A Probabilistic Approach to Language Change

Alexandre Bouchard-Côté, Percy Liang, Thomas L. Griffiths, Dan Klein

University of California at Berkeley

1. Phono-.• Basis of the Swim..• Phonological rules more regular than morphological or syntactic ones
• Basis of the comparative method:
- 1a: Classical Latin
- 1b: Vulgar Latin
- 2b: Proto-Ibero Romance

2. Model: assume for now that the tree topology is known
- Types of operations:
- Context:
- Example:

3. Prior:
- A log-linear model
- Standard L2 regularization
- Various context granularities

4. Inference: stochastic EM (exact E step is intractable)
- We approximate E step based on Gibbs sampling

5. Experiments
- Task 1: reconstruction of Latin
- Task 2: inference of phonological rules
- Task 3: selection of phylogenies

6. Conclusion and future work:
- A probabilistic approach to diachronic phonology
- Log-linear prior yields better reconstructions; interesting connection with stochastic optimality theory
- Enables reconstruction of ancient and modern word forms, phonological rules and tree topologies

7. Word/verb
- In practice, the ancient words and/or the evolutionary tree are unknown
- Methodology: manually inspecting the data