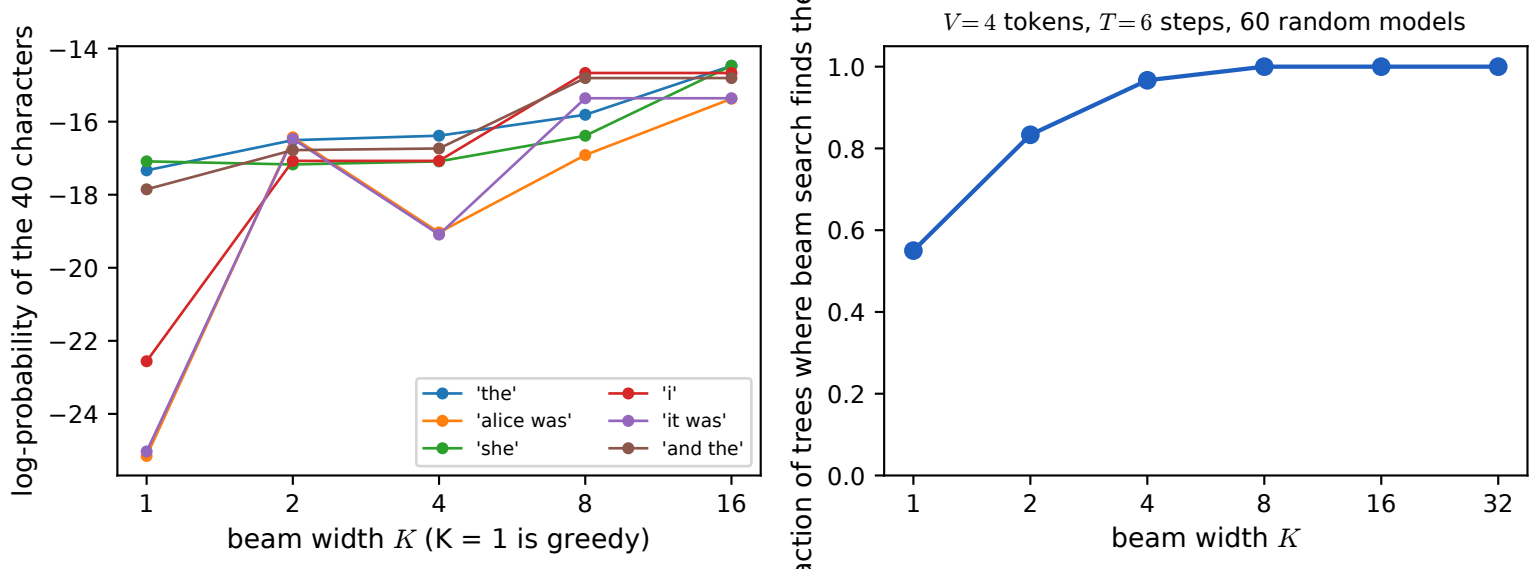

log-probability of the 40 characters
−14
−16
−18
−20
−22
−24
1
2
4
8
16
beam width K (K = 1 is greedy)
'the'
'alice was'
'she'
'i'
'it was'
'and the'
$V = 4$ tokens, $T = 6$ steps, 60 random models
fraction of trees where beam search finds the
0.0
0.2
0.4
0.6
0.8
1.0
1
2
4
8
16
32
beam width K