Friedrich-Alexander-Universität
Technische Fakultät

Pattern Recognition Lab

Modeling Hippocampus in context of language

Philipp Rost

Friedrich-Alexander-Universität Erlangen-Nürnberg

July 1, 2022

Agenda

1. Hippocampus

2. Successor Representation

3. Goal of the thesis

Hippocampus

Hippocampus

- Has the shape of a seahorse

Hippocampus

- Has the shape of a seahorse
- Key factor in forming memories (not preserving them!)

Hippocampus

- Has the shape of a seahorse
- Key factor in forming memories (not preserving them!)
- Related to emotions

Hippocampus

- Has the shape of a seahorse
- Key factor in forming memories (not preserving them!)
- Related to emotions
- Responsible for any kind of navigation (e.g. ranking stuff like danger of animals)

Hippocampus

- Has the shape of a seahorse
- Key factor in forming memories (not preserving them!)
- Related to emotions
- Responsible for any kind of navigation (e.g. ranking stuff like danger of animals)
  - Crafts a **cognitive room** of the “surroundings” by using **place cells** and **grid cells**

Hippocampus

- Has the shape of a seahorse
- Key factor in forming memories (not preserving them!)
- Related to emotions
- Responsible for any kind of navigation (e.g. ranking stuff like danger of animals)
  - Crafts a **cognitive room** of the “surroundings” by using **place cells** and **grid cells**
  - Place cell: irregular arranged, fires at specific positions in space

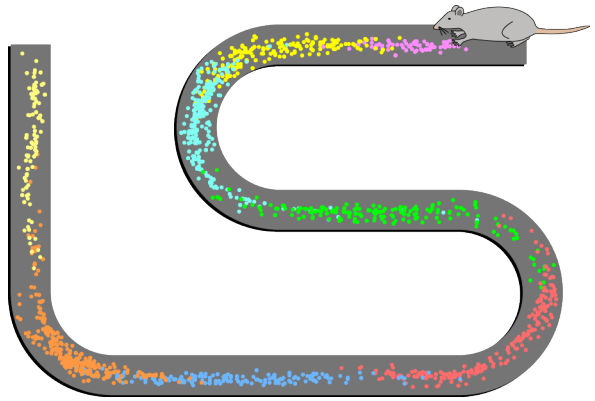


Figure: Different colors represent different place cells (= states)

Hippocampus

- Has the shape of a seahorse
- Key factor in forming memories (not preserving them!)
- Related to emotions
- Responsible for any kind of navigation (e.g. ranking stuff like danger of animals)
  - Crafts a **cognitive room** of the “surroundings” by using **place cells** and **grid cells**
  - Place cell: irregular arranged, fires at specific positions in space
  - Grid cell: lattice-like arranged, fires continuously

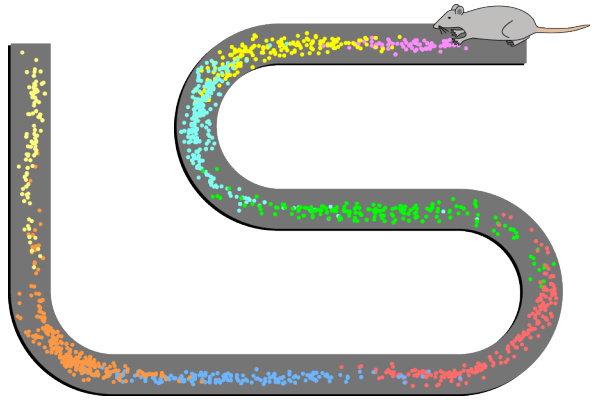


Figure: Different colors represent different place cells (= states)

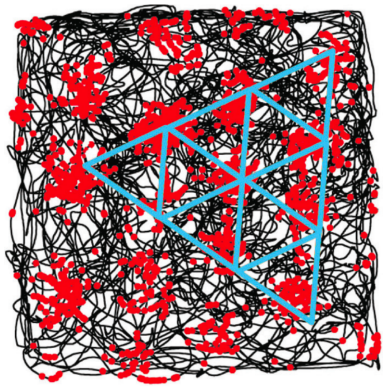


Figure: Grid cells can describes as triangulation

Successor Representation

Successor Representation

- Claim: Hippocampus learns a map representing each state as a successor state which is encoded via place cell

Successor Representation

- Claim: Hippocampus learns a map representing each state as a successor state which is encoded via place cell
  - Where does the claim originate?

Successor Representation

- Claim: Hippocampus learns a map representing each state as a successor state which is encoded via place cell
  - Where does the claim originate? The proposed technique works fine for spatial navigation.

Successor Representation

- Claim: Hippocampus learns a map representing each state as a successor state which is encoded via place cell
  - Where does the claim originate? The proposed technique works fine for spatial navigation.
- Mathematical principle roots in reinforcement learning (and transition probability matrices) and can be computed like

$$M_a = \sum_{t=0}^{a} \gamma^t T^t,$$

with discount factor $\gamma \in (0, 1)$ and transition probability matrix T

Successor Representation

- Claim: Hippocampus learns a map representing each state as a successor state which is encoded via place cell
  - Where does the claim originate? The proposed technique works fine for spatial navigation.
- Mathematical principle roots in reinforcement learning (and transition probability matrices) and can be computed like

$$M_a = \sum_{t=0}^{a} \gamma^t T^t,$$

  with discount factor $\gamma \in (0, 1)$ and transition probability matrix T
- The policy is encoded in matrix $T \Longrightarrow$ the SR is policy-dependent (it is based on RL)

Successor Representation

- Claim: Hippocampus learns a map representing each state as a successor state which is encoded via place cell
  - Where does the claim originate? The proposed technique works fine for spatial navigation.
- Mathematical principle roots in reinforcement learning (and transition probability matrices) and can be computed like

$$M_a = \sum_{t=0}^{a} \gamma^t T^t,$$

  with discount factor $\gamma \in (0, 1)$ and transition probability matrix T
- The policy is encoded in matrix $T \Longrightarrow$ the SR is policy-dependent (it is based on RL)
- Row i: resembles (all) successor states of state i (the higher the value the higher the chance to be in this state after the next step)

Successor Representation

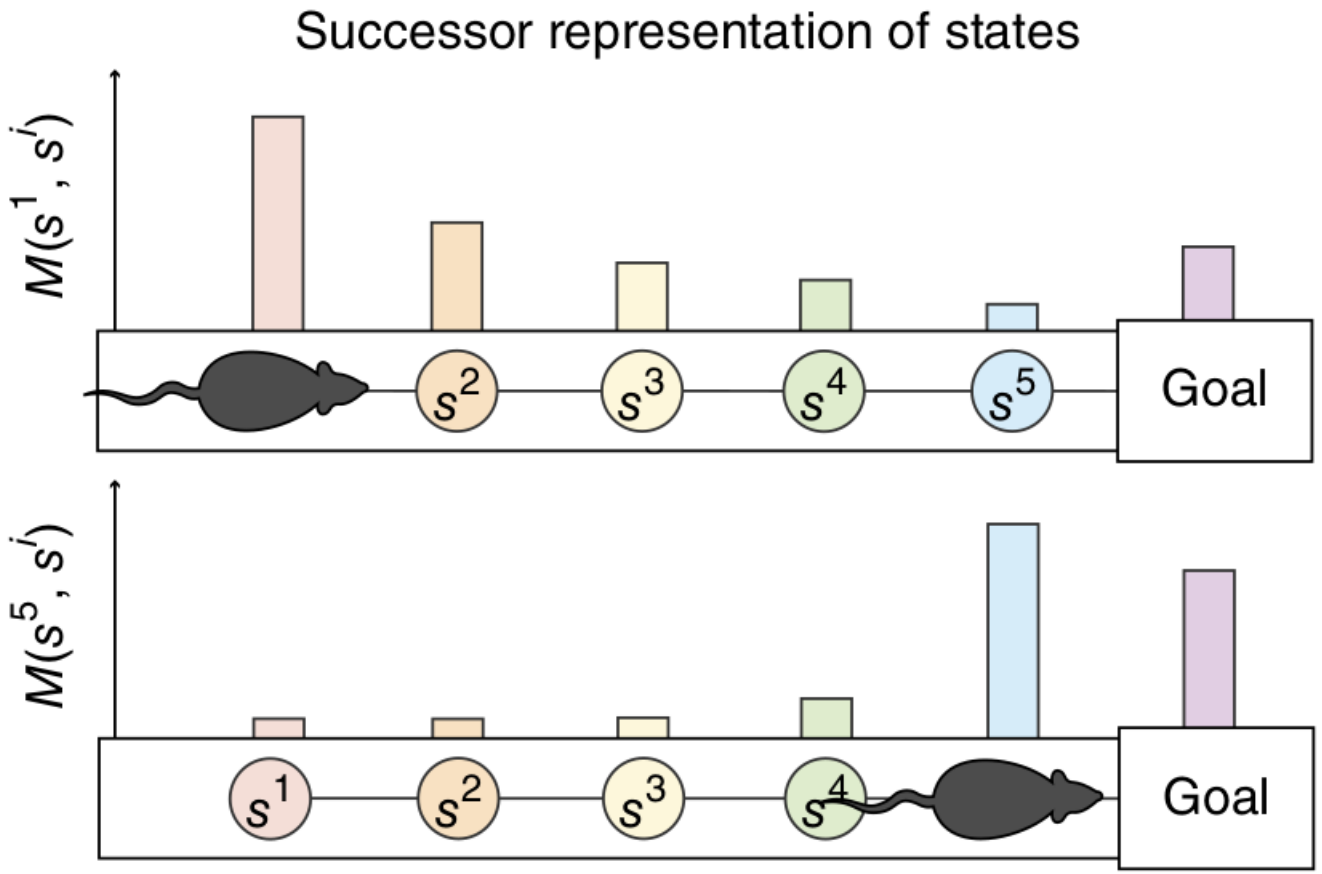
- Claim: Hippocampus learns a map representing each state as a successor state which is encoded via place cell
  - Where does the claim originate? The proposed technique works fine for spatial navigation.
- Mathematical principle roots in reinforcement learning (and transition probability matrices) and can be computed like

$$M_a = \sum_{t=0}^{a} \gamma^t T^t,$$

  with discount factor $\gamma \in (0, 1)$ and transition probability matrix T
- The policy is encoded in matrix $T \Longrightarrow$ the SR is policy-dependent (it is based on RL)
- Row *i*: resembles (all) successor states of state *i* (the higher the value the higher the chance to be in this state after the next step)
- Column *j*: High value in component *i* means state *j* is visited regularly after starting at state *i*

Successor Representation

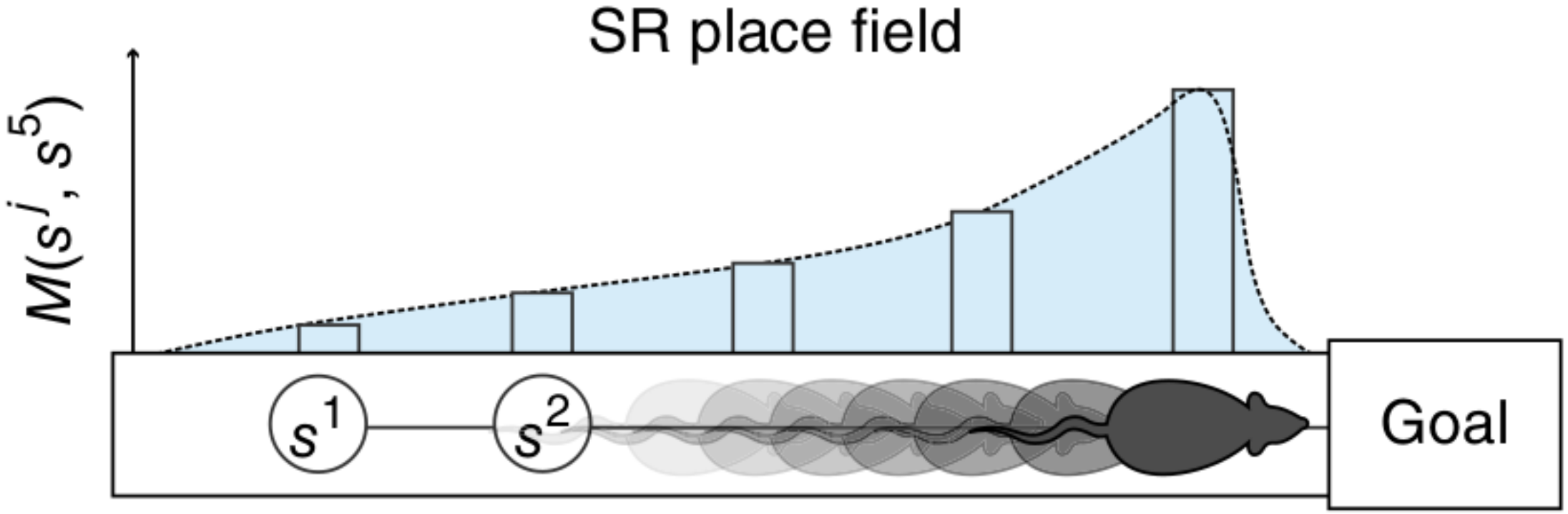
Example 1/2 of interpreting a SR matrix *M*



Superscript means index not power

Successor Representation

Example 2/2 of interpreting a SR matrix *M*



Goal of the thesis

Goal of the thesis

I try to answer: *Does the Hippocampus, i.e. place (and grid) cells, generate a cognitive room of languages?*

Goal of the thesis

I try to answer: *Does the Hippocampus, i.e. place (and grid) cells, generate a cognitive room of languages?*

Approach(es): Training a (shallow, supervised) neural network by...

Goal of the thesis

I try to answer: *Does the Hippocampus, i.e. place (and grid) cells, generate a cognitive room of languages?*

Approach(es): Training a (shallow, supervised) neural network by...

1. building a graphical model based on word classes

Goal of the thesis

I try to answer: *Does the Hippocampus, i.e. place (and grid) cells, generate a cognitive room of languages?*

Approach(es): Training a (shallow, supervised) neural network by...

1. building a graphical model based on word classes
   - Word classes with made up rules like “Pronoun $\rightarrow$ Verb”

Goal of the thesis

I try to answer: *Does the Hippocampus, i.e. place (and grid) cells, generate a cognitive room of languages?*

Approach(es): Training a (shallow, supervised) neural network by...

1. building a graphical model based on word classes
   - Word classes with made up rules like “Pronoun $\rightarrow$ Verb”
   - Building a cognitive room: List of all words

Goal of the thesis

I try to answer: *Does the Hippocampus, i.e. place (and grid) cells, generate a cognitive room of languages?*

Approach(es): Training a (shallow, supervised) neural network by...

1. building a graphical model based on word classes
   - Word classes with made up rules like “Pronoun $\rightarrow$ Verb”
   - Building a cognitive room: List of all words
   - For each rule two 1 hot encoded vectors related to the cognitive room are used for training (first vector as input, second as preferable output)

Goal of the thesis

I try to answer: *Does the Hippocampus, i.e. place (and grid) cells, generate a cognitive room of languages?*

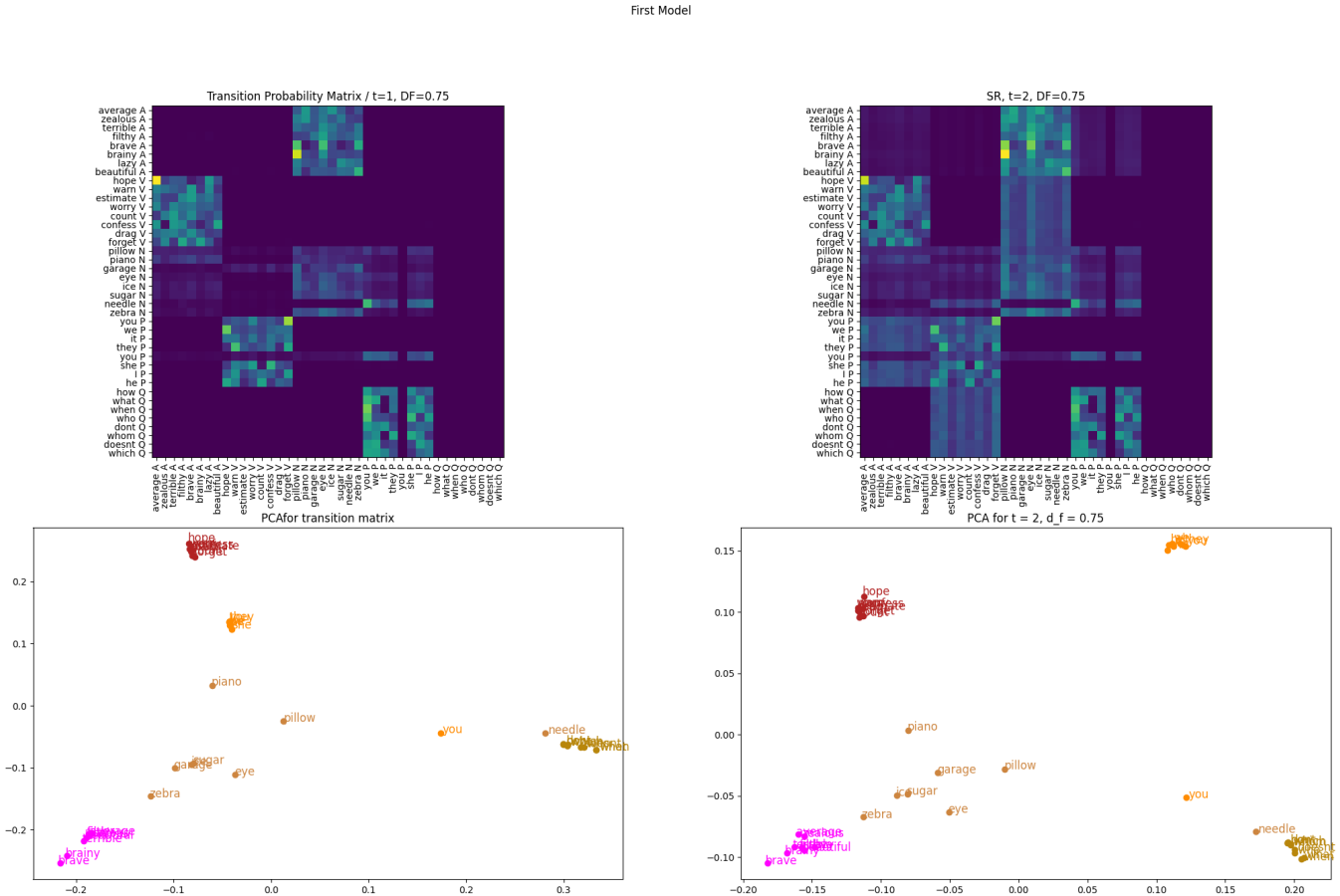
Approach(es): Training a (shallow, supervised) neural network by...

1. building a graphical model based on word classes
   - Word classes with made up rules like “Pronoun $\rightarrow$ Verb”
   - Building a cognitive room: List of all words
   - For each rule two 1 hot encoded vectors related to the cognitive room are used for training (first vector as input, second as preferable output)

   $\Rightarrow$ Works well due to graph environment

Example results for graphical model

Rules used: Question Word → Personal Pronoun, Adjective → Noun, Personal Pronoun → Verb & Verb → Adjective



Goal of the thesis

We try to answer the question: *Does the Hippocampus, i.e. place (and grid) cells, generate a cognitive map of languages?*

Approach(es): Training a (shallow, supervised) neural network by...

1. building a graphical model based on word classes
   - Word classes with made up rules like “Pronoun $\rightarrow$ Verb”
   - Building a cognitive room: List of all words
   - For each rule two 1 hot encoded vectors related to the cognitive room are used for training (first vector as input, second as preferable output)

   $\Rightarrow$ Works well due to graph environment

Goal of the thesis

We try to answer the question: *Does the Hippocampus, i.e. place (and grid) cells, generate a cognitive map of languages?*

Approach(es): Training a (shallow, supervised) neural network by...

1. building a graphical model based on word classes
   - Word classes with made up rules like “Pronoun $\rightarrow$ Verb”
   - Building a cognitive room: List of all words
   - For each rule two 1 hot encoded vectors related to the cognitive room are used for training (first vector as input, second as preferable output)

   $\Rightarrow$ Works well due to graph environment
2. Parsing text to use two following words as training data by...

Goal of the thesis

We try to answer the question: *Does the Hippocampus, i.e. place (and grid) cells, generate a cognitive map of languages?*

Approach(es): Training a (shallow, supervised) neural network by...

1. building a graphical model based on word classes
   - Word classes with made up rules like “Pronoun $\rightarrow$ Verb”
   - Building a cognitive room: List of all words
   - For each rule two 1 hot encoded vectors related to the cognitive room are used for training (first vector as input, second as preferable output)

   $\Rightarrow$ Works well due to graph environment
2. Parsing text to use two following words as training data by...
   - Using 1 hot encoded vectors but without made up rules. Input & output are determined by running order

Goal of the thesis

We try to answer the question: *Does the Hippocampus, i.e. place (and grid) cells, generate a cognitive map of languages?*

Approach(es): Training a (shallow, supervised) neural network by...

1. building a graphical model based on word classes
   - Word classes with made up rules like “Pronoun $\rightarrow$ Verb”
   - Building a cognitive room: List of all words
   - For each rule two 1 hot encoded vectors related to the cognitive room are used for training (first vector as input, second as preferable output)

   $\Rightarrow$ Works well due to graph environment
2. Parsing text to use two following words as training data by...
   - Using 1 hot encoded vectors but without made up rules. Input & output are determined by running order
   - Calculating word vectors using `spacy`

Goal of the thesis

We try to answer the question: *Does the Hippocampus, i.e. place (and grid) cells, generate a cognitive map of languages?*

Approach(es): Training a (shallow, supervised) neural network by...

1. building a graphical model based on word classes
   - Word classes with made up rules like “Pronoun $\rightarrow$ Verb”
   - Building a cognitive room: List of all words
   - For each rule two 1 hot encoded vectors related to the cognitive room are used for training (first vector as input, second as preferable output)

   $\Rightarrow$ Works well due to graph environment
2. Parsing text to use two following words as training data by...
   - Using 1 hot encoded vectors but without made up rules. Input & output are determined by running order
   - Calculating word vectors using `spacy`
     - `spacy` can (bidirectionally) transfer between words and vectors

Goal of the thesis

We try to answer the question: *Does the Hippocampus, i.e. place (and grid) cells, generate a cognitive map of languages?*

Approach(es): Training a (shallow, supervised) neural network by...

1. building a graphical model based on word classes
   - Word classes with made up rules like “Pronoun $\rightarrow$ Verb”
   - Building a cognitive room: List of all words
   - For each rule two 1 hot encoded vectors related to the cognitive room are used for training (first vector as input, second as preferable output)

   $\Rightarrow$ Works well due to graph environment
2. Parsing text to use two following words as training data by...
   - Using 1 hot encoded vectors but without made up rules. Input & output are determined by running order
   - Calculating word vectors using `spacy`
     - `spacy` can (bidirectionally) transfer between words and vectors

   $\Rightarrow$ Works ...not really

Example of word vector model