# Automating Sound Change Prediction for Phylogenetic Inference

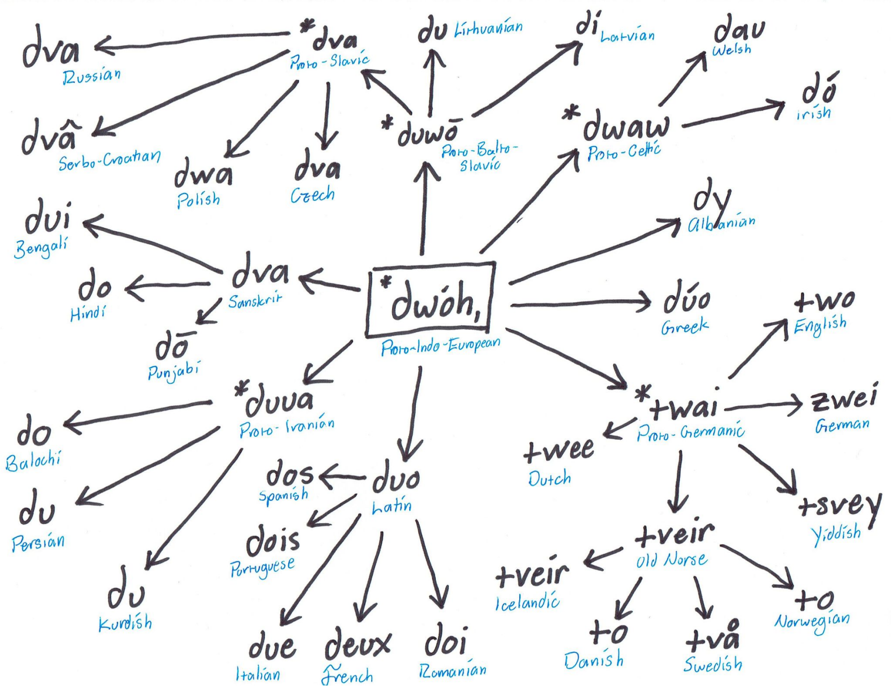
## A Tukanoan Case Study

Kalvin Chang*, Nathaniel Robinson*, Anna Cai*, Ting Chen, Annie Zhang, David Mortensen

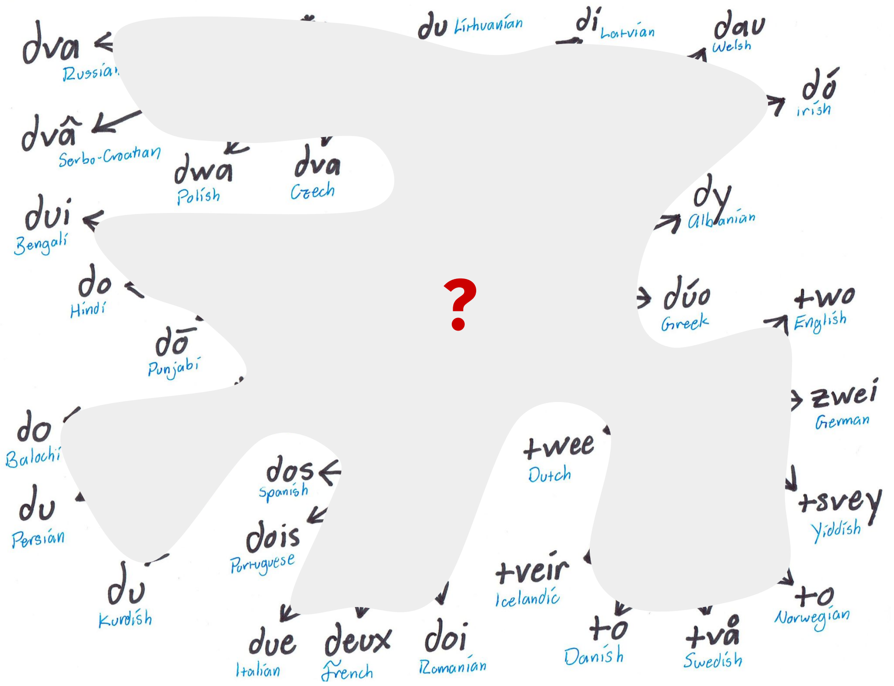
Carnegie Mellon University
Language Technologies Institute

LLab
Language Technologies Institute
Carnegie Mellon University

# Many modern languages are descended from a common ancestor.



# How can we deduce the protoforms and evolutionary history?



# Laws governing sound changes are regular and exceptionless!

| | | | | |
|---|---|---|---|---|
| | kaprə | karo | kapos | … |
| a > ɛ | kɛprə | kɛro | kɛpos | … |
| k > ʃ / # _ | ʃɛprə | ʃɛro | ʃɛpos | … |
| … | | | | |

# Comparative method (Campbell 2013)

1. Assemble cognate sets
2. Find corresponding sounds
3. Propose proto-sounds and sound laws for each correspondence
4. Iterate previous step
    a. sound laws must be logically consistent and probabilistically likely
5. Reconstruct protoforms using proto-sounds
6. Reconstruct evolutionary tree using sound laws (phylogenetic inference)

# Comparative method (Campbell 2013)

Automate?

1. Assemble cognate sets
2. Find corresponding sounds
3. Propose proto-sounds and sound laws for each correspondence
4. Iterate previous step
   a. sound laws must be logically consistent and probabilistically likely
5. Reconstruct protoforms using proto-sounds
6. Reconstruct evolutionary tree using sound laws (phylogenetic inference)

# Towards automating the comparative method

Automate?

1. Assemble cognate sets
2. Find corresponding sounds
3. Propose proto-sounds and sound laws for each correspondence
4. Iterate previous step
    a. sound laws must be logically consistent and probabilistically likely
5. Reconstruct protoforms using proto-sounds
6. Reconstruct evolutionary tree using sound laws (phylogenetic inference)

# Prior work: Chacon & List (2016)

Parsimony-based algorithm

1. Align protoforms with reflexes (expert)
2. Learn sound laws from aligned reflexes (expert)
3. Create sound change transition matrix (expert)
   a. Identify intermediate sound changes 💭
   b. Assign weight to intermediate sound transitions 🏋
4. Infer phylogeny via maximum parsimony
5. Obtain consensus tree

# Our realization

Parsimony-based algorithm

1. Align protoforms with reflexes (expert)
2. Learn sound laws from aligned reflexes (expert)
3. Create sound change transition matrix (expert)
   a. Identify intermediate sound changes 💭
   b. Assign weight to intermediate sound transitions 🏋
4. Infer phylogeny via maximum parsimony
5. Obtain consensus tree

Replace with Automatic intermediate sound change prediction (AISCP)

# Our realization

Parsimony-based algorithm

1. Align protoforms with reflexes (expert)
2. Learn sound laws from aligned reflexes (expert)
3. Create sound change transition matrix (expert)
    a. Identify intermediate sound changes 💭
    b. Assign weight to intermediate sound transitions 🏋
4. Infer phylogeny via maximum parsimony
5. Obtain consensus tree

Replace with Automatic sound law induction (ASLI)

Replace with Automatic intermediate sound change prediction (AISCP)

# AISCP: Intermediate sound changes 💥

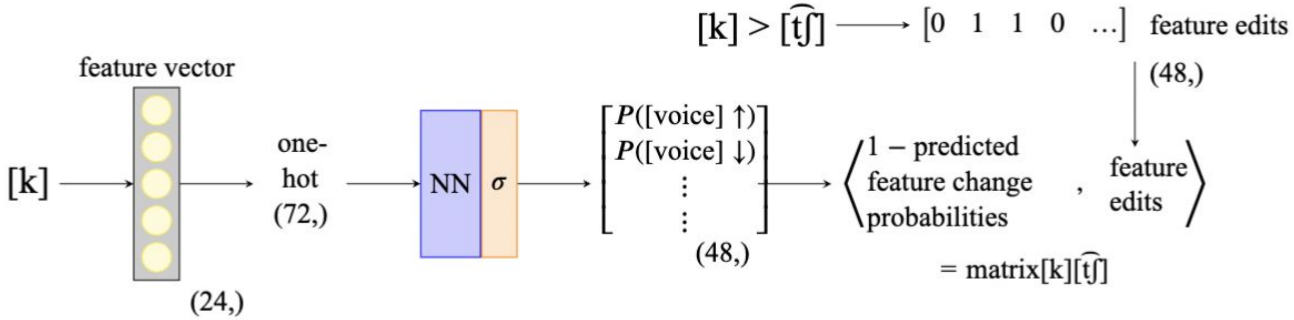
- Need: mapping from phones to articulatory features
  - $f: s \rightarrow \{-1, 0, 1\}^N$
- Create fully connected graph of phones
  - Edges weighted by feature edit distance (FED)
- Encodes similarity of sounds
  - [d] and [t] differ only in one feature (voice)
  - [k] and [t] differ in four
  - I.e. *FED*([k], [t]) = 4 * *FED*([d], [t])

# AISCP: Intermediate sound changes 💭

- Need: mapping from phones to articulatory features
  - $f: s \rightarrow \{-1, 0, 1\}^N$
- Create fully connected graph of phones
  - Edges weighted by feature edit distance (FED)
- Encodes similarity of sounds
  - [d] and [t] differ only in one feature (voice)
  - [k] and [t] differ in four
  - I.e. *FED*([k], [t]) = 4 * *FED*([d], [t])
- However, this is not directional!

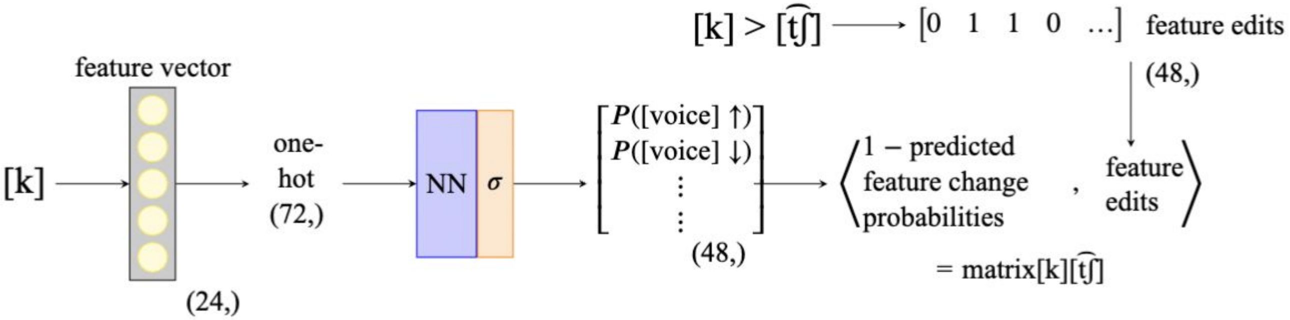
# AISCP: Directional weighted FED 🏋️

- Need to model *P*(voicing) ≠ *P*(devoicing)
- Neural network $M: \{0, 1\}^{3N} \rightarrow \{0, 1\}^{2N}$
  - Prediction of each feature’s direction of change given the source phone’s features



# AISCP: Directional weighted FED 🏋

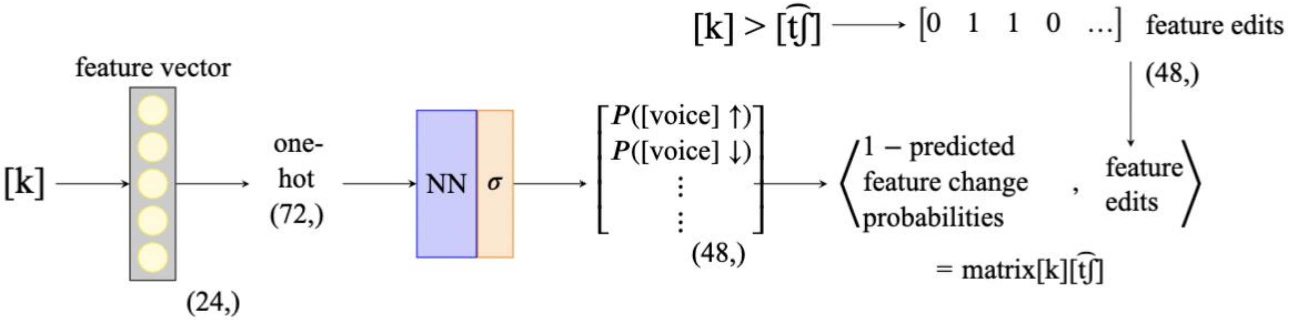
- Need to model *P*(voicing) ≠ *P*(devoicing)
- Neural network $M: \{0, 1\}^{3N} \rightarrow \{0, 1\}^{2N}$
  - Prediction of each feature's direction of change given the source phone's features




Trained on database of attested sound changes (e.g. [k] > [tʃ])

# AISCP: Directional weighted FED 🏋️

- Fit NN on multilingual sound changes from the *Index Diachronica*
- Predicts realistic intermediate paths:
  - k > c > tɕ > tʃ
  - p > f > h



# AISCP: Results

- Generalized Quartet Distance (GQD) = 0.12
  - reproduces 88% of valid quartets
- Recovers major Tukanoan subgroups from Chacon (2014)

| **Experiment** | **GQD (Min)** $\downarrow$ | **GQD (Mean** $\pm\sigma$**)** $\downarrow$ |
|---|---|---|
| Baseline: cognacy | 0.533 | 0.533 |
| Baseline: shared innovations | 0.355 | 0.355 |
| C+L w/ AISCP, 1 layer NN | **0.120** | **0.295** ±0.118 |
| C+L w/ AISCP, 4 layer NN | 0.191 | 0.309 ±0.096 |
| C+L w/ AISCP, 8 layer NN | 0.402 | 0.439 ±0.021 |
| C+L w/ AISCP, 16 layer NN | 0.248 | 0.435 ±0.080 |

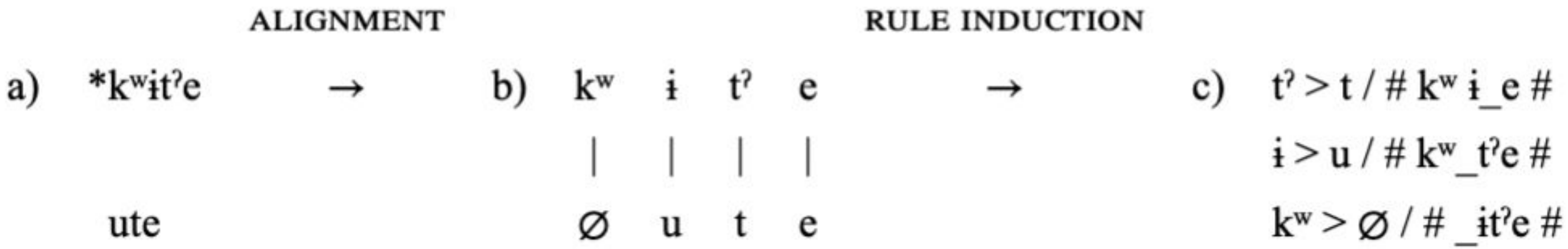
# Ablations

- Standard *FED* (non directional weights)
- Direct paths (no intermediate sound changes)

| Experiment | GQD (Min) $\downarrow$ | GQD (Mean $\pm\sigma$) $\downarrow$ |
|---|---|---|
| C+L, w/ AISCP (*standard FED* ablation) | 0.325 | 0.440 ±0.062 |
| C+L, w/ AISCP (*direct paths* ablation) | 0.281 | 0.397 ±0.072 |
| C+L w/ AISCP, 1 layer NN | **0.120** | **0.295** ±0.118 |
| C+L w/ AISCP, 4 layer NN | 0.191 | 0.309 ±0.096 |
| C+L w/ AISCP, 8 layer NN | 0.402 | 0.439 ±0.021 |
| C+L w/ AISCP, 16 layer NN | 0.248 | 0.435 ±0.080 |

# Automatic Sound Law Induction (ASLI)

- Automatically generate correspondences from cognate data
- Alignment
  - Minimize FED
- Minimal generalization (Albright & Hayes 2002, Wilson & Li 2021)
  - iteratively generalize rules from base rules and keep most applicable



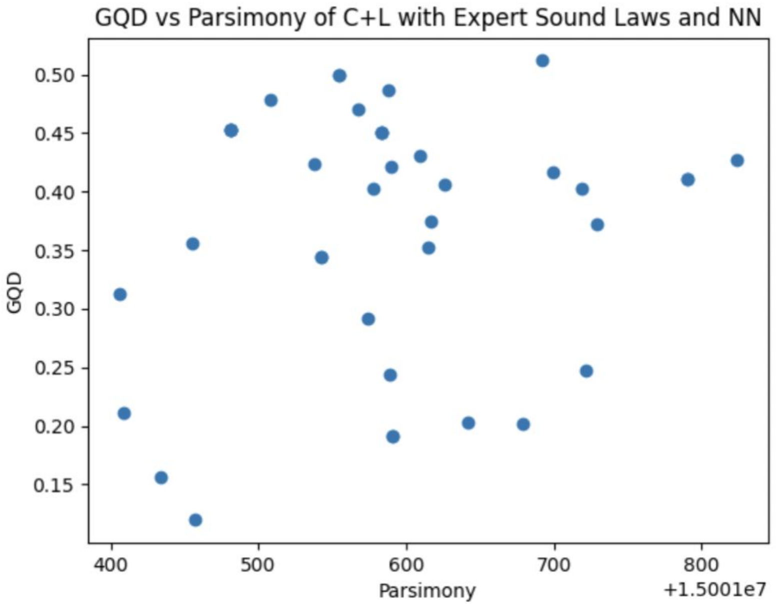
# ASLI: Results

- Many generated sound laws are hyper-specific
  - e.g. p > m / # (pˀ | ˀp) o _ a #
- However, overall their phylogenetic signal may still be sufficient
- Note: the numbers in the proceedings are incorrect

| Experiment | GQD (Min) ↓ | GQD (Mean $\pm\sigma$) ↓ |
|---|---|---|
| C+L w/ AISCP + ASLI 1 layer NN | **0.124** | **0.224** ±0.076 |
| C+L w/ AISCP + ASLI, 4 layer NN | 0.354 | 0.461 ±0.070 |
| C+L w/ AISCP + ASLI, 8 layer NN | 0.237 | 0.433 ±0.092 |
| C+L w/ AISCP + ASLI, 16 layer NN | 0.396 | 0.483 ±0.084 |

# Parsimony is not correlated with GQD

- Spearman's correlation: -0.04
- Optimizing over parsimony may not yield optimal trees!
- Should consider probabilistic methods instead
  - Bayesian inference




Carnegie Mellon University
Language Technologies Institute

LLab
Language Technologies Institute
Carnegie Mellon University

# Conclusion

- Novel method for modeling diachronic intermediate sound changes for phylogenetic inference
- Predicted tree with 0.12 GQD for Tukanoan language family
- Proposed intermediate sound changes capture expert intuitions on phonetic naturalness

# Thank you for listening!

# Questions?

https://github.com/cmu-llab/aiscp

Carnegie Mellon University
Language Technologies Institute

LLab
Language Technologies Institute
Carnegie Mellon University