Q.1 a) i) $B=y$ is independent of $A=y$ if
$P(B=y|A=y) != P(B=y)$

$P(B=y|A=y) = \frac{\#B=y, A=y}{\#A=y} = \frac{3}{4}$ $P(B=y) = \frac{\#B=y}{\#d} = \frac{4}{6} = \frac{2}{3}$

$P(B=y|A=y) != P(B=y)$ and therefore $B=y$ is not independent of $A=y$.

ii) $\delta =$

d	A	B
1	y	y
2	y	n
3	n	y
4	n	n

$P(B=y|A=y) = \frac{1}{2}$ $P(B=y) = \frac{2}{4} = \frac{1}{2}$

$P(B=y|A=y) = P(B=y)$ and $B=y$ is therefor independent of $A=y$ for δ.

b) i) $\frac{dL(x)}{dx} = \frac{\#h}{x} - \frac{\#t}{1-x}$

Let $\frac{\#h}{x} - \frac{\#t}{1-x} = 0$

$\frac{\#h}{x} = \frac{\#t}{1-x}$

$\#h - \#hx = \#tx$

let $\#t = D - \#h$

$\#h - \#hx = x(D - \#h)$

$\#h - \#hx = Dx - \#hx$

$\#h = Dx$

$\therefore x = \frac{\#h}{D}$ when $\frac{dL(x)}{dx} = 0$

ii) The above proves this to be the case as by differentiating with respect to x and find the value of x that makes this 0, we are finding the value for x (the probability of h) that is most likely, the above showed this to be $\frac{\#h}{D}$, which is the relative frequency, and is therefor the maximum likelihood estimator, while this is for $L(x)$, the value that maximizes $L(x)$ is the same one that maximizes $P(C^{1:D})$.

c) i) $\gamma_d = P(S=b \mid V^d)$

$$= \frac{P(S=b)\,P(V^d \mid S=b)}{P(V^d)}$$

Let $P(V^d) = P(S=b \mid V^d) \cup P(S=w \mid V^d) \cup P(S=g \mid V^d)$

$$\gamma_d(b) = \frac{P(S=b)\,P(V^d \mid S=b)}{P(S=b \mid V^d) + P(S=w \mid V^d) + P(S=g \mid V^d)}$$

$$\gamma_d(b) = \frac{P(S=b)\,P(V^d \mid S=b)}{(P(S=b)\,P(V^d \mid S=b)) + (P(S=w)\,P(V^d \mid S=b)) + (P(S=g)\,P(V^d \mid S=b))}$$

Let $P(S=b) = \Theta(b)$, $P(S=w) = \Theta(w)$; $P(S=g) = \Theta(g)$

$$\gamma_d(b) = \frac{\Theta(b) \times P(S=b \mid V^d)}{(\Theta(b) \times P(S=b \mid V^d)) + (\Theta(w) \times P(S=w \mid V^d)) + (\Theta(g) + P(S=g \mid V^d))}$$

ii) $\cancel{P(V^d \mid S=b) = \prod_{v \in V^d} (\Psi[v,b])}$

$$P(V^d \mid S=b) = \prod_{v \in V^d} (\Psi[v,b])$$

iii) $$\Theta[b] = \frac{\sum_{d \in D}[\gamma_d(b)]}{\sum_{d \in D}[\gamma_d(b) + \gamma_d(w) + \gamma_d(g)]} = \frac{\sum_{d \in D}[\gamma_d(b)]}{\sum_{d \in D}[1]} = \frac{\sum_{d \in D}[\gamma_d(b)]}{D}$$

$\Theta(b)$ is the sum of all $\gamma_d(b)$'s in the range D, divided by all the probabilities in each d added together, however, for each d, the sum of these γ's will equal 1, meaning we are summing 1 D times, which is just D.

$$\Psi[b,v] = \frac{\sum_{d \in D} \gamma_d(b)\,\#(d,v)}{|V^d| \sum_d \gamma_d(b)}$$

~~As #(d,v) is either 0 or 1, the numerator sums the $\gamma_d(b)$ of all d where the v we are concerned with occurs, and the sum of the counts of the virtual corpus multiplied~~

The numerator is the sum of all ~~datapoints where~~ virtual coun times the occurence of the sound v, and the denomenator is the sum of all virtual counts of b, times the ~~size~~ number of sequences.

3. ~~v. $P(V^d, \Theta, \Psi) = \prod_d f$~~

~~v $P(V^d, \Theta, \Psi) = \Theta_b \Theta_{v|b}^{\#(d,v)} (1-\Theta_{v|b})^{|V^d| - \#(d,v)}$~~

~~$+ \Theta_w \Theta_{v|w}^{\#(d,v)} (1-\Theta_{v|w})^{|V^d| - \#(d,v)} + \Theta_g \Theta_{v|g}^{\#(d,v)} (1-\Theta_{v|g})^{|V^d| - \#(d,v)}$~~

$$v.\ P(V^d; \Theta, \Psi) = \Theta_b \Theta_{v|b}^{\#(d,v)} (1-\Theta_{v|b})^{|V^d| - \#(d,v)} +$$
$$\Theta_w \Theta_{v|w}^{\#(d,v)} (1-\Theta_{v|w})^{|V^d| - \#(d,v)} + \Theta_g \Theta_{v|g}^{\#(d,v)} (1-\Theta_{v|g})^{|V^d| - \#(d,v)}$$

~~Becaug~~ Because of the addition of the sub-expression you cannot break them apart with logs, making the calculus almost impossible.

2. a) i) The number of potential alignments is $l_s^{l_o}$, ~~which is expont~~ so by using brute force through all alignments, ~~we are to~~ ours algorithm is going to ~~reach~~ be exponentially dependent on l_o.

ii. $\left(\frac{\text{paisti, kids}}{\text{paisti}}\right)$ (number of times of alignments)

$$\frac{\frac{1}{2}}{\frac{1}{2}+\frac{1}{4}}(1)$$

$$= \frac{2}{3}$$

~~iii.~~

ii.	the	kids
na	$\frac{1}{2}/\frac{1}{2}+\frac{1}{2}$	$\frac{1}{2}/\frac{1}{2}+\frac{1}{2}$
paistí	$\frac{1}{4}/\frac{1}{4}+\frac{1}{2}$	$\frac{1}{2}/\frac{1}{4}+\frac{1}{2}$

γ_d	the	kids
na	$\frac{1}{2}$	$\frac{1}{2}$
paístí	$\frac{1}{3}$	$\frac{2}{3}$

E_1	the	kids
na	$\frac{1}{2}$	$\frac{1}{2}$
paístí	$\frac{1}{3}$	$\frac{2}{3}$

iv)

γ_1	the	kids
na	$\frac{1}{2}/\frac{1}{2}+\frac{1}{2}$	$\frac{1}{2}/\frac{1}{2}+\frac{1}{2}$
paistí	$\frac{1}{4}/\frac{1}{4}+\frac{1}{2}$	$\frac{1}{2}/\frac{1}{4}+\frac{1}{2}$

γ_2	the	pets
na	$\frac{1}{2}/\frac{1}{2}+\frac{1}{2}$	$\frac{1}{2}/\frac{1}{2}+\frac{1}{2}$
peataí	$\frac{1}{4}/\frac{1}{2}+\frac{1}{4}$	$\frac{1}{2}/\frac{1}{2}+\frac{1}{4}$

γ_1	the	kids
na	$\frac{1}{2}$	$\frac{1}{2}$
paistí	$\frac{1}{3}$	$\frac{2}{3}$

γ_2	the	pets
na	$\frac{1}{2}$	$\frac{1}{2}$
peataí	$\frac{1}{3}$	$\frac{2}{3}$

v)

$E(o,s)$	the	kids	pets
na	$\frac{1}{2}+\frac{1}{2}=1$	$\frac{1}{2}$	$\frac{1}{2}$
paistí	$\frac{1}{3}$	$\frac{2}{3}$	0
peataí	$\frac{1}{3}$	0	$\frac{2}{3}$

To get $tr(o|s)$, normalize each element by the sum of the counts in their column.

$tr(o|s)$	the	kids	pets
na	$\frac{3}{5}$	$\frac{3}{7}$	$\frac{3}{7}$
paistí	$\frac{1}{5}$	$\frac{4}{7}$	0
peataí	$\frac{1}{5}$	0	$\frac{4}{7}$

b)