
Cathcart & Wandl (2020)	✓	✓	✓	✗	✗
Auderset (2025)	✗	✓	✗	✗	✗
Desired	✓	✓	✓	✓	✓

None of the former computational models of sound change model all aspects of local segmental change.

Sources: Greenhill et al., 2017; Bouchard-Côté et al., 2013; Hruschka et al., 2015; Moran et al., 2021; He and Zhao, 2024; Auderset, 2025; Cathcart and Wandl, 2020

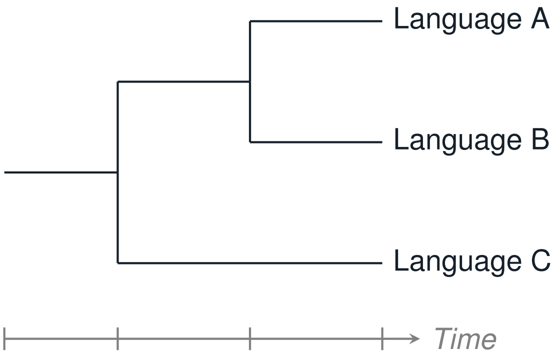
This talk

Goal: Build a probabilistic model that infers *rates of segmental change* while encoding *phonetic similarity*, *directional biases*, *local context*, and *regularity*.

Data

Southern California Annual Meeting on Phonology 2026

Input Data 1: Phylogeny

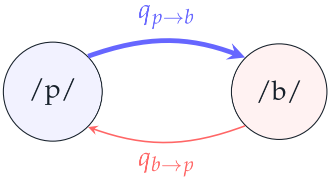


Input Data 2: Aligned Cognates

	Site 2	Site 4
L1	a	e
L2	a	e
L3	a	e

Homologous Column

Input Data 3: Segment Features



[+cons]	[+cons]
[−son]	[−son]
[−cont]	[−cont]
[−voice]	**[+voice]**
[−nas]	[−nas]

State space and edit types

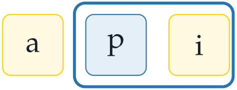
Illustrative inventory

$\{p, t, k, b, d, g, f, s, h, m, n, l, r, a, i, u, \varnothing\}$
16 segments
1 gap state __
1 boundary marker #

Three transition types

$x > y$	substitution
$\varnothing > y$	insertion
$x > \varnothing$	deletion

Local context as part of the rate structure



target + neighbors = context

- edge vs. medial position
- neighboring features from parent states

Feature engineering

PanPhon-style feature vectors as covariates

Representation

Segment $x \rightarrow$ feature vector $\mathbf{f}(x)$

- **Consonantal:** place, manner, voicing
- **Vocalic:** height, backness, rounding

Feature engineering

PanPhon-style feature vectors as covariates

Representation

Segment $x \rightarrow$ feature vector $\mathbf{f}(x)$

- **Consonantal:** place, manner, voicing
- **Vocalic:** height, backness, rounding

Features as Predictors

Features act as covariates to estimate weight vector $\mathbf{w}$:

- **Symmetric:** weights on shared/differing features
- **Directional:** weights on specific transitions (e.g., $+ \rightarrow -$)

Sources: Mortensen et al., 2016; Ohala, 1993; Blevins, 2004; Garrett and Johnson, 2013

Feature Vectors in Practice

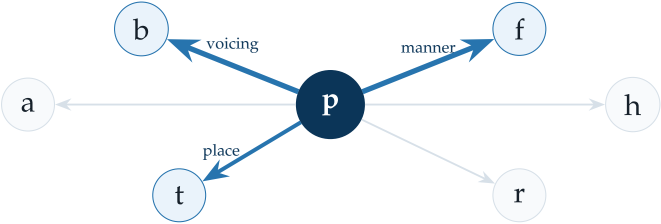
/m/

$$\mathbf{f}(/\mathrm{m}/) \in \{1, 0, -1\}^n$$

Covariates and Weights

Feature	Value	Weight
Syllabic (SYL)	-1	w_{syl}
Consonantal (CONS)	1	w_{cons}
Sonorant (SON)	1	w_{son}
Coronal (COR)	-1	w_{cor}
Labial (LAB)	1	w_{lab}
Nasal (NAS)	1	w_{nas}
Voiced (VOI)	1	w_{voi}

Feature distance shapes transition probability



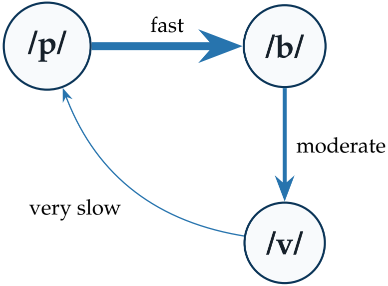
thick = small distance → higher rate

thin = large distance → lower rate

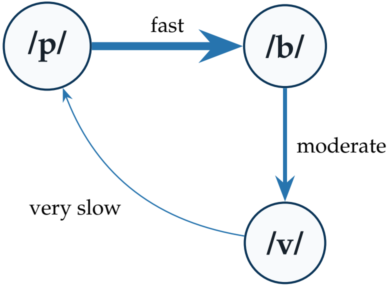
Methods

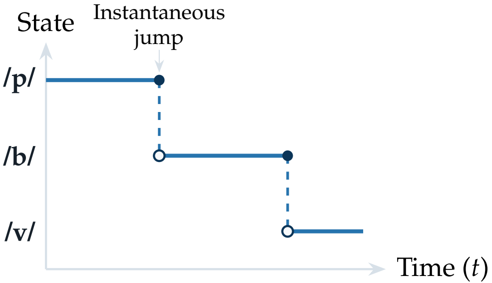
Southern California Annual Meeting on Phonology 2026

Continuous-Time Markov Chains

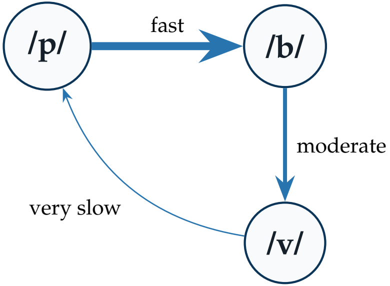


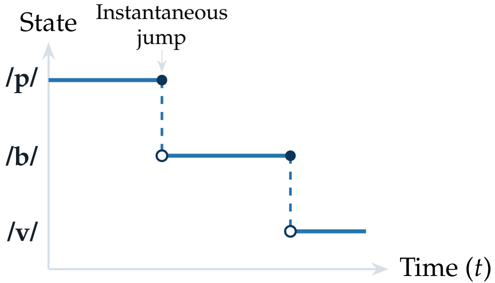
Continuous-Time Markov Chains





Continuous-Time Markov Chains





Instantaneous Propensity

- **Rates (Q):** Innate phonetic biases.
- **Evolutionary Process:** Continuous time, discrete categories.
- **Segmental Change:** Instantaneous jumps (e.g., /p/ > /b/).

Regularity encoded in contextual modeling

Same context ⇒ same generator

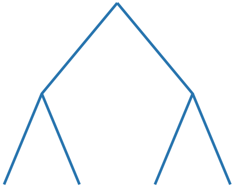
a	p	i	same context
a	p	i	same context
a	p	i	same context

All sites with /p/ between /a/ and /i/ share $Q(a\_\_i)$.

This models Neo-grammarian regularity in probabilistic form.

Strict vs. relaxed clock

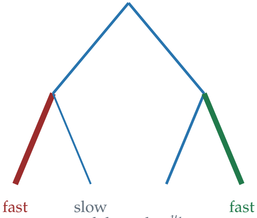
Strict clock



all branches same rate

vs

Relaxed clock

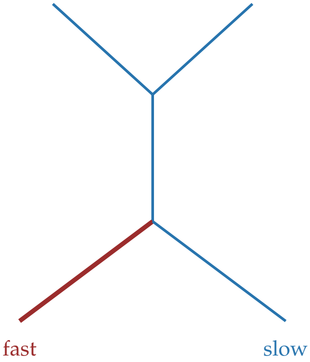


each branch: e^{u_b}

Language families routinely violate the strict clock.

Sources: Bromham et al., 2015; Greenhill et al., 2017; Bowern and Atkinson, 2012

Branch-specific rate heterogeneity



- multiplicative rate effect e^{u_b} per branch
- hierarchical shrinkage priors over branch-rate heterogeneity

Sources: cf. Bromham et al., 2015; Greenhill et al., 2017; Chang et al., 2015; Yang, 2000

What determines the rate of change ($x > y$)?

$$Q_{xy}(\mathbf{c}) \propto \pi_y \exp\{\eta_{xy} + \kappa(\mathbf{c}) + u_b\}$$

Phonetic & Structural Drivers

π_y **Stationary frequency**

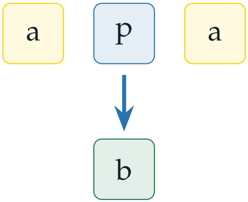
η_{xy} **Similarity & Directionality**
Phonetic similarity and directional biases (e.g., lenition) encoded via covariates.

Context & Phylogeny

$\kappa(\mathbf{c})$ **Local context**
Effect of surrounding environment.

u_b **Branch multiplier**
Lineage-specific rate variation.

Example: intervocalic lenition



$\eta_{p,b}$: small feature distance (voicing)

β_{dir}: lenition bias

$\kappa(V\_V)$: intervocalic boost

All add up $\Rightarrow$ high $Q_{p,b}(V\_V)$

$Q_{p,r}(V\_V)$ stays low: no similarity or bias support

Continuous-Time Evolution: Probabilities

The Outcome Over Time ($P(t)$)

Ancestor

/p/

Time duration (t)

Descendant

/p/ : 75%

/b/ : 20%

/v/ : 5%

Accumulation of Change

- **Process:** Integrate rates (Q) over time (t).
- **Result:** A probability distribution over jumps in discrete states.
- **Impact:** Longer branches = more accumulated change.

Simulation analysis

1. simulate tree and branch lengths

Simulation analysis

1. simulate tree and branch lengths
2. set ground-truth parameters

Simulation analysis

1. simulate tree and branch lengths
2. set ground-truth parameters
3. generate cognate-site data forward

Simulation analysis

1. simulate tree and branch lengths
2. set ground-truth parameters
3. generate cognate-site data forward
4. hide internal states; keep tips only

Simulation analysis

1. simulate tree and branch lengths
2. set ground-truth parameters
3. generate cognate-site data forward
4. hide internal states; keep tips only
5. run full inference pipeline

Simulation analysis

1. simulate tree and branch lengths
2. set ground-truth parameters
3. generate cognate-site data forward
4. hide internal states; keep tips only
5. run full inference pipeline
6. compare recovered vs. true values

Simulation analysis

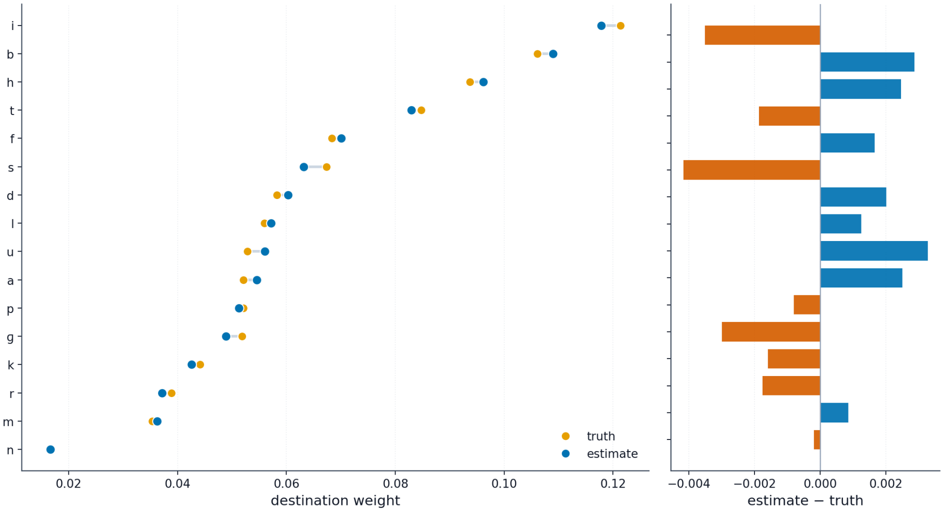
1. simulate tree and branch lengths
2. set ground-truth parameters
3. generate cognate-site data forward
4. hide internal states; keep tips only
5. run full inference pipeline
6. compare recovered vs. true values

Inference goal: show that the model developed here can recover parameters defining the probabalistic sound change process *in principle*.

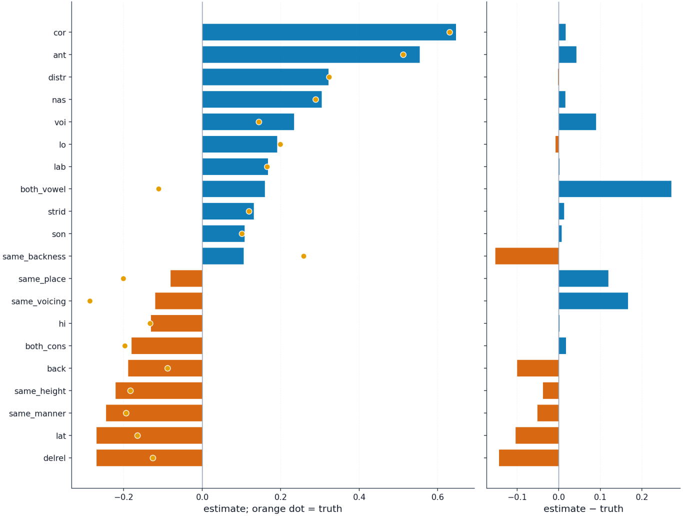
Results

Southern California Annual Meeting on Phonology 2026

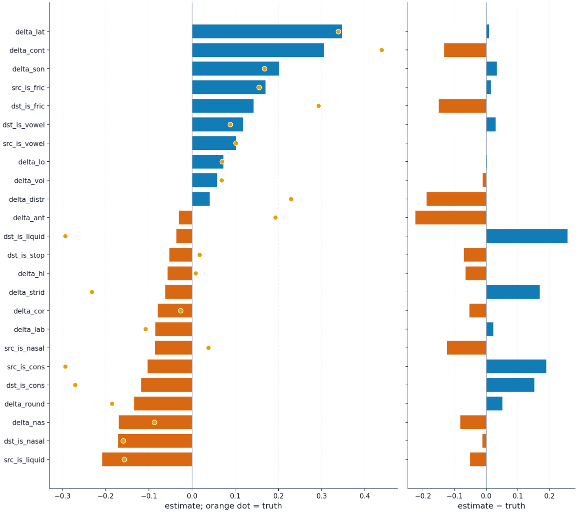
Stationarity probabilities



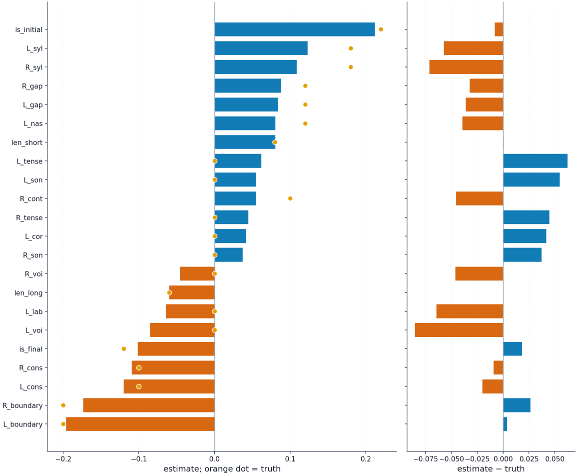
Symmetric coefficients



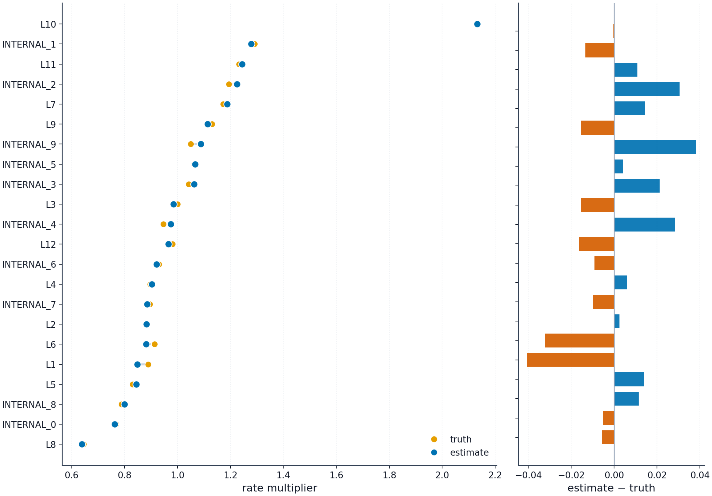
Directional coefficients



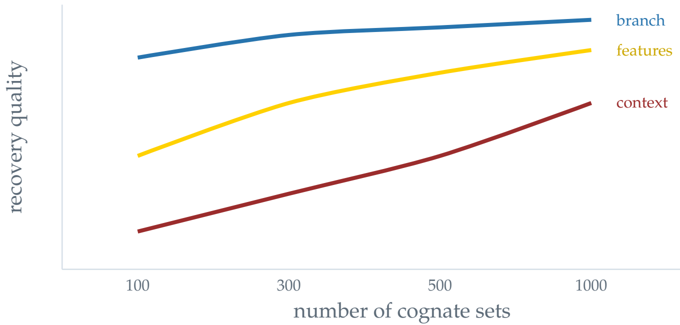
Context coefficients



Branch rate heterogeneity



How identifiability scales with data



Context coefficients benefit most from larger datasets.

Discussion

Southern California Annual Meeting on Phonology 2026

Main contributions

1. **Context-sensitive CTMC.** Generator depends on local environment.
2. **Feature-structured rates.** Similarity and directionality taken into account.
3. **Regularity by design.** Shared contexts share transition matrices
4. **Unified indels.** Insertions and deletions via ghost sites.
5. **Branch heterogeneity.** Separate local conditioning from tempo shifts.

Applications

Extending the probabilistic framework

What affects the rate of sound change?

- **Regularity vs. Diffusion**
  Test the validity of the lexical diffusion hypothesis.
- **Phonetic Grounding**
  Check the adequacy of model with phonetic grounding in generating the attested data.
- **Functional Load**
  Regress estimated rates to test for communicative efficiency.
- **Typological Validation**
  Compare inferred rates against databases (PHOIBLE/BDPROTO).

Conclusion

- Rich history of qualitative accounts of segmental stability and change

Conclusion

- Rich history of qualitative accounts of segmental stability and change
- Unified quantitative framework for estimating rates of change while representing sound structure and conditioning environments missing

Conclusion

- Rich history of qualitative accounts of segmental stability and change
- Unified quantitative framework for estimating rates of change while representing sound structure and conditioning environments missing
- A context-sensitive CTMC with feature-structured rates one plausible step in that direction.

Conclusion

- Rich history of qualitative accounts of segmental stability and change
- Unified quantitative framework for estimating rates of change while representing sound structure and conditioning environments missing
- A context-sensitive CTMC with feature-structured rates one plausible step in that direction.
- Parameter recovery in simulation studies provide a solid footing for future inference

Future steps

- The model now needs to be fit to real historical data.

Future steps

- The model now needs to be fit to real historical data.
- Once we have inferred rates from attested diachronic data, we can use them to ask which factors systematically condition phonological change.

Future steps

- The model now needs to be fit to real historical data.
- Once we have inferred rates from attested diachronic data, we can use them to ask which factors systematically condition phonological change.
- In particular, the estimated rates would let us test the relative contribution of functional load, phonetic similarity, lexical frequency, structural markedness, and contact-related pressures on segmental evolution.

Future steps

- The model now needs to be fit to real historical data.
- Once we have inferred rates from attested diachronic data, we can use them to ask which factors systematically condition phonological change.
- In particular, the estimated rates would let us test the relative contribution of functional load, phonetic similarity, lexical frequency, structural markedness, and contact-related pressures on segmental evolution.

Thank you

Questions?

References I

S. Auderset. Mixtec sound change database 2.0: Integrating tone change. *Journal of Open Humanities Data*, 11:37, 2025. doi: 10.5334/johd.322.

E. Baković. Local assimilation and constraint interaction. In P. de Lacy, editor, *The Cambridge Handbook of Phonology*, pages 335–352. Cambridge University Press, Cambridge, 2007.

D. E. Blasi, S. Moran, S. R. Moisik, P. Widmer, D. Dediu, and B. Bickel. Human sound systems are shaped by post-Neolithic changes in bite configuration. *Science*, 363(6432):eaav3218, 2019. doi: 10.1126/science.aav3218.

J. Blevins. *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge University Press, Cambridge, 2004.

A. Bouchard-Côté, D. Hall, T. L. Griffiths, and D. Klein. Automated reconstruction of ancient languages using probabilistic models of sound change. *Proceedings of the National Academy of Sciences*, 110(11):4224–4229, 2013. doi: 10.1073/pnas.1204678110.

C. Bowern and Q. Atkinson. Computational phylogenetics and the internal structure of Pama-Nyungan. *Language*, 88(4):817–845, 2012.

L. Bromham, X. Hua, T. G. Fitzpatrick, and S. J. Greenhill. Rate of language evolution is affected by population size. *Proceedings of the National Academy of Sciences*, 112(7):2097–2102, 2015. doi: 10.1073/pnas.1419704112.

J. Bybee. *Phonology and Language Use*. Cambridge University Press, Cambridge, 2001.

J. Bybee. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, 14(3): 261–290, 2002. doi: 10.1017/S0954394502143018.

J. Bybee. Grammatical and lexical factors in sound change: A usage-based approach. *Language Variation and Change*, 29(3):273–300, 2017. doi: 10.1017/S0954394517000199.

C. Cathcart and F. Wandl. In search of isoglosses: continuous and discrete language embeddings in Slavic historical phonology. In G. Nicolai, K. Gorman, and R. Cotterell, editors, *Proceedings of the 17th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 233–244, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.sigmorphon-1.28. URL `https://aclanthology.org/2020.sigmorphon-1.28/`.

W. Chang, C. Cathcart, D. Hall, and A. Garrett. Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language*, 91(1):194–244, 2015. doi: 10.1353/lan.2015.0005.

M. Y. Chen and W. S.-Y. Wang. Sound change: Actuation and implementation. *Language*, 51(2):255–281, 1975. doi: 10.2307/412854.

References II

D. Dediu and S. R. Moisik. Pushes and pulls from below: Anatomical variation, articulation and sound change. *Glossa*, 4(1):7, 2019.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, 39(1):1–38, 1977.

J. Felsenstein. Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution*, 17(6):368–376, 1981. doi: 10.1007/BF01734359.

A. Garrett. Sound change. In *The Routledge Handbook of Historical Linguistics*, pages 227–248. Routledge, 2015.

A. Garrett and K. Johnson. Phonetic bias in sound change. In A. C. L. Yu, editor, *Origins of Sound Change: Approaches to Phonologization*, pages 51–97. Oxford University Press, Oxford, 2013.

C. J. Geyer. Markov chain Monte Carlo maximum likelihood. *Computing Science and Statistics*, pages 156–163, 1991.

S. J. Greenhill, C.-H. Wu, X. Hua, M. Dunn, S. C. Levinson, and R. D. Gray. Evolutionary dynamics of language systems. *Proceedings of the National Academy of Sciences*, 114(42):E8822–E8829, 2017. doi: 10.1073/pnas.1700388114.

J. Hay and P. Foulkes. The evolution of medial /t/ over real and remembered time. *Language*, 92(2):298–330, 2016.

B. Hayes and D. Steriade. Introduction: The phonetic bases of phonological markedness. In B. Hayes, R. Kirchner, and D. Steriade, editors, *Phonetically Based Phonology*, pages 1–33. Cambridge University Press, Cambridge, 2004.

B. Hayes and J. White. Saltation and the p-map. *Phonology*, 32(2):267–302, 2015. doi: 10.1017/S0952675715000159.

B. Hayes and C. Wilson. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, 39(3):379–440, 2008. doi: 10.1162/ling.2008.39.3.379.

S. He and W. Zhao. Exploring sound change over time: A review of computational and human perception, 2024. URL `https://arxiv.org/abs/2407.05092`.

H. H. Hock. *Principles of Historical Linguistics*. Mouton de Gruyter, Berlin, 2 edition, 1991.

D. J. Hruschka, S. Branford, E. D. Smith, J. Wilkins, A. Meade, M. Pagel, and T. Bhattacharya. Detecting regular sound changes in linguistics as events of concerted evolution. *Current Biology*, 25(1):1–9, 2015. doi: 10.1016/j.cub.2014.10.064.

References III

J. P. Huelsenbeck, R. Nielsen, and J. P. Bollback. Stochastic mapping of morphological characters. *Systematic Biology*, 52(2):131–158, 2003. doi: 10.1080/10635150390192780.

P. Kiparsky. Labov, sound change, and phonological theory. *Journal of Sociolinguistics*, 20(4):464–488, 2016. doi: 10.1111/josl.12196.

J. Kirby. The role of probabilistic enhancement in phonologization. In *Origins of Sound Change: Approaches to Phonologization*. Oxford University Press, 2013. doi: 10.1093/acprof:oso/9780199573745.003.0011.

W. Labov. *Principles of Linguistic Change. Volume 1: Internal Factors*. Blackwell, Oxford, 1994.

W. Labov. The regularity of regular sound change. *Language*, 96(1):42–59, 2020. doi: 10.1353/lan.2020.0001.

I. Maddieson. *Patterns of Sounds*. Cambridge University Press, Cambridge, 1984.

A. Martinet. Function, structure, and sound change. *Word*, 8(1):1–32, 1952.

C. Mayer. Capturing gradience in long-distance phonology using probabilistic tier-based strictly local grammars. In *Proceedings of the Society for Computation in Linguistics 2021*, pages 39–50, Online, 2021. Association for Computational Linguistics.

L. Milroy. *Language and Social Networks*. Blackwell, Oxford, 2 edition, 1987.

C. Moore-Cantwell. *The Representation of Probabilistic Phonological Patterns: Neurological, Behavioral, and Computational Evidence from the English Stress System*. PhD thesis, University of Massachusetts Amherst, 2016.

S. Moran and D. McCloy. PHOIBLE 2.0, 2019.

S. Moran, N. A. Lester, and E. Grossman. Inferring recent evolutionary changes in speech sounds. *Philosophical Transactions of the Royal Society B*, 376 (1824):20200198, 2021. doi: 10.1098/rstb.2020.0198.

E. Moreton. Analytic bias and phonological typology. *Phonology*, 25(1):83–127, 2008. doi: 10.1017/S0952675708001413.

D. R. Mortensen, P. Littell, A. Bharadwaj, K. Goyal, C. Dyer, and L. Levin. Panphon: A resource for mapping IPA segments to articulatory feature vectors. In *Proceedings of COLING 2016*, pages 3475–3484, 2016.

R. Nielsen. Mapping mutations on phylogenies. *Systematic Biology*, 51(5):729–739, 2002.

References IV

J. J. Ohala. The listener as a source of sound change. In C. S. Masek, R. A. Hendrick, and M. F. Miller, editors, *Papers from the Parasession on Language and Behavior*, pages 178–203. Chicago Linguistic Society, Chicago, 1981.

J. J. Ohala. The phonetics of sound change. *Speech Communication*, 13(1–2):153–161, 1993.

M. E. Paster. *Phonological Conditions on Affixation*. PhD thesis, University of California, Berkeley, 2006.

J. Pater. Universal grammar with weighted constraints. *Harmonic Grammar and Harmonic Serialism*, pages 1–46, 2016.

B. S. Phillips. Word frequency and the actuation of sound change. *Language*, 60(2):320–342, 1984.

M. Ramsammy. The life cycle of phonological processes: Accounting for dialectal microtypologies. *Language and Linguistics Compass*, 9(1):33–54, 2015. doi: 10.1111/lnc3.12102.

S. Rose. Long-distance assimilation of consonants. In M. van Oostendorp, C. J. Ewen, E. Hume, and K. Rice, editors, *The Blackwell Companion to Phonology*, pages 1–27. Wiley-Blackwell, Malden, MA, 2011.

S. Rose and R. Walker. A typology of consonant agreement as correspondence. *Language*, 80(3):475–531, 2004.

E. R. Round. Evolution and trade-off dynamics of functional load. *Entropy*, 24(4):507, 2022. doi: 10.3390/e24040507.

S. S. Shih. Constraint conjunction in weighted probabilistic grammar. *Phonology*, 34(2):243–268, 2017.

M. Sóskuthy. Understanding change through stability: A computational study of sound change actuation. *Lingua*, 163:40–60, 2015.

S. Tavaré. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences*, 17:57–86, 1986.

S. J. Todd, J. B. Pierrehumbert, and J. B. Hay. Word frequency effects in sound change as a consequence of perceptual asymmetries: An exemplar-based model. *Cognition*, 185:1–20, 2019. doi: 10.1016/j.cognition.2019.01.004.

W. S.-Y. Wang. Competing changes as a cause of residue. *Language*, 45(1):9–25, 1969.

A. Wedel, A. Kaplan, and S. Jackson. High functional load inhibits phonological contrast loss: A corpus study. *Cognition*, 128(2):179–186, 2013. doi: 10.1016/j.cognition.2013.03.002.

References V

U. Weinreich, W. Labov, and M. I. Herzog. Empirical foundations for a theory of language change. In W. P. Lehmann and Y. Malkiel, editors, *Directions for Historical Linguistics*, pages 95–195. University of Texas Press, Austin, 1968.

Z. Yang. Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus A. *Journal of Molecular Evolution*, 51(5):423–432, 2000.

Z. Yang. *Computational Molecular Evolution*. Oxford University Press, Oxford, 2006.

K. Zuraw. A model of lexical variation and the grammar with application to tagalog nasal substitution. *Natural Language and Linguistic Theory*, 28(2): 417–472, 2010. doi: 10.1007/s11049-010-9095-z.

Lexical diffusion

Lexical diffusion

Some changes propagate unevenly through the lexicon; frequency and neighborhood density affect which words change first.

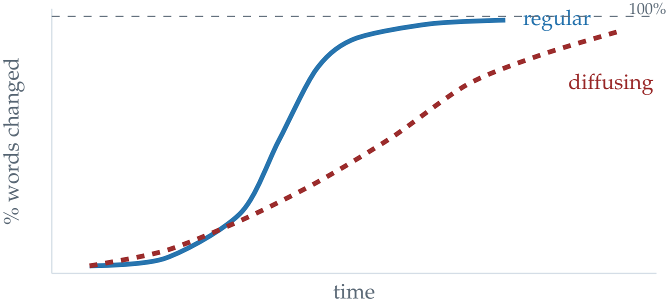
- Wang (1969): competing changes create residue
- Bybee (2002): sound change shows usage and frequency effects

A quantitative model should treat regularity as a *testable baseline*, but not enforce it.

Sources: Wang, 1969; Chen and Wang, 1975; Phillips, 1984; Bybee, 2001; Bybee, 2002; Todd et al., 2019

Diffusion dynamics: the S-curve

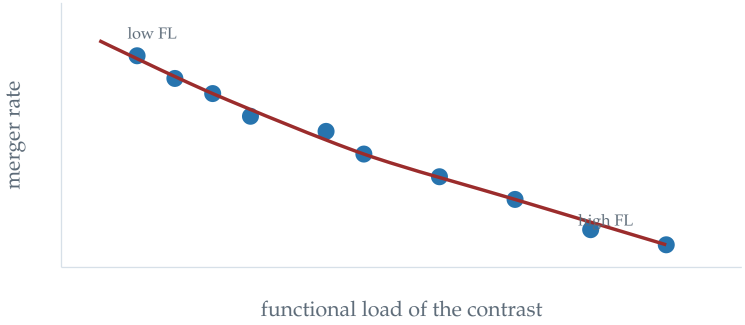
Both regularity and diffusion predict S-shaped adoption



The empirical question: do lexical items differ in their intrinsic rate of adopting a change?

Sources: Labov, 1994; Bybee, 2002; Sóskuthy, 2015; Hay and Foulkes, 2016

Functional load as a rate inhibitor



High-FL contrasts resist merger. This is absent from version 1 but is a natural extension.

Sources: Martinet, 1952; Wedel et al., 2013; Round, 2022

External and demographic pressures

- network density shapes diffusion
- contact and isolation influence the tempo of change
- anatomical and climatic factors may bias inventories

Actuation problem
Why this change, here, now? A rate model should represent *heterogeneity across branches*.

Sources: Weinreich et al., 1968; Milroy, 1987; Bromham et al., 2015; Greenhill et al., 2017; Dediu and Moisik, 2019

Defining the Context

Locality, similarity, and phonological domains

1. Strictly Local

Assimilation, lenition, and boundary effects driven by **adjacent** material.

2. Long-Distance

Harmony and correspondence sensitive to **non-local** similarity.

3. Domain-Sensitive

Rules behaving differently inside roots vs. across affixes (**morphological** domains).

Modeling Assumption

The local-context approach taken here is a principled **first approximation**, not a theoretical claim that all phonological structure is strictly local, or should be modeled as such.

Sources: Baković, 2007; Rose and Walker, 2004; Rose, 2011; Mayer, 2021; Paster, 2006

Continuous-Time Markov Chains

At a single site

A segment evolves as a continuous-time Markov chain, a probabilistic model for discrete jumps between states.

Sources: Felsenstein, 1981; Bouchard-Côté et al., 2013; Yang, 2006

Continuous-Time Markov Chains

The Transition Rate Matrix Q

$$Q = \begin{pmatrix} -\sum_{j\neq 1} q_{1j} & q_{12} & \cdots & q_{1n} \\ q_{21} & -\sum_{j\neq 2} q_{2j} & \cdots & q_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ q_{n1} & q_{n2} & \cdots & -\sum_{j\neq n} q_{nj} \end{pmatrix}$$

- **Off-diagonal entries ($q_{ij} \geq 0$):** Represent the instantaneous rate of transitioning from state i to state j.
- **Diagonal entries ($q_{ii} \leq 0$):** Defined such that each row sums to zero.
- **Transition Probabilities:** Calculated via the matrix exponential $P(t) = e^{Qt}$.

Sources: Felsenstein, 1981; Bouchard-Côté et al., 2013; Yang, 2006

Instantaneous rates of change > probabilities of change

Generator Q

instantaneous rates Q_{xy}: propensity to move $x > y$

$$\xrightarrow{e^{Qt}}$$

Transition $P(t)$

probability of state y after time t, given started in x

Longer branches $\Rightarrow$ more change

Sources: Felsenstein, 1981; Yang, 2006; Tavaré, 1986

APPENDIX

Mathematical details

A1. Site expansion and root priors

Lexical root prior

$P(X_r{=}x \mid \text{lex}) = (1{-}\varepsilon_{\text{lex}})\pi_x$ for $x \in \mathcal{S}_{\text{real}}$; $\quad \varepsilon_{\text{lex}}$ for $x{=}\varnothing$.

Ghost root prior

$P(X_r{=}x \mid \text{ghost}) = \varepsilon_{\text{ghost}}\pi_x$ for $x \in \mathcal{S}_{\text{real}}$; $\quad 1{-}\varepsilon_{\text{ghost}}$ for $x{=}\varnothing$.

A2. Generator matrix

$$Q_{xy} = \exp(\log \mu + \kappa + \eta_{xy})\, \pi_y, \qquad Q_{xx} = -\sum_{y \neq x} Q_{xy}$$

$$\eta_{xy} = \langle \mathbf{X}_{xy}^{\mathrm{sym}}, \boldsymbol{\beta}_{\mathrm{sym}} \rangle + \langle \mathbf{X}_{xy}^{\mathrm{dir}}, \boldsymbol{\beta}_{\mathrm{dir}} \rangle, \qquad \kappa = \langle \mathbf{c}, \boldsymbol{\beta}_{\mathrm{ctx}} \rangle$$

A3. Branch-rate heterogeneity

$$u_b^{\text{raw}} = \tau\lambda_b z_b, \;\; z_b \sim \mathcal{N}(0,1); \qquad \nu_b = u_b^{\text{raw}} - \overline{u}; \qquad \tilde{\ell}_b = \ell_b e^{\nu_b}$$

Centering makes branch effects relative.

A4. Likelihood and pruning

Pattern-compressed likelihood

$\log p(\mathbf{y} \mid \theta) = \sum_u n_u \log P(\mathbf{y}^{(u)} \mid \theta)$

Felsenstein recursion

$L_v(k) = \prod_{c \in \text{children}(v)} \sum_j P_{b_c}(k, j)\, L_c(j);\quad \text{root: } \sum_k \rho(k)\, L_{\text{root}}(k).$

A5. Parallel tempering

Tempered targets $p(\theta \mid \mathbf{y})^{1/T_m}$; swap acceptance:

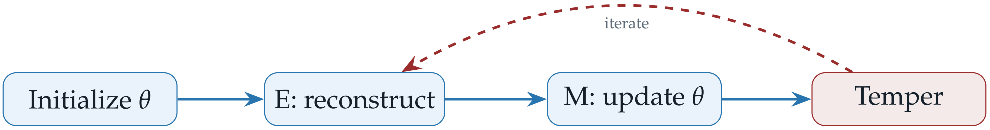
$$\log a_{\text{swap}} = \left(\frac{1}{T_m} - \frac{1}{T_{m+1}}\right)\left[\log p(\theta_{m+1}) - \log p(\theta_m)\right]$$

Sources: Geyer, 1991

Inference

Core difficulty

Rates depend on *local context*, but context is determined by *latent ancestral states*.



Sources: Felsenstein, 1981; Bouchard-Côté et al., 2013; Dempster et al., 1977; Geyer, 1991

A6. Stochastic mapping

Set $\omega \geq \max_k(-Q_{kk})$, define $B = I + Q/\omega$. Conditional uniformization samples Poisson event counts and skeleton paths to estimate posterior event histories.

Sources: Huelsenbeck et al., 2003; Nielsen, 2002

A7. Ancestral reconstruction

$$P(X_r{=}i \mid \mathbf{y}) \propto \rho(i)\, L_r(i); \qquad P(X_c{=}j \mid X_p{=}i, \mathbf{y}) \propto P_{ij}^{(g,b)}\, L_c(j)$$

Simulation design

Tree

20-tip balanced tree; log-normal branch lengths; one fast clade, one slow.

Data

500 cognate sets, 3–8 sites each; 17-state inventory; ground truth mimics known asymmetries.

A8. Regularization

- Gaussian priors on coefficients to cause shrinkage
- Half-normal priors on branch scales
- Structured generator prevents combinatorial explosion

A9. Relation to MaxEnt / Harmonic Grammar

Shared	Different
Weighted structures preserving gradience and restricting hypothesis space.	MaxEnt/HG: output probabilities. CTMC: transition intensities over time.

Sources: Hayes and Wilson, 2008; Moore-Cantwell, 2016; Shih, 2017; Pater, 2016

A10. Why rates matter for theory

Different theories make different predictions about *where* change should be fast or slow:

- **Phonetic-bias:** rates highest in articulatorily/perceptually pressured contexts
- **Analytic-bias:** rates relate to learnability, even with comparable phonetic grounding
- **Usage-based:** rates relate to frequency and predictability
- **Grammar-internal:** rates cluster by domain and life-cycle stage

Sources: Blevins, 2004; Moreton, 2008; Garrett and Johnson, 2013; Bybee, 2017; Kiparsky, 2016; Ramsammy, 2015