Assignment 2

L245
Writers’s aids
Due Thursday, September 19

1. Do question #2a in chapter 2 of the textbook (p. 65).

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>j</th>
<th>k</th>
<th>l</th>
</tr>
</thead>
<tbody>
<tr>
<td>j</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>k</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>l</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 textbook (p. 66).

   • **Bonus:** Do question #4 in chapter 2 of the textbook (p. 66).

4. Consider the misspelling *wierd*, and assume our edit distance calculations have insertions, deletions, substitutions, and transpositions. Describe how probabilities are used to rank *weird*, *wired*, and *wield* (and any other words you may think of). Do you have an intuition as to which should be highest, and why?

5. Pick a search engine of your choice, as long as the search engine provides spelling suggestions. Do two sets of searches, one based around the keywords *rocky* and *raccoon* and the other based around the keywords *stop* and *sign*.

   (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? Is there a difference between the two sets of keywords? 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 textbook (p. 66) ... but, instead of evaluating spelling errors, evaluate grammar errors.