Lecture 30: Diachronic Models

Dan Klein – UC Berkeley
Work with Alex Bouchard-Cote and Tom Griffiths

http://andromeda.rutgers.edu/~jlynch/language.html
Language Evolution

Latin  camera /kamera/

Deletion: /e/

Change of place: /k/ .. /tʃ/ .. /ʃ/ 

Insertion: /b/

French  chambre /ʃambʁ/

Eng, camera from Latin, “camera obscura”

Eng, chamber from Old Fr. before the initial /t/ dropped

---

Diachronic Evidence

Yahoo! Answers

Yahoo! Answers Appendix Probi

tonight not tonite
	onitru non tonotru

- Spelling (orthography) can reflect old pronunciation
- Corrections show when orthography hasn’t kept up!
Example: Great Vowel Shift

(Simplified!)

“time” = teem → “time” = taim

This is why the letter “i” is spoken as “ee” by many other languages, etc.

Where’s It Going?

- Language isn’t going anywhere in particular
- In fact, it’s basically going everywhere
  - Over time, languages drift around
  - Related languages diverge
  - Eventually, results say more about the human language system than about history [Griffiths and Kalish 2007]
- Examples of tradeoffs
  - More consonant clusters vs. more syllables
  - More morphology vs. more rigid word order
  - Stress vs. tones vs. vowel variety
### Synchronic (Comparative) Evidence

<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><em>verbum</em></td>
<td><em>verbo</em></td>
<td><em>verbo</em></td>
<td><em>verbu</em></td>
</tr>
<tr>
<td>Fruit</td>
<td><em>fructus</em></td>
<td><em>frutta</em></td>
<td><em>fruta</em></td>
<td><em>fruta</em></td>
</tr>
<tr>
<td>Laugh</td>
<td><em>ridere</em></td>
<td><em>ridere</em></td>
<td><em>reir</em></td>
<td><em>rir</em></td>
</tr>
<tr>
<td>Center</td>
<td><em>centrum</em></td>
<td><em>centro</em></td>
<td><em>centro</em></td>
<td><em>centro</em></td>
</tr>
<tr>
<td>August</td>
<td><em>augustus</em></td>
<td><em>agosto</em></td>
<td><em>agosto</em></td>
<td><em>agosto</em></td>
</tr>
<tr>
<td>Swim</td>
<td><em>natare</em></td>
<td><em>nuotare</em></td>
<td><em>nadar</em></td>
<td><em>nadar</em></td>
</tr>
</tbody>
</table>

---

### A Mini-Romance Phylogeny

[Diagram showing the phylogeny of Romance languages]
A Probabilistic Model

Model Parameters
Local Mutation along Tree

Ancient to Modern Forms

<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>verbo</td>
</tr>
</tbody>
</table>
Ancient to Modern Forms

Learned Rules / Mutations

coluber non colober
passim non passi
Learned Rules / Mutations

- $u \rightarrow \emptyset$ / many environments
- $v \rightarrow b$ / init. or intervocal.
- $t \rightarrow t e / ALv_{-\#}$

Oceanic Languages

Proto-Oceanic
### Oceanic Data

<table>
<thead>
<tr>
<th>Gloss</th>
<th>Hawai’ian</th>
<th>Maori</th>
<th>Samoan</th>
<th>Tongan</th>
<th>ProtoOceanic</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘break’</td>
<td>haki</td>
<td>whati</td>
<td>fati</td>
<td>fasi</td>
<td>*fati</td>
</tr>
<tr>
<td>‘house’</td>
<td>hale</td>
<td>whare</td>
<td>fale</td>
<td>fale</td>
<td>*fale</td>
</tr>
<tr>
<td>‘yam’</td>
<td>uhi</td>
<td>uhi</td>
<td>ufi</td>
<td>ufi</td>
<td>*ufi</td>
</tr>
<tr>
<td>‘woman’</td>
<td>wahine</td>
<td>wahine</td>
<td>fafine</td>
<td>fefine</td>
<td>*wafine</td>
</tr>
<tr>
<td>‘moon’</td>
<td>mahina</td>
<td>mahina</td>
<td>masina</td>
<td>mahina</td>
<td>*masiana</td>
</tr>
</tbody>
</table>

### POc Reconstruction Results

<table>
<thead>
<tr>
<th>Condition</th>
<th>Edit dist.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full system</td>
<td>1.87</td>
</tr>
<tr>
<td>-FAITHFULNESS</td>
<td>2.02</td>
</tr>
<tr>
<td>-MARKEDNESS</td>
<td>2.18</td>
</tr>
<tr>
<td>-Sharing</td>
<td>1.99</td>
</tr>
<tr>
<td>-Topology</td>
<td>2.06</td>
</tr>
</tbody>
</table>
Learned Phonological Shifts

Prior Weights

Posterior Weights

Example Parameters
Conclusion

- Languages undergo evolutionary processes
- Can model as regular edits along a tree
- Using modern forms ONLY:
  - We can determine the historical phylogeny
  - We can reconstruct ancient forms (though inherently less accurate for older forms)
- A lot still left to do!

Thank You!