Input:

1. α: 1×3, operation hyperparameter
2. ρ: $1 \times k$, insert bottom hyperparameter
3. ϕ: $k \times k$, substitution hyperparameter
4. T: set of top-level strings

Algorithm 1: noisy channel generative model

draw the operation probability $\beta \sim Dir(\alpha)$
draw the insert bottom probability $\omega \sim Dir(\rho)$
for $i \in \{1, \ldots, k\}$ **do**
  draw $\lambda_i \sim Dir(\phi_i)$
for *utterance* $u \in T$ **do**
  for *top PLU* $t \in u$ **do**
    draw an operation $o \sim Cat(\beta)$
    if *o is insert bottom* **then**
      draw bottom-level PLU $x \sim Cat(\omega)$
      do not advance t
    if *o is substitute* **then**
      draw bottom-level PLU $x \sim Cat(\lambda_t)$
    if *o is insert top* **then**
      do nothing