

"Observations"

I have made a table to compare the prediction accuracy for 2-grams and 3-grams. From looking at the results, I noticed that the prediction for a language depends on many factors such as training data, input text and the semantics of the target language. Here are some results for each language and my observations.

For English:

When comparing 2-grams vs 3-grams, I noticed that 2-grams were giving more accurate results as compared to 3-grams. On average, there is a 15 – 18 % improvement in the accuracy of results when using 2-grams. One reason is that most joining words in English are 2-letter words such as "is", "am", "is", "to", etc.

Words	2-Gram (%)	3-Gram (%)
"The weather is very pleasant today"	76.024	62.124
"Language is very important for communication"	62.234	69.807
"Can you come around the corner"	73.026	57.352
"the sun looks good"	69.009	43.770
"everything happens for a reason"	79.285	66.768
"you have to come here"	82.765	74.255
"the suit looks good on you"	69.822	51.272

For German:

From the results, My observation is that since German language contains a lot of 3-letter words, So the accuracy on using 3-grams increases in contrast to English language.

Words	2-Gram (%)	3-Gram (%)
"Das Wetter ist heute sehr angenehm"	78.844	99.0
"Sprache ist für die Kommunikation sehr wichtig",	99.0	99.0
"Kannst du um die Ecke kommen"	76.292	48.239
"Die Sonne sieht gut aus"	89.859	99.0
"alles geschieht aus einem Grund"	80.337	99.0
"du musst hierher kommen"	74.677	81.875
"Der Anzug sieht dir gut an"	74.034	99.0

For Spanish:

Spanish also contains many 2-letter words, so using 2-grams increases the accuracy.

Words	2-Gram (%)	3-Gram (%)
"El clima es muy agradable hoy",	72.751	79.584
"El lenguaje es muy importante para la comunicación"	99.0	99.0
"¿Puede venir vuelta de la esquina"	82.821	99.0
"El sol se ve bien"	84.858	53.279
"Todo sucede por una razón"	99.0	99.0
"Tienes que venir aquí"	79.962	63.695
"El traje que hace el bien"	88.737	73.944

For Italian:

Words	2-Gram (%)	3-Gram (%)
"Il clima è molto piacevole oggi"	99.0	99.0
"lingua è molto importante per la comunicazione"	99.0	99.0
"venire dietro l'angolo"	73.822	99.0
"Il sole sembra buono"	66.233	66.366
"Tutto accade per una ragione"	79.149	74.252
"Devi venire qui"	71.609	58.221
"L'abito fa bene"	61.565	99.0