Distributional Semantics in the wild

ESERCITAZIONI DI LINGUISTICA APPLICATA A.A. 2024/2025

Mattia Proietti

Topics of the lab

- **Vector semantics**
  - Static Embeddings
  - Contextual Embedding
- **Word2Vec**
- **Transformers Language Models**
  - **BERT**
  - **Generative Language Models** (GPT family)

Topic of the day

- **Word2Vec**
  - What is it?
  - How to train it?
  - What can be used used for?
- Visualize a vector **semantic space** and the **training process**
- **Embeddings** manipulations
  - Vector operations
  - Intruder detection
  - Bias detection
- **Contextual vs static embeddings**
  - **BERT** vs **W2V**

Word2Vec

- Neural Language Model
  - A.k.a. a ***predict model***
- A single-layer NN trained with a "fake" task of classification
- The learned weights are used as word representations
- Repres. are dense vectors we call **embeddings**

- Embeddings are **statics**
- Each word is associated with a single embedding
- They form a **semantic space**

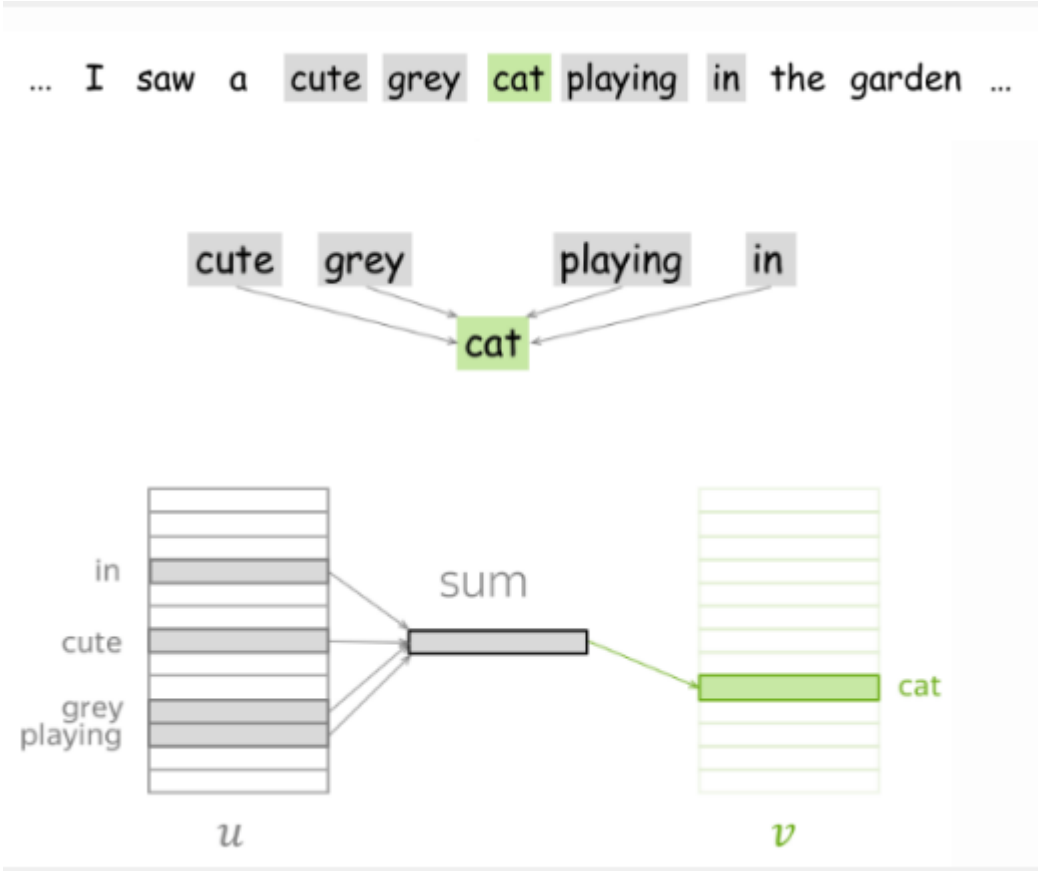
Two flavours:

- **CbOW**
- **Skip-Gram**

CBoW

From context to target words

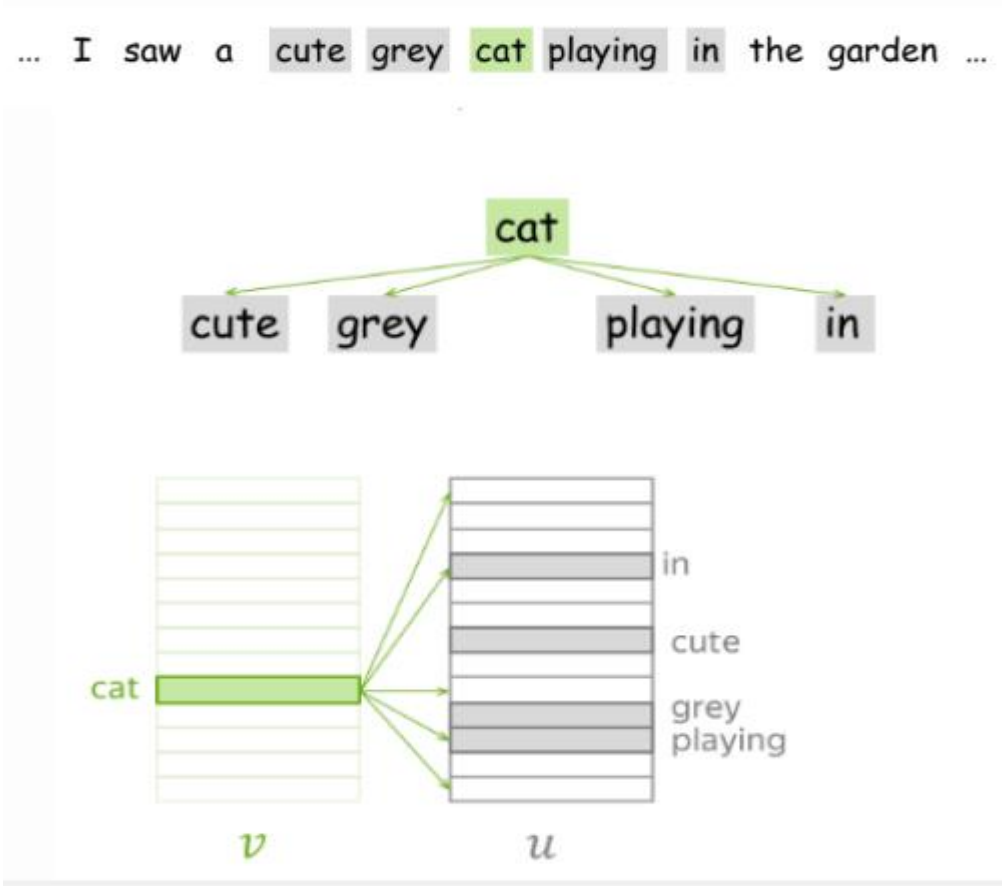
- Learns to predict a target word leveraging its neighbourhood.
- The sum of the context vectors is used to predict the target word.
- Context window size can be chosen arbitrarily



Skip-Gram

From target words to context

- The skip-gram model predicts context words given a central word.
- The context of a word can be represented through a set of skip-gram pairs of (target_word, context_word) where context_word appears in the neighbouring context of the target word.



CBoW vs Skip-Gram

The final goal is the same! → Create coherent and informative semantic spaces

In a vector semantic space **similar/related** words are grouped together and occupy similar portions

Examples

We can put our hands on a Word2Vec model and see how it works!

Google Colab notebook

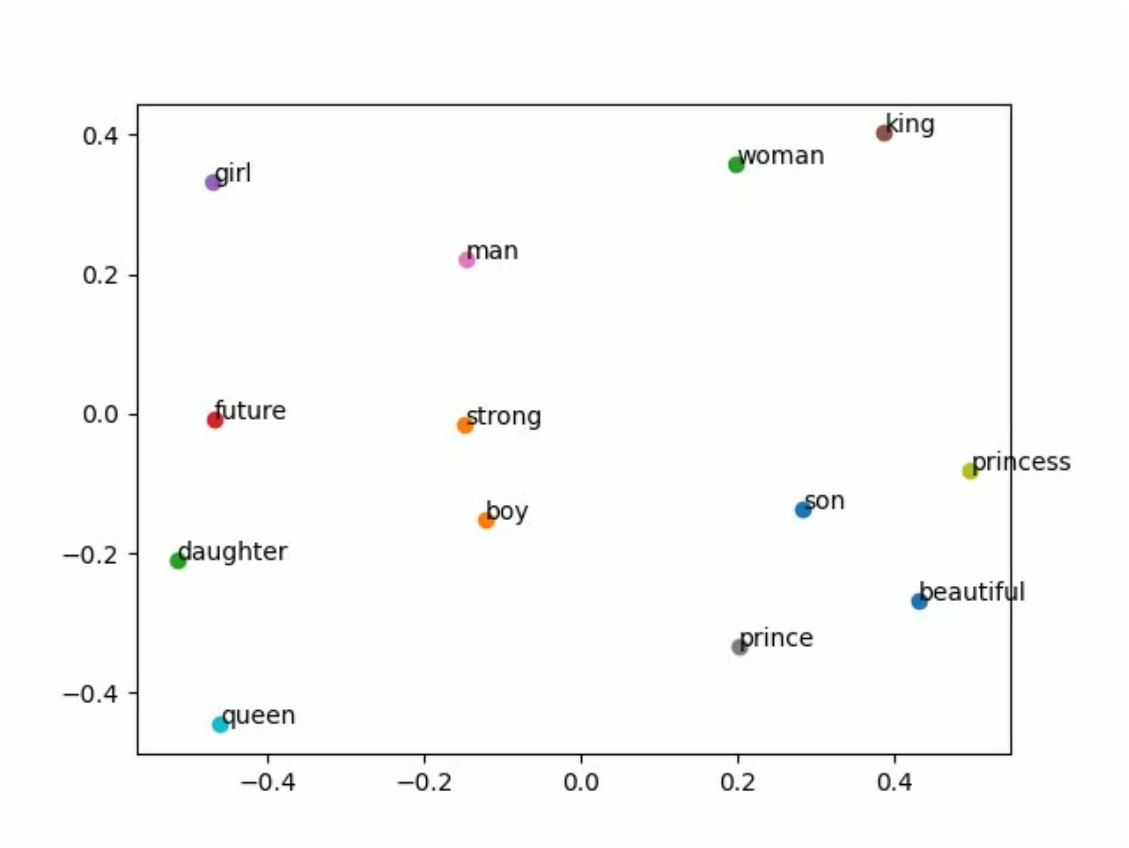
Summary of the content:

- Data preparation and model training (toy example)
- Train W2V with a line of code!
- Operations between (word) vectors
- Visualize words in a vector space

A closer look inside the training process and the formation of a semantic space

In a vector semantic space related words are progressively rushed together

Here you can see how a toy model trained on a simple corpus learns its semantic space from the data



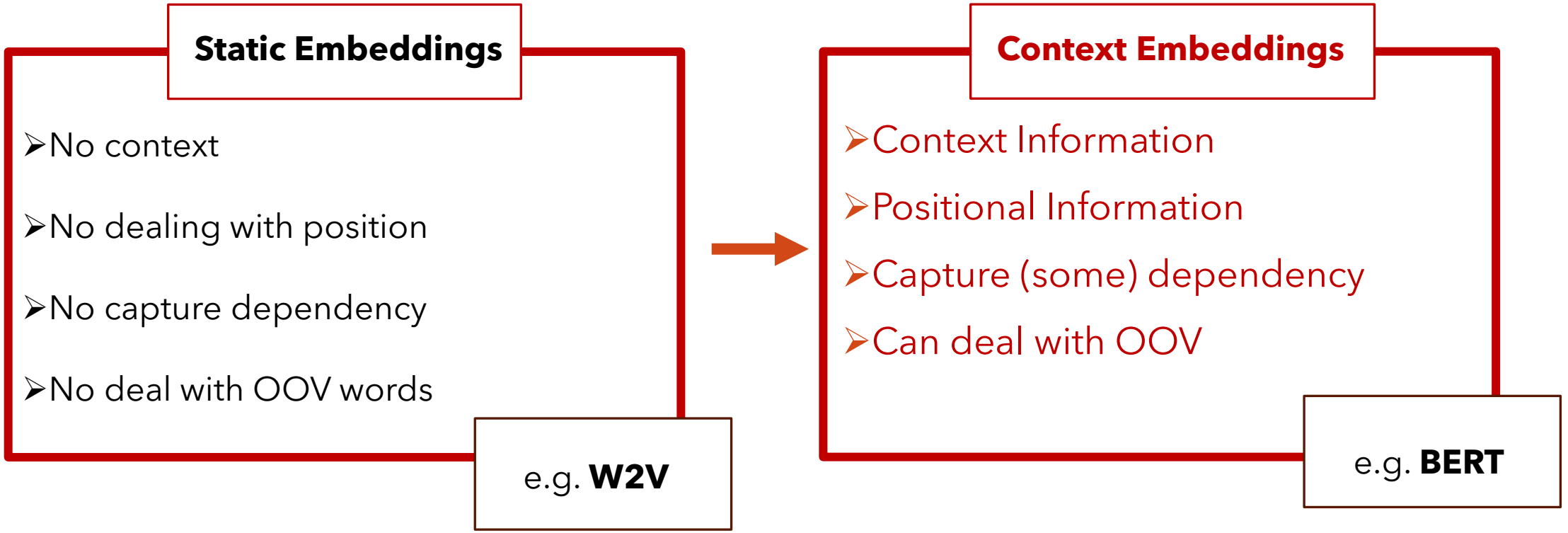
Limits of the static embeddings

Static Embeddings

- No context
- No dealing with position
- No capture dependency
- No deal with OOV words

e.g. **W2V**

From **static** to **contextual** embeddings



To know more

- Mikolov et. Al (2014), Efficient estimation of word representations in a vector space
- Jay Alammar, The illustrated Word2Vec
- Jurafsky & Martin, cap. 6), Speech and Language Processing
- Lena-Voita blog post, Word Embeddings
- McCormick Blog post on Skip-Gram