Supplementary Information for CS4LL5

**IBM Model 1 formula**

\[ P(o, a, \ell_o | s) = \kappa(\ell_o, \ell_s) \times \prod_j [p(o_j | s_{a(j)})] \]  

(1)

where \( \kappa(\ell_o, \ell_s) \) is \( \frac{p(\ell_o | \ell_s)}{p(\ell_o)} \) and isolates a factor which is fixed for a given \( s \) and \( o \).

**Phrase-based formula**

\[ P(\bar{o}_{1:K}, \tau, \bar{s}_{1:K}) = \prod_{k=1}^{K} \left[ \text{tr}(\bar{o}_{\tau(k)}|\bar{s}_k)d(\bar{o}_{\tau(k-1)}, \bar{o}_{\tau(k)}) \right] \times \Phi_{LM}(s) \]  

(2)

where \( s \) is the word sequence underlying the phrase sequence \( \bar{s}_{1:K} \).

**Definitions for HMMs**

- "forward probability" \( \alpha_t(i) \)  
  \[ = P(o_1 \ldots o_t, s_t = i) \]  
  = joint prob of being in state \( i \) at time \( t \) and emitting the observation symbols \( o_1 \ldots o_t \)

- "backward probability" \( \beta_t(i) \)  
  \[ = P(o_{t+1} \ldots o_T|s_t = i) \]  
  = conditional prob of emitting the observation symbols \( o_{t+1} \ldots o_T \) given being in state \( i \) at time \( t \)

- \( \alpha_t(i)\beta_t(i) \)  
  \[ = P(o_1 \ldots o_t, s_t = i, o_{t+1} \ldots o_T) \]  
  = joint prob of being in state \( i \) at time \( t \) and emitting the observation symbols \( o_1 \ldots o_T \)