de Marcken’s version of EM

John A Goldsmith
Expected counts (soft counts)

Let’s calculate the soft counts in a particular string of the words that happen to be in our lexicon. (“happen” here means that we will talk later about deciding which words should be there.)
A distribution

Counts summed to 1,000,000.

<table>
<thead>
<tr>
<th>word</th>
<th>count</th>
<th>frequency</th>
<th>plog</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>15 600</td>
<td>.0156</td>
<td>6</td>
</tr>
<tr>
<td>B</td>
<td>15 600</td>
<td>.0156</td>
<td>6</td>
</tr>
<tr>
<td>C</td>
<td>15 600</td>
<td>.0156</td>
<td>6</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HE</td>
<td>62 500</td>
<td>0.0625</td>
<td>4</td>
</tr>
<tr>
<td>HER</td>
<td>62 500</td>
<td>0.0625</td>
<td>4</td>
</tr>
<tr>
<td>THE</td>
<td>125 000</td>
<td>0.125</td>
<td>3</td>
</tr>
<tr>
<td>HERE</td>
<td>31 125</td>
<td>0.03125</td>
<td>5</td>
</tr>
<tr>
<td>THERE</td>
<td>31 125</td>
<td>0.03125</td>
<td>5</td>
</tr>
<tr>
<td>RENT</td>
<td>7 810</td>
<td>0.00781</td>
<td>7</td>
</tr>
<tr>
<td>IS</td>
<td>62 500</td>
<td>0.0625</td>
<td>4</td>
</tr>
<tr>
<td>TIS</td>
<td>3 906</td>
<td>0.003906</td>
<td>8</td>
</tr>
<tr>
<td>DUE</td>
<td>7 810</td>
<td>0.00781</td>
<td>7</td>
</tr>
</tbody>
</table>
Compute $\alpha$ (alpha)

<table>
<thead>
<tr>
<th>t</th>
<th>term 1</th>
<th>prob</th>
<th>term 2</th>
<th>prob</th>
<th>partial</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>T</td>
<td>0.015 6</td>
<td>6</td>
<td>0.015 6</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>T</td>
<td>H</td>
<td>0.015 6</td>
<td>6</td>
<td>.000 244</td>
<td>.000 244</td>
</tr>
<tr>
<td>3</td>
<td>TH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.000 244</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>THE</td>
<td>0.125</td>
<td>6</td>
<td>0.125</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>T</td>
<td>HE</td>
<td>0.062 5</td>
<td>5</td>
<td>0.000 975</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>TH</td>
<td>E</td>
<td>0.015 6</td>
<td>6</td>
<td>3.80 x $10^{-5}$</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>THE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.125 978</td>
</tr>
<tr>
<td>5</td>
<td>T</td>
<td>HER</td>
<td>0.062 5</td>
<td>5</td>
<td>0.000 937</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>TH</td>
<td>ER</td>
<td></td>
<td></td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>THE</td>
<td>R</td>
<td>0.015 6</td>
<td>6</td>
<td>0.001 96</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>THER</td>
<td></td>
<td></td>
<td></td>
<td>0.002 902</td>
<td></td>
</tr>
</tbody>
</table>
## Compute $\alpha$ (alpha)

<table>
<thead>
<tr>
<th>t</th>
<th>term 1</th>
<th>prob</th>
<th>term 2</th>
<th>prob</th>
<th>partial</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>T</td>
<td>0.015</td>
<td>HER</td>
<td>0.0625</td>
<td>0.000937</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>TH</td>
<td></td>
<td>ER</td>
<td></td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>THE</td>
<td>0.125</td>
<td>R</td>
<td>0.0156</td>
<td>0.00196</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>THER</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.002902</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>1.0</td>
<td>THERE</td>
<td>1</td>
<td>0.3125</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>T</td>
<td>0.0156</td>
<td>HERE</td>
<td>0.03125</td>
<td>0.000975</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>TH</td>
<td></td>
<td>ERE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>THE</td>
<td></td>
<td>RE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>THER</td>
<td>0.00290</td>
<td>E</td>
<td>0.0156</td>
<td>4.52 \times 10^{-4}</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>THERE</td>
<td></td>
<td></td>
<td></td>
<td>0.313</td>
<td></td>
</tr>
</tbody>
</table>
### Compute $\alpha$ (alpha)

<table>
<thead>
<tr>
<th>W t</th>
<th>term 1</th>
<th>prob</th>
<th>term 2</th>
<th>prob</th>
<th>partial</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>THERE</td>
<td>0.313</td>
<td>N</td>
<td>0.0156</td>
<td>0.00489</td>
</tr>
<tr>
<td></td>
<td>THEREN</td>
<td></td>
<td>total:</td>
<td>0.00489</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>THE</td>
<td>0.0029</td>
<td>RENT</td>
<td>0.00781</td>
<td>0.000015</td>
</tr>
<tr>
<td>W8</td>
<td>THEREN</td>
<td>0.00489</td>
<td>T</td>
<td>0.0156</td>
<td>0.000076</td>
</tr>
<tr>
<td></td>
<td>THERENT</td>
<td></td>
<td>total:</td>
<td>0.000091</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>THERENT</td>
<td>0.000091</td>
<td>I</td>
<td>0.0156</td>
<td>0.0000014</td>
</tr>
<tr>
<td></td>
<td>THERENTI</td>
<td></td>
<td>total:</td>
<td>0.0000014</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>THERENT</td>
<td>0.000091</td>
<td>IS</td>
<td>0.0625</td>
<td>0.000005688</td>
</tr>
<tr>
<td></td>
<td>THERENTI</td>
<td>0.0000014</td>
<td>S</td>
<td>0.0156</td>
<td>0.000000022</td>
</tr>
<tr>
<td></td>
<td>THERENTIS</td>
<td></td>
<td>total:</td>
<td>0.00000571</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>THERENTIS</td>
<td></td>
<td>D</td>
<td>0.000000089</td>
<td></td>
</tr>
<tr>
<td></td>
<td>THERENTISD</td>
<td></td>
<td>total:</td>
<td>0.000000089</td>
<td></td>
</tr>
</tbody>
</table>
Compute $\alpha$ (alpha)

<table>
<thead>
<tr>
<th>$t$</th>
<th>term 1</th>
<th>prob</th>
<th>term 2</th>
<th>prob</th>
<th>partial</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>THERENTIS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>THERENTISD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>D</td>
<td>0.0156</td>
<td>0.000 000 089</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>total:</td>
<td></td>
<td>0.000 000 089</td>
</tr>
<tr>
<td>12</td>
<td>THERENTISD</td>
<td>0.000 000 089</td>
<td>U</td>
<td>0.0156</td>
<td>1.38 x 10^{-9}</td>
</tr>
<tr>
<td></td>
<td>THERENTISDU</td>
<td></td>
<td>total:</td>
<td></td>
<td>1.38 x 10^{-9}</td>
</tr>
<tr>
<td>13</td>
<td>THERENTIS</td>
<td>0.00000571</td>
<td>DUE</td>
<td>0.00781</td>
<td>4.45 x 10^{-8}</td>
</tr>
<tr>
<td></td>
<td>THERENTISDU</td>
<td>1.38 x 10^{-9}</td>
<td>E</td>
<td>0.0156</td>
<td>2.15 * 10^{-8}</td>
</tr>
<tr>
<td></td>
<td>THERENTISDUE</td>
<td></td>
<td>total:</td>
<td></td>
<td>4.46 x 10^{-8}</td>
</tr>
</tbody>
</table>
Soft count of *due*

\[
\begin{align*}
\alpha(\text{therentis}) &= 0.000\,005\,71 \\
\beta(\text{due}) &= 0.007\,81 \\
4.46 \times 10^{-8} &
\end{align*}
\]
Soft count of is

\[\text{alpha}\text{(therent)} = 0.000\,091\]
\[\text{beta}\text{(due)} = 0.007\,81\text{ (sum of 0.007\,81 and 0.000\,000\,38)}\]
\[\text{pr}\text{(is)} = 0.062\,5\]
\[0.000091 \times 0.0625 \times 0.00781 = 4.44 \times 10^{-8}\]
\[\text{soft count of is} = \frac{4.44 \times 10^{-8}}{4.46 \times 10^{-8}}\]
\[4.46 \times 10^{-8} = 0.995\,51\]
Viterbi parse is different

<table>
<thead>
<tr>
<th>t</th>
<th>term 1</th>
<th>prob</th>
<th>term 2</th>
<th>prob</th>
<th>partial</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>T</td>
<td>0.015 6</td>
<td>0.015 6</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Best parse is just T</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>T</td>
<td>H</td>
<td>0.015 6</td>
<td>0.000 244</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Best parse is T-H</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>THE</td>
<td>0.125</td>
<td>0.125</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>T</td>
<td>HE</td>
<td>0.062 5</td>
<td>0.000 975</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>TH</td>
<td>E</td>
<td>0.015 6</td>
<td>3.80 x 10^{-5}</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Best parse is THE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>T</td>
<td>HER</td>
<td>0.062 5</td>
<td>0.000 937</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>TH</td>
<td>ER</td>
<td>0.015 6</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>THE</td>
<td>R</td>
<td>0.015 6</td>
<td>0.001 96</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Best parse is T-HER</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Viterbi 2

<table>
<thead>
<tr>
<th>t</th>
<th>term 1</th>
<th>prob</th>
<th>term 2</th>
<th>prob</th>
<th>partial</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>T</td>
<td>0.015</td>
<td>HER</td>
<td>0.062</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>TH</td>
<td></td>
<td>ER</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>THE</td>
<td>0.125</td>
<td>R</td>
<td>0.015</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Best parse is T-HER</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| 6 | 1.0 | THERE | 1 | 0.312 | 5 |
| 6 | T   | 0.015 | HERE | 0.031 | 25 | 0.000 | 975 |
| 6 | TH  |      | ERE  |       |     |       |     |
| 6 | THE |      | RE   |       |     |       |     |
| 6 | THER| 0.002 | E    | 0.015 | 6 | 4.52x10^{-1} |
| 6 | Best parse is THERE |</p>
<table>
<thead>
<tr>
<th>t</th>
<th>term 1</th>
<th>prob</th>
<th>term 2</th>
<th>prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>THERE</td>
<td>0.313</td>
<td>N</td>
<td>0.0156</td>
</tr>
<tr>
<td></td>
<td>Best parse is THERE-N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>THE</td>
<td>0.0029</td>
<td>RENT</td>
<td>0.0078</td>
</tr>
<tr>
<td></td>
<td>THEREN</td>
<td>0.0049</td>
<td>T</td>
<td>0.0156</td>
</tr>
<tr>
<td></td>
<td>Best parse is THE-RENT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>THERENT</td>
<td>0.0001</td>
<td>I</td>
<td>0.0156</td>
</tr>
<tr>
<td></td>
<td>THERENTI</td>
<td>0.000014</td>
<td>S</td>
<td>0.0156</td>
</tr>
<tr>
<td></td>
<td>Best parse is [THE-RENT]-I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>THERENT</td>
<td>0.0001</td>
<td>IS</td>
<td>0.0625</td>
</tr>
<tr>
<td></td>
<td>THERENTI</td>
<td>0.000014</td>
<td>S</td>
<td>0.0156</td>
</tr>
<tr>
<td></td>
<td>Best parse is [THE-RENT]-IS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>THERENTIS</td>
<td>D</td>
<td>0.000000089</td>
<td></td>
</tr>
<tr>
<td></td>
<td>THERENTISD</td>
<td></td>
<td></td>
<td>0.000000089</td>
</tr>
<tr>
<td>t</td>
<td>term 1</td>
<td>prob</td>
<td>term 2</td>
<td>prob</td>
</tr>
<tr>
<td>---</td>
<td>----------------------</td>
<td>---------</td>
<td>--------</td>
<td>---------</td>
</tr>
<tr>
<td>11</td>
<td>THERENTIS</td>
<td></td>
<td>D</td>
<td>0.0156</td>
</tr>
<tr>
<td></td>
<td>Best parse is [THE-RENT-IS]-D</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>THERENTISD</td>
<td>0.000000089</td>
<td>U</td>
<td>0.0156</td>
</tr>
<tr>
<td></td>
<td>Best parse is [THE-RENT-IS-D]-U</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>THERENTIS</td>
<td>0.00000571</td>
<td>DUE</td>
<td>0.00781</td>
</tr>
<tr>
<td></td>
<td>THERENTISD</td>
<td>1.38 x 10^{-9}</td>
<td>E</td>
<td>0.0156</td>
</tr>
<tr>
<td></td>
<td>Best parse is [THE-RENT-IS]-DUE</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>