Обозначения:

1) Количество предложений в корпусе: R
2) Длина j-го исходного предложения: n_j
3) Длина j-го перевода: m_j
4) s^j — j-ое предложение (исходное)
5) t_i^j — i-ое слово j-го ~~предложения~~ перевода
6) s_k^j — k-ое слово j-го ~~перевода~~ исходного предл.
7) $q(A)$ — распределение на скрытых переменных
8) a_i^j — выравнивание (скрытая переменная) i-го слова в j-ом предложении-переводе

E-шаг: вычислим $q(A) = p(A \mid T, S)$

$$q(a_i^j = k) = p(a_i^j = k \mid t_i^j, s^j) =$$

$$= \frac{p(a_i^j = k, t_i^j \mid s^j)}{p(t_i^j \mid s^j)} = \frac{p(a_i^j = k, t_i^j \mid s^j)}{\sum_{l=1}^{n_j} p(a_i^j = l, t_i^j \mid s^j)} =$$

$= \{$учтём вид нашей модели$\} =$

$$= \frac{p(a_i^j = k)\, p(t_i^j \mid a_i^j = k, s^j)}{\sum_{l=1}^{n_j} p(a_i^j = l)\, p(t_i^j \mid a_i^j = l \mid s^j)} = \frac{\frac{1}{n_j}\theta(t_i^j \mid s_k^j)}{\sum_{l=1}^{n_j} \frac{1}{n_j}\theta(t_i^j \mid s_l^j)} =$$

$$= \frac{\theta(t_i^j \mid s_k^j)}{\sum_{l=1}^{n_j} \theta(t_i^j \mid s_l^j)}$$

Итак, $q(a_i^j = k) = \dfrac{\theta(t_i^j \mid s_k^j)}{\sum_{l=1}^{n_j} \theta(t_i^j \mid s_l^j)}$

Получим выражение для подсчёта <u>нижней оценки правдоподобия</u>:

$$L(q, S) = \sum_A q(A) \log \frac{p(T, A \mid S)}{q(A)} =$$

$$= \sum_{j=1}^{R} \sum_{i=1}^{m_j} \sum_{k=1}^{n_j} q(a_i^j = k) \cdot \log \frac{p(a_i^j = k, t_i^j \mid s^j)}{q(a_i^j = k)} =$$

$$= \sum_{j=1}^{R} \sum_{i=1}^{m_j} \sum_{k=1}^{n_j} q(a_i^j = k) \log \frac{\theta(t_i^j \mid s_k^j)}{n_j \cdot q(a_i^j = k)}$$

M-шаг: найдём аналитический максимум по параметрам.

$$L(q,S)=\sum_{j=1}^{R}\sum_{i=1}^{m_j}\sum_{k=1}^{n_j} q(a_i^j=k)\log\frac{\theta(t_i^j\mid s_k^j)}{n_j\cdot q(a_i^j=k)}=$$

$$=\sum_{j,i,k}\Big[q(a_i^j=k)\log\theta(t_i^j\mid s_k^j)-\underbrace{q(a_i^j=k)\log n_j q(a_i^j=k)}\Big]$$

не зависит от $\theta(y|x)$ $\Rightarrow$ можно отбросить при нахождении максимума по $\theta(y|x)$

Тогда, нужно $\sum\limits_{j,i,k} q(a_i^j=k)\log\theta(t_i^j\mid s_k^j)\to\max\limits_{\theta}$

Учтём следующее ограничение: $\sum\limits_{y}\theta(y|x)=1$, где суммирование идёт по всему множеству слов "y" из целевого языка. Причём равенство верно для любых слов "x" из исходного языка.

Обозначим функцию Лагранжа через $\mathcal{L}$. Тогда:

$$\mathcal{L}=\sum_{i,j,k} q(a_i^j=k)\log\theta(t_i^j\mid s_k^j)+\sum_x\lambda_x\Big(\sum_y\theta(y|x)-1\Big)$$

$$\begin{cases}\dfrac{\partial\mathcal{L}}{\partial\theta(y|x)}=\sum\limits_{i,j,k} q(a_i^j=k)\dfrac{1}{\theta(y|x)}[t_i^j=y][s_k^j=x]+\lambda_x=0\\[2ex]\dfrac{\partial\mathcal{L}}{\partial\lambda_x}=\sum\limits_y\theta(y|x)-1=0\end{cases}$$

Из первого уравнения системы $\Rightarrow$

$$\theta(y|x)=-\lambda_x^{-1}\sum_{i,j,k} q(a_i^j=k)[t_i^j=y][s_k^j=x]\quad(*)$$

Подставим это выражение для $\theta(y|x)$ во второе уравнение системы:

$$\sum_y \left(-\lambda_x^{-1}\right) \sum_{i,j,k} q(a_i^j = k)\,[t_i^j = y]\,[s_k^j = x] = 1$$

Поменяем порядок суммирования:

$$-\lambda_x = \sum_{i,j,k} q(a_i^j = k) \underbrace{\sum_y [t_i^j = y]}_{1} [s_k^j = x] =$$

$$= \sum_{i,j,k} q(a_i^j = k)\,[s_k^j = x]$$

Подставим выражение для $(-\lambda_x)$ в $(*)$:

$$\theta(y|x) = \frac{\sum_{i,j,k} q(a_i^j = k)\,[t_i^j = y]\,[s_k^j = x]}{\sum_{i,j,k} q(a_i^j = k)\,[s_k^j = x]}$$

Запишем выражение в полном виде:

$$\theta(y|x) = \frac{\sum_{j=1}^{R} \sum_{i=1}^{m_j} \sum_{k=1}^{n_j} q(a_i^j = k)\,[t_i^j = y]\,[s_k^j = x]}{\sum_{j=1}^{R} \sum_{i=1}^{m_j} \sum_{k=1}^{n_j} q(a_i^j = k)\,[s_k^j = x]}$$

Покажем, что данная точка максимум:

$$d^2 \mathcal{L} = -\sum_{i,j,k} \underbrace{q(a_i^j = k)}_{\geq 0} \underbrace{\frac{1}{\theta^2(y|x)}}_{\geq 0} \underbrace{[t_i^j = y]}_{\geq 0} \underbrace{[s_k^j = x]}_{\geq 0} \cdot$$

$\cdot\, d[\theta(y|x)]^2 \Rightarrow d^2\mathcal{L} < 0$, т.к. в данной

сумме встретится слагаемое, строго

больше нуля $\Rightarrow$ точка максимума.