Summary

The Great Eskimo Vocabulary Hoax

Submitted by:
Pallavit Aggarwal
Intelligent Systems – 22333721

The great Eskimo vocabulary hoax, brings forth the problem of mapping different words or their inflectional forms to a singular entity. It goes beyond synonyms, and discusses different nouns and verbs, and their morphologically altered forms and tries to posit the challenges that might occur in bringing them all onto the same root point. This compounded by the ever-evolving vocabulary and new words, makes the problem even more challenging for researchers.

The paper takes as an example, Eskimos and their usage of word snow. Some researchers and anthropologists state that there are around a doze to a hundred (or even 400) different words for the root word "Snow" in the Inuit and Yupik languages (Eskimo languages). The author however disagrees with this and calls out these researchers for their sloppiness and flags them as "intellectual-sloths". The problem lies in the fact that there are multiple versions of different Eskimo languages spoken across from Siberia to Greenland, one is the urbanized one spoken by the young and the other is the old version, so there's no particular version to base the facts on. The other problem goes deeper into the interpretation of languages and highlights the fact that a word can have different synonyms or inflectional forms to it. For example, the word water may also be referred to as liquid, lake, river, brook, rain, dew, wave and foam. For somebody researching the term liquid, it would be hard to ignore its usage in different contexts for an exhaustive and genuine data gathering.

Similarly, Snow can often times be referred to as sleet, slush, blizzard in different contexts. Another problem is about taking the root word like Snow in this case, and transforming into other similar meaning words like a snowball, snowflake, snowstorm, snowy. Should a researcher ignore these words while researching the different words that are used to describe "Snow" in English language? Should nouns like snow, snowing, snowed also be counted differently?

Overall, the author takes a critical and often humorous approach to the complexities of languages and examining their origins and debunking myths. He finally settles with the fact that there seem to be a dozen words (a relatively short number) describing snow in Eskimo language.

From a text analytics point of view, the paper highlights the importance of avoiding overgeneralization and stereotypes when it comes to languages. Text analytics, which involves the use of computational methods to extract insights from large volumes of text data, relies heavily on language-based models and algorithms. However, the models used in text analytics are only as good as the data they are trained on. If the data used to train these models is biased or oversimplified, then the insights generated will also be biased and inaccurate. For example, if the models are trained on data that perpetuates the myth that the Eskimo language has a large number of words for snow, then the models will also generate outputs that reinforce this stereotype. Thus these models need to be accurate in their culmination of the complexities of the language that they're modelling.