A Probabilistic Approach to Diachronic Phonology

Alexandre Bouchard-Côté    Percy Liang
Tom Griffiths    Dan Klein
# Languages evolve

| Gloss     | Latin     | Italian   | Spanish   | Portuguese |
|-----------|-----------|-----------|-----------|------------|
| Word/verb | verbum    | verbo     | verbo     | verbu      |
| Fruit     | fructus   | frutta    | fruta     | fruta      |
| Laugh     | ridere    | ridere    | reir      | rir        |
| Center    | centrum   | centro    | centro    | centro     |
| August    | augustus  | agosto    | agosto    | agosto     |
| Swim      | natare    | nuotare   | nadar     | nadar      |
## Language evolution

| Gloss  | Latin   | Italian  | Spanish  | Portuguese |
|--------|---------|----------|----------|------------|
| Word/verb | verbum  | verbo    | verbo    | verbu      |
| Fruit   | fructus | frutta   | fruta    | fruta      |
| Laugh   | ridere  | ridere   | reir     | rir        |
| Center  | centrum | centro   | centro   | centro     |
| August  | augustus| agosto   | agosto   | agosto     |
| Swim    | natare  | nuotare  | nadar    | nadar      |

• Phonological rules more **regular** than morphological or syntactic ones

• basis of the **comparative method**
Example of a mutation process as seen by the comparative method

- la
- vl
- it
- es
- pt

• ib : Proto-ibero Romance
• vl : Vulgar Latin
Example of a mutation process as seen by the comparative method

- Deterministic re-write rules at each branch
- Activated by some context
Example of a mutation process as seen by the comparative method

| Gloss | Latin | Italian | Spanish | Portuguese |
|-------|-------|---------|---------|------------|
| Word/verb | verbum | verbo | verbo | verbu |
Example of a mutation process as seen by the comparative method

| Gloss | Latin   | Italian | Spanish | Portuguese |
|-------|---------|---------|---------|------------|
| Word/verb | verbum | verbo   | verbo   | verbu      |
| Center    | centrum | centro  | centro  | centro     |
Example of a mutation process as seen by the comparative method

- In practice, the ancient words and/or the evolutionary tree are unknown
- Methodology: manually inspecting the data
Our work:

- A probabilistic model that captures phonological aspects of language change.

- Many usages:

Reconstruction of word forms (ancient and modern)
Our work:

- A probabilistic model that captures phonological aspects of language change.

- Many usages:

```
/kwintam/
/kinta/
/kinto/
/kimtu/
/kwinto/
```

Inference of phonological rules
Our work:

- A probabilistic model that captures phonological aspects of language change.
- Many usages:

Selection of phylogeny
Our work:

- A probabilistic model that captures phonological aspects of language change.

- Many usages:
  - Reconstruction of word forms (ancient and modern)
  - Inference of phonological rules
  - Selection of phylogenies

- An inference procedure and experiments on all three applications

- A new task and evaluation framework
The model
• Assume for now that the tree topology is known
• Assume for now that the tree topology is known
• Track individual words
• Let’s look at how a single word evolves along one of the edges of the tree

• Mutation of Latin *FOCUS* (/fokus/) into Italian *fuoco* (/fwɔko/) (fire)
Stochastic edit model: operations

- Substitution
Stochastic edit model: operations

- Substitution (incl. self-substitution)
Stochastic edit model: operations

- Substitution (incl. self-substitution)
- Insertion
Stochastic edit model: operations

- Substitution (incl. self-substitution)
- Insertion
- Deletion
Stochastic edit model: context

- Distribution over operations conditioned on adjacent phonemes
Stochastic edit model: generation process

# f o k u s #
Stochastic edit model: generation process

# f o k u s #

?
Stochastic edit model: generation process

- \( P(f \rightarrow f \, w \mid \# \, \# \, V) = 0.05 \)
Stochastic edit model: generation process

- \( \mathbb{P}(f \rightarrow f w / \# \_ V) = 0.05 \)
Stochastic edit model: generation process

- $P(f \rightarrow f\,w / \# \_ V) = 0.05$
- $P(o \rightarrow o / C \_ V) = 0.1$
Stochastic edit model: generation process

- P(f → f w / # V) = 0.05
- P(o → o / C V) = 0.1
- . . .
- P(/fokus/ → /fwcko/)) = 0.05 × 0.1 × . . .
Edit parameters

Diagram showing relationships between words and their transformations.
• One set of parameter $\theta_{A \rightarrow B}$ for each edge $A \rightarrow B$ in the tree

• Shared across all word forms evolving along this edge
- $\theta_{A \to B}$ specifies $P(\text{operation}|\text{context})$

| context | operation                  | $P(\text{operation}|\text{context})$ |
|---------|----------------------------|-------------------------------------|
| u m #   | deletion                   | 0.1                                 |
| u m #   | substitution to /m/        | 0.8                                 |
| u m #   | substitution to /b/        | 0.1                                 |
| a c b   | deletion                   | 0.8                                 |
| a c b   | insertion of c             | 0.1                                 |
| :       | :                          | :                                   |
Distribution on the edit parameters

- Too many parameters

- Addressed by:
  - Sparsity prior: independent Dirichlet priors (one for each context)
  - Group context distributions. Example:

| context | operation               | \( \mathbb{P}(\text{operation}|\text{context}) \) |
|---------|-------------------------|---------------------------------------------------|
| V m #   | deletion                | 0.1                                               |
| V m #   | substitution to /a/     | 0.8                                               |
| V m #   | substitution to /b/     | 0.1                                               |
| V c C   | deletion                | 0.8                                               |
| V c C   | insertion of c          | 0.1                                               |
| ...     | ...                     | ...                                               |
Inference and experiments
Inference: EM

• Exact E step is intractable
  – We use a stochastic E step based on Gibbs sampling

• E: fix the edit parameters, resample the derivations

• M: update the edit parameters from expected edit counts
Automatic extraction of a Romance corpus

Wiktionary $\rightarrow$ XML dump

Bible $\rightarrow$ Align. $\rightarrow$ Closure $\rightarrow$ Cognate detector

Europarl $\rightarrow$ Align.

- Noisier than manually curated cognate lists
- More data available
- Our model overcomes this noise

Data available online:
http://nlp.cs.berkeley.edu/pages/historical.html
Reconstruction of ancient word forms

• Task: reconstruction of Latin given all of the Spanish and Italian words, and some of the Latin words
• Evaluation: uniform cost edit distance on held-out data
• Baseline: pick one of the modern languages at random
Reconstruction of ancient word forms

- Task: reconstruction of Latin given all of the Spanish and Italian words, and some of the Latin words
- Example: “teeth”, nearly correctly reconstructed

\[
\text{/dʒɛntɪs/}
\]

\[
i \rightarrow \varepsilon
\]
\[
\varepsilon \rightarrow j\varepsilon
\]

\[
\text{/dʒɛntes/}
\]
\[
s \rightarrow
\]

\[
\text{/dɛnti/}
\]

- Numbers:

| Language | Baseline | Model | Improvement |
|----------|----------|-------|-------------|
| Latin    | 2.84     | 2.34  | 9%          |
Reconstruction of word forms

- Evaluation: uniform cost edit distance on held-out data
- Baseline: pick one of the modern languages at random
- Example: “teeth”, nearly correctly reconstructed

\[
/d\text{\'entis}/
\]

\[
i \rightarrow \varepsilon \\
\varepsilon \rightarrow j \varepsilon \\
s \rightarrow
\]

\[
/d\text{\'entis}/ \\
/d\text{\'entes}/ \\
/d\text{\'enti}/
\]

- Numbers:

| Language  | Baseline | Model | Improvement |
|-----------|----------|-------|-------------|
| Latin     | 2.84     | 2.34  | 9%          |
| Spanish   | 3.59     | 3.21  | 11%         |
Inference of phonological rules

- $ib$: Proto-ibero Romance
- $vl$: Vulgar Latin
Inference of phonological rules

- Reconstruct the internal nodes
- Focus on the rules used most often during the last E step
Hypothesized derivation for “word” along with top rules

\[
/\text{werbum}/ \ (1a) \\
\quad m \rightarrow \ _ \\
\quad u \rightarrow o \\
\quad w \rightarrow v \\
\downarrow \\
/\text{verbo}/ \ (vl) \\
\quad r \rightarrow r \\
\quad e \rightarrow \varepsilon \\
\quad \ldots \\
\quad \ldots \\
\]

- m → / _ #
- u → o / _
- w → v / many environments
- ...

• Comparison with historical evidence: the *Appendix Probi*

  coluber  non colober
  passim  non passi
Hypothesized derivation for “word” along with top rules

- /v/ to /b/ fortition
- /s/ to /z/ voicing in Italian
Selection of phylogenies
Inference of topology

? 

la

es it pt
Example of previous approaches

- Gray and Atkinson, 2003

- Coarse encoding:

| Language | Meaning            | Eat | Cognate set |
|----------|--------------------|-----|-------------|
| Latin    | mandere (to chew)  | 1   | 2           |
| French   | manger             | 0   | 1           |
| Italian  | mangiare           | 1   | 0           |

| Language | Meaning            | Eat | Cognate set |
|----------|--------------------|-----|-------------|
| Latin    | comedere (to consume) | 1   | 0           |
| Spanish  | comer              | 0   | 1           |
| Portuguese | comer           | 0   | 1           |

- These characters evolve independently in their model

- Lots of information discarded
Our samples look like this
What we did

- Present good vs. bad topologies and compute the likelihood ratio

- this can be turned into a full topology inference algorithm using the quartet method [Erdos et al., 1996]
Conclusion

- Introduced a probabilistic approach to diachronic phonology
- Enables reconstruction of ancient and modern word forms, phonological rules and tree topologies
- Future work:
  - We are scaling it up to larger phylogenies
  - We are working on an extension using a log-linear parametrization of the contexts, reminiscent of stochastic OT
- Data available online:
  [http://nlp.cs.berkeley.edu/pages/historical.html](http://nlp.cs.berkeley.edu/pages/historical.html)