Vocabulary Size vs OOV Rate: Why Subword Tokenization Wins