To handle unknown words, I add “smoothing method” when computing the probability of each words.

For example, when computing P (when| <s>) by adding smoothing

P (when| <s>) = (number of times “when” appear given <s>+1)/ (number of times <s> appears + length of the sentence

Therefore, when there is an unknown word, the numerator would be 1 and the probability would not be 0, so that it prevent the sentence probability becomes 0