Problem 2. Named entity recognition (NER) is the problem of identifying the names of persons, organizations, locations etc. In this problem you will construct a naive Bayes classifier to identify named entities in Czech. The table below is a snapshot of the data set, where phrases are labeled as to whether or not they represent a named entity. Each phrase is followed by The number of times it appears in the data.

Named entities	Not named entities
Nové Město (3)	Nové Auto (1)
Nové Dillí (5)	Kostel (9)
Kostel Panny Marie (2)	Červený (7)
Pan Červený (1)	Staré Auto (3)
Marie (4)	Nové (12)
	Červený Muž (3)

a. (2 points) Identify the priors for each class:

Named entity: ______ Not named entity: ______

b. (5 points) You will be constructing two types of features: *first word*, and *any word*. The *first word* feature of a phrase is the first word of the phrase; the *any word* feature of a phrase will have multiple occurrences – one for each word, including the first (so a three-word phrase, for example, will have three *any word* features).

Start by tabulating the number of instances of each feature, for each class.

	First word		Any word	
	Named Entity	Not Named Entity	Named Entity	Not Named Entity
Červený				
Kostel				
Marie				
Nové				
Pan				
Staré				
Auto				
Dillí				
Město				
Muž				
Panny				