$$\mathcal{L}_{supervised} = -\mathbb{E}_{x,y \sim p_{data}(x,y)} \log p_{model}(y|x, y < n+1)$$