Algorithm 1 Dynamic Programming (DP) for Exact Marginalization

Input: $\boldsymbol{y}$ is a sequence of T characters, V is a subword vocabulary, m is the maximum subword length

Output: $\log p(\boldsymbol{y})$ marginalizing out different subword segmentations.

$\alpha_0 \leftarrow 0$

for $k = 1$ **to** T **do**

  $\alpha_k \leftarrow \log \sum_{j=k-m}^{k-1} \mathbb{1}[\boldsymbol{y}_{j,k} \in V] \exp\left(\alpha_j + \log P_\theta(\boldsymbol{y}_{j,k} | y_1, .., y_j)\right)$

end for

return α_T ▷ the marginal probability $\log p(\boldsymbol{y}) = \log \sum_{\boldsymbol{z} \in \mathcal{Z}_y} p(\boldsymbol{y}, \boldsymbol{z})$