# Accessible Tutorial - Skip-Gram Model for code-switched Telugu - English Language

## Tool/ Process being explored:

Skip-Gram word2vec model using Gensim for code-switched NLP, specifically trained on Telugu - English data.

## Skip-Gram Model:

The Skip-Gram model is a neural network architecture used in word2vec to learn word embeddings based on the surrounding context of words. It captures semantic similarity and context in a low-dimensional vector space.

It aims at translating words to embedding vectors. The algorithm maps words to vectors based on the spatial proximity to other words. The assumption is, if the words are placed near they are semantically similar. It iterates over all words in the text input and optimizes vector representations distance based on the context.

In this project, I will create an accessible tutorial for Skip-gram model for code-switched data, Telugu - English YouTube comments.

**GitHub Link:** https://github.com/AnkithaNamala/LING539-project-AnkithaNamala/

**Reference:** https://leshem-ido.medium.com/skip-gram-word2vec-algorithm-explained-85cd67a45ffa