Lecture 30: Diachronic Models

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

Language Evolution

Latin: camera /kamera/
- Deletion: /e/
- Change of place: /k/ → /ʃ/...
- Insertion: /b/

French: chambre /ʃambr/ (Old Fr. chambre, before initial /t/ dropped)

Eng. camera from Latin, “camera obscura”
Eng. chamber from Old Fr.

Diachronic Evidence

Yahoo! Answers
- tonight not tonite
- tonitru non tonotru

Appendix Probi
- Spelling (orthography) can reflect old pronunciation
- Corrections show when orthography hasn’t kept up!

Example: Great Vowel Shift

“time” = teem → “time” = taim

(Simplified!)

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>verbum</td>
<td>verbo</td>
<td>verbo</td>
<td>verbo</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>rir</td>
<td></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>

A Mini-Romance Phylogeny

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

Oceanic Languages

Oceanic Data

POc Reconstruction Results

<table>
<thead>
<tr>
<th>Gloss</th>
<th>Hawaiian</th>
<th>Maori</th>
<th>Samoan</th>
<th>Tongan</th>
<th>Proto-Oceanic</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘break’</td>
<td>haki</td>
<td>whati</td>
<td>fasi</td>
<td>*fati</td>
<td></td>
</tr>
<tr>
<td>‘house’</td>
<td>hale</td>
<td>whare</td>
<td>fale</td>
<td>*fale</td>
<td></td>
</tr>
<tr>
<td>‘yam’</td>
<td>uhi</td>
<td>uhi</td>
<td>ufi</td>
<td>ufi</td>
<td></td>
</tr>
<tr>
<td>‘woman’</td>
<td>wahine</td>
<td>wahine</td>
<td>fatine</td>
<td>fatine</td>
<td>*fatine</td>
</tr>
<tr>
<td>‘moon’</td>
<td>mahina</td>
<td>mahina</td>
<td>masina</td>
<td>mahina</td>
<td>*masiana</td>
</tr>
</tbody>
</table>

Condition | Edit dist. |
---|---|
Full system | 1.87 |
-FAITHFULNESS | 2.02 |
-MARKEDNESS | 2.18 |
-Sharing | 1.99 |
-Topology | 2.06 |
Learned Phonological Shifts

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!