

$$p(w_t | w_{1:t-1})$$

$$p(w_{t+1} | w_{1:t})$$

$$p(w_{t+2} | w_{1:t+1})$$

next-utterance
prediction

context
representation

