3(a) $\dfrac{\partial(CE(\mathbf{y},\widehat{\mathbf{y}}))}{\partial \mathbf{v_c}}$

=	$-\dfrac{\partial}{\partial \mathbf{v_c}}\sum_{w=1}^{W}\left(y_w \log(\widehat{y_w})\right)$	(1)
=	$-\dfrac{\partial}{\partial \mathbf{v_c}}\left(\left(y_1 \log \widehat{y_1}\right)+\left(y_2 \log \widehat{y_2}\right)+\ldots+\left(y_o \log \widehat{y_o}\right)+\ldots+\left(y_W \log \widehat{y_W}\right)\right)$	(2)
=	$-\dfrac{\partial}{\partial \mathbf{v_c}}\left(\left(y_1 \log\left(\dfrac{\exp(\mathbf{u_1}^{\mathbf{T}}\mathbf{v_c})}{\sum_{m=1}^{m=W}\exp(\mathbf{u_m}^{\mathbf{T}}\mathbf{v_c})}\right)\right)+\ldots+\left(y_o \log\left(\dfrac{\exp(\mathbf{u_o}^{\mathbf{T}}\mathbf{v_c})}{\sum_{m=1}^{m=W}\exp(\mathbf{u_m}^{\mathbf{T}}\mathbf{v_c})}\right)\right)+\ldots+\left(y_W \log\left(\dfrac{\exp(\mathbf{u_W}^{\mathbf{T}}\mathbf{v_c})}{\sum_{m=1}^{m=W}\exp(\mathbf{u_m}^{\mathbf{T}}\mathbf{v_c})}\right)\right)\right)$	(3)

If we look at the i^{th} term of Equation (3), it is of the form below:

=	$-\dfrac{\partial}{\partial \mathbf{v_c}}\left(y_i \log\left(\dfrac{\exp(\mathbf{u_i}^{\mathbf{T}}\mathbf{v_c})}{\sum_{m=1}^{m=W}\exp(\mathbf{u_m}^{\mathbf{T}}\mathbf{v_c})}\right)\right)$	(4)
=	$-\dfrac{\partial}{\partial \boldsymbol{v_c}}\left(y_i\left(\log\left(\exp(\mathbf{u_i}^{\mathbf{T}}\mathbf{v_c})\right)-\log\sum_{m=1}^{m=W}\exp(\mathbf{u_m}^{\mathbf{T}}\mathbf{v_c})\right)\right)$	(5)
=	$-y_i\left(\dfrac{\partial\left(\mathbf{u_i}^{\mathbf{T}}\mathbf{v_c}\right)}{\partial \mathbf{v_c}}-\dfrac{\partial}{\partial \mathbf{v_c}}\left(\log\sum_{m=1}^{m=W}\exp(\mathbf{u_m}^{\mathbf{T}}\mathbf{v_c})\right)\right)$	(6)
=	$-y_i\left(\mathbf{u_i}-\left(\dfrac{1}{\sum_{m=1}^{m=W}\exp(\mathbf{u_m}^{\mathbf{T}}\mathbf{v_c})}\right)\dfrac{\partial}{\partial \mathbf{v_c}}\left(\sum_{x=1}^{x=W}\exp(\mathbf{u_x}^{\mathbf{T}}\mathbf{v_c})\right)\right)$	(7)
=	$-y_i\left(\mathbf{u_i}-\sum_{x=1}^{x=W}\dfrac{\frac{\partial}{\partial \mathbf{v_c}}\left(\exp(\mathbf{u_x}^{\mathbf{T}}\mathbf{v_c})\right)}{\sum_{m=1}^{m=W}\exp(\mathbf{u_m}^{\mathbf{T}}\mathbf{v_c})}\right)$	(8)
=	$-y_i\left(\mathbf{u_i}-\sum_{x=1}^{x=W}\dfrac{\exp(\mathbf{u_x}^{\mathbf{T}}\mathbf{v_c})\mathbf{u_x}}{\sum_{m=1}^{m=W}\exp(\mathbf{u_m}^{\mathbf{T}}\mathbf{v_c})}\right)$	(9)

=	$-y_i\left(\mathbf{u}_{\boldsymbol{i}} - \sum_{x=1}^{x=W} \hat{y}_x \mathbf{u}_{\boldsymbol{x}}\right)$	(10)

Substituting in Equation (3):

$\frac{\partial(CE(\mathbf{y},\hat{\mathbf{y}}))}{\partial \mathbf{v_c}} =$	$-y_1\left(\mathbf{u}_{\boldsymbol{1}} - \sum_{x=1}^{x=W} \hat{y}_x \mathbf{u}_{\boldsymbol{x}}\right) - \ldots - y_o\left(\mathbf{u}_{\boldsymbol{o}} - \sum_{x=1}^{x=W} \hat{y}_x \mathbf{u}_{\boldsymbol{x}}\right) - \ldots - y_W\left(\mathbf{u}_{\boldsymbol{W}} - \sum_{x=1}^{x=W} \hat{y}_x \mathbf{u}_{\boldsymbol{x}}\right)$	(11)
=	$y_1\left(-\mathbf{u}_{\boldsymbol{1}} + \sum_{x=1}^{x=W} \hat{y}_x \mathbf{u}_{\boldsymbol{x}}\right) + \ldots + y_o\left(-\mathbf{u}_{\boldsymbol{o}} + \sum_{x=1}^{x=W} \hat{y}_x \mathbf{u}_{\boldsymbol{x}}\right) + \ldots + y_W\left(-\mathbf{u}_{\boldsymbol{W}} + \sum_{x=1}^{x=W} \hat{y}_x \mathbf{u}_{\boldsymbol{x}}\right)$	(12)

Since only **o** is the expected word (only $y_o = 1$), Equation (12) becomes:

=	$-\mathbf{u}_{\boldsymbol{o}} + \sum_{x=1}^{x=W} \hat{y}_x \mathbf{u}_{\boldsymbol{x}}$	(13)