IBM Model 1

$\bar{a} = (a_1, \ldots, a_n)$ $\bar{t} = (t_1, \ldots, t_n)$ $\bar{s} = (s_1, \ldots, s_m, \text{NULL})$

each word in $\bar{t}$ aligns to one word in $\bar{s}$

placeholder for unaligned words

$$P(\bar{t}, \bar{a} | \bar{s}) = P(\bar{a}) \, P(\bar{t} | \bar{a}, \bar{s}) = \prod_{i=1}^{n} P(a_i) \, P(t_i | s_{a_i})$$

Model params: translation prob. matrix

includes NULL → $|V_s| \times |V_t|$ $P(\text{target} | \text{source})$

$P(t_i | s_{a_i})$: look up the prob of t_i given source word s_{a_i} a_i is a "pointer"

$P(a_i)$ = uniform over $\{1, \ldots, m, \text{NULL}\}$ $\frac{1}{m+1}$

Inference in Model 1

$$P(\bar{a} \mid \bar{s}, \bar{t}) = \frac{P(\bar{a}, \bar{t} \mid \bar{s})}{P(\bar{t} \mid \bar{s})} = \frac{\prod_{i=1}^{n} P(a_i)^{\frac{1}{m+1}} P(t_i \mid s_{a_i})}{\cancel{P(\bar{t} \mid \bar{s})}}$$

ignore

$$P(\bar{a} \mid \bar{s}, \bar{t}) \propto \prod_i P(t_i \mid s_{a_i})$$

$$P(a_i \mid \bar{s}, \bar{t}) \propto P(t_i \mid s_{a_i})$$

$\bar{s}$		I	like	eat	$\bar{t}$
	Je	0.8	0.1	0.1	
	J'	0.8	0.1	0.1	
	mange	0	0	1.0	
	aime	0	1.0	0	
	NULL	0.4	0.3	0.4	

J' aime NULL

I like

$P(a_1 \mid \bar{s}, \bar{t})$

$$\propto \begin{cases} 0.8 & J' \\ 0 & aime \\ 0.4 & NULL \end{cases} \rightarrow \begin{matrix} 2/3 & J' \\ 0 & aime \\ 1/3 & NULL \end{matrix}$$

HMM Model (Vogel, 1996)

$$P(\bar{a}) = \prod_{i=1}^{n} P(a_i \mid a_{i-1})$$

" Categorical $(a_i - a_{i-1})$

" moving the alignment pointer by +1 is most likely "

Learning No labeled alignment data

Optimize : $\sum_{i=1}^{D} \log \underbrace{P(\bar{f} \mid \bar{s})}_{\text{model: } P(\bar{f}, \bar{a} \mid \bar{s})} = \sum_{i=1}^{D} \log \overline{\sum_{\bar{a}} P(\bar{f}, \bar{a} \mid \bar{s})}$

marginal log likelihood

Expectation maximization (EM)