Assignment 2

L245

Writers’s aids

Due Monday, February 6

1. Do question #1a in chapter 2 of the draft textbook (p. 83).

2. Pretend we have a bigram array, as in the given table, where the first letter of the bigram is given in the vertical letters (i.e., down the side), and the second letter is given in the horizontal ones (i.e., across the top)

<table>
<thead>
<tr>
<th></th>
<th>i</th>
<th>j</th>
<th>k</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>j</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>k</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

Make this into a positional bigram array, namely one which captures the position “middle of word”. Provide a word which justifies each 1 you put in the chart.

3. Do question #3 in chapter 2 of the draft textbook (p. 84).

   • **Bonus:** Do question #4 in chapter 2 of the draft textbook (p. 84).

4. Consider the misspelling forg, and assume our edit distance calculations have insertions, deletions, substitutions, and transpositions. Describe how probabilities are used to rank frog vs. forge. Do you have an intuition as to which should be higher, and why?

5. Pick a search engine of your choice, as long as the search engine provides spelling suggestions. Search with the keywords ghouls and ghosts, and do not use quotes.

   (a) Try as many sensible misspellings as you can think of (and vary the order of the words, too). Which ones does the search engine catch and which doesn’t it catch? Try using with and without quotation marks.

   (b) Given what we talked about in class, can you explain why certain variations were flagged as misspellings and others were not? If the search engine always correctly caught your misspellings, it will help to try some even more varied misspellings, in order to find the “boundary” where the correct suggestion is no longer made.

6. Do question #6 in chapter 2 of the draft textbook (p. 84–85) ... but, instead of evaluating spelling errors, evaluate grammar errors.
7. Look at the following examples of mistakes made by smartphone auto-correct programs.

- “Sitting at the dentist waiting for my son...they said he has NO favorites!”\(^1\) cavities → favorites
- “I’m like $300 short on my MOTTSAAPPLESAUCE payment due the 15th.”\(^2\) mortgage payment → MOTTSAAPPLESAUCE
- “I’ll get back with you by tomorrow. We’re planning my starvation”\(^3\) staycation → starvation
- “Thanks for letting me centipede to you today by the way”\(^4\) vent → centipede
- “I have to go get my kids. my parrots watched them last night”\(^5\) parents → parrots
- “Please bring anvils if you can”\(^6\) a coke → anvils
- “Sry I was teething”\(^7\) tweeting → teething
- “don’t worry, you look super affordable”\(^8\) adorable → affordable
- “T.G.I.Fibromyalgia”\(^9\) FRIDAYYYYY → Fibromyalgia

(a) Try typing a few misspellings for each of the original words in Microsoft Word, Hunspell/ispell/aspell, or a similar word processing spelling checker. Are the suggested corrections similar to the ones made by the smartphone auto-correct program?

(b) What do you think might be happening to cause the smartphones to make such different corrections?