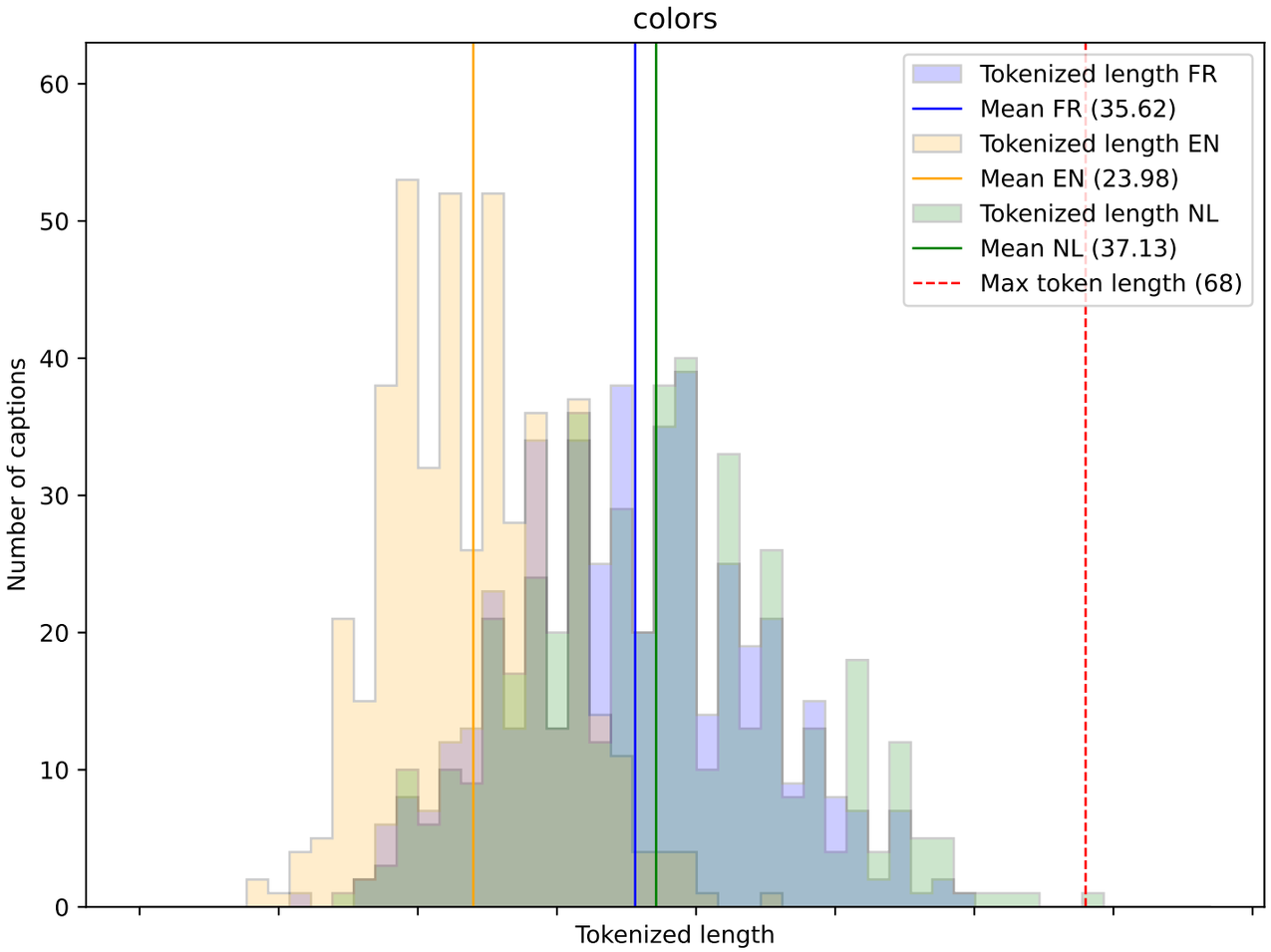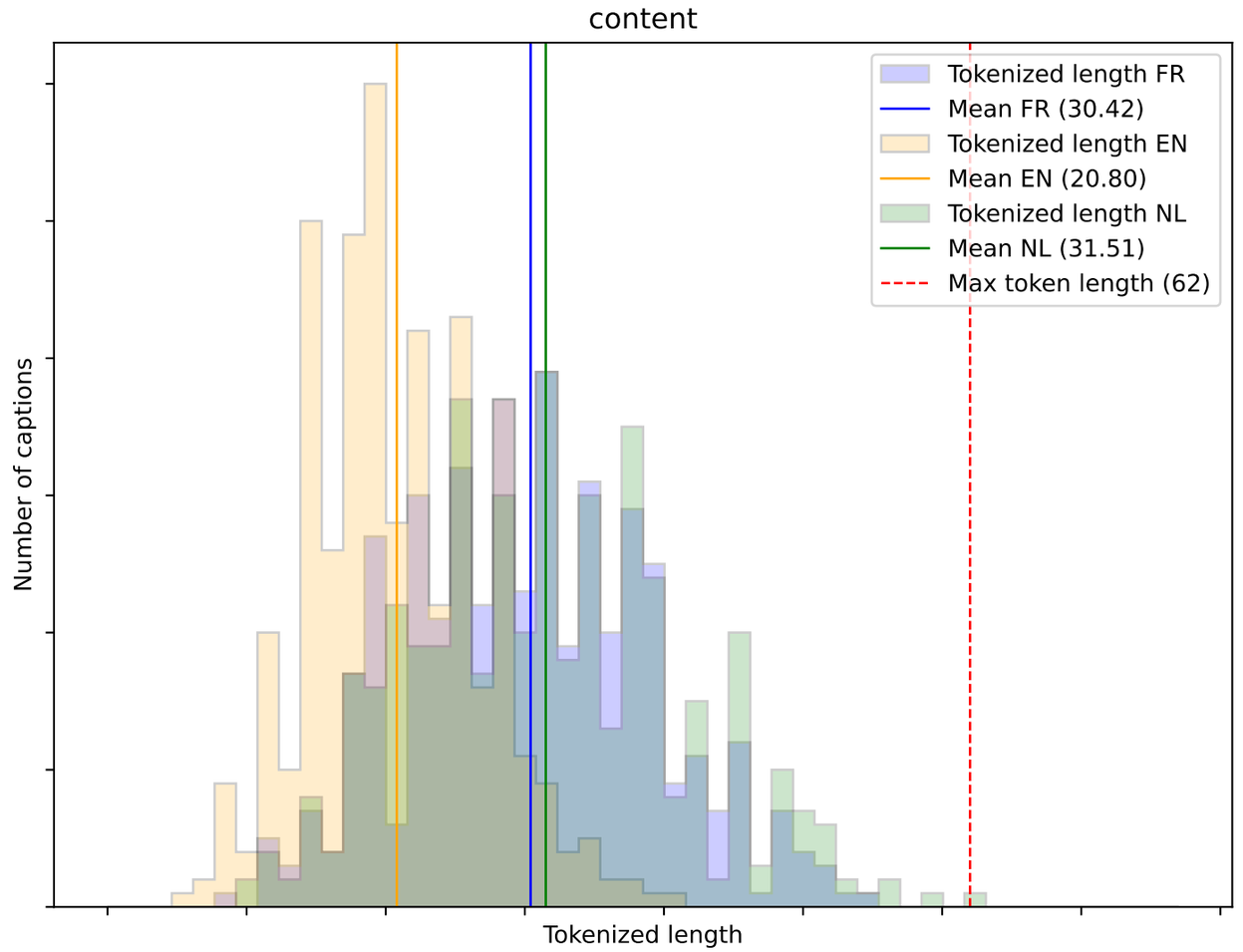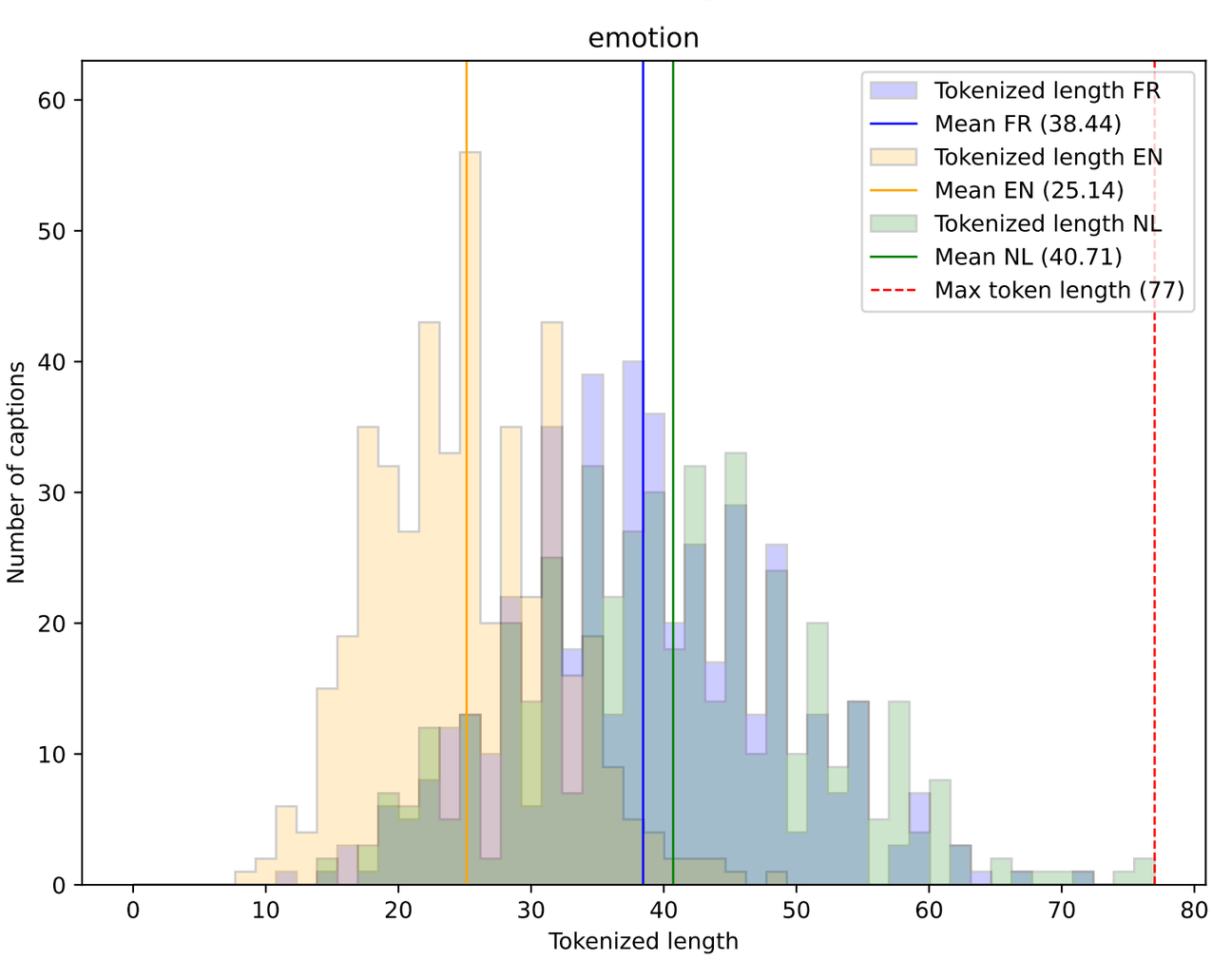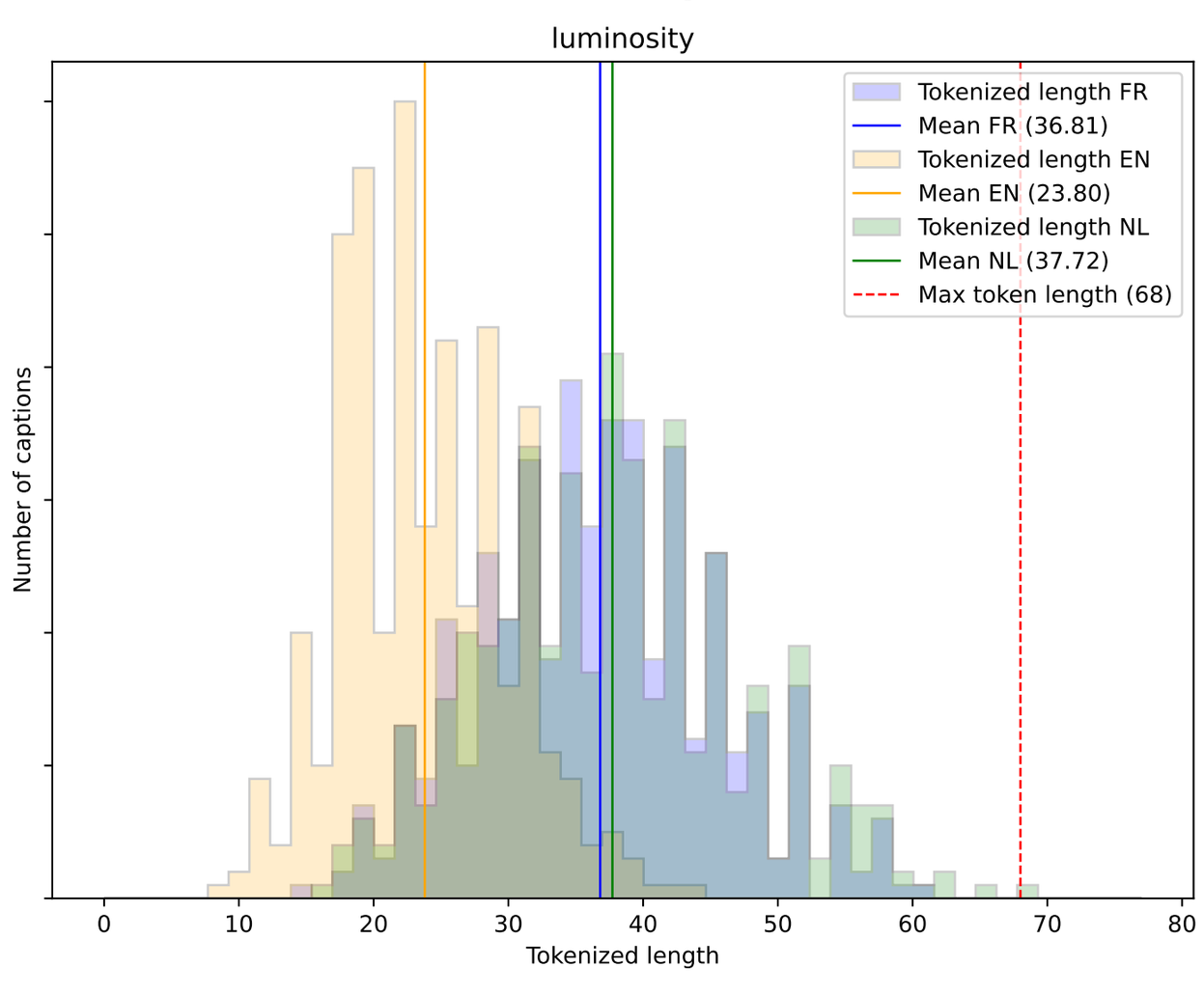Tokenized length of captions per focus