Algorithm 2 Dynamic Programming Encoding (DPE) for Subword Segmentation

Input: $\boldsymbol{y}$ is a sequence of T characters, V is a subword vocabulary, m is the maximum subword length
Output: Segmentation $\boldsymbol{z}$ with highest posterior probability.

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

$\beta_k \leftarrow \max_{\{j \in [k-m,k-1] \,|\, \boldsymbol{y}_{j,k} \in V\}} \beta_j + \log P_\theta(\boldsymbol{y}_{j,k} | y_1, .., y_j)$

$b_k \leftarrow \text{argmax}_{\{j \in [k-m,k-1] \,|\, \boldsymbol{y}_{j,k} \in V\}} \beta_j + \log P_\theta(\boldsymbol{y}_{j,k} | y_1, .., y_j)$

end for

$\boldsymbol{z} \leftarrow \text{backtrace}(b_1, .., b_T)$ ▷ backtrace the best segmentation using $\boldsymbol{b}$