A Probabilistic Approach to Diachronic Phonology

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

<table>
<thead>
<tr>
<th>Gloss</th>
<th>Latin</th>
<th>Italian</th>
<th>Spanish</th>
<th>Portuguese</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word/verb</td>
<td>verbum</td>
<td>verbo</td>
<td>verbo</td>
<td>verbu</td>
</tr>
<tr>
<td>Fruit</td>
<td>fructus</td>
<td>frutta</td>
<td>fruta</td>
<td>fruta</td>
</tr>
<tr>
<td>Laugh</td>
<td>ridere</td>
<td>ridere</td>
<td>reir</td>
<td>rir</td>
</tr>
<tr>
<td>Center</td>
<td>centrum</td>
<td>centro</td>
<td>centro</td>
<td>centro</td>
</tr>
<tr>
<td>August</td>
<td>augustus</td>
<td>agosto</td>
<td>agosto</td>
<td>agosto</td>
</tr>
<tr>
<td>Swim</td>
<td>natare</td>
<td>nuotare</td>
<td>nadar</td>
<td>nadar</td>
</tr>
</tbody>
</table>
## Language evolution

<table>
<thead>
<tr>
<th>Gloss</th>
<th>Latin</th>
<th>Italian</th>
<th>Spanish</th>
<th>Portuguese</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word/verb</td>
<td>verbum</td>
<td>verbo</td>
<td>verbo</td>
<td>verbu</td>
</tr>
<tr>
<td>Fruit</td>
<td>fructus</td>
<td>frutta</td>
<td>fruta</td>
<td>fruta</td>
</tr>
<tr>
<td>Laugh</td>
<td>ridere</td>
<td>ridere</td>
<td>reir</td>
<td>rir</td>
</tr>
<tr>
<td>Center</td>
<td>centrum</td>
<td>centro</td>
<td>centro</td>
<td>centro</td>
</tr>
<tr>
<td>August</td>
<td>augustus</td>
<td>agosto</td>
<td>agosto</td>
<td>agosto</td>
</tr>
<tr>
<td>Swim</td>
<td>natare</td>
<td>nuotare</td>
<td>nadar</td>
<td>nadar</td>
</tr>
</tbody>
</table>

- 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

- **ib**: Proto-iberic 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

<table>
<thead>
<tr>
<th>Gloss</th>
<th>Latin</th>
<th>Italian</th>
<th>Spanish</th>
<th>Portuguese</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word/verb</td>
<td>verbum</td>
<td>verbo</td>
<td>verbo</td>
<td>verbu</td>
</tr>
</tbody>
</table>
Example of a mutation process as seen by the comparative method

<table>
<thead>
<tr>
<th>Gloss</th>
<th>Latin</th>
<th>Italian</th>
<th>Spanish</th>
<th>Portuguese</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word/verb</td>
<td>verbum</td>
<td>verbo</td>
<td>verbo</td>
<td>verbu</td>
</tr>
<tr>
<td>Center</td>
<td>centrum</td>
<td>centro</td>
<td>centro</td>
<td>centro</td>
</tr>
</tbody>
</table>
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:

  ![Diagram showing phonological rules](image)

  Inference of phonological rules
Our work:

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

- Many usages:

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

VS.

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

Selection of phylogenies
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 words evolve 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

\[ P(f \rightarrow fw / \# - V) = 0.05 \]
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

\[
\begin{align*}
\# & \rightarrow f & o & k & u & s & \rightarrow # \\
& \rightarrow f & w & \circ & k & o
\end{align*}
\]

- \( \mathbb{P}(f \rightarrow f \ w \ / \ # \ _ \ V) = 0.05 \)
- \( \mathbb{P}(o \rightarrow \circ \ / \ C \ _ \ V) = 0.1 \)
- \( \ldots \)
- \( \mathbb{P}(/fokus/ \rightarrow /fw\circko/)) = 0.05 \times 0.1 \times \ldots \)
• 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
• $\theta_{A \rightarrow B}$ 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               | $P(operation|context)$ |
|---------|-------------------------|------------------------|
| $V_m$ # | 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

\[ /\text{dɛntis/} \]

\[
\begin{aligned}
i &\rightarrow \varepsilon \\
\varepsilon &\rightarrow j \varepsilon \\
s &\rightarrow
\end{aligned}
\]

\[ /\text{djɛntes/} \quad /\text{dɛnti/} \]

- Numbers:

<table>
<thead>
<tr>
<th>Language</th>
<th>Baseline</th>
<th>Model</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latin</td>
<td>2.84</td>
<td>2.34</td>
<td>9%</td>
</tr>
</tbody>
</table>
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

\[
/d'entis/ \\
\text{i} \rightarrow \varepsilon \\
\varepsilon \rightarrow j\varepsilon \\
\text{s} \rightarrow \\
/d'jentes/ \\
/d'enti/ \\
\]

- Numbers:

<table>
<thead>
<tr>
<th>Language</th>
<th>Baseline</th>
<th>Model</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latin</td>
<td>2.84</td>
<td>2.34</td>
<td>9%</td>
</tr>
<tr>
<td>Spanish</td>
<td>3.59</td>
<td>3.21</td>
<td>11%</td>
</tr>
</tbody>
</table>
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

/werbum/ (la)
  m → u
  u → o
  w → v

/verbo/ (vl)
  r → r
  e → ε

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

- \(u \rightarrow o\) / many environments
- \(v \rightarrow b\) / init. or intervocal.
- \(t \rightarrow t e / ALV _ #\)

\(\ldots\)

\(/\text{verbo}/ (ib)\)

\(v \rightarrow b\)

\(\ldots\)

\(/\text{berbo}/ (es)\)

\(u \rightarrow o\)

\(/\text{verbu}/ (pt)\)

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

![Diagram showing the inference of topology with nodes labeled 'la', 'es', 'it', and 'pt'].

The diagram illustrates the relationships and connections among these nodes, possibly representing different languages or concepts in the field of topology.
Example of previous approaches

- Gray and Atkinson, 2003

- Coarse encoding:

<table>
<thead>
<tr>
<th></th>
<th>Latin</th>
<th>French</th>
<th>Italian</th>
</tr>
</thead>
<tbody>
<tr>
<td>mandere (to chew)</td>
<td>manger</td>
<td>mangiare</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Meaning</th>
<th>Eat</th>
<th>…</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognate set</td>
<td>1</td>
<td>2</td>
<td>…</td>
</tr>
<tr>
<td>Latin</td>
<td>1</td>
<td>1</td>
<td>…</td>
</tr>
<tr>
<td>French</td>
<td>1</td>
<td>0</td>
<td>…</td>
</tr>
<tr>
<td>Italian</td>
<td>1</td>
<td>0</td>
<td>…</td>
</tr>
<tr>
<td>Spanish</td>
<td>0</td>
<td>1</td>
<td>…</td>
</tr>
<tr>
<td>Portuguese</td>
<td>0</td>
<td>1</td>
<td>…</td>
</tr>
</tbody>
</table>

- These characters evolve independently in their model

- Lots of information discarded
Our samples look like this
Comparison

Atkinson’s
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