

Word-Piece dictionary

starboard ... skate
board

Substring
Levenshtein
Distance

sktaebaord

puzzle
together

Segmentation hypothesis

72%: single token
(most likely: starboard)
28%: 2 five letter tokens
(most likely: skate + board)