e. (6 points) Why do we construct the feature as "any word" rather than "word other than first"? (Hint: how would we classify *Dillí* with such features?)

f. (7 points) The first word of each phrase contributes two features for classification (*first word* and *any word*), so in effect it is counted twice. Is this justified? What would happen to *Pan Auto* ("Mr. Auto"), *Nové Marie* ("New Mary"), and *Staré Dillí* ("Old Delhi") if the first word only contributed one feature?