This paper introduces two efficient methods for learning word embeddings: **Continuous Bag-of-Words (CBOW)** and **Skip-gram**. These are simple models that can train on extremely large datasets

and still produce high-quality word vectors faster than previous neural network models like NNLMs and RNNs.

The main idea is to represent words as vectors in such a way that similar words end up being close together in this vector space. CBOW predicts a word based on the words around it, while Skip-gram does the opposite, as it predicts the surrounding words based on the current one. Skip-gram especially stands out because it captures complex relationships between words using simple math, like: **“king” - “man” + “woman” = “queen”**.

Even though these models are simpler, they still outperform more complex ones in many tasks, especially when trained on large corpora like Google News. The paper also created a new test set to measure how well word vectors capture both **syntactic** and **semantic** relationships like verb forms or country-capital pairs, and these models still did really well.

Overall, the paper shows that simpler models, when scaled properly, can be very powerful for understanding language and can be used in NLP tasks like translation or question answering.