

## A Supplementary Material

---

### Algorithm 1 AGENT with prediction mechanism

---

**Require:** NMT system, Policy  $\pi_\theta$ ,  $m_{\text{MAX}}$ , input buffer  $X$ , output buffer  $Y$ .

```

1: Init  $x_1 \leftarrow X, h_1 \leftarrow f_{\text{ENC}}(x_1), H_1^0 \leftarrow \{h_1\},$ 
    $z_0 \leftarrow f_{\text{INIT}}(H_1^0), w_0 \leftarrow \langle s \rangle, m \leftarrow 0, n \leftarrow 1, \rho \leftarrow 0,$ 
    $t \leftarrow 0$ 
2: while  $m < m_{\text{MAX}}$  do
3:    $t \leftarrow t + 1$ 
4:    $y_t, s_t, o_t \leftarrow f_{\text{DEC}}(z_m, w_m, H_n^\rho)$ 
5:    $a_t \sim \pi_\theta(a_t; a_{<t}, o_{<t})$ 
6:   if  $a_t = \text{READ}$  and  $x_n \neq \langle /s \rangle$  then
7:      $n \leftarrow n + 1, \rho \leftarrow 0$ 
8:      $x_n \leftarrow X, h_n \leftarrow f_{\text{ENC}}(h_{n-1}, x_n)$ 
9:      $H_n^\rho \leftarrow H_{n-1}^\rho \cup \{h_n\}$ 
10:    if  $|Y| = 0$  then
11:       $z_0 \leftarrow f_{\text{INIT}}(H_n^\rho)$ 
12:    else if  $a_t = \text{WRITE}$  then
13:       $m \leftarrow m + 1$ 
14:       $z_m \leftarrow s_t, w_m \leftarrow y_t$ 
15:       $Y \leftarrow w_m$ 
16:      if  $w_m = \langle /s \rangle$  then
17:        break
18:    else if  $a_t = \text{PREDICT}$  then
19:       $\text{pw} \leftarrow x_n$ 
20:      if  $\rho > 0$  then
21:         $\text{pw} \leftarrow x'_\rho$ 
22:       $\rho \leftarrow \rho + 1, x'_\rho \leftarrow f_{\text{PRED}}(\text{pw})$ 
23:       $h_{n+\rho} \leftarrow f_{\text{ENC}}(h_{n+\rho-1}, x'_\rho)$ 
24:       $H_n^\rho \leftarrow H_n^{\rho-1} \cup \{h_{n+\rho}\}$ 

```

---

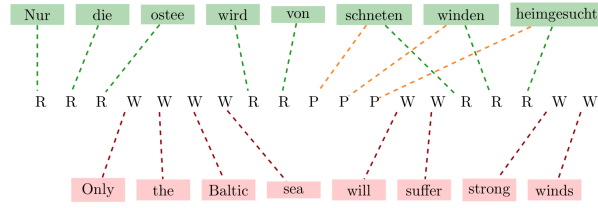


Figure 5: A sample translation scenario with an action sequence. After reading "von" the AGENT decides to predict and after three predicts, "will" is written to the output buffer.

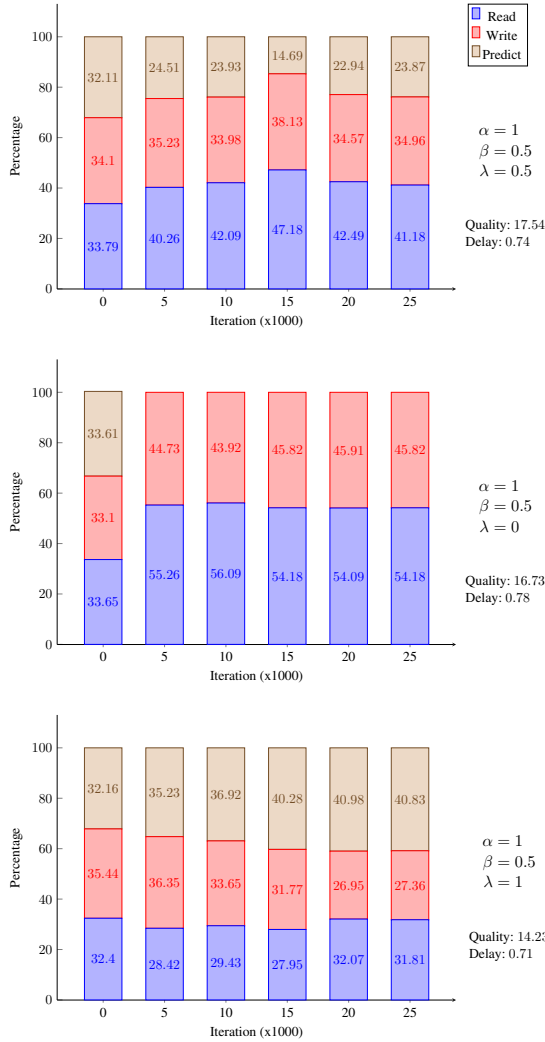


Figure 6: Action distribution in the first 25000 iterations for varying  $\lambda$  values. The numbers in each bar are the average action percentage over the previous 5000 iterations. All these graphs are for English to German translation; However, very similar behaviour was also observed for German to English translation.