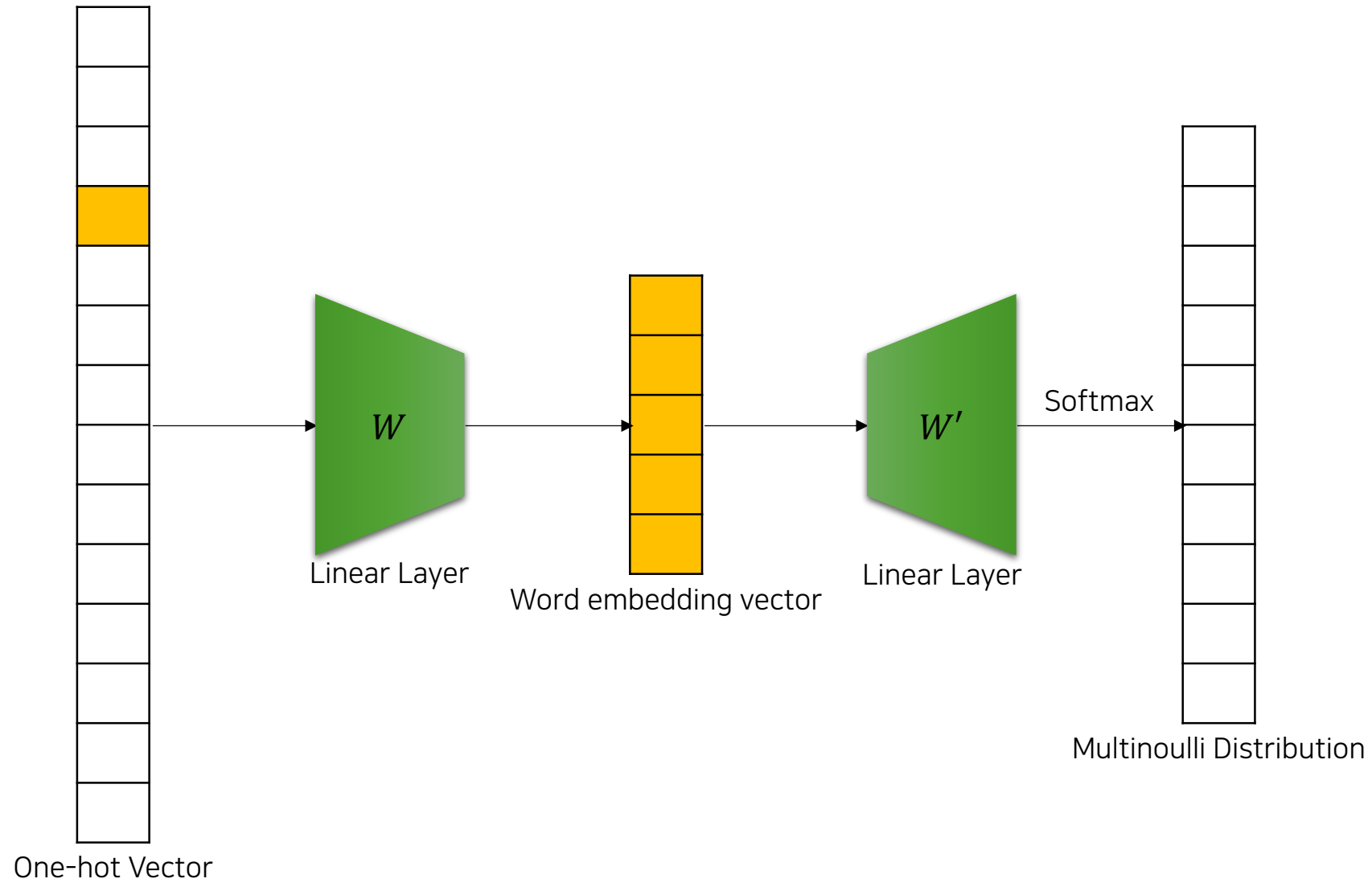


# FastText

Ki Hyun Kim

[nlp.with.deep.learning@gmail.com](mailto:nlp.with.deep.learning@gmail.com)

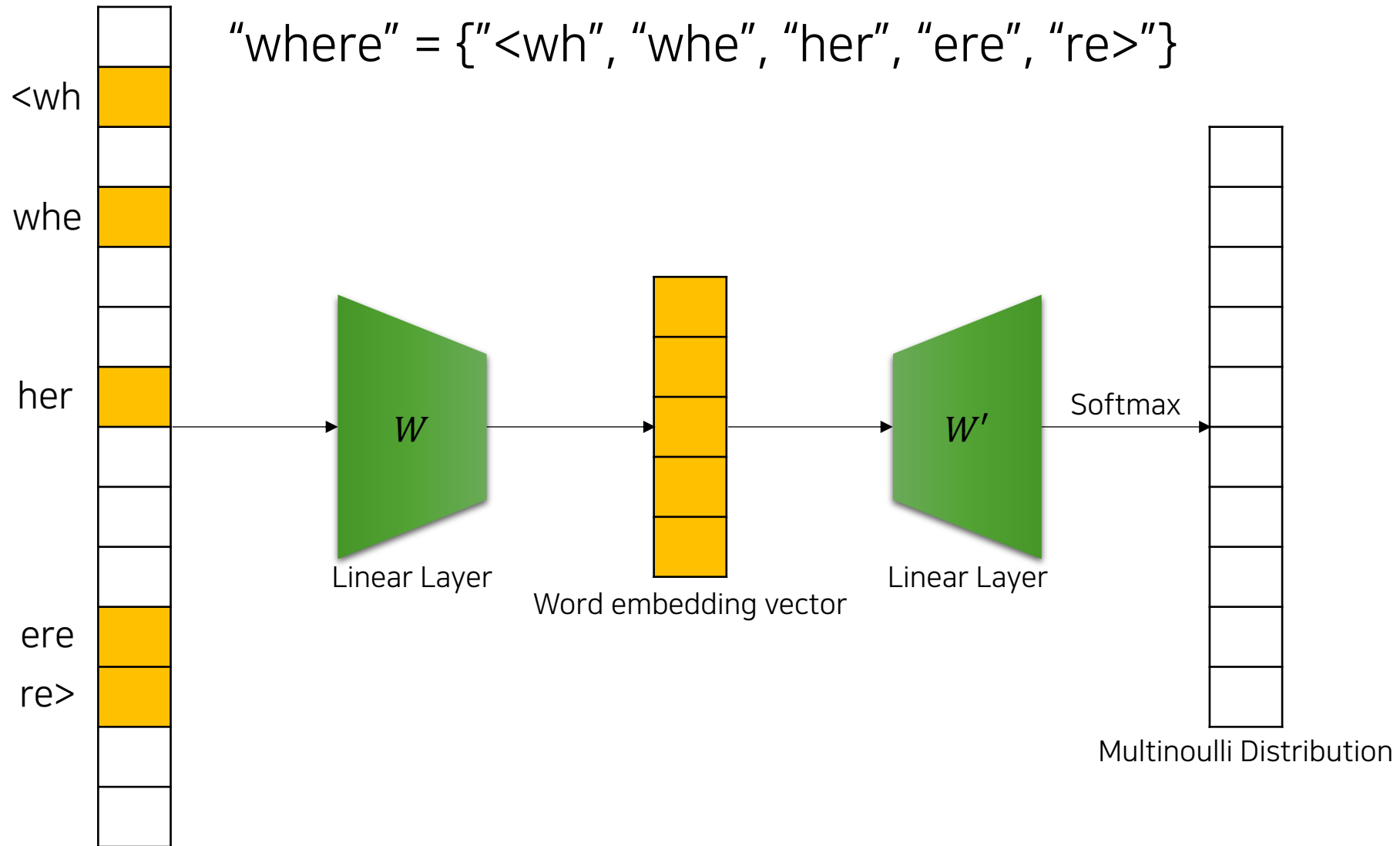
# Review: Skip-gram



# FastText: Upgrade Version of Skip-gram

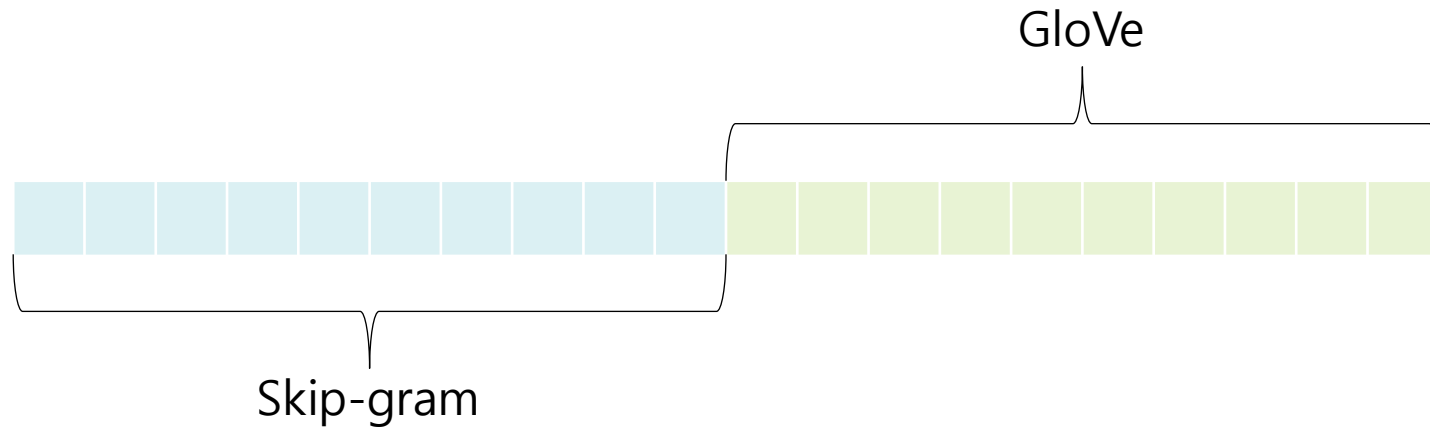
- Enriching Word Vectors with Subword Information  
[Bojanowski and Grave et al., 2016]
- Motivation:
  - 기존의 word2vec은 저빈도 단어에 대한 학습과 OoV에 대한 대처가 어려웠음
- FastText는 학습시,
  - 1) 단어를 subword로 나누고,
  - 2) Skip-gram을 활용하여, 각 subword에 대한 embedding vector에 주변 단어의 context vector를 곱하여 더한다.
  - 3) 이 값이 최대가 되도록 학습을 수행한다.
- 최종적으로 각 subword에 대한 embedding vector의 합이 word embedding vector가 된다.

# Example: where



# Conclusion: Word Embedding

- 딱히 어떤 알고리즘이 더 뛰어나다고는 할 수 없다.
  - 구현이 쉽고 빠른 오픈소스를 사용하는 것이 낫다.
- 두 개의 다른 알고리즘 결과물을 concat하여 사용하기도



# Open-source

- Gensim
  - <https://radimrehurek.com/gensim/install.html>
- GloVe
  - <https://github.com/stanfordnlp/GloVe>
- FastText
  - <https://github.com/facebookresearch/fastText/>
  - [http://bit.ly/fasttext\\_win](http://bit.ly/fasttext_win)