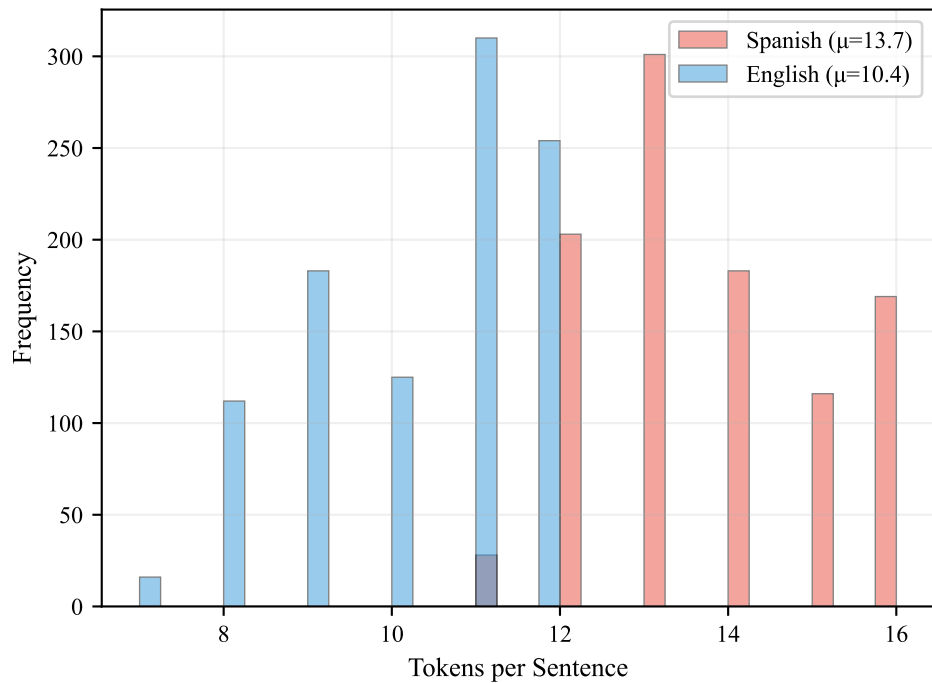


(a) Token Distribution



(b) Subword Tokenization

