transformers and beyond

GESIS Fall Seminar 2023

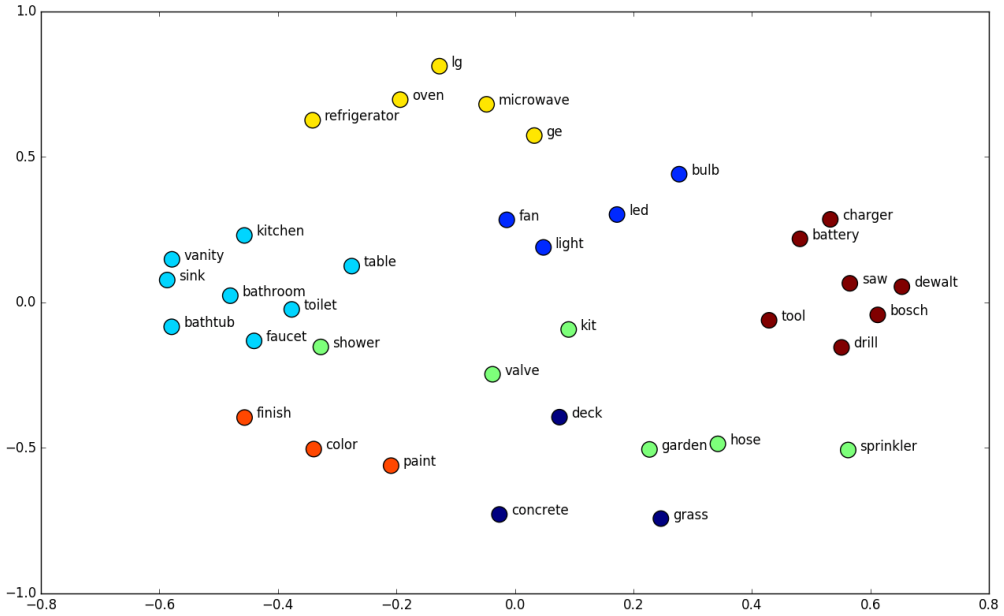
"From Embeddings to Transformers: Advanced Text Analysis in Python"

[day 3, afternoon: contextualization]

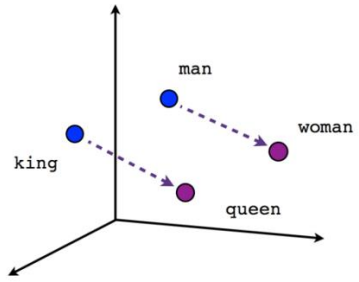
word embedding

the *vector* representing a word

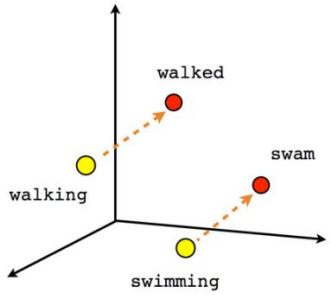
in a *d*-dimensional, real-valued space



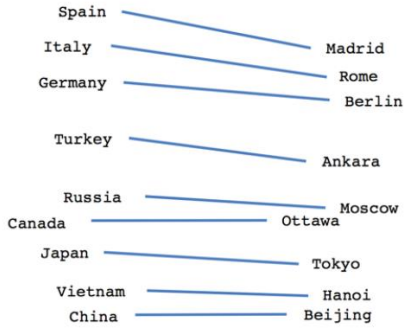
[Mikolov et. al. (2013)]



Male-Female



Verb tense



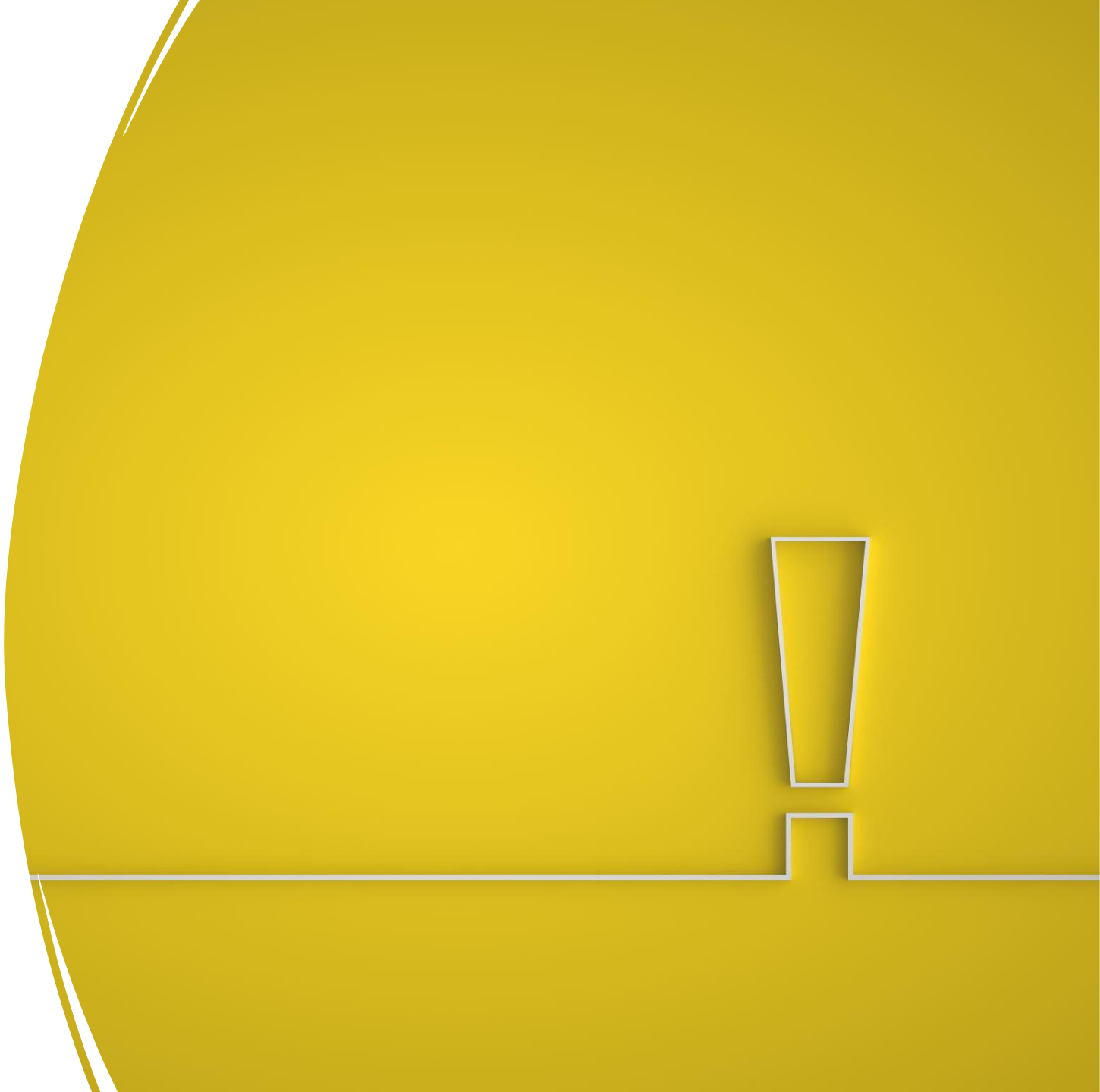
Country-Capital

[Creating Word Embeddings | TDS]

limitations of […] word embeddings

[classic, traditional, static]

- lack of context
- semantic similarity challenges
- out-of-vocabulary words
- language dependency



what’s the problem here?

BANK

OPEN

BAT

polysemy

[…is a common linguistic phenomenon where a single word can have several senses or interpretations depending on the *context* in which it is used.]

BANK

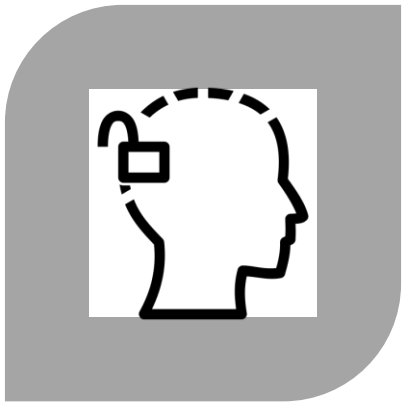
OPEN

BAT

polysemy

[…is a common linguistic phenomenon where a single word can have several senses or interpretations depending on the *context* in which it is used.]

BANK



OPEN

BAT

what’s the solution?

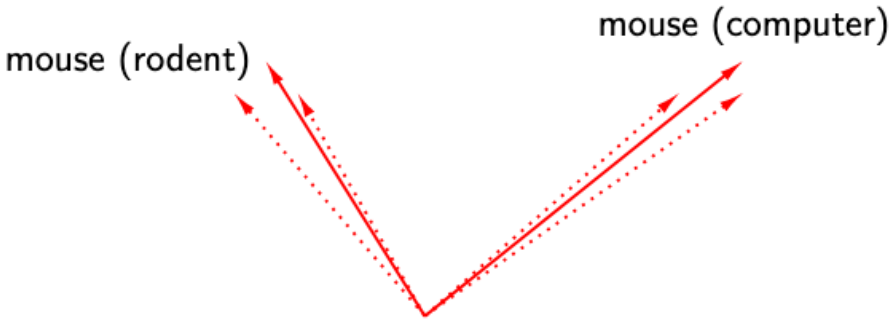
Words don’t appear in isolation.

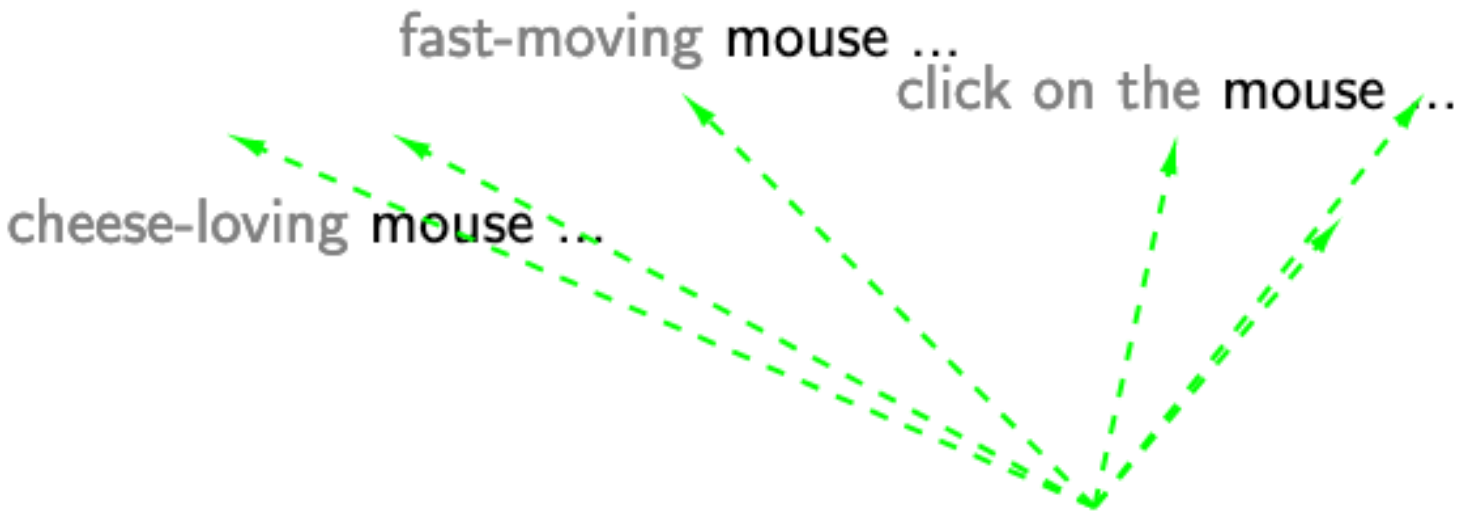
The word use (syntax and semantics) depends on its context.

Why not learn the representations for each word in its context?

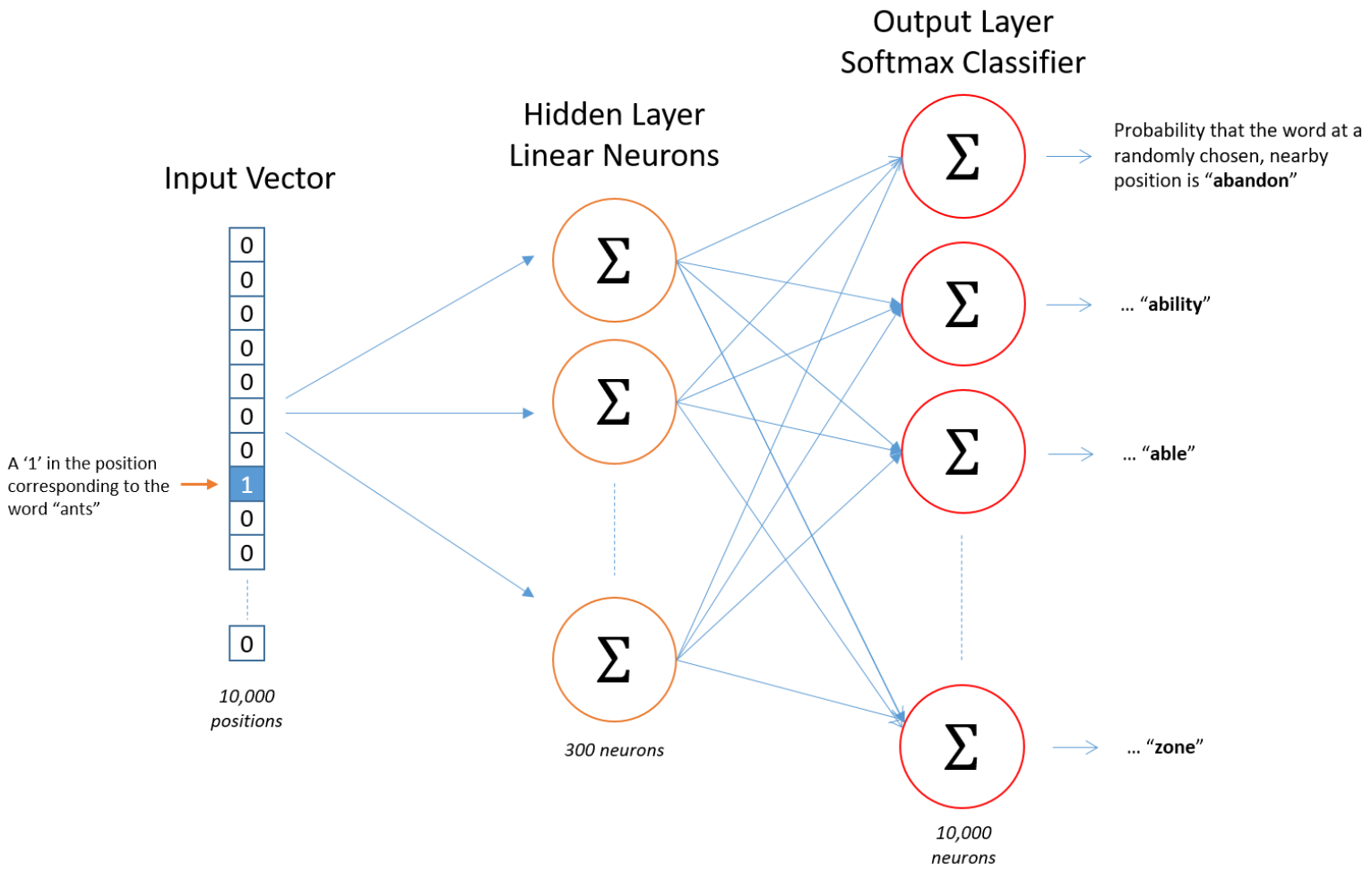
[dynamic]

contextualized word embeddings





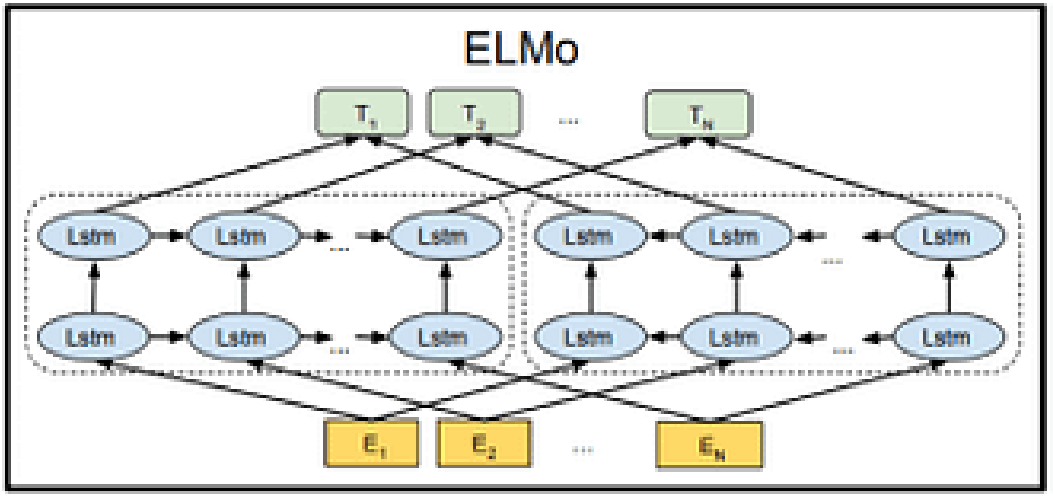
[The Standford AI Lab Blog]



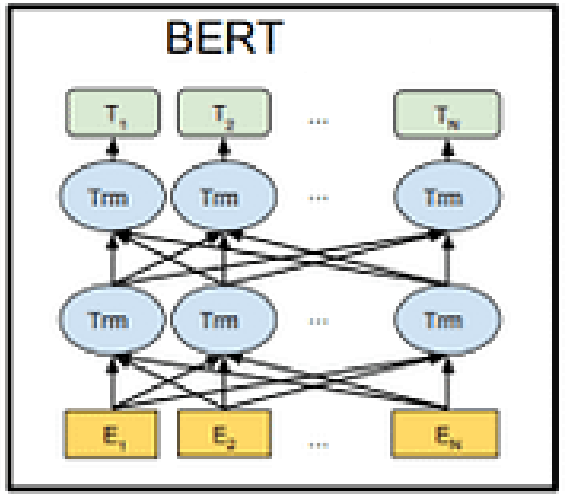
how?

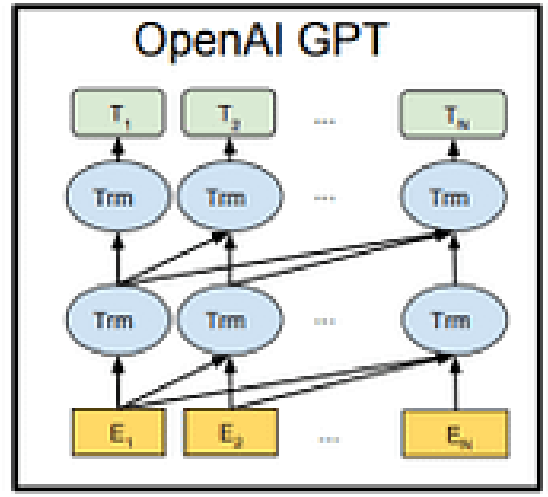
[McCormick 2016]

contextualized word embeddings | how?



LSTM





Transformers

[Devlin et al. 2018]

notebook:
03_BERT_word_embeddings

a comparison

	word embeddings [Word2Vec, GloVe]	contextualized word embeddings [ELMo, BERT, GPT]
representation	static	dynamic
context-	agnostic	aware
models	pre-trained, non-adaptable	finetuning