

$$\dots + R(y_5) \nabla \ln p(\text{an} | y_{<5}, x) + R(y_6) \nabla \ln p(\text{American} | y_{<6}, x) + R(y_7) \nabla \ln p(\text{computer} | y_{<7}, x) + \dots$$

Decoder
Hidden
States

