University of Zurich[UZH]

The evolution of similarity avoidance

A phylogenetic approach to phonotactic change

Chundra Cathcart

Department of Comparative Language Science
Center for the Interdisciplinary Study of Language Evolution
University of Zurich

Annual Meeting on Phonology, UCLA, 22 October 2022

Similarity avoidance

- Gradient avoidance of adjacent consonants agreeing in place of articulation is common cross-linguistically (Berkley 2000, Pozdniakov and Segerer 2007, Gallagher and Coon 2009, Graff and Jaeger 2009, Wilson and Obdeyn 2009, a.o.)
- Participants in lexical decision tasks are slower to accept words and faster to reject non-words containing identical consonants at any distance (van de Weijer 2005)
- Forms containing consonants at any distance are infrequent in corpora (van de Weijer 2003)

Origins of similarity avoidance

What drives the patterns we observe? No consensus on the issue.

- Sound change? Not clear that dissimilation in place of articulation is more frequent than assimilation in place of articulation (Kümmel 2007)
- Lexical usage? Forms with a sub-optimal pattern may be avoided, or phased out of use (Frisch et al. 2004, Martin 2007, Pozdniakov and Segerer 2007)

Goal of this talk: use phylogenetic modeling to investigate the diachronic pressures that underlie the synchronic distribution of forms with similar place avoidance.

Origins of similarity avoidance

What drives the patterns we observe? No consensus on the issue.

- Sound change? Not clear that dissimilation in place of articulation is more frequent than assimilation in place of articulation (Kümmel 2007)
- Lexical usage? Forms with a sub-optimal pattern may be avoided, or phased out of use (Frisch et al. 2004, Martin 2007, Pozdniakov and Segerer 2007)

Goal of this talk: use phylogenetic modeling to investigate the diachronic pressures that underlie the synchronic distribution of forms with similar place avoidance.

Origins of similarity avoidance

What drives the patterns we observe? No consensus on the issue.

- Sound change? Not clear that dissimilation in place of articulation is more frequent than assimilation in place of articulation (Kümmel 2007)
- Lexical usage? Forms with a sub-optimal pattern may be avoided, or phased out of use (Frisch et al. 2004, Martin 2007, Pozdniakov and Segerer 2007)

Goal of this talk: use phylogenetic modeling to investigate the diachronic pressures that underlie the synchronic distribution of forms with similar place avoidance.

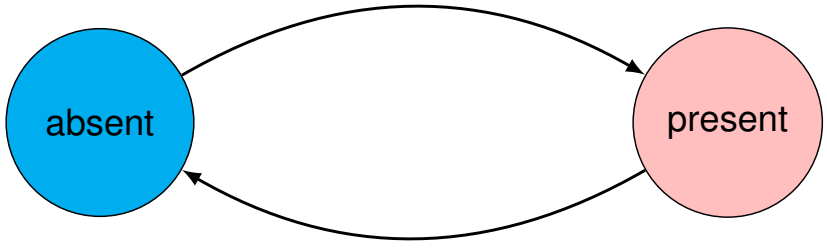
Phylogenetic modeling: preliminaries

Popular model: discrete features evolve over a phylogeny according to a continuous-time Markov process (stochastic process parameterized by **transition rates**)

Ingredients: linguistic data, phylogeny $\rightarrow$ (posterior) rate estimates

Phylogenetic modeling: preliminaries

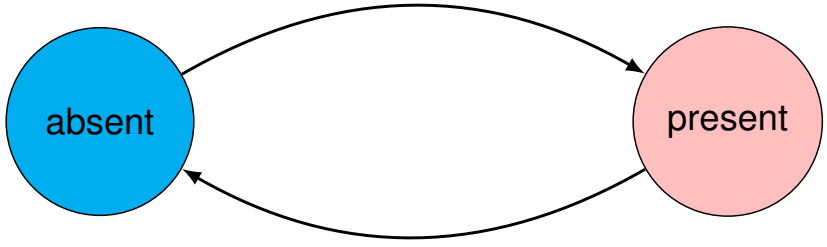
Popular model: discrete features evolve over a phylogeny according to a continuous-time Markov process (stochastic process parameterized by **transition rates**)

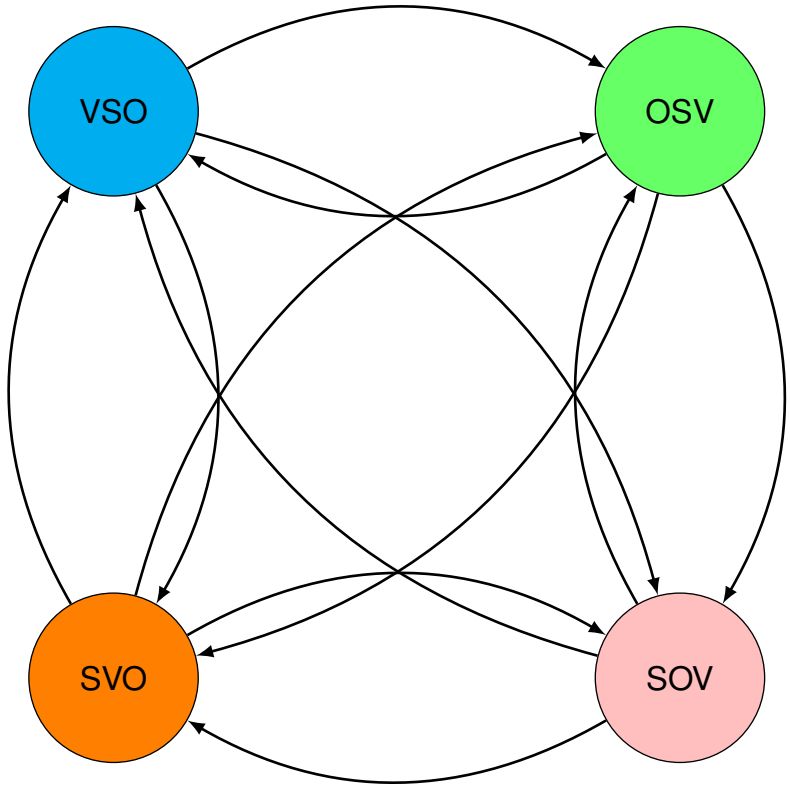


Ingredients: linguistic data, phylogeny $\rightarrow$ (posterior) rate estimates

Phylogenetic modeling: preliminaries

Popular model: discrete features evolve over a phylogeny according to a continuous-time Markov process (stochastic process parameterized by **transition rates**)

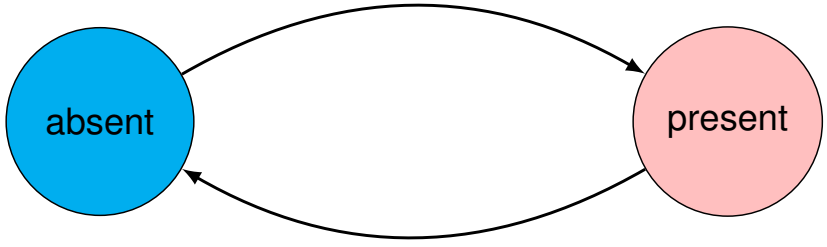


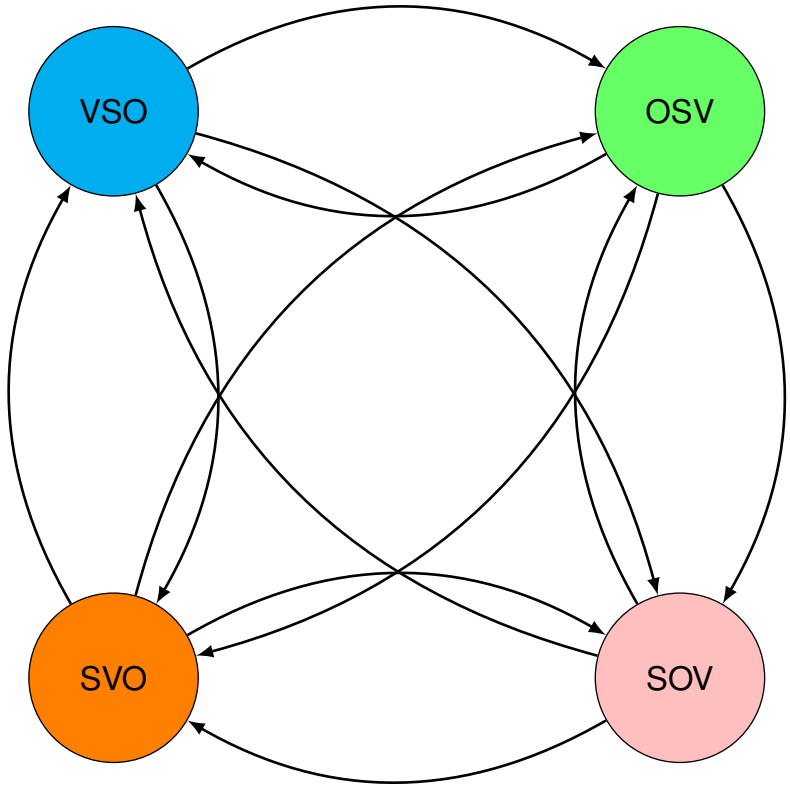


Ingredients: linguistic data, phylogeny $\rightarrow$ (posterior) rate estimates

Phylogenetic modeling: preliminaries

Popular model: discrete features evolve over a phylogeny according to a continuous-time Markov process (stochastic process parameterized by **transition rates**)





Ingredients: linguistic data, phylogeny $\rightarrow$ (posterior) rate estimates

Phylogenetic modeling: applications

- Estimating preferences for individual features
- Assessing interdependent evolution between features (Dunn et al. 2011, Jäger and Wahle 2021), constrained pathways of change (Shirtz et al. 2021)
- Ancestral state reconstruction

Limitations:

- Change only involves attested states (ascertainment bias)
- Use restricted to large, well-studied phylogenies
- Difficult to explicitly model areal dynamics
- Simplification of feature representations required for tractability

Not widely applied to questions regarding sound patterns (with exceptions, e.g., Macklin-Cordes and Round 2020, Urban and Moran 2021, Babinski 2022)

Phylogenetic modeling: applications

- Estimating preferences for individual features
- Assessing interdependent evolution between features (Dunn et al. 2011, Jäger and Wahle 2021), constrained pathways of change (Shirtz et al. 2021)
- Ancestral state reconstruction

Limitations:

- Change only involves attested states (ascertainment bias)
- Use restricted to large, well-studied phylogenies
- Difficult to explicitly model areal dynamics
- Simplification of feature representations required for tractability

Not widely applied to questions regarding sound patterns (with exceptions, e.g., Macklin-Cordes and Round 2020, Urban and Moran 2021, Babinski 2022)

Phylogenetic modeling: applications

- Estimating preferences for individual features
- Assessing interdependent evolution between features (Dunn et al. 2011, Jäger and Wahle 2021), constrained pathways of change (Shirtz et al. 2021)
- Ancestral state reconstruction

Limitations:

- Change only involves attested states (ascertainment bias)
- Use restricted to large, well-studied phylogenies
- Difficult to explicitly model areal dynamics
- Simplification of feature representations required for tractability

Not widely applied to questions regarding sound patterns (with exceptions, e.g., Macklin-Cordes and Round 2020, Urban and Moran 2021, Babinski 2022)

Rationale

- Here, I use phylogenetic methods to assess support for the LEXICAL VARIATION hypothesis
- If a phonotactically dispreferred pattern arises in a lexical item (e.g., through regular sound change), will the form become less frequent?
  - Due to logistical reasons, BASIC/NON-BASIC vocabulary distinction used: e.g., Latin *pellis* ‘pelt, hide’ (non-basic) $>$ French *peau* ‘skin’ (basic)

Other possible scenarios

- “Sporadic” sound change (or analogical change) serves to remove dispreferred patterns in basic vocabulary (PHONOLOGICAL VARIATION hypothesis; cf. Blevins and Wedel 2009)
- Both mechanisms are active (LEXICAL+PHONOLOGICAL VARIATION hypothesis)

Rationale

- Here, I use phylogenetic methods to assess support for the LEXICAL VARIATION hypothesis
- If a phonotactically dispreferred pattern arises in a lexical item (e.g., through regular sound change), will the form become less frequent?
  - Due to logistical reasons, BASIC/NON-BASIC vocabulary distinction used: e.g., Latin *pellis* 'pelt, hide' (non-basic) $>$ French *peau* 'skin' (basic)

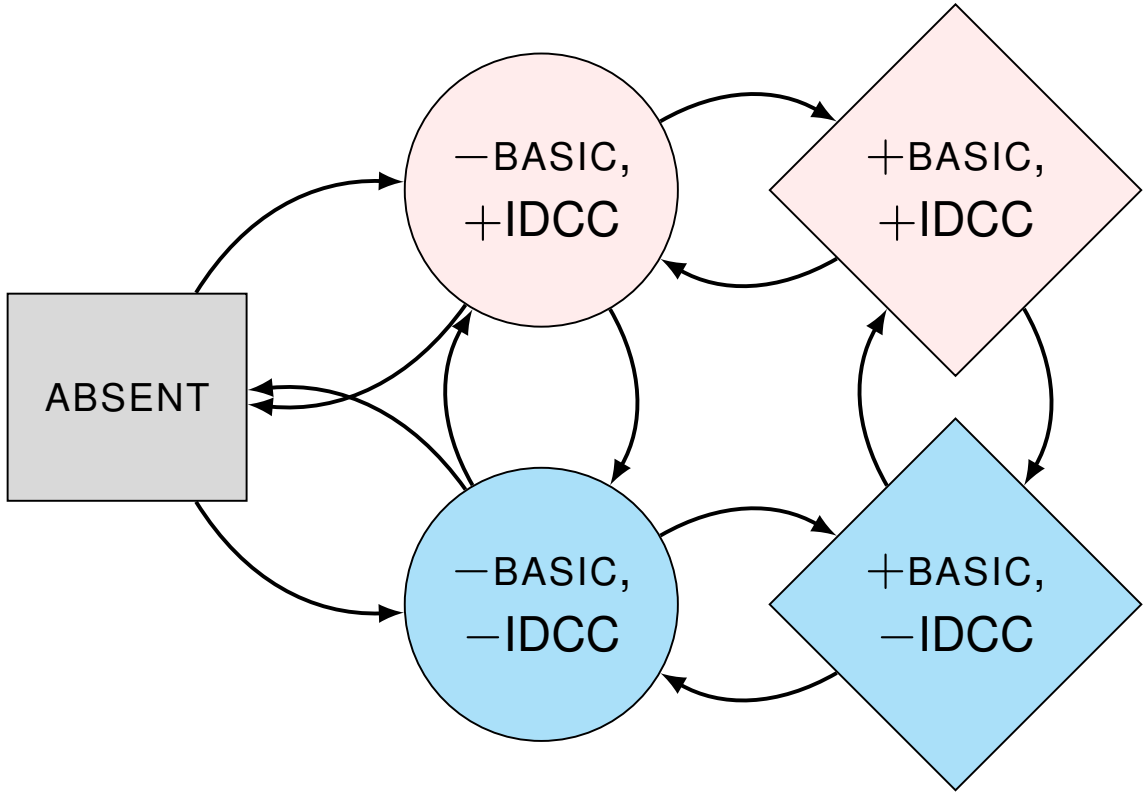
Other possible scenarios

- "Sporadic" sound change (or analogical change) serves to remove dispreferred patterns in basic vocabulary (PHONOLOGICAL VARIATION hypothesis; cf. Blevins and Wedel 2009)
- Both mechanisms are active (LEXICAL+PHONOLOGICAL VARIATION hypothesis)

Data

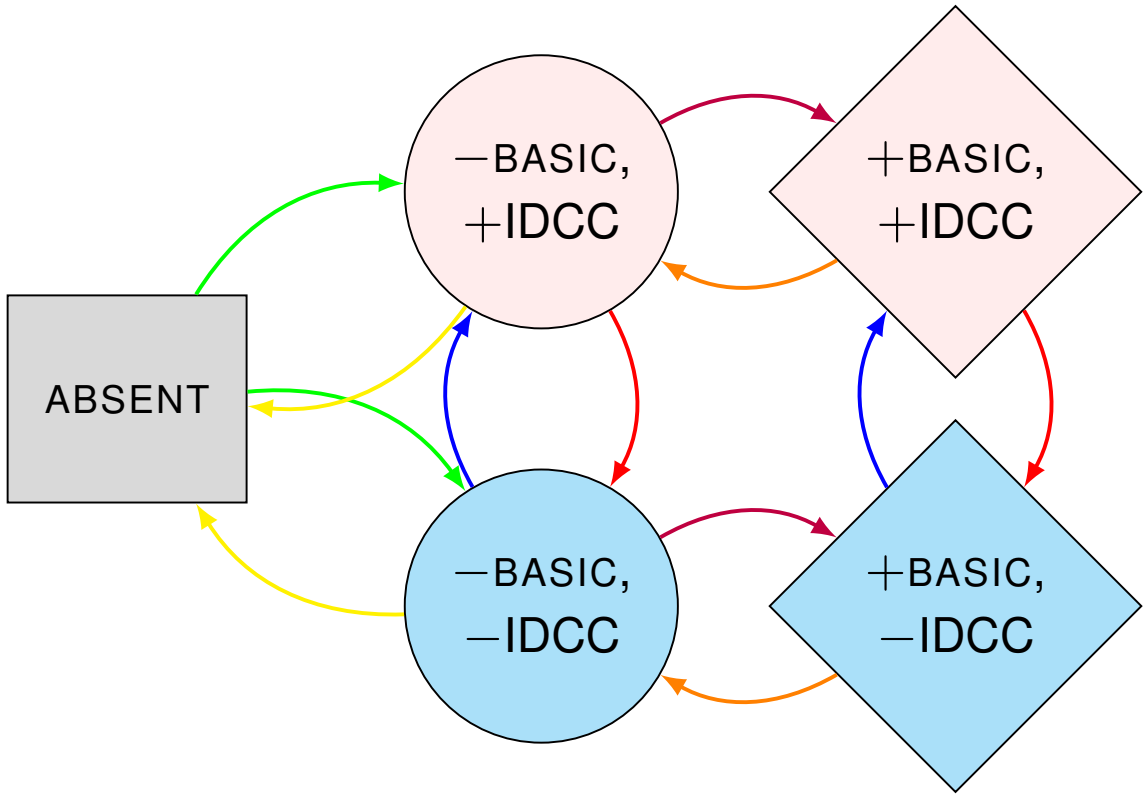
- Data taken from the Austronesian Comparative Dictionary (Blust and Trussel 2013), which provides forms and meanings for cognate lexical items
- Austronesian Basic Vocabulary Database (Greenhill et al. 2008) consulted in order to determine which lexical items in the ACD are basic vocabulary items
- Forms labeled according to presence/absence of tautomorphemic identical consonants separated by a vowel — reliant on morphemic analyses found in ACD
  - E.g., Aklanon *ba-bayi* 'woman' (original *babayi* in Zorc 1969) based on pAN **ba-b<in>ahi*
- 47 cognate sets attest all possible combinations of ±BASIC, ±IDCC in 306 languages found in reference phylogeny (Gray et al. 2009)

Basic model



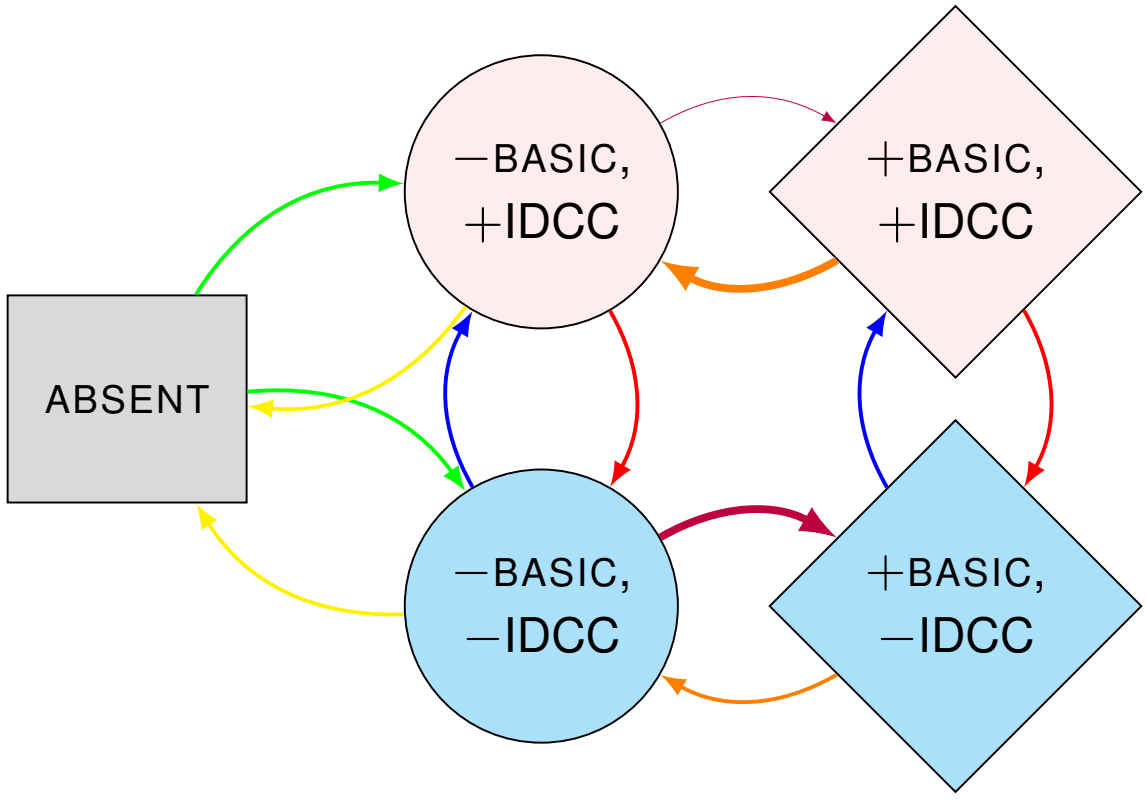
- An etymon can visit several states during its evolution

Uniform model



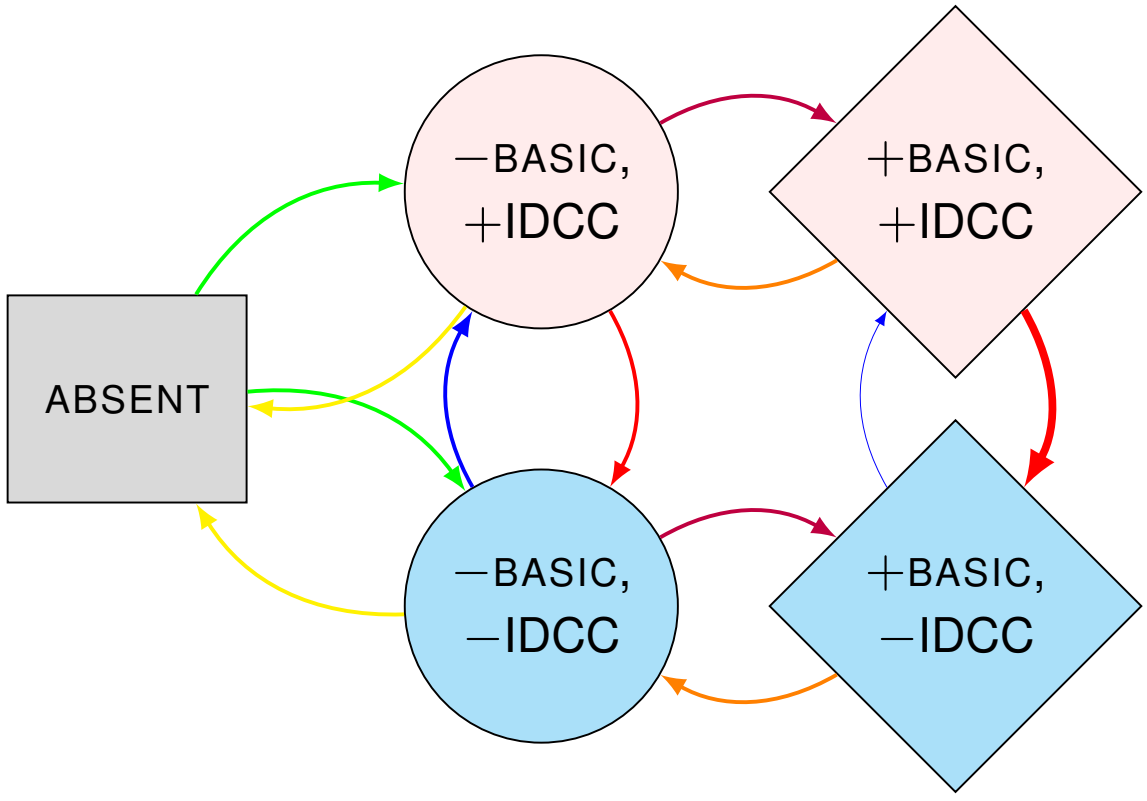
- Identical consonants are gained/lost with equal frequency in basic and non-basic vocabulary items
- Items with identical consonants are as likely to become basic/non-basic as items lacking identical consonants

Lexical variation model



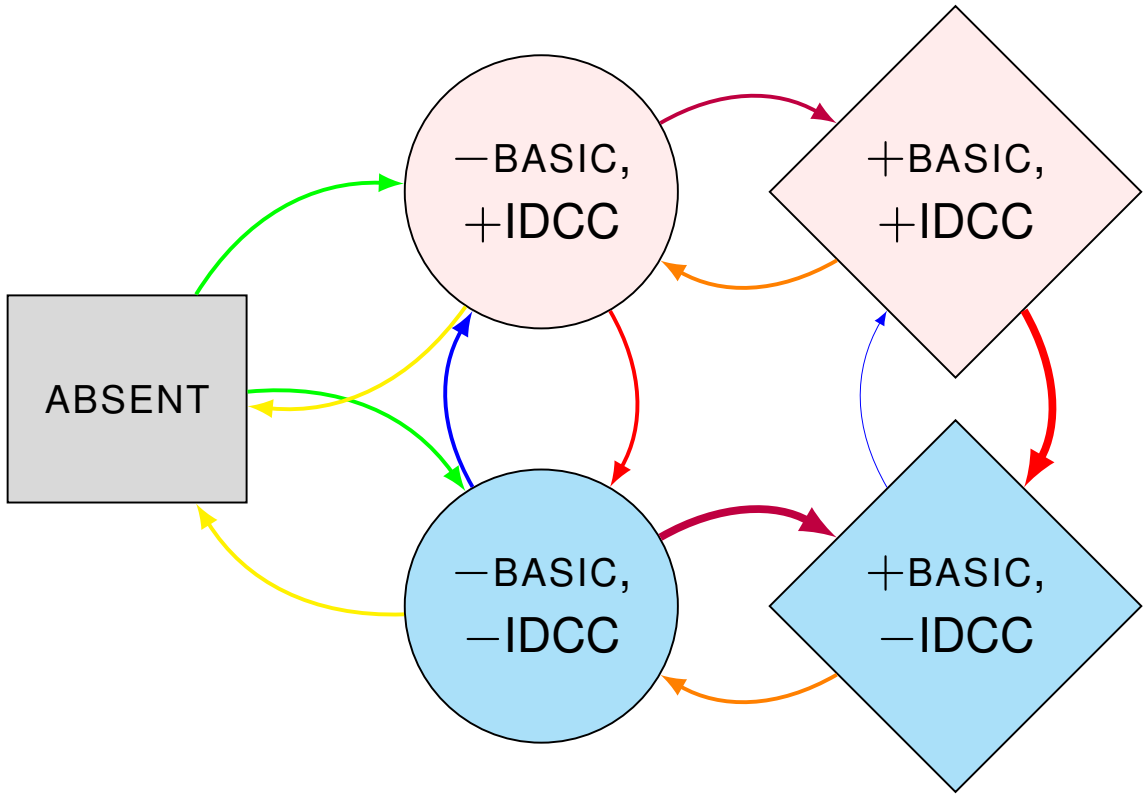
- Identical consonants are gained/lost with equal frequency in basic and non-basic vocabulary items
- Items with identical consonants are less likely to become basic than items lacking identical consonants

Phonological variation model



- Identical consonants are less likely to arise in basic than non-basic vocabulary items
- Items with identical consonants are as likely to become basic as items lacking them

Lexical+phonological variation model



- Identical consonants arise at different rates in basic and non-basic vocabulary items
- Change between basic and non-basic occurs at varying rates across the two phonological conditions

Results

- All models fitted using RStan (Carpenter et al. 2017)
- Gamma$(1, 1)$ priors placed over transition rates
- Model comparison carried out via model stacking (Yao et al. 2017)

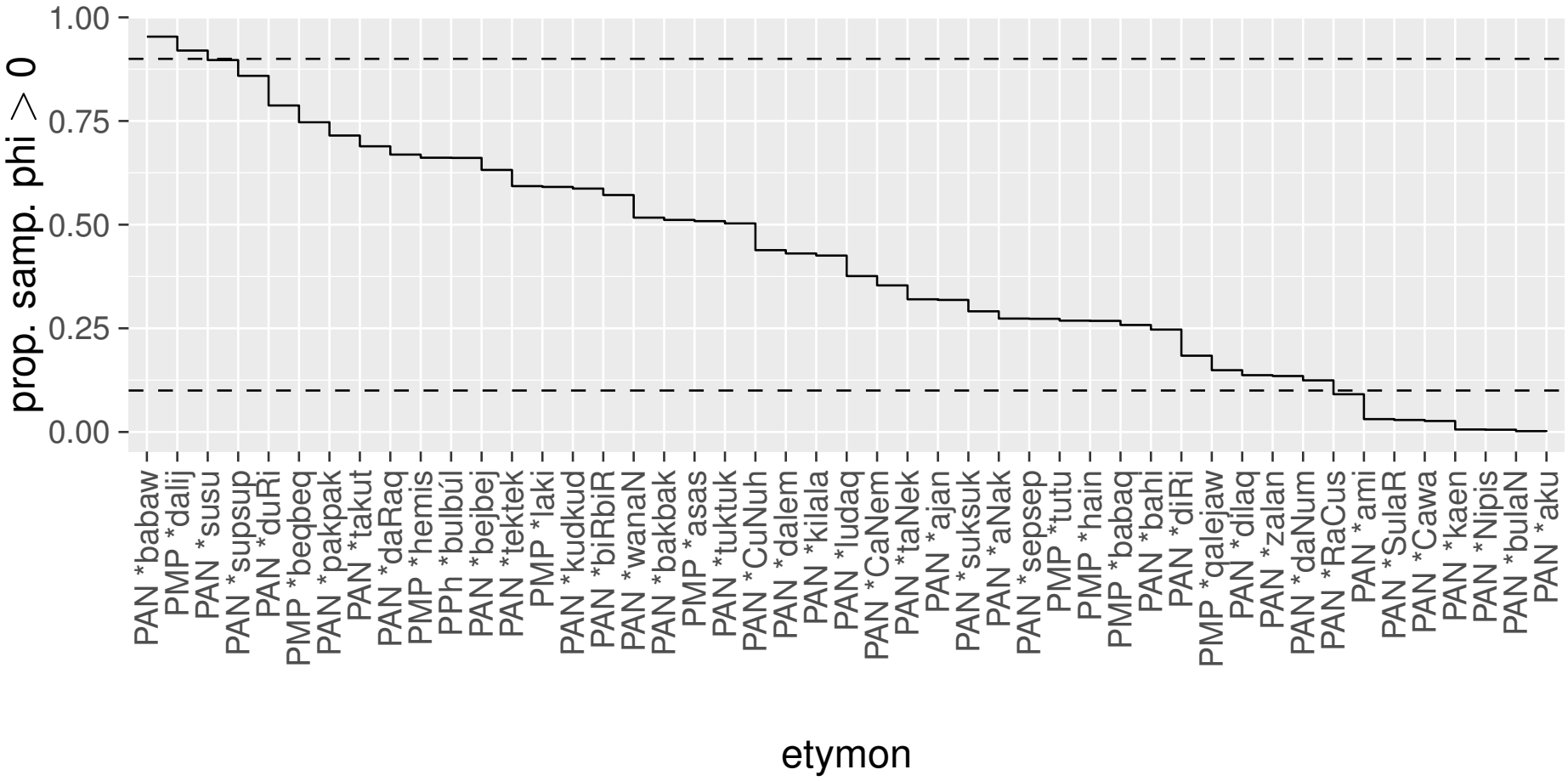
Model	Stacking weight
Uniform	0.435
Lexical variation	0.470
Phonological variation	0.095
Lexical $+$ phonological variation	0.000

The lexical variation model shows best fit, but not decisively so. Possible reasons:

- Reliance on morphological segmentation in ACD undercounts tautomorphemic identical consonants
- Different pressures at work in different lexical items

Inspecting posterior parameters

Following Jäger and Wahle (2021), ϕ coefficient calculated for posterior stationary probabilities for each etymon; positive values indicate a long-term association between basicness and identical consonants



Next steps

Found at least some support for the idea that lexical change serves to phase words with dispreferred phonotactic patterns out of usage, but results mixed

- More phylogenies? Requires databases organized around cognate classes, not concepts
- Finer-grained models capable of incorporating information regarding different phonological features
- Reversible-Jump MCMC (Pagel and Meade 2008)
- Move beyond basic/non-basic vocabulary distinction (Dellert and Buch 2018)

Regardless of degree of success of current study, phylogenetic methods a promising avenue for investigating the diachrony of static sound patterns in lexical items

Thank you!

- Questions/comments appreciated (chundra.cathcart@uzh.ch)
- Slides/code available at https://github.com/chundrac/amp-2022-conf

Bibliography I

Babinski, S. (2022). *Archival Phonetics and Prosodic Typology in Sixteen Australian Languages*. Ph. D. thesis, Yale University.

Berkley, D. M. (2000). *Gradient obligatory contour principle effects*. Ph. D. thesis, Northwestern University.

Blevins, J. and A. Wedel (2009). Inhibited sound change: An evolutionary approach to lexical competition. *Diachronica 26*(2), 143–183.

Blust, R. and S. Trussel (2013). The Austronesian comparative dictionary: a work in progress. *Oceanic Linguistics 52*(2), 493–523.

Carpenter, B., A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell (2017). Stan: A probabilistic programming language. *Journal of statistical software 76(1)*(1), 1–32.

Dellert, J. and A. Buch (2018). A new approach to concept basicness and stability as a window to the robustness of concept list rankings. *Language Dynamics and Change 8*(2), 157–181.

Bibliography II

Dunn, M., S. J. Greenhill, S. C. Levinson, and R. D. Gray (2011). Evolved structure of language shows lineage-specific trends in word-order universals. *Nature 473*(7345), 79–82.

Frisch, S. A., J. B. Pierrehumbert, and M. B. Broe (2004). Similarity avoidance and the OCP. *Natural Language & Linguistic Theory 22*(1), 179–228.

Gallagher, G. and J. Coon (2009). Cooccurrence restrictions, similarity, and correspondence in chol (mayan). In *Proceedings of NELS 38*. GLSA Publications.

Graff, P. and T. Jaeger (2009). Locality and feature specificity in OCP effects: Evidence from Aymara, Dutch, and Javanese. In *Proceedings from the annual meeting of the Chicago linguistic society*, Volume 45, pp. 127–141. Chicago Linguistic Society.

Gray, R. D., A. J. Drummond, and S. J. Greenhill (2009). Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science 323*(5913), 479–483.

Bibliography III

Greenhill, S. J., R. Blust, and R. D. Gray (2008). The Austronesian basic vocabulary database: from bioinformatics to lexomics. *Evolutionary Bioinformatics 4*, EBO–S893.

Jäger, G. and J. Wahle (2021). Phylogenetic typology. *Frontiers in Psychology 12*.

Kümmel, M. (2007). *Konsonantenwandel*. Wiesbaden: Dr. Ludwig Reichert Verlag.

Macklin-Cordes, J. L. and E. R. Round (2020). Re-evaluating phoneme frequencies. *Frontiers in psychology*, 3181.

Martin, A. T. (2007). *The evolving lexicon*. Ph. D. thesis, University of California, Los Angeles.

Pagel, M. and A. Meade (2008). Modelling heterotachy in phylogenetic inference by reversible-jump Markov chain Monte Carlo. *Phil. Trans. R. Soc. B 363*, 3955–3964.

Bibliography IV

Pozdniakov, K. and G. Segerer (2007). Similar place avoidance: A statistical universal.

Shirtz, S., L. Talamo, and A. Verkerk (2021). The evolutionary dynamics of negative existentials in indo-european. *Front. Commun. 6: 661862. doi: 10.3389/fcomm*.

Urban, M. and S. Moran (2021). Altitude and the distributional typology of language structure: Ejectives and beyond. *Plos one 16*(2), e0245522.

van de Weijer, J. (2003). Consonant variation within words. In *Proceedings of the Corpus Linguistics 2003 Conference. University Centre for Computer Corpus Research on Language Technical Papers*, Volume 16, pp. 184–190. Citeseer.

van de Weijer, J. (2005). Listeners' sensitivity to consonant variation within words. *Lund University Working Papers in Linguistics 51*, 225–238.

Wilson, C. and M. Obdeyn (2009). Simplifying subsidiary theory: statistical evidence from arabic, muna, shona, and wargamay. *Ms, Johns Hopkins University*.

Bibliography V

Yao, Y., A. Vehtari, D. Simpson, and A. Gelman (2017). Using stacking to average Bayesian predictive distributions. *Bayesian Analysis*.

Zorc, R. D. (1969). A study of the aklanon dialect, volume two: Dictionary (of root words and derivations), aklanon to english.