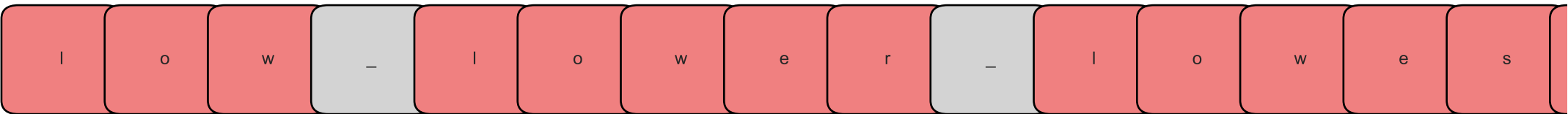


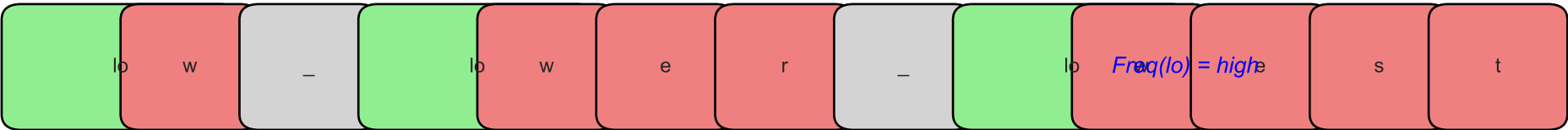
Byte Pair Encoding: Learning Subwords

Corpus: "low lower lowest"

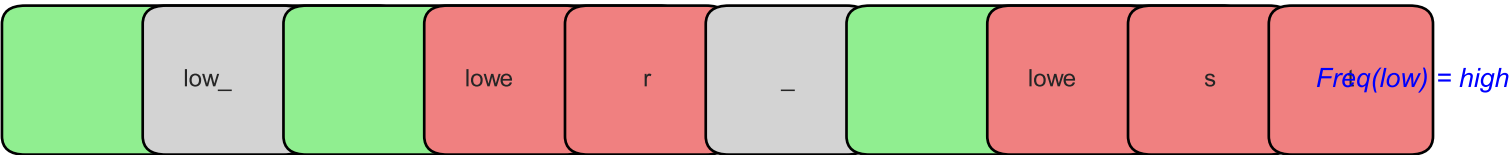
Initial



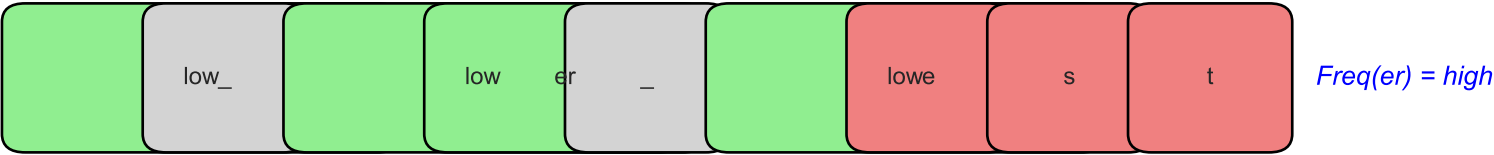
Merge "lo"



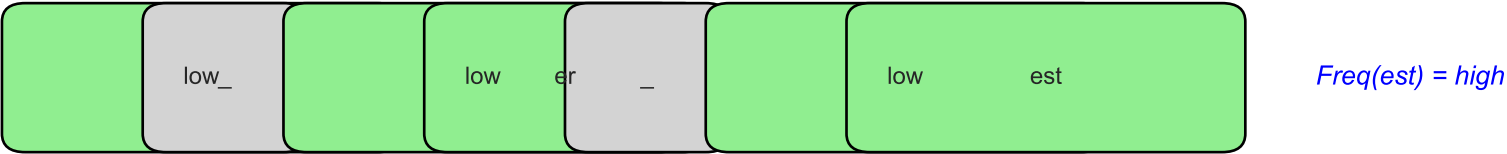
Merge "low"



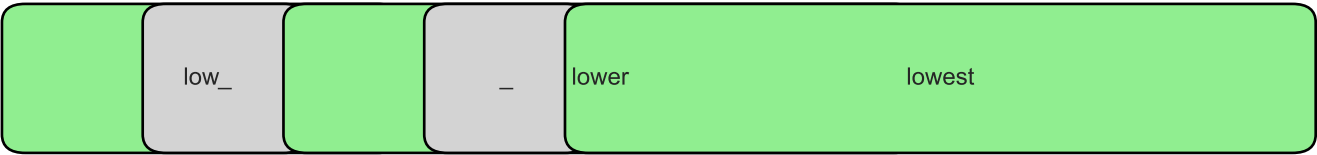
Merge "er"



Merge "est"



Final



BPE discovers:
• "low" = root
• "er" = comparative
• "est" = superlative



Character



Subword



Space