# Siamese recurrent networks can learn first-order logic reasoning and exhibit zero-shot generalization to novel expressions

Mathijs Mul  Jelle Zuidema

May 3, 2019

# Introduction

- Entailment Recognition

# Introduction

- Entailment Recognition
- What is the logical relation between a pair of expressions?

# Introduction

- Entailment Recognition
- What is the logical relation between a pair of expressions?
  - Equivalence?
    - *Some people are no logicians*
    - *Not all people are logicians*

# Introduction

- Entailment Recognition
- What is the logical relation between a pair of expressions?
  - Equivalence?
    - *Some people are no logicians*
    - *Not all people are logicians*
  - Contradiction?
    - *Some people are no logicians*
    - *All people are logicians*

# Introduction

- Entailment Recognition
- What is the logical relation between a pair of expressions?
  - Equivalence?
    - *Some people are no logicians*
    - *Not all people are logicians*
  - Contradiction?
    - *Some people are no logicians*
    - *All people are logicians*
  - Independence?
    - *Some people are no logicians*
    - *The moon is made of green cheese*

# Introduction

- Entailment Recognition
- What is the logical relation between a pair of expressions?
  - Equivalence?
    - *Some people are no logicians*
    - *Not all people are logicians*
  - Contradiction?
    - *Some people are no logicians*
    - *All people are logicians*
  - Independence?
    - *Some people are no logicians*
    - *The moon is made of green cheese*
  - ...

# Motivation

- Are recurrent neural networks capable of inferring entailment relations?

# Motivation

- Are recurrent neural networks capable of inferring entailment relations?
- How do they perform in relation to symbolically guided models?
  - In particular: the recursive neural networks used by Bowman e.a. in 'Recursive neural networks can learn logical semantics' (2015)

# Motivation

- Are recurrent neural networks capable of inferring entailment relations?
- How do they perform in relation to symbolically guided models?
  - In particular: the recursive neural networks used by Bowman e.a. in 'Recursive neural networks can learn logical semantics' (2015)
- Do they apply compositional generalization, or just memorization?

# Approach

1. Define an artificial language $\mathcal{L}$

# Approach

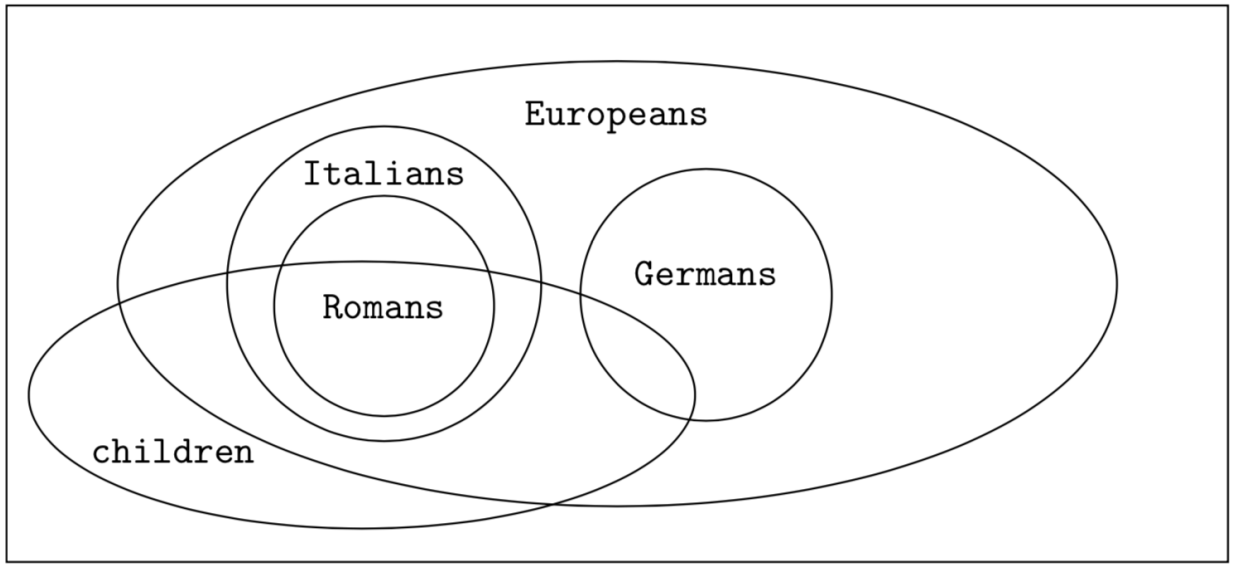
1. Define an artificial language $\mathcal{L}$
2. Deduce entailment relations between random pairs of sentences in $\mathcal{L}$, using first-order logic

# Approach

1. Define an artificial language $\mathcal{L}$
2. Deduce entailment relations between random pairs of sentences in $\mathcal{L}$, using first-order logic
3. Test the models on the data thus generated

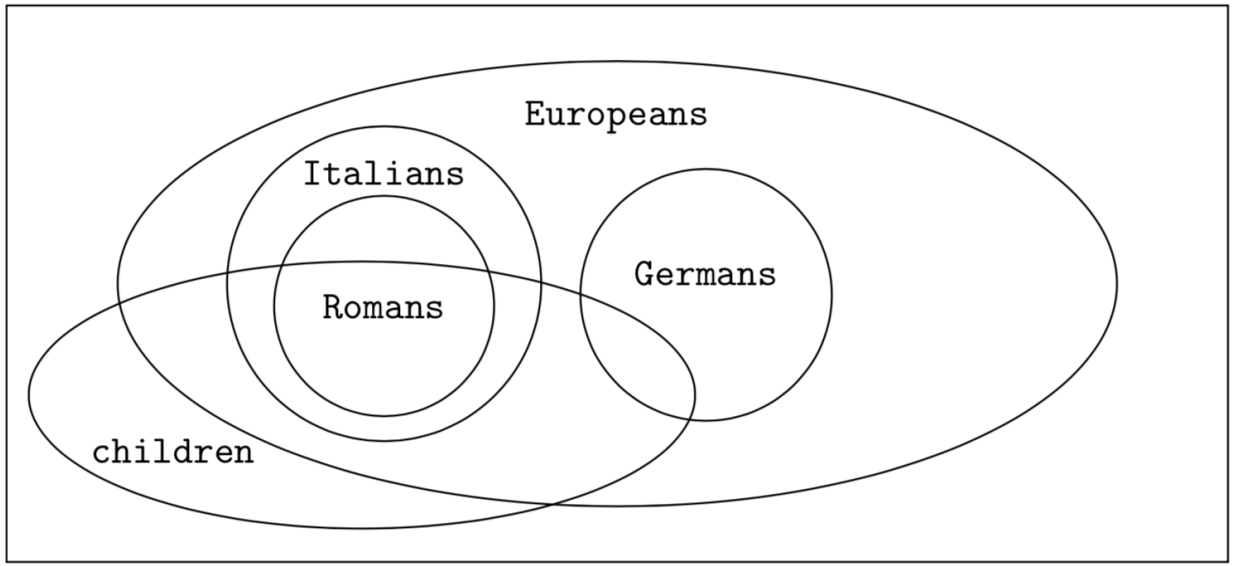
# Language $\mathcal{L}$

- Asymmetric taxonomy of nouns:

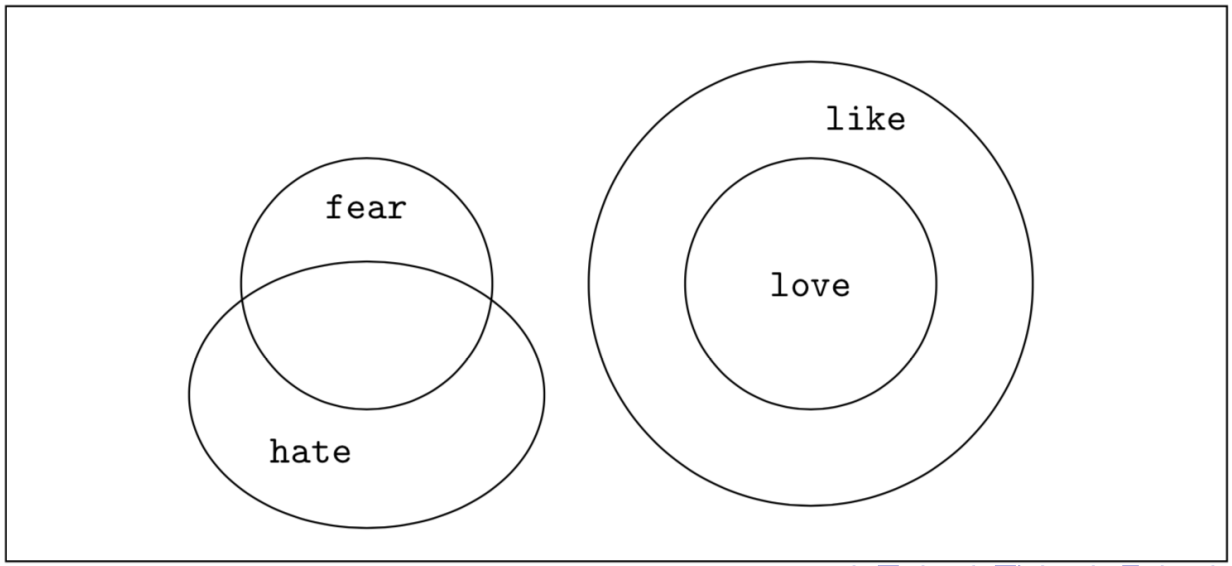


# Language $\mathcal{L}$

- Asymmetric taxonomy of nouns:




- Binary predicates:



# Language $\mathcal{L}$

- Quantifiers: $\{\texttt{all}, \texttt{some}\}$

# Language $\mathcal{L}$

- Quantifiers: $\{\texttt{all}, \texttt{some}\}$
- Adverbs: $\{\texttt{not}, \epsilon\}$

# Data sample

```
< ( ( all Europeans ) ( hate ( all Germans ) ) )
      ( ( all Romans ) ( hate ( some Europeans ) ) )
```

# Data sample

```
< ( ( all Europeans ) ( hate ( all Germans ) ) )
      ( ( all Romans ) ( hate ( some Europeans ) ) )

> ( ( some children ) ( like ( all Germans ) ) )
      ( ( all children ) ( love ( all Germans ) ) )
```

# Data sample

```
< ( ( all Europeans ) ( hate ( all Germans ) ) )
      ( ( all Romans ) ( hate ( some Europeans ) ) )

> ( ( some children ) ( like ( all Germans ) ) )
      ( ( all children ) ( love ( all Germans ) ) )

= ( ( some Italians ) ( ( not like ) ( all Romans ) ) )
      ( ( ( not all ) Italians ) ( like ( all Romans ) ) )
```

# Data sample

```
< ( ( all Europeans ) ( hate ( all Germans ) ) )
     ( ( all Romans ) ( hate ( some Europeans ) ) )

> ( ( some children ) ( like ( all Germans ) ) )
     ( ( all children ) ( love ( all Germans ) ) )

= ( ( some Italians ) ( ( not like ) ( all Romans ) ) )
     ( ( ( not all ) Italians ) ( like ( all Romans ) ) )

# ( ( ( not some ) Romans ) ( fear ( all children ) ) )
     ( ( all Germans ) ( like ( some ( not Europeans ) ) ) )
```

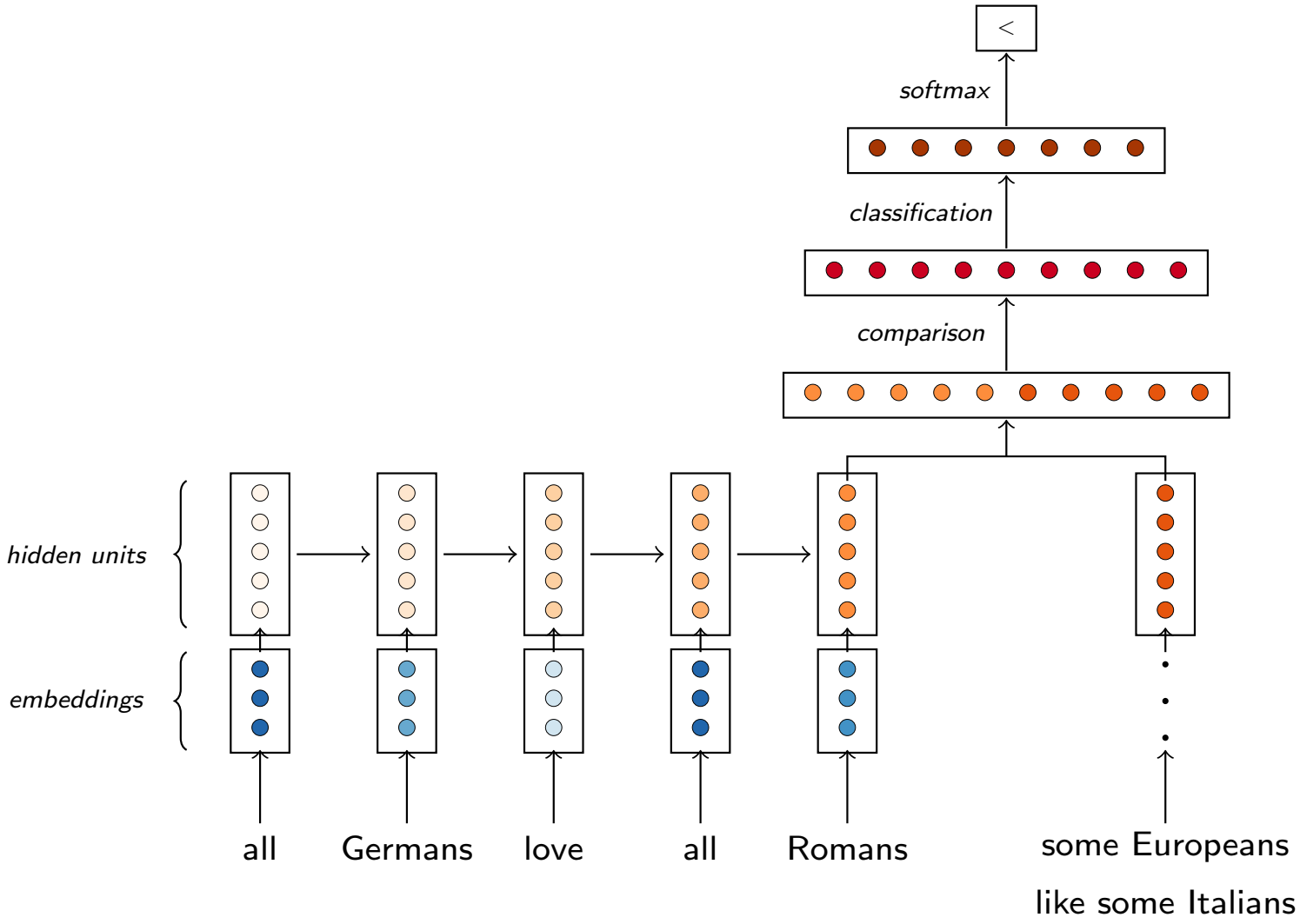
# Data sample

```
< ( ( all Europeans ) ( hate ( all Germans ) ) )
     ( ( all Romans ) ( hate ( some Europeans ) ) )

> ( ( some children ) ( like ( all Germans ) ) )
     ( ( all children ) ( love ( all Germans ) ) )

= ( ( some Italians ) ( ( not like ) ( all Romans ) ) )
     ( ( ( not all ) Italians ) ( like ( all Romans ) ) )

# ( ( ( not some ) Romans ) ( fear ( all children ) ) )
     ( ( all Germans ) ( like ( some ( not Europeans ) ) ) )
```

*Only 0.07% of the data space is seen during training.*

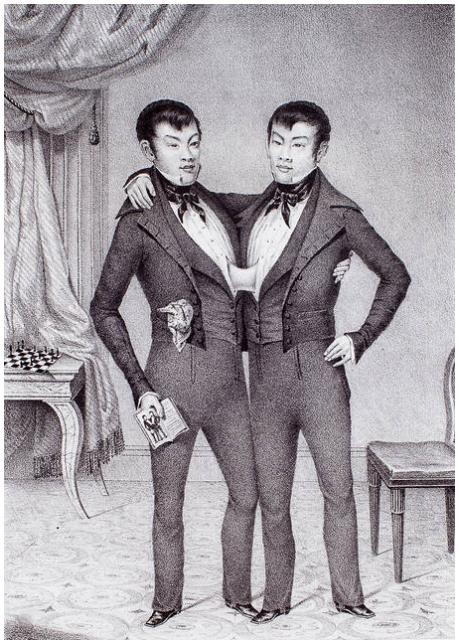
# Models

- Recurrent network



# Models

- Recurrent network

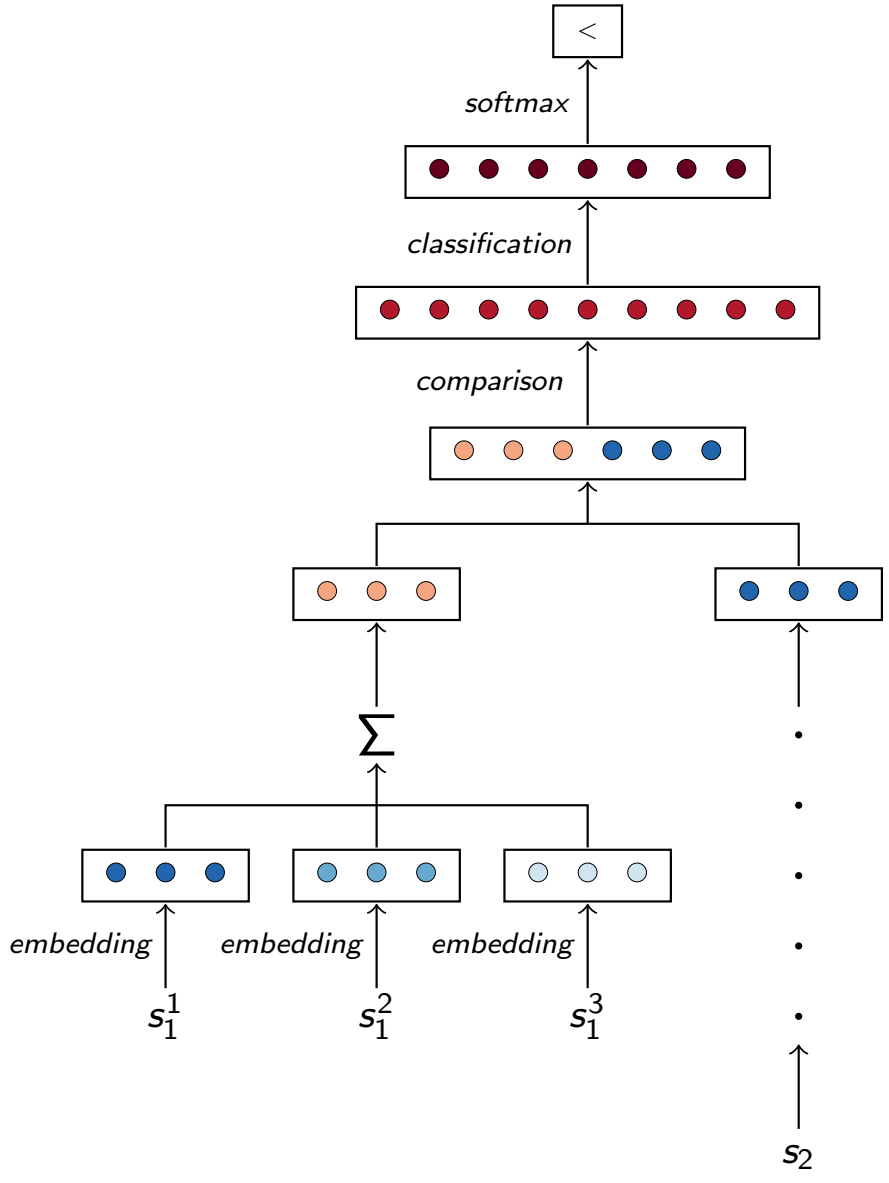


# Models

- Recurrent network
- Three types:
  - Simple Recurrent Network (SRN)
  - Gated Recurrent Unit (GRU)
  - Long Short-Term Memory (LSTM)

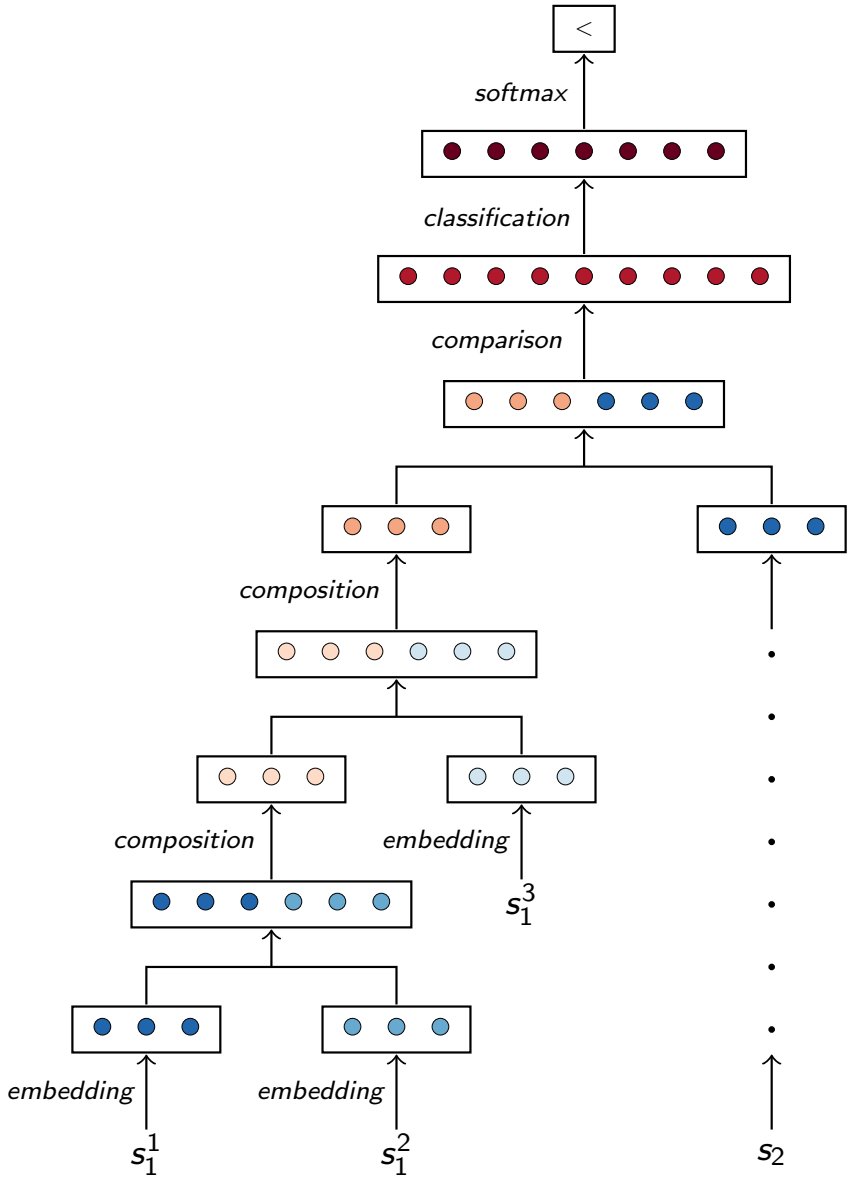
# Models

- Baseline 1: Summing Neural Network (sumNN)



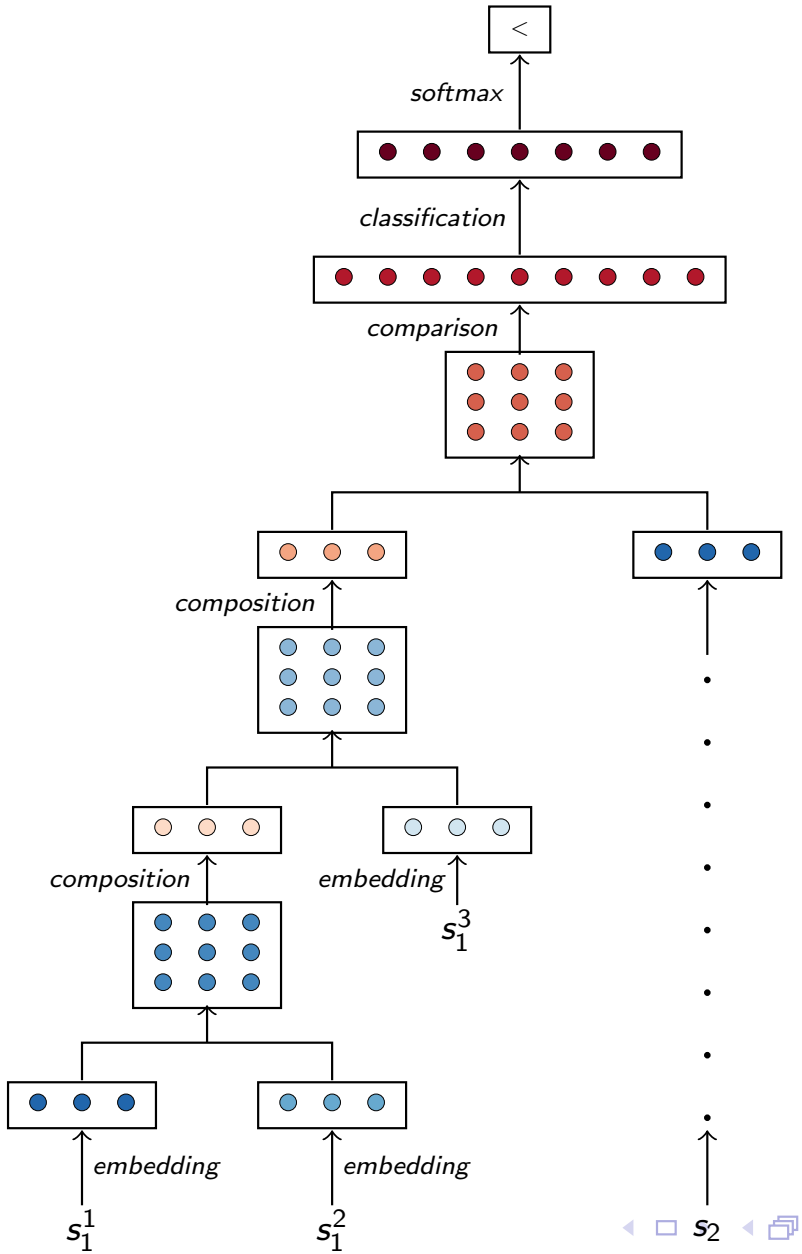
# Models

- Baseline 2: Tree-Shaped Neural (Matrix) Networks (tRNN)

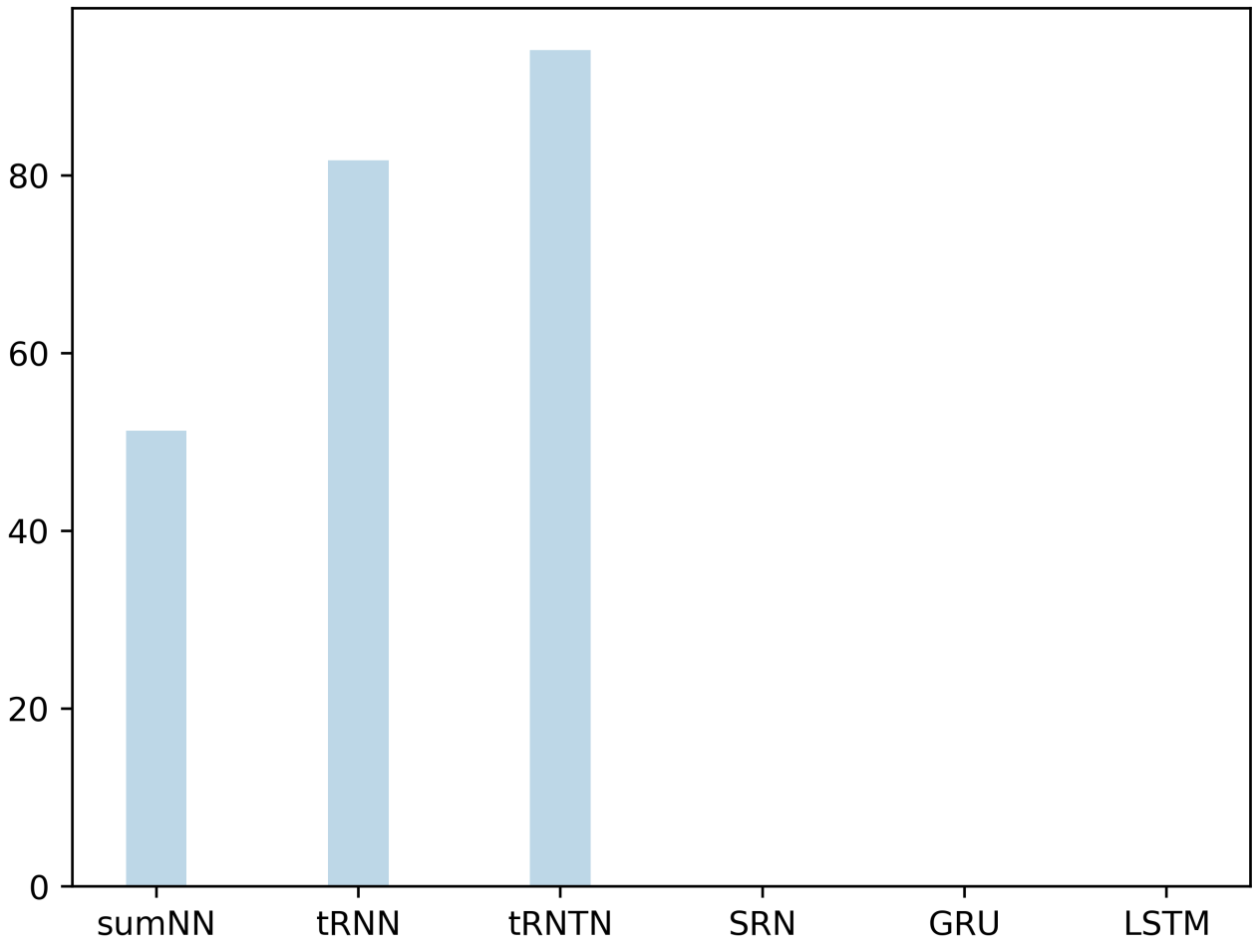


# Models

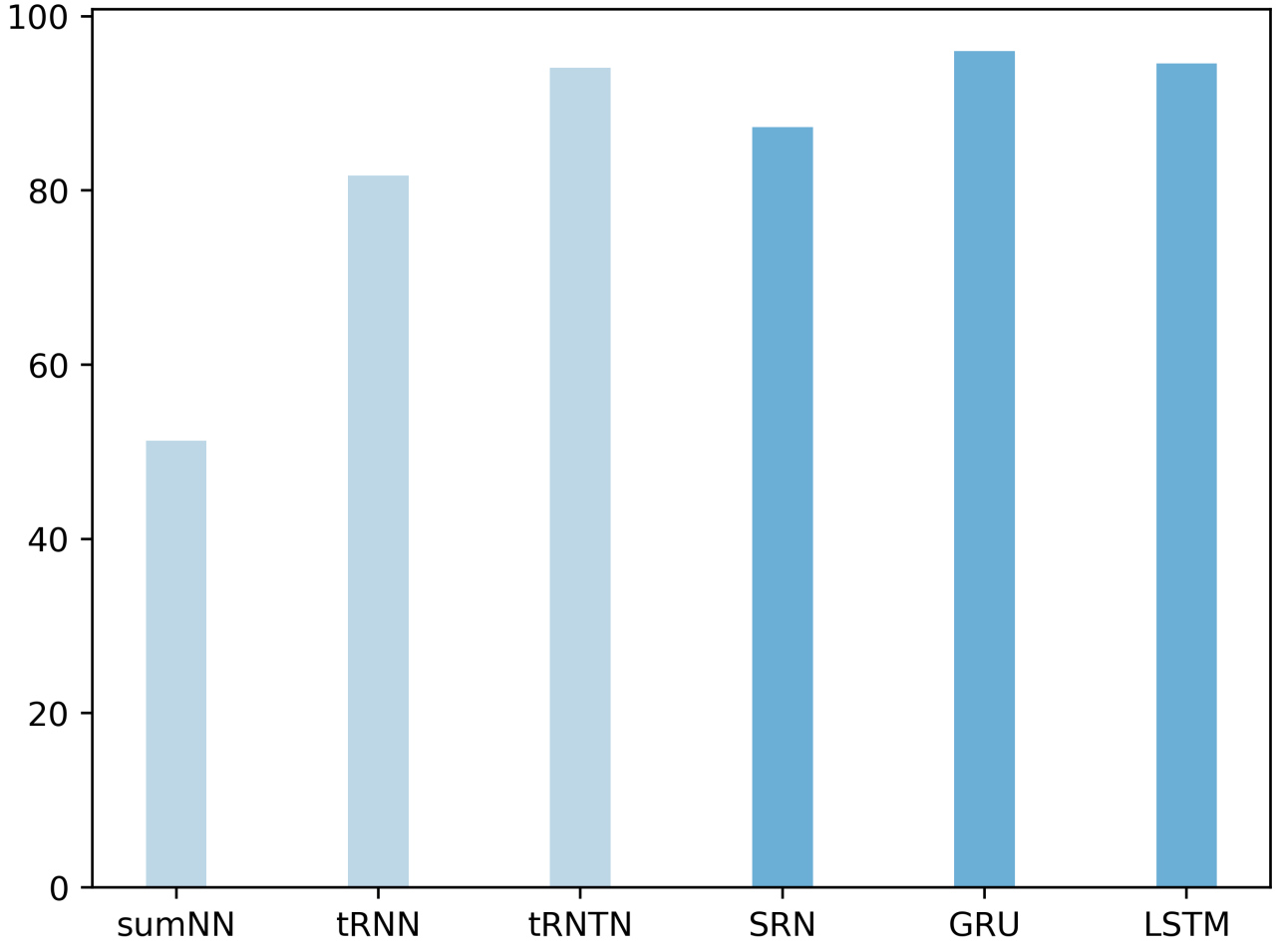
- Baseline 3: Tree-Shaped Neural Tensor Networks (tRNTN)



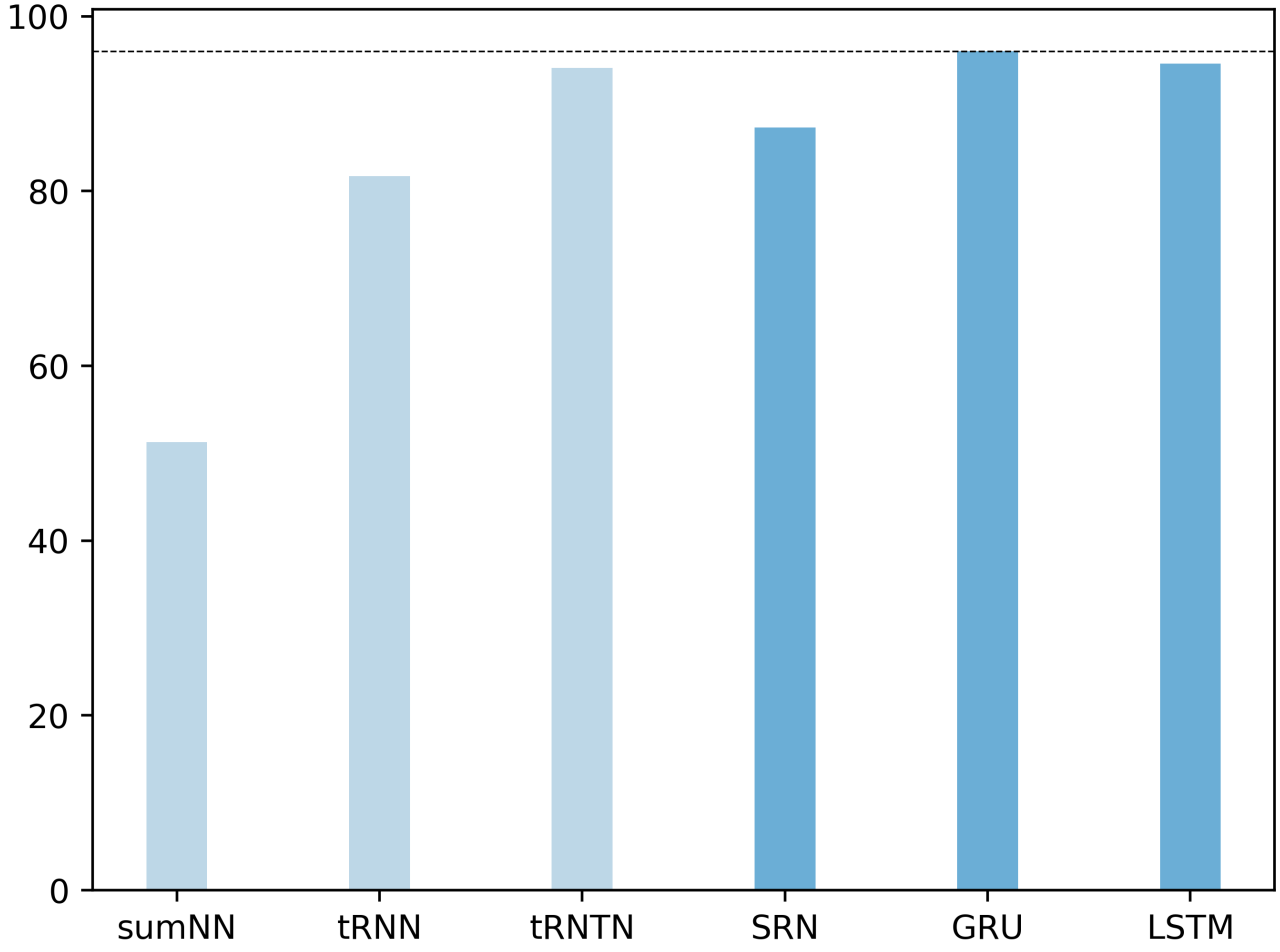
# Testing accuracy



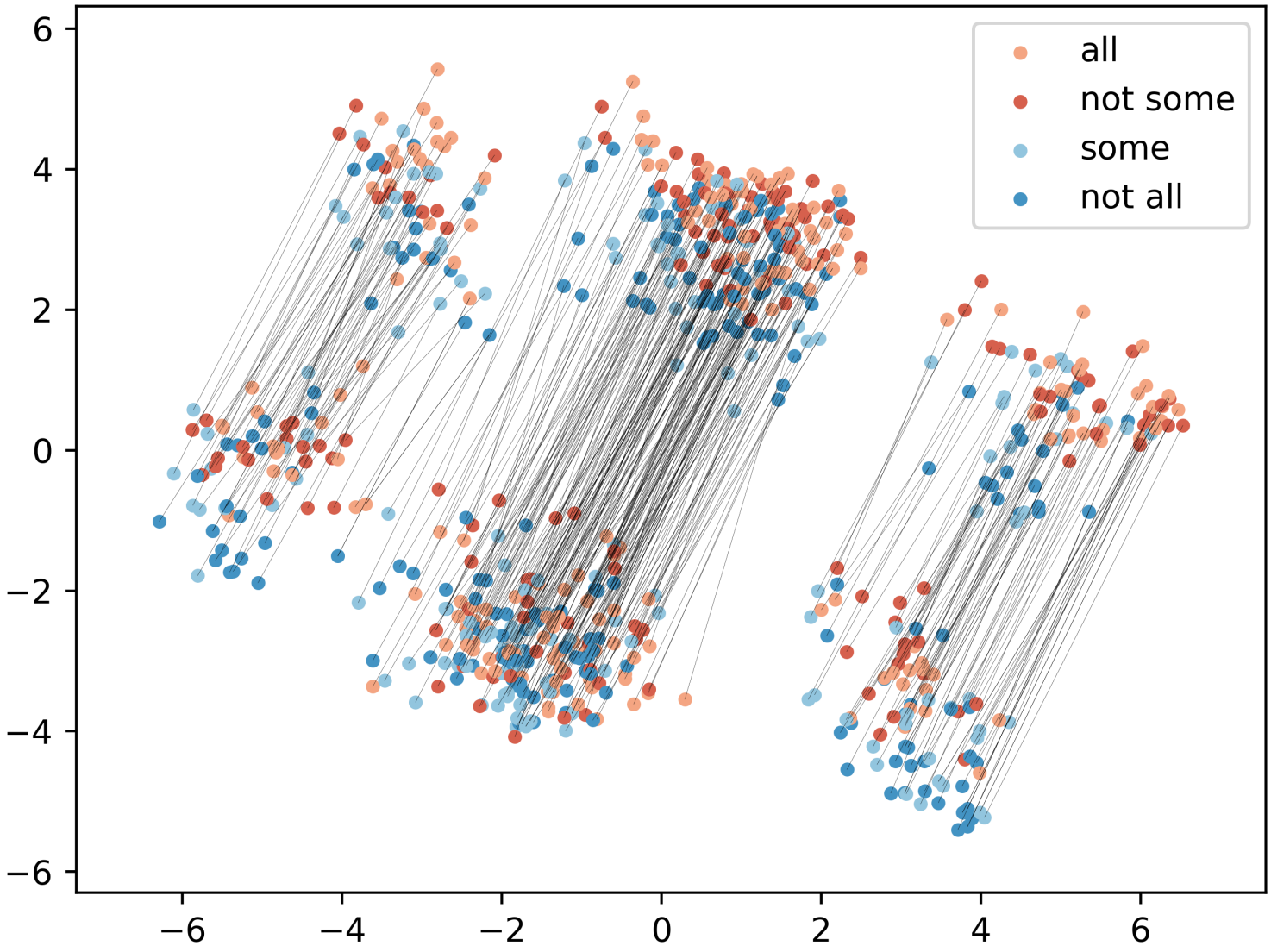
# Testing accuracy



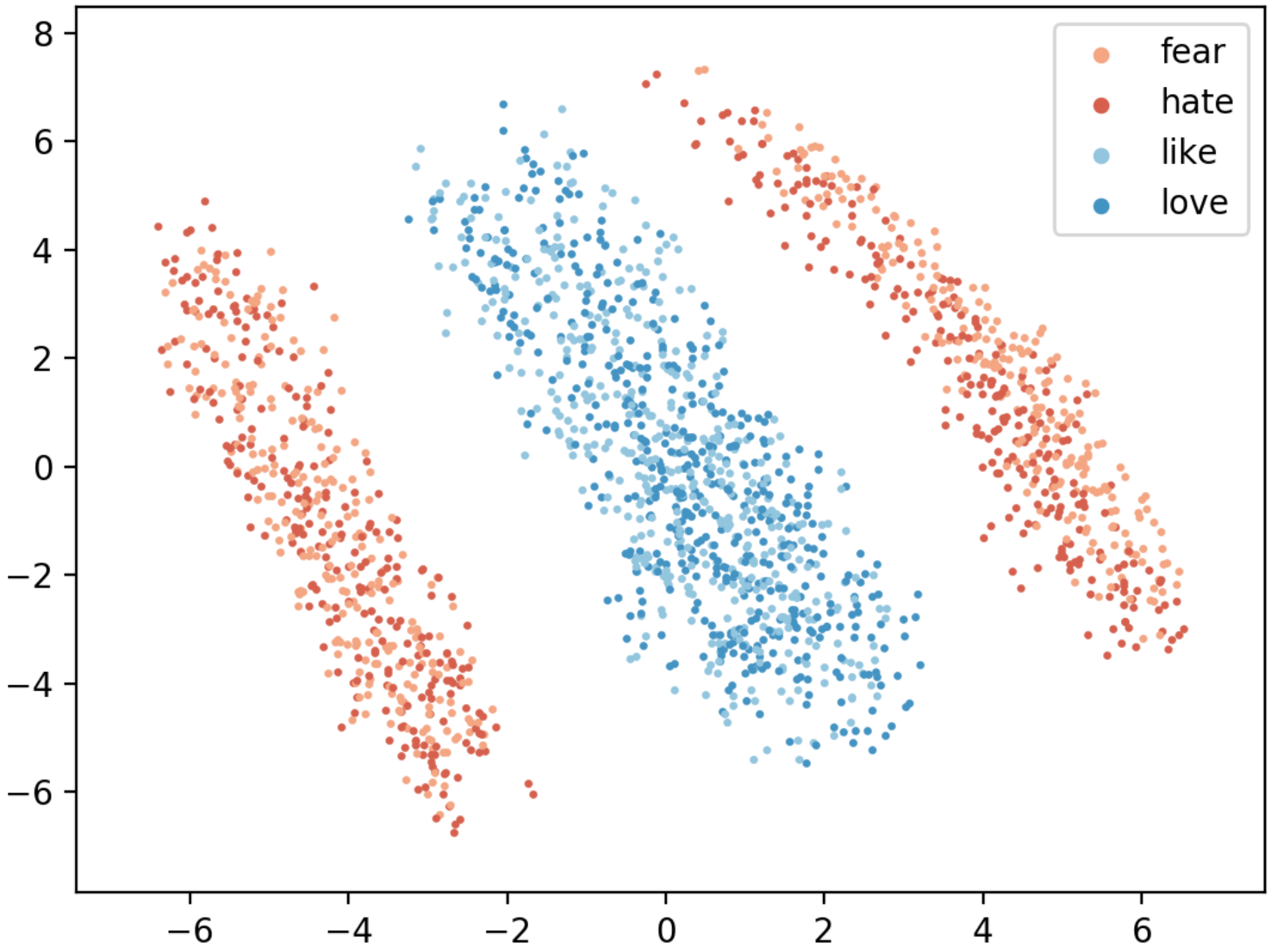
# Testing accuracy



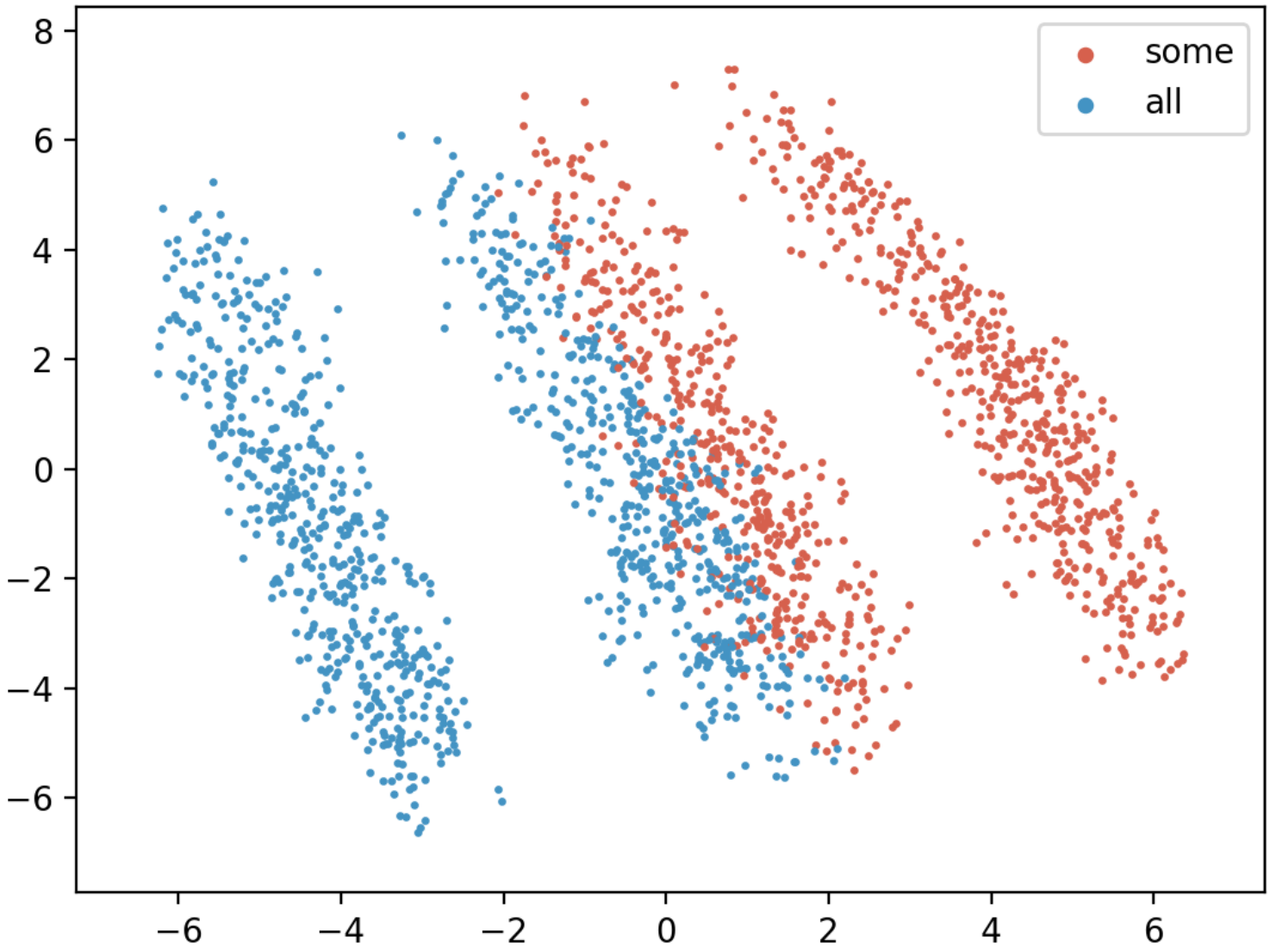
# Negated sentence vectors (GRU)



# Sentence vectors cluster according to verb...



# ... and second quantifier



# Interpretation

- What is happening in the hidden units?

# Interpretation

- What is happening in the hidden units?
- Diagnostic classification on best-performing GRU suggests awareness of:

# Interpretation

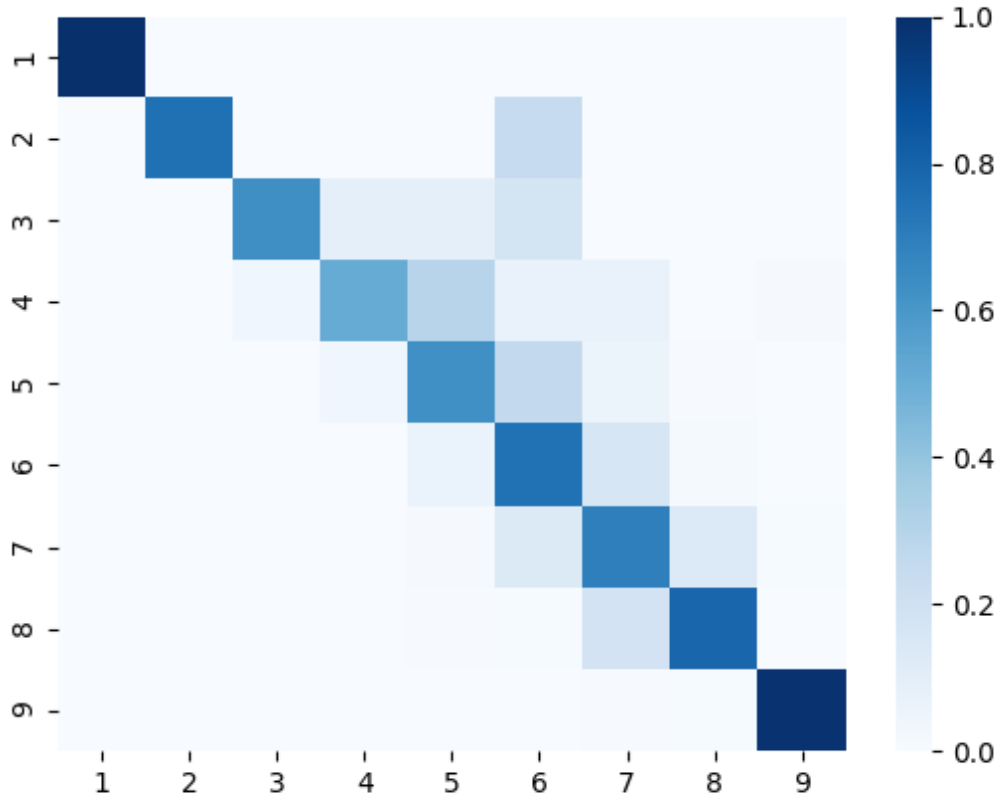
- What is happening in the hidden units?
- Diagnostic classification on best-performing GRU suggests awareness of:
  - Semantic type

# Interpretation

- What is happening in the hidden units?
- Diagnostic classification on best-performing GRU suggests awareness of:
  - Semantic type
  - Recursive depth

# Interpretation

- What is happening in the hidden units?
- Diagnostic classification on best-performing GRU suggests awareness of:
  - Semantic type
  - Recursive depth
  - Position in sentence



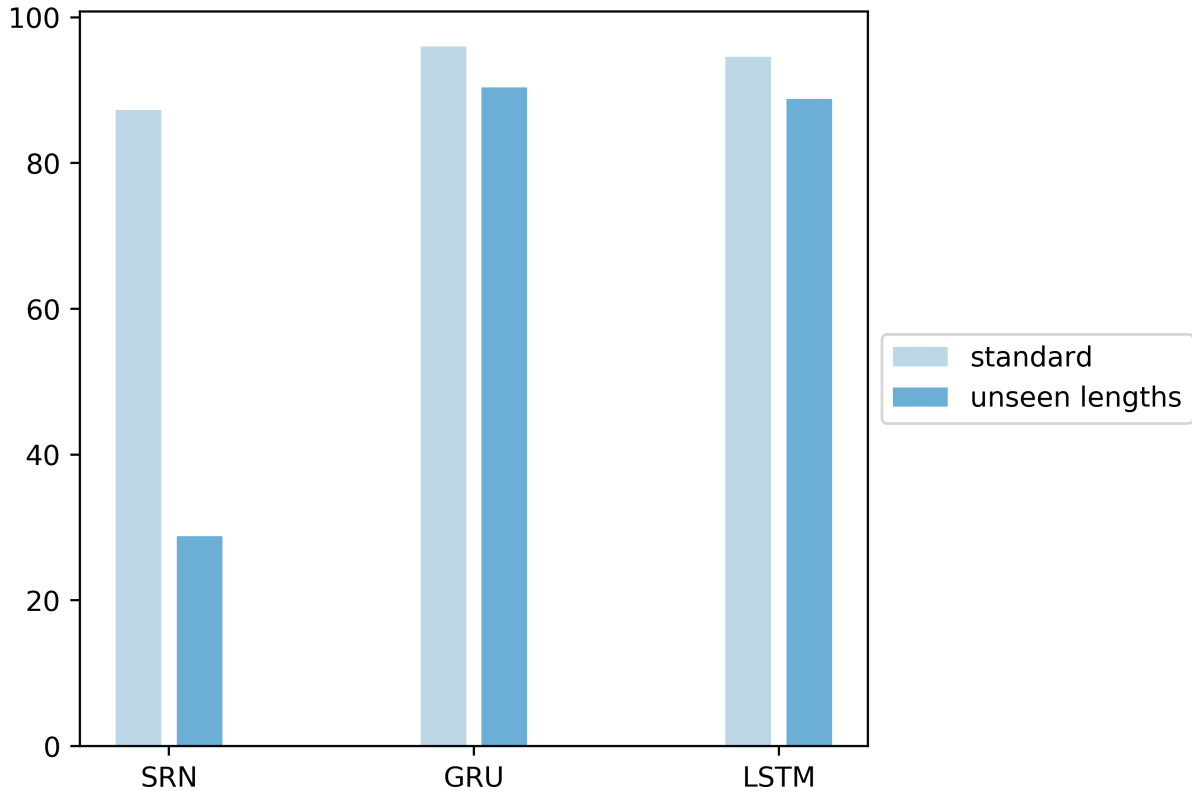
# Compositional learning

# Compositional learning

- Generalization to unseen lengths
  - Training on sentences of lengths 5, 7 or 8
  - Testing on sentences of lengths 6 or 9

# Compositional learning

- Generalization to unseen lengths
  - Training on sentences of lengths 5, 7 or 8
  - Testing on sentences of lengths 6 or 9
  - Testing accuracy:



# Compositional learning

# Compositional learning

- One-shot learning
  - Train GRU with fixed GloVe embeddings
  - At testing time, replace words in data with unseen ones, and add corresponding word embeddings to models

# Compositional learning

- One-shot learning
  - Synonyms:

# Compositional learning

- One-shot learning
  - Synonyms:
    - children $\rightarrow$ kids

# Compositional learning

- One-shot learning
  - Synonyms:
    - children $\rightarrow$ kids
    - love $\rightarrow$ adore

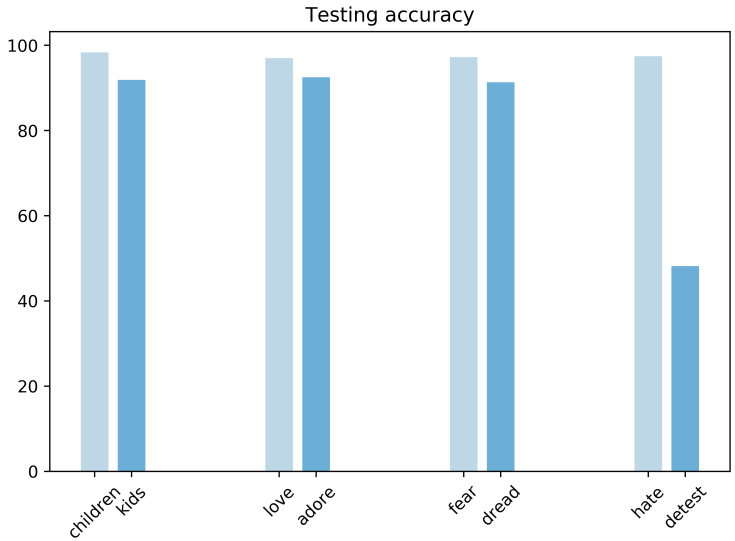
# Compositional learning

- One-shot learning
  - Synonyms:
    - children $\rightarrow$ kids
    - love $\rightarrow$ adore
    - fear $\rightarrow$ dread

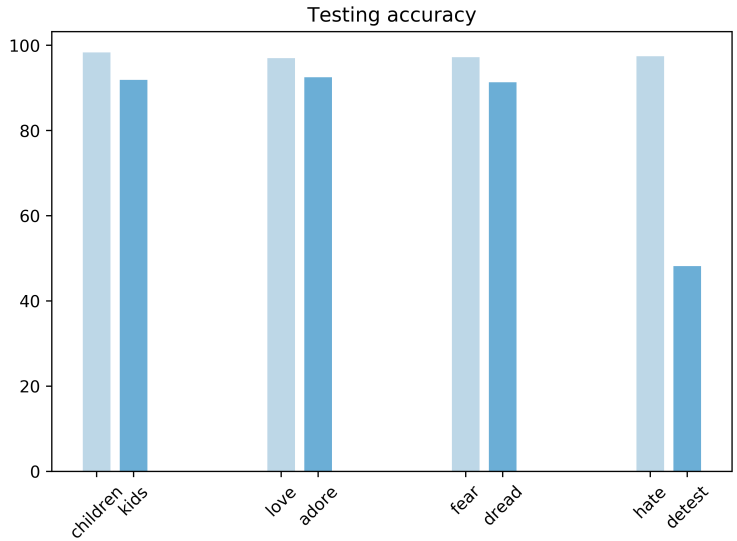
# Compositional learning

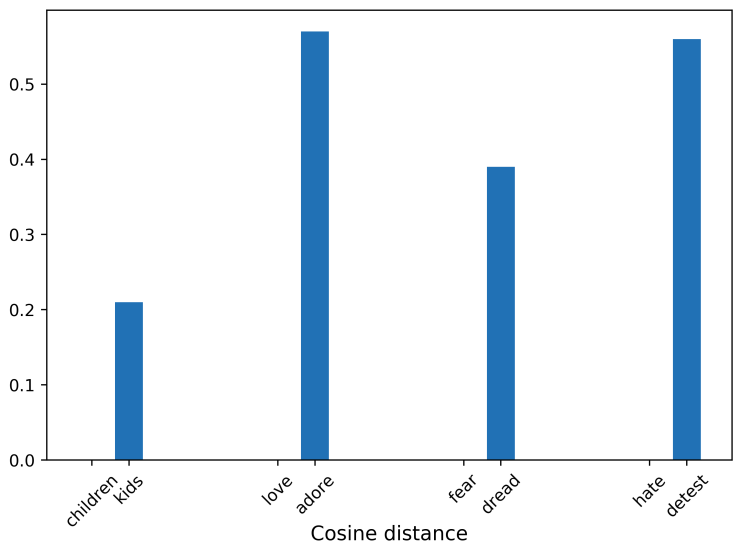
- One-shot learning
  - Synonyms:
    - children $\rightarrow$ kids
    - love $\rightarrow$ adore
    - fear $\rightarrow$ dread
    - hate $\rightarrow$ detest

# Synonyms


Testing accuracy
100
80
60
40
20
0
children
kids
love
adore
fear
dread
hate
detest

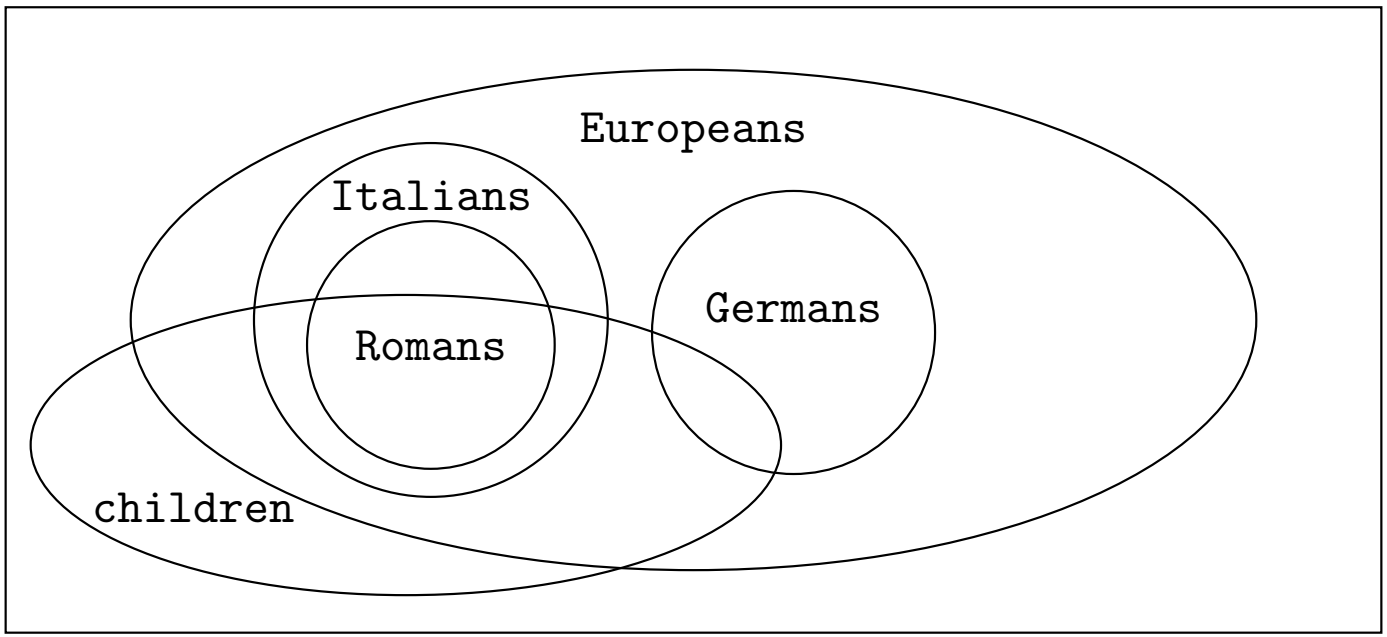
# Synonyms


Testing accuracy
100
80
60
40
20
0
children
kids
love
adore
fear
dread
hate
detest


0.5
0.4
0.3
0.2
0.1
0.0
children
kids
love
adore
fear
dread
hate
detest
Cosine distance

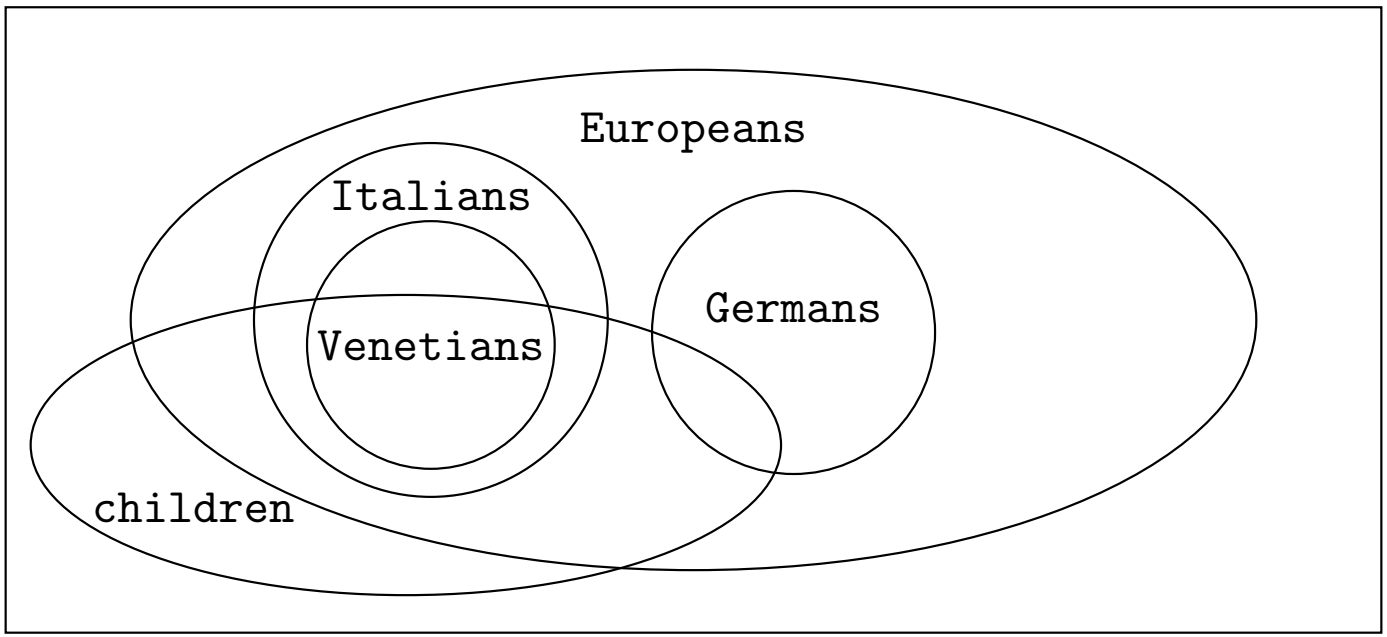
# Compositional learning

- One-shot learning
  - Ontological twins



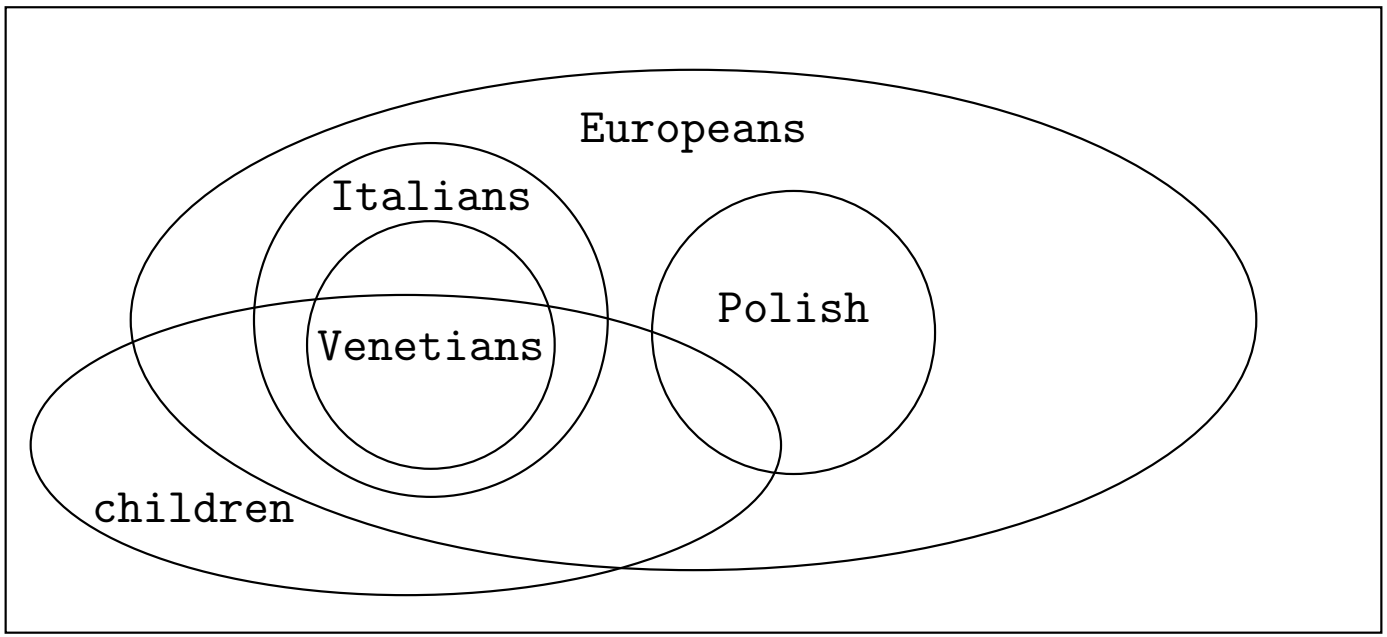
# Compositional learning

- One-shot learning
  - Ontological twins



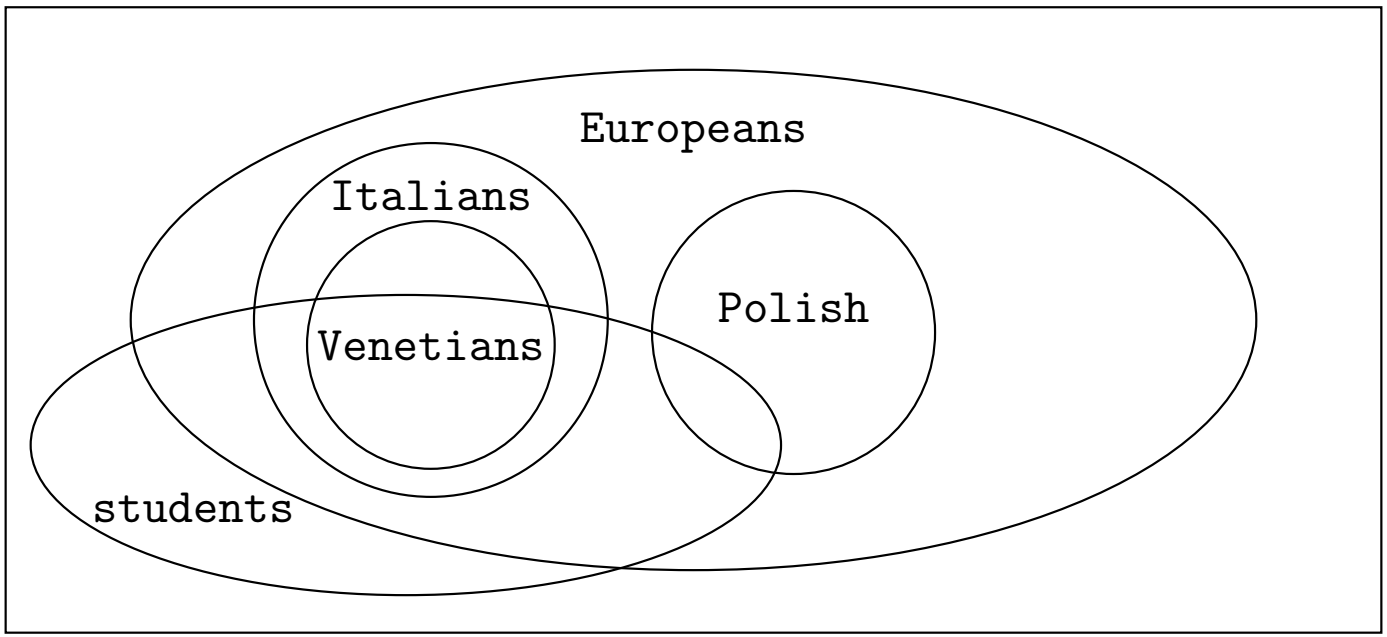
# Compositional learning

- One-shot learning
  - Ontological twins

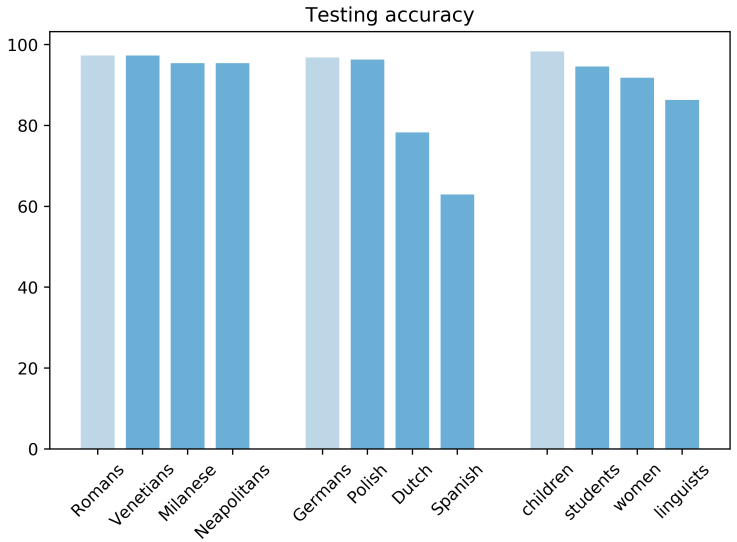


# Compositional learning

- One-shot learning
  - Ontological twins



# Ontological twins



# Ontological twins