

Losses computed during training

**Prefix
tokens**

\hat{s}_1 \hat{s}_2 \hat{s}_3 \hat{s}_4 \cdots \hat{s}_l EOS

Encoder



Decoder

x_1 x_2 \cdots x_n

BOS s_1 s_2 s_3 s_4 \cdots s_l

Tokens given during decoding