
# 3 Algorithms for Sense Embedding

Thomas Bonald

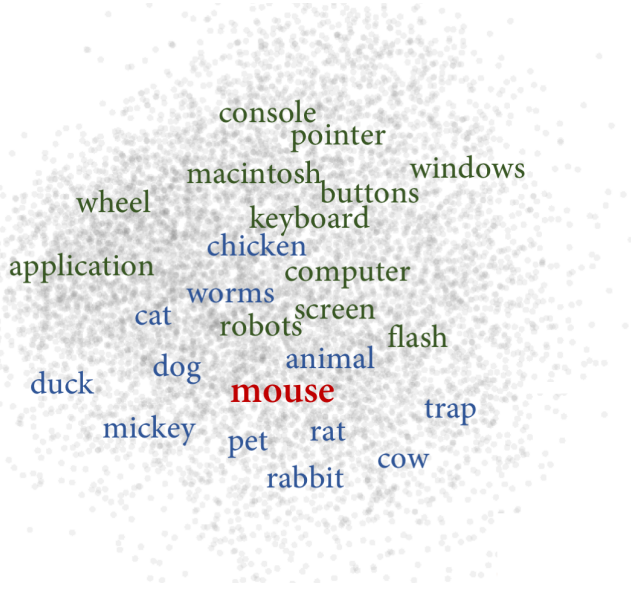
**TALia seminar**
February 4, 2022

# Bibliography


- From word to sense embeddings: A survey on vector representations of meaning[1]
  Camacho-Collados & Pilehvar 2018
- Multi-prototype vector-space models of word meaning
  Reisinger & Mooney 2010
- Efficient non-parametric estimation of multiple embeddings per word in vector space
  Neelakantan, Shankar, Passos & McCallum 2014
- A unified model for word sense representation and disambiguation
  Chen, Liu & Sun 2014

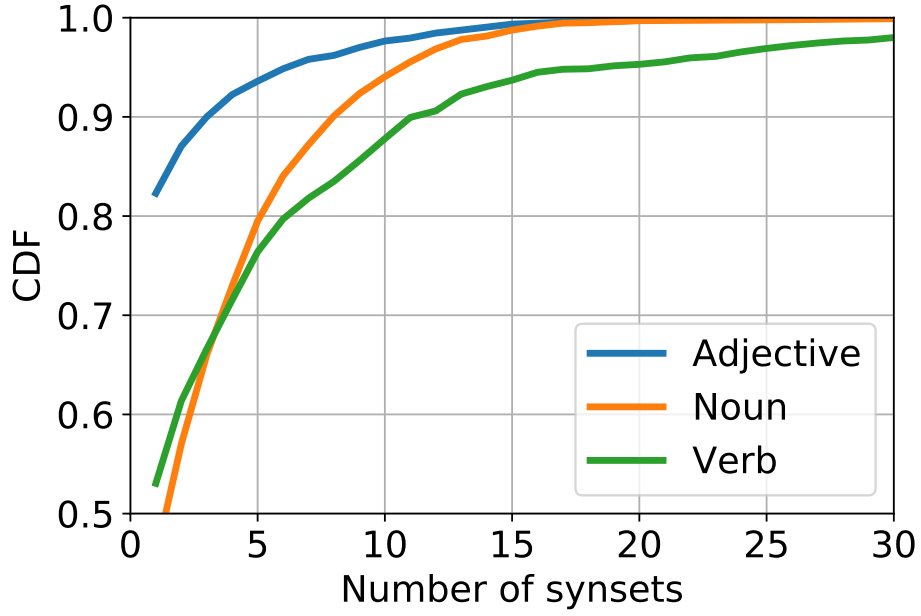

[1] Source for illustrations

# Motivation




Confusion due to the polysemy of the word **mouse**

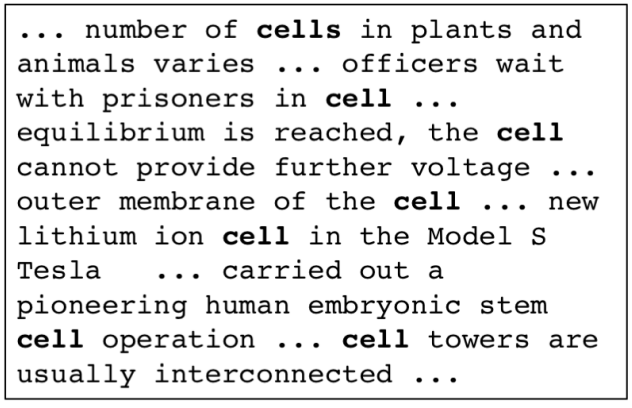
# Polysemy



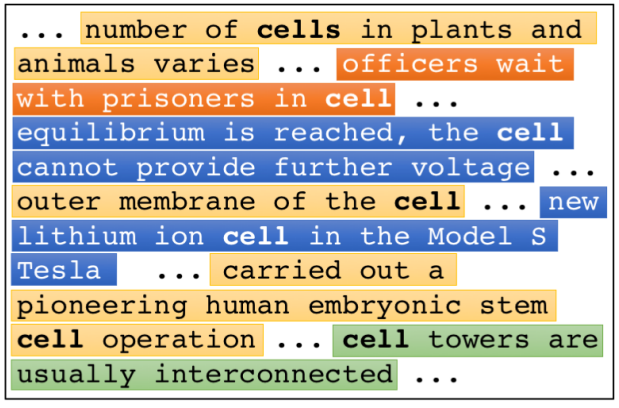

Distribution of the number of synsets per word
(Wordnet, 2,000 most frequent words)

# Sense embedding

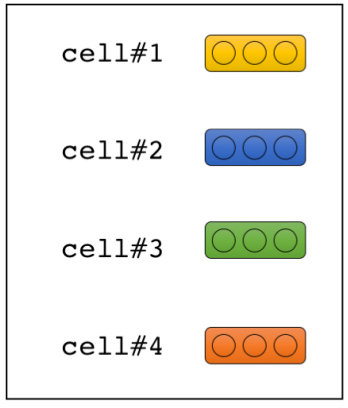
Learn the representation of **senses** instead of **words**




(1) Get occurrences of a word from text corpora




(2) Analyze contexts and induce senses of the word




(3) Compute sense representation

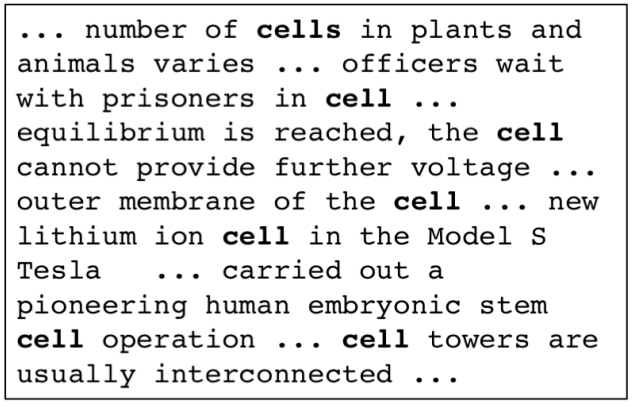
## Note

A **single** vector per sense

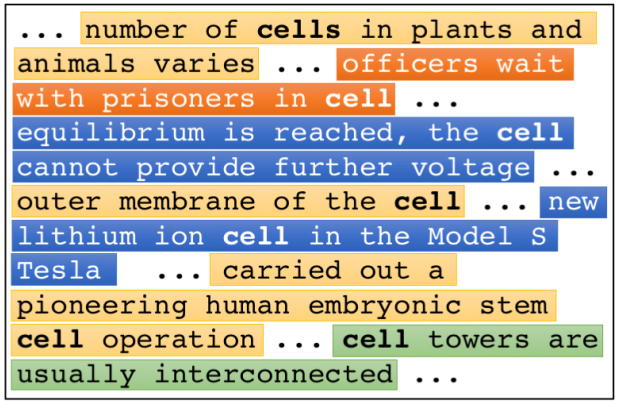
Different from contextual embedding (e.g., BERT)

# 2 main issues

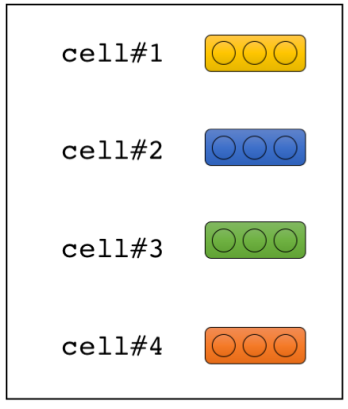
1. The **number** of senses is not well defined
2. The **context** itself becomes ambiguous...




(1) Get occurrences of a word from text corpora




(2) Analyze contexts and induce senses of the word




(3) Compute sense representation

# Outline

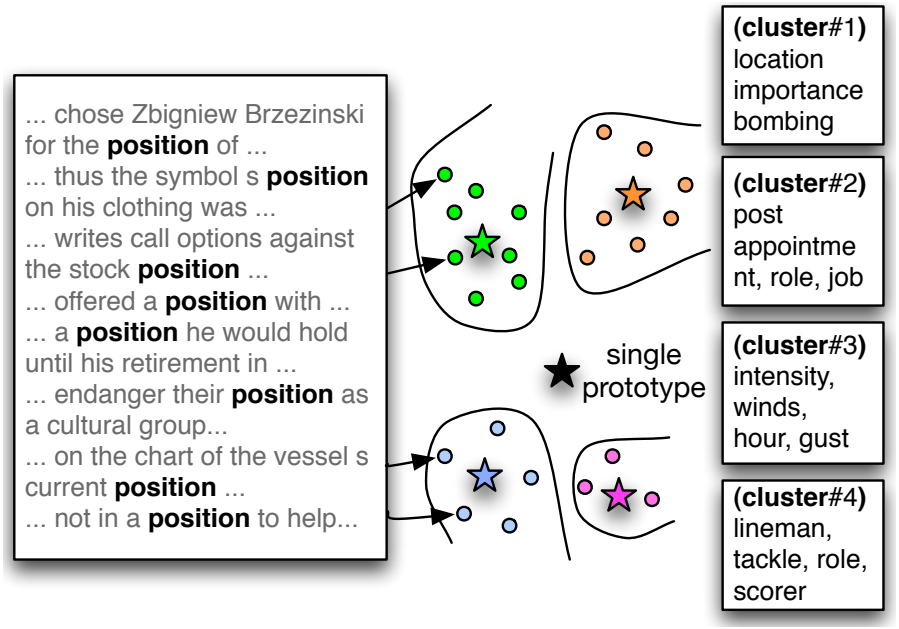
Focus on 3 algorithms:

1. Clustering
   Schutze 1998, Reisinger & Mooney 2010
2. Online matching
   Neelakantan, Shankar, Passos & McCallum 2014
3. Knowledge
   Chen, Liu & Sun 2014

# Clustering

2 steps:

1. For each word, cluster the **contexts** in which this word appears (cosine-similarity in the word space)
2. Learn the representation of each **sense** (= context cluter)

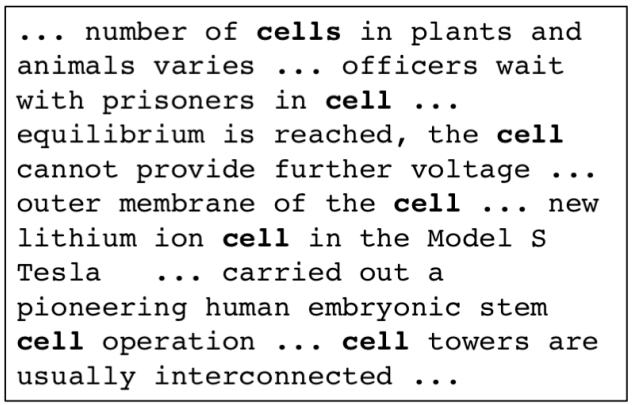



**Idea:** Hierarchical clustering of the word-context bipartite graph

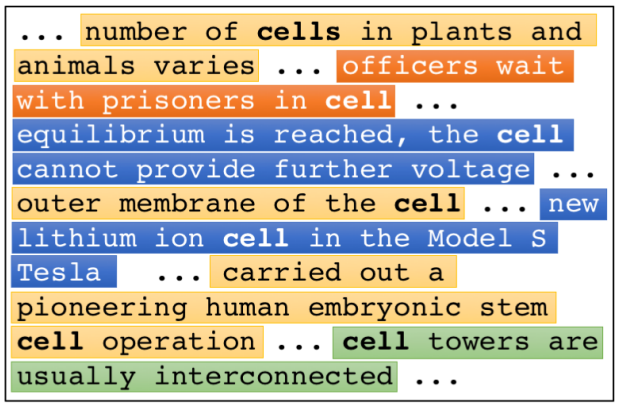
# Online matching

Joint learning of word & sense representations:

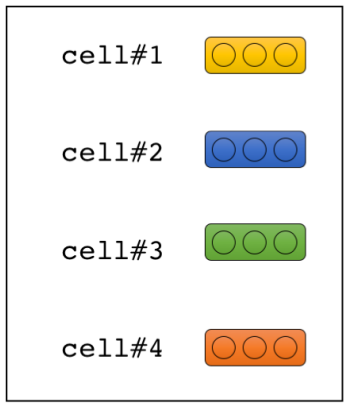
- **Context** represented by the average of **word** vectors
- **Sense** induced by the closest **context**
  ($k$ senses per word)




(1) Get occurrences of a word from text corpora

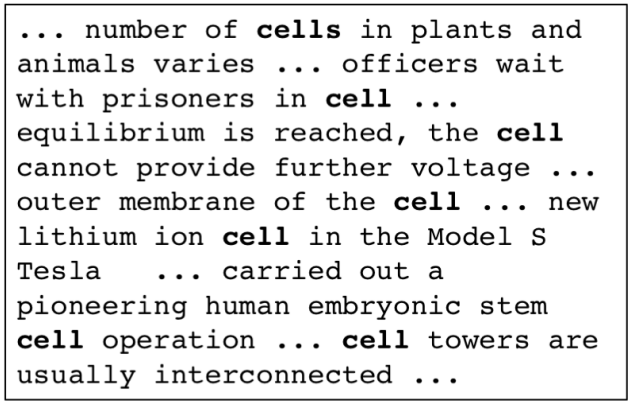



(2) Analyze contexts and induce senses of the word

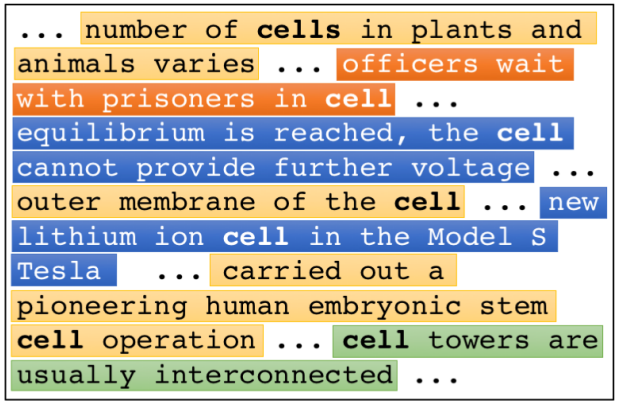



(3) Compute sense representation

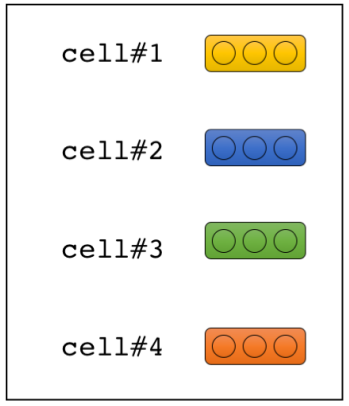
# Example




(1) Get occurrences of a word from text corpora




(2) Analyze contexts and induce senses of the word

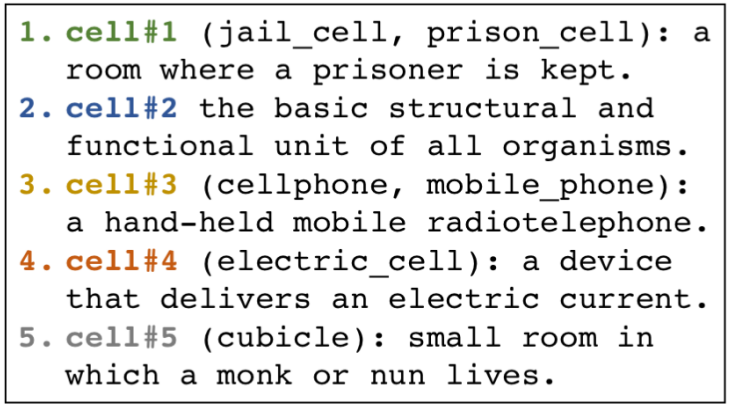



(3) Compute sense representation

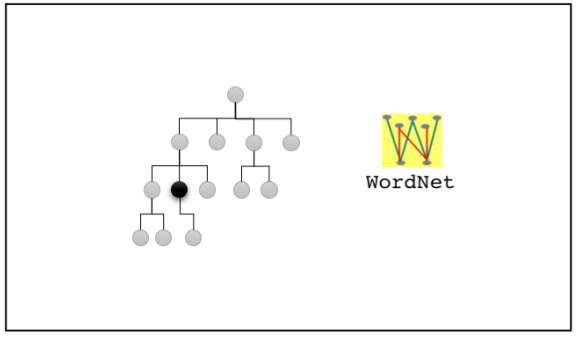
# Knowledge

Use of a knowledge base (e.g., Wordnet, Babelnet, Conceptnet)

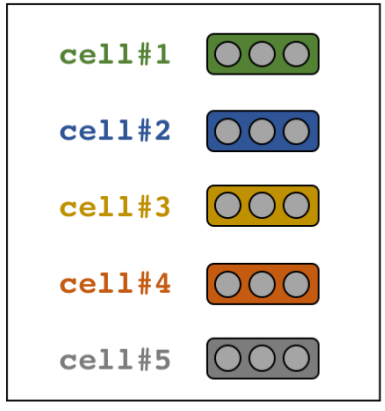
1. Initialize each **sense** vector as the average of **word** vectors used in the textual definition of this sense
2. **Disambiguate** each word of a large corpus on this basis
3. Joint learning of **word** & **sense** representations




(1) Get senses as defined by a sense inventory (e.g., WordNet)




(2) Gather information for each sense (e.g., by exploiting the structural properties of sense inventory's semantic network, and (optionally) then from text corpora)

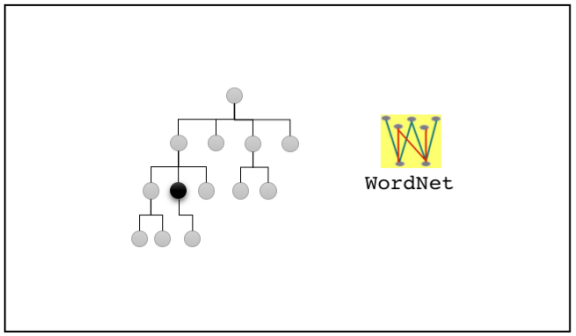



(3) Compute sense representation

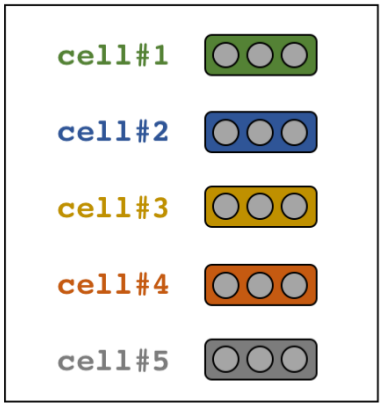
# Example

1. **cell#1** (jail_cell, prison_cell): a room where a prisoner is kept.
2. **cell#2** the basic structural and functional unit of all organisms.
3. **cell#3** (cellphone, mobile_phone): a hand-held mobile radiotelephone.
4. **cell#4** (electric_cell): a device that delivers an electric current.
5. **cell#5** (cubicle): small room in which a monk or nun lives.

(1) Get senses as defined by a sense inventory (e.g., WordNet)



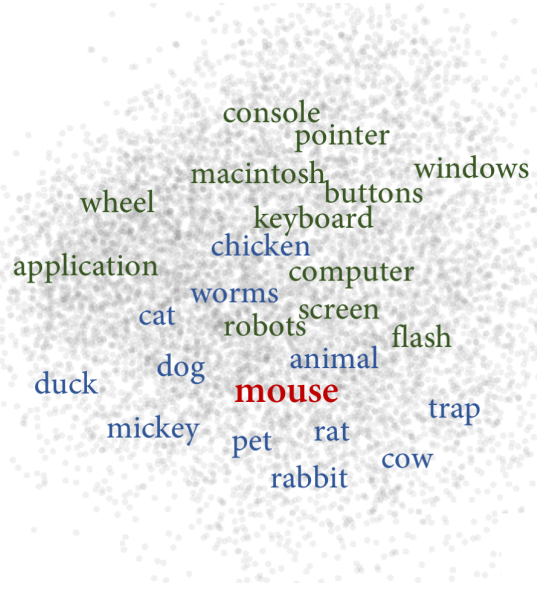

(2) Gather information for each sense (e.g., by exploiting the structural properties of sense inventory's semantic network, and (optionally) then from text corpora)




(3) Compute sense representation

# Summary

- Focus on 3 approaches to **sense** embedding
- Potential improvement by **graph techniques**
- Other approaches exist (e.g., multilingual)




"My mouse is broken"