The *“Efficient Estimation of Word Representations in Vector Space”* paper (Word2Vec) marked a breakthrough in natural language processing. It introduced two novel and efficient architectures: Continuous Bag of Words (CBOW) and Skip-Gram, both designed to learn dense vector representations (embeddings) of words from large corpora efficiently.

Before this work, word representations were typically derived using computationally expensive models like neural language models or statistical techniques like LSA, which were slow and struggled to scale. Word2Vec changed that by introducing fast, scalable training methods using techniques such as negative sampling and hierarchical softmax, enabling training on billions of words in just a few hours.

The real power of Word2Vec lies in the semantic richness of the learned vectors. The model captured both syntactic and semantic regularities in language, allowing for arithmetic operations on vectors that reflected meaningful linguistic relationships

(“King - Man + Woman = Queen”). This discovery showcased a new way to represent meaning numerically.

The paper significantly influenced the entire field, making word embeddings a core component of nearly all NLP systems. It paved the way for the next generation of embedding techniques such as GloVe, ELMo, and eventually contextual models like BERT and GPT, which build on Word2Vec's foundation but add contextual awareness and deeper language understanding.