$S = [S_1, \dots, S_R]$

$T = [T_1, \dots, T_R]$, $A = [A_1, \dots, A_R]$

$S_i = (s_1^i, \dots, s_{n_i}^i)$ – исходное предложение

$T_i = (t_1^i, \dots, t_{m_i}^i)$ – перевод , $i = \overline{1,R}$, R – число предложений

$A_i = (a_1^i, \dots, a_{m_i}^i)$ – выравнивания , $a_j^i \in \{1, \dots, n^i\}$ $\forall j$

Θ – матрица условных вероятностей $\in R^{k \times l}$

k – кол-во слов в исходном языке

l – кол-во слов в целевом языке

$\Theta_{ij} = p(y_j / x_i)$ – вер-ть того, что переводом слова x_i является y_j, $i = \overline{1,k}$, $j = \overline{1,l}$

$$P(A_k, T_k \mid S_k) = \prod_{i=1}^{m_k} p(a_i^k)\, p(t_i^k / a_i^k, S_k) = \prod_{i=1}^{m_k} \frac{1}{n_k} \Theta(t_i^k \mid S_{a_i^k}^k)$$

Нижняя оценка правдоподобия

$$\mathcal{L}(q,\theta) = \int q(A) \log \frac{P(A,T \mid S,\theta)}{q(A)} = \int q(A) \log p(A,T \mid S,\theta)\, dA - \int q(A) \log q(A)\, dA$$

$$1)\ \int q(A) \log p(A,T \mid S,\theta)\, dA = \int q(A) \log \prod_{k=1}^{R} p(A_k, T_k \mid S_k, \theta)\, dA =$$

$$= \sum_{k=1}^{R} \int q_k(A_k) \log p(A_k, T_k \mid S_k, \theta)\, dA_k = \sum_{k=1}^{R} \int q_k(A_k) \log \prod_{i=1}^{m_k} p(a_i^k)\, p(t_i^k \mid a_i^k, S_k)\, dA_k =$$

$$= \sum_{k=1}^{R} \sum_{i=1}^{m_k} \int q_{ki}(a_i^k) \log[p(a_i^k)\, p(t_i^k \mid a_i^k, S_k)]\, da_i^k = \sum_{k=1}^{R} \sum_{i=1}^{m_k} \int q_{ki}(a_i^k) \left[\log \frac{1}{n_k} + \log \Theta(t_i^k \mid S_{a_i^k}^k)\right] da_i^k$$

$$= -\sum_{k=1}^{R} m_k \log n_k \underbrace{\int q_{ki}(a_i^k)\, da_i^k}_{=1} + \sum_{k=1}^{R} \sum_{i=1}^{m_k} \int q_{ki}(a_i^k) \log \Theta(t_i^k \mid S_{a_i^k}^k)\, da_i^k = -\sum_{k=1}^{R} m_k \log n_k +$$

$$+ \sum_{k=1}^{R} \sum_{i=1}^{m_k} \sum_{j=1}^{n_k} q_{ki}(j) \log \Theta(t_i^k / S_j^k)$$

$$2)\ \int q(A) \log q(A)\, dA = \sum_{k=1}^{R} \int q_k(A_k) \log q(A_k)\, dA_k = \sum_{k=1}^{R} \sum_{i=1}^{m_k} \int q_{ki}(a_i^k) \log q_{ki}(a_i^k)\, da_i^k =$$

$$= \sum_{k=1}^{R} \sum_{i=1}^{m_k} \sum_{j=1}^{n_k} q_{ki}(j) \log q_{ki}(j)$$

$$\mathcal{L}(q,\theta) = \sum_{k=1}^{R} \sum_{i=1}^{m_k} \sum_{j=1}^{n_k} q_{ki}(j) \log \Theta(t_i^k / S_j^k) - \sum_{k=1}^{R} m_k \log n_k - \sum_{k=1}^{R} \sum_{i=1}^{m_k} \sum_{j=1}^{n_k} q_{ki}(j) \log q_{ki}(j)$$

E-step

$$q^*(A) = p(A \mid T, S, \theta) \quad ; \qquad q(A) = [q_1(A_1), \dots, q_R(A_R)]$$

$$q_k(A_k) = [q_{k1}(a_1^k), \dots, q_{km_k}(a_{m_k}^k)], \quad a_j^k \in \{1, \dots, n_i\}$$

$$q_{ki}(a_i^k) = p(a_i^k \mid t_i^k, S_k, \theta) = \frac{p(a_i^k, t_i^k \mid S_k, \theta)}{p(t_i^k \mid S_k, \theta)} = \frac{p(a_i^k)\, p(t_i^k \mid a_i^k, S_k, \theta)}{\sum_{j=1}^{n_k} p(j, t_i^k \mid S_k, \theta)} =$$

$$= \frac{p(a_i^k)\, p(t_i^k \mid a_i^k, S_k, \theta)}{\sum_{j=1}^{n_k} p(j)\, p(t_i^k \mid j, S_k, \theta)} = \left\{p(j) = \frac{1}{n_k}\ \forall j\right\} = \frac{\Theta(t_i^k \mid S_{a_i^k}^k)}{\sum_{j=1}^{n_k} \Theta(t_i^k \mid S_j^k)}$$

M-step

$$\theta^* = \underset{\theta}{\operatorname{argmax}}\, \mathcal{L}(q, \theta)$$

$$\mathcal{L}(q,\theta) = \sum_{k=1}^{R}\sum_{i=1}^{m_k}\sum_{j=1}^{n_k} q_{ki}(j)\log\theta(t_i^k | S_j^k) - \sum_{k=1}^{R} m_k \log n_k - \sum_{k=1}^{R}\sum_{i=1}^{m_k}\sum_{j=1}^{n_k} q_{ki}(j)\log q_{ki}(j) =$$

$$= \sum_{k=1}^{R}\sum_{i=1}^{m_k}\sum_{j=1}^{n_k} q_{ki}(j)\log\theta(t_i^k | S_j^k) + \text{const}$$

$$\theta \in R^{b\times c}, \quad \sum_{j=1}^{c}\theta_{ij} = 1, \ \forall i \in [1,b]$$

$$\theta_{ij} \geq 0, \ \forall i \in [1,b], \forall j \in [1,c]$$

Основная задача

$$\begin{cases} -\sum_{k=1}^{R}\sum_{i=1}^{m_k}\sum_{j=1}^{n_k} q_{ki}(j)\log\theta(t_i^k | S_j^k) \to \min_\theta \\ \sum_{j=1}^{c}\theta_{ij} = 1, \ \forall i \in [1,b] \\ -\theta_{ij} \leq 0, \ \forall i \in [1,b], \forall j \in [1,c] \end{cases}$$

$$L(\theta,\lambda) = -\sum_{k=1}^{R}\sum_{i=1}^{m_k}\sum_{j=1}^{n_k} q_{ki}(j)\log\theta(t_i^k | S_j^k) - \sum_{i=1}^{b}\sum_{j=1}^{c}\lambda_{ij}\theta_{ij} + \sum_{i=1}^{b}\nu_i\Big(\sum_{j=1}^{c}\theta_{ij} - 1\Big)$$

$$\frac{dL(\theta,\lambda)}{d\theta_{hg}} = -\sum_{k=1}^{R}\sum_{i=1}^{m_k}\sum_{j=1}^{n_k} q_{ki}(j)\cdot\frac{1}{\theta_{hg}}[t_i^k == g][S_j^k == h] - \lambda_{hg} + \nu_h$$

Условия Куна-Таккера:

$$\begin{cases} -\sum_{k=1}^{R}\sum_{i=1}^{m_k}\sum_{j=1}^{n_k} q_{ki}(j)\cdot\frac{1}{\theta_{hg}}[t_i^k == g][S_j^k == h] - \lambda_{hg} + \lambda_h = 0 \quad (1), \ h,g \in [1,b]\times[1,c] \\ -\theta_{ij} \leq 0, \forall i \in [1,b], \forall j \in [1,c]; \quad \sum_{j=1}^{c}\theta_{ij} = 1, \forall i \in [1,b] \\ \lambda_{ij} \geq 0, \ \forall i \in [1,b], \forall j \in [1,c] \\ \lambda_{ij}\theta_{ij} = 0, \forall i \in [1,b], \forall j \in [1,c] \quad (2) \end{cases}$$

Из (1) и (2) $\Rightarrow$

$$\begin{cases} \lambda_{hg} = 0 \\ -\frac{1}{\theta_{hg}}\sum_{k=1}^{R}\sum_{i=1}^{m_k}\sum_{j=1}^{n_k} q_{ki}(j)[t_i^k == g][S_j^k == h] - \lambda_{hg} + \lambda_h = 0 \end{cases} \quad \forall h,g \in [1,b]\times[1,c]$$

$$\Rightarrow \theta_{hg} = \frac{1}{\lambda_h}\sum_{k=1}^{R}\sum_{i=1}^{m_k}\sum_{j=1}^{n_k} q_{ki}(j)[t_i^k == g][S_j^k == h]$$

$$\sum_{f=1}^{C} \theta_{hf} = \sum_{f=1}^{C} \frac{1}{\lambda_h} \sum_{K=1}^{R} \sum_{i=1}^{m_K} \sum_{j=1}^{n_K} q_{x_i}(j) \, [t_i^K == f] [s_j^K == h] = 1$$

$$\lambda_h = \sum_{f=1}^{C} \sum_{K=1}^{R} \sum_{i=1}^{m_K} \sum_{j=1}^{n_K} q_{x_i}(j) \, [t_i^K == f] [s_j^K == h]$$

$$\Rightarrow \theta_{hg}^{*} = \frac{\sum_{K=1}^{R} \sum_{i=1}^{m_K} \sum_{j=1}^{n_K} q_{x_i}(j) \, [t_i^K == g] [s_j^K == h]}{\sum_{f=1}^{C} \sum_{K=1}^{R} \sum_{i=1}^{m_K} \sum_{j=1}^{n_K} q_{x_i}(j) \, [t_i^K == f] [s_j^K == h]}$$