The Phrase-Based Translation Model

- Simplest assumption is that all segmentations are equally likely, so there are no explicit probability terms for choice of segmentation.
- For any particular segmentations of \( o \) and \( s \) into \( K \) segments, the model of \( p(\bar{o}, \tau, \bar{s}) \) is:

\[
p(\bar{o}, \tau, \bar{s}) = p(\bar{o}, \tau|\bar{s}) \times p(\bar{s})
\]

\[
= \prod_{k=1}^{K} tr(\delta_{r(k)}|\bar{s}_k) \ d(\delta_{r(k-1)}, \delta_{r(k)}) \ LM(s)
\]

- **Phrase translation**: there’s a term \( tr(\delta_{r(k)}|\bar{s}_k) \) for the phrase-translation probabilities for an observed phrase \( \delta_{r(k)} \) to be generated from source phrase \( \bar{s}_k \)
- **Reordering**: there is re-ordering term concerning how likely the destination for a \( \bar{s}_k \) phrase is given destination for previous \( \bar{s}_{k-1} \) phrase: \( d(\delta_{r(k-1)}, \delta_{r(k)}) \)
- **Language model**: the probability of the source phrases \( \bar{s} \) is equated to simply the probability of the source sequence \( s \) as given by an \( n \)-gram model \( LM(s) \)
The Phrase Translation Model contd

- using (2), preferred translation $s_{\text{best}}$ is defined as the source part of $\langle \bar{s}, \tau \rangle_{\text{best}} = \arg \max_{\bar{s}, \tau} p(\bar{o}, \tau, \bar{s})$

- note you could seek to define $s_{\text{best}}$ as $s_{\text{best}} = \arg \max_{\bar{s}} \sum_\tau p(\bar{o}, \tau, \bar{s})$

but this is not standardly done; the above is regarded as a 'Viterbi' approximation of the sum

The distortion term

- The distortion term $d(\bar{o}_{\tau(k-1)}, \bar{o}_{\tau(k)})$ is not learned via EM
- instead just standardly defined as an exponentially decaying function of the 'distance' $x$ between end of $\bar{o}_{\tau(k-1)}$ and start of $\bar{o}_{\tau(k)}$
- in particular where $x = \lvert \text{fst}(\bar{o}_{\tau(k)}) - \text{lst}(\bar{o}_{\tau(k-1)}) - 1 \rvert$, the $d$ term is $\alpha^x$ for some $\alpha < 1$.  

illustration of Distance-Based Reordering

```
src   |   |   |   |
\downarrow
obs   | 1 2 3 | 4 5 | 6 7 |
```

<table>
<thead>
<tr>
<th>phrase</th>
<th>translated to</th>
<th>change $\bar{o}<em>{\tau(i-1)}$ to $\bar{o}</em>{\tau(i)}$</th>
<th>'displacement' $x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{o}_1$</td>
<td>$\bar{o}_1 : o_3$</td>
<td>start at beginning</td>
<td>0</td>
</tr>
<tr>
<td>$\bar{o}_2$</td>
<td>$\bar{o}_2 : o_6$</td>
<td>$o_3$ to $o_6$</td>
<td>$+2 = \text{fst}(o_6 : o_6) - \text{lst}(o_1 : o_3) - 1$</td>
</tr>
<tr>
<td>$\bar{o}_3$</td>
<td>$\bar{o}_3 : o_5$</td>
<td>$o_6$ to $o_4$</td>
<td>$-3 = \text{fst}(o_4 : o_5) - \text{lst}(o_6 : o_6) - 1$</td>
</tr>
<tr>
<td>$\bar{o}_4$</td>
<td>$\bar{o}_4 : o_7$</td>
<td>$o_5$ to $o_7$</td>
<td>$+1 = \text{fst}(o_7 : o_7) - \text{lst}(o_4 : o_5) - 1$</td>
</tr>
</tbody>
</table>

Scoring function: $\alpha^{|x|} =$ exponential with distance