

Result Table

<i>Perpl</i>	Unigram (no smoothing)	Bigram (no smoothing)	Trigram (no smoothing)	Bigram (smoothed)	Bigram (Add-one smoothed)	Bigram (Katz backoff smoothed)
English	49.69407533	inf	inf	36.2706049	93.37432686	39.77277337
French	74.65342427	inf	inf	46.204344	150.7925373	53.653989
German	30.63471311	inf	inf	22.426784	53.14784201	24.17793953
<i>Acc</i>	Unigram (no smoothing)	Bigram (no smoothing)	Trigram (no smoothing)	Bigram (smoothed)	Bigram (Add-one smoothed)	Bigram (Katz backoff smoothed)
English	0.613333333	0.48	0.426666667	0.6	0.6	0.606666667
French	0.38	0.533333333	0.54	0.406666667	0.393333333	0.033333333
German	0.006666667	0.24	0.326666667	0.02	0.006666667	0.413333333
All-three	0	0.126666667	0.146666667	0.01333333	0	0.026666667

Part 4

As we can see in above table, if the model is not smoothed, only the unigram one have finite perplexity, both bigram and trigram goes to infinity. Also, we can observe that the normal smoothing and katz backoff smoothing models have smaller total perplexity compared to Add-one smoothing method, which is not performed very well. As we learned in the class, the Additive Smoothing method is inaccurate in practice. Moreover, we can conclude the smoothing does remove unseen word (maybe add-one not).

In the analysis of accuracy, overall speaking, all of the accuracies of all-three classes using different n-gram model are not very good. However, if we look at the English / Not English, French / Not French, German / Not German classification, some are relatively good, such as 0.61 in Unigram (no smoothing) model for English, 0.53 in Bigram (no smoothing) model for French, 0.54 in Trigram (no smoothing) model for french and around 0.6 in both Bigram (smoothed) & Bigram (Katz backoff smoothed) models for English. We can say it is different from English to French and German. German has relatively lowest accuracy, the reason might be its grammar structure and the complexity of vocabulary are hard to learn. In the meanwhile, English is the most accurate one comparing to other two language, proving English is easier to learn.