Considering Named Entity Recognition with max-margin loss function:

$$J = max(1 + s_c - s, 0)$$

where $s = U^T f(z) = U^T f(Wx + b)$ and $s_c = U^T f(Wx_c + b)$;
$x \in \mathbb{R}^{n\times1}, W \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^{m\times1}$.

$$a = f(Wx + b) = f(z) = [f(z_1)...f(z_m)]^T$$

Assume $J = 1 + s_c - s$.

$$\frac{\partial J}{\partial W} = \frac{\partial J}{\partial s}\frac{\partial s}{\partial W}$$

$$\frac{\partial J}{\partial s} = -\frac{\partial J}{\partial s_c} = -1$$

$$\frac{\partial s}{\partial W_{ij}} = \frac{\partial(U^T f(Wx + b))}{\partial W_{ij}} = \frac{\partial(\sum_{k=1}^m U_k a_k)}{\partial W_{ij}} = U_i\frac{\partial a_i}{\partial W_{ij}} = U_i\frac{\partial a_i}{\partial z_i}\frac{\partial z_i}{\partial W_{ij}}$$

$$z_i = \sum_{k=1}^n W_{ik}x_k \Rightarrow \frac{\partial z_i}{\partial W_{ij}} = x_j$$

So $\frac{\partial s}{\partial W_{ij}} = U_i f'(z_i)x_j \Rightarrow \frac{\partial s}{\partial W} = diag(f'(z))Ux^T$.