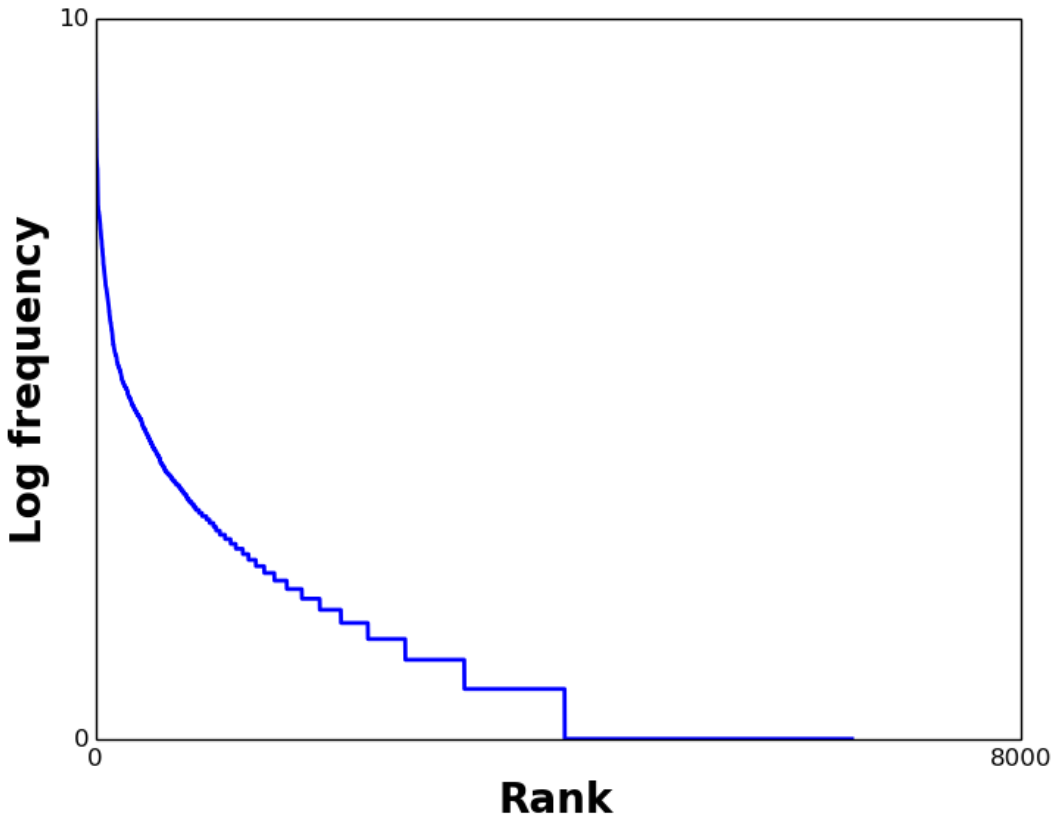
How may neural networks process hierarchical structure? Insights from recursive and recurrent networks learning arithmetics

Dieuwke Hupkes

Institute for Logic, Language and Computation
University of Amsterdam

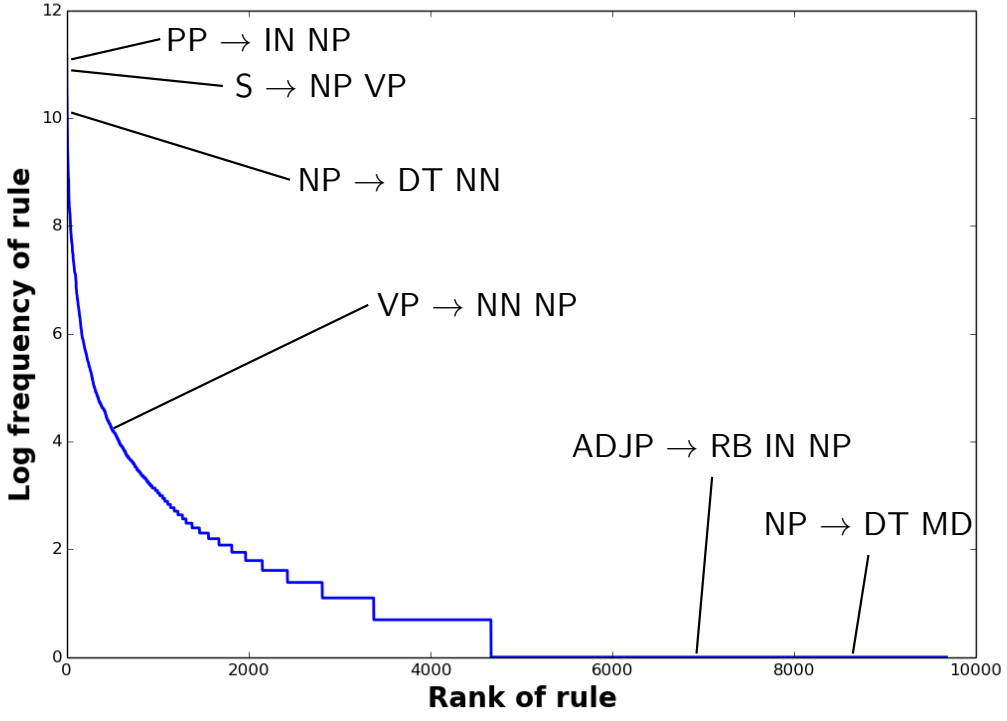
May 4, 2017

Few words are extremely frequent, many words are extremely infrequent



Corpus: CHILDES, Brown: Mother of Adam, Eve and Sara

The distribution of combination rules is extremely skewed



It is difficult to decide which rule to use

- a **library card** is a card for a library
- a **toy car** is a toy and a car
- an **apple pie** is a pie containing apples
- a **flat tire** is a tire that is flat
- a **willow forest** is a forest of willows

The same word order might correspond to different structures

- **Structural ambiguity**

  They bought coffee with coins

  They bought coffee with milk

- **Scope ambiguity**

  All men carry a piano

  All men carry a chair

Many different orders might convey the same message

Who did what to whom

The boy gave the flowers to the girl

The girl received the flowers from the boy.

The flowers were given to the girl by the boy.

The boy gave the girl the flowers.

The girl was given flowers by the boy.

Desiderata for a model

- **Scalability**
  - Deal with massive structural ambiguity
  - Process a diversity of constructions
  - Compositional as well as non-compositional bits of language

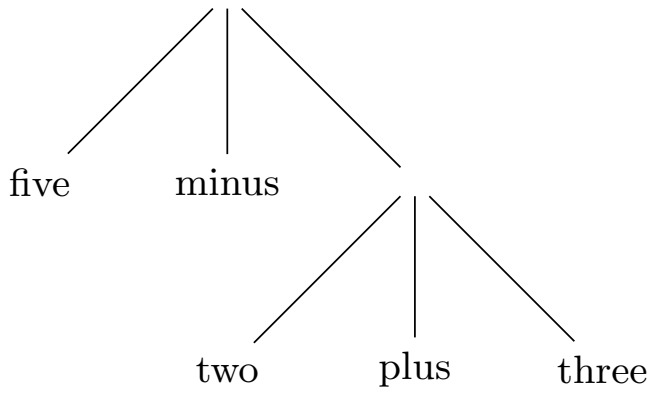
Desiderata for a model

- **Scalability**
  - Deal with massive structural ambiguity
  - Process a diversity of constructions
  - Compositional as well as non-compositional bits of language
- **(towards) Neural Plausibility**
  - Neural network architecture
  - No external intervention
  - Incremental Processing
  - Matches psycholinguistic/neurobiological findings about language

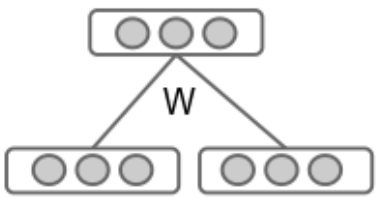
Arithmetic Language

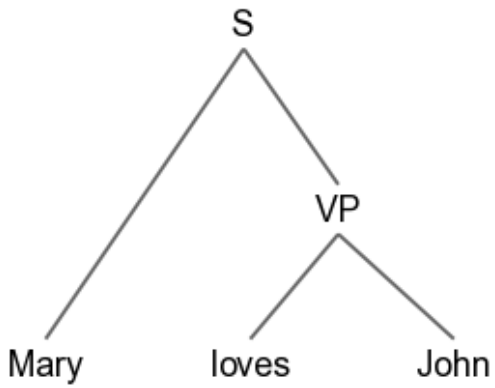
Name	*Numeric leaves*	*Example*
$L1$	1	-3
$L2$	2	(5 + 7)
$L3$	3	(3 - (1 + -2))
...		
$L5R$	5	((((-9 + 6) + 7) + 5) - -7)
$L5L$	5	(8 + (6 - (2 - (10 + 9))))

Arithmetic Language



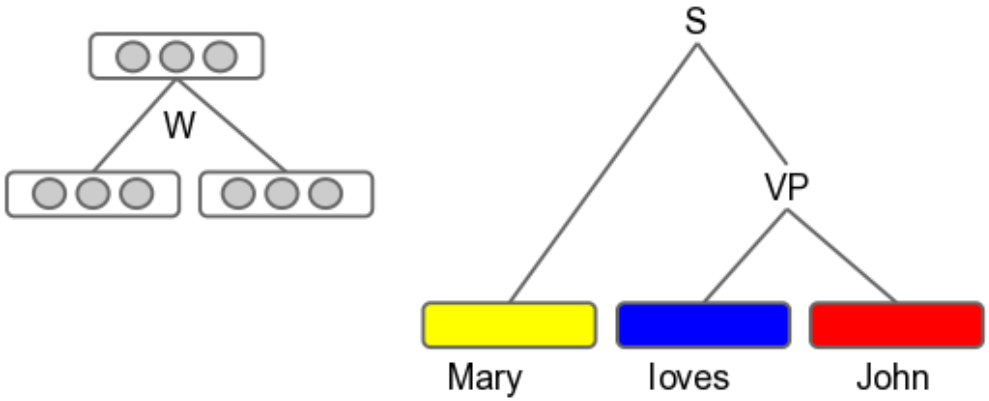
Recursive neural networks (TreeRNN)





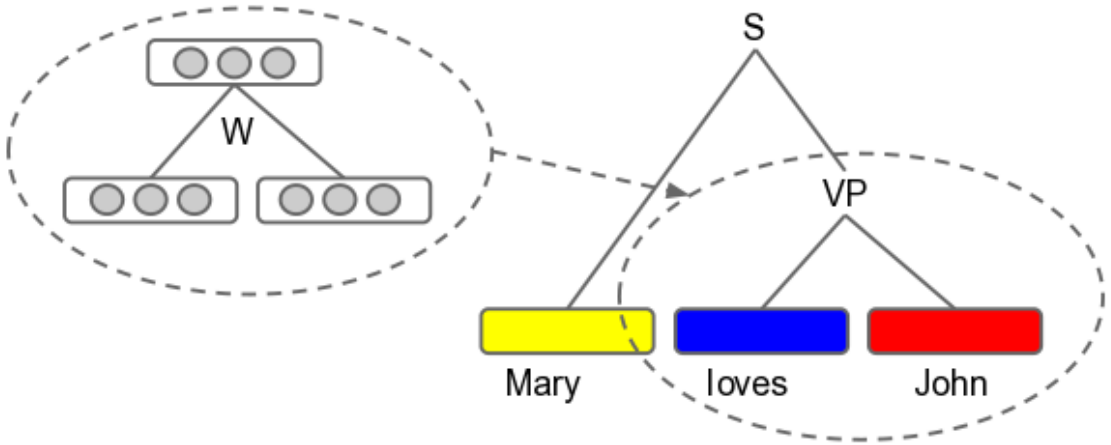
Socher et al. (2011, 2010)

Recursive neural networks (TreeRNN)



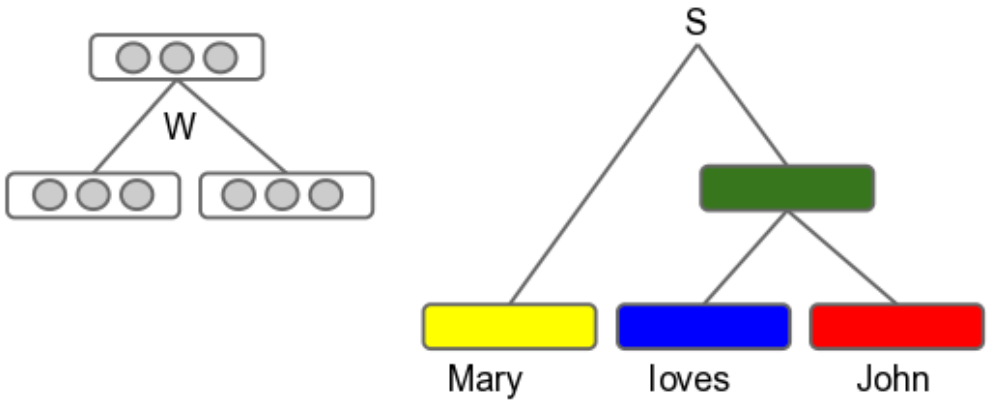
Socher et al. (2011, 2010)

Recursive neural networks (TreeRNN)



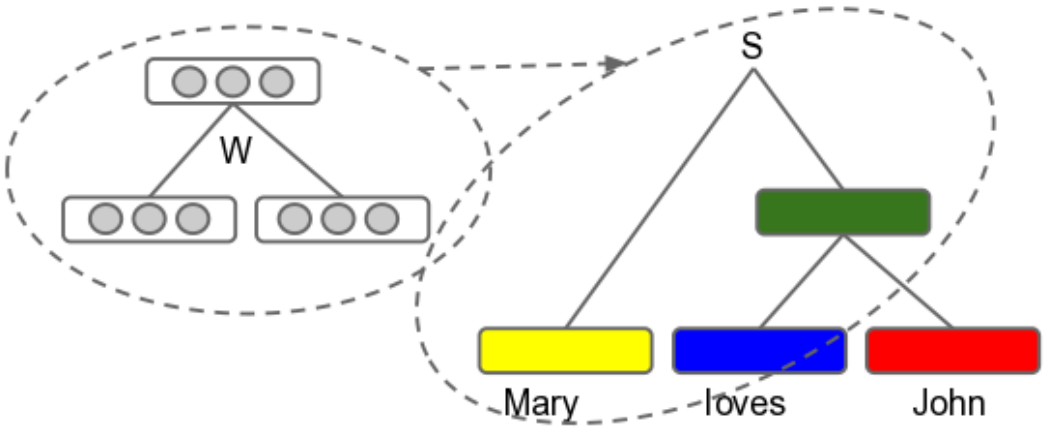
Socher et al. (2011, 2010)

Recursive neural networks (TreeRNN)



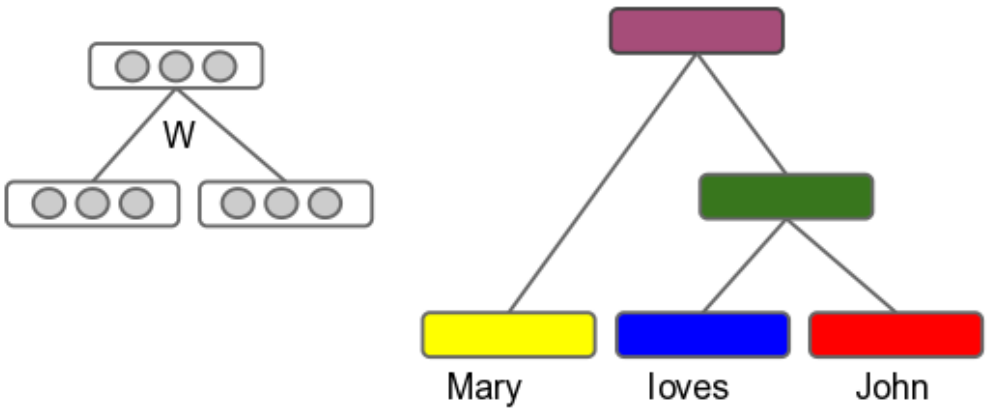
Socher et al. (2011, 2010)

Recursive neural networks (TreeRNN)



Socher et al. (2011, 2010)

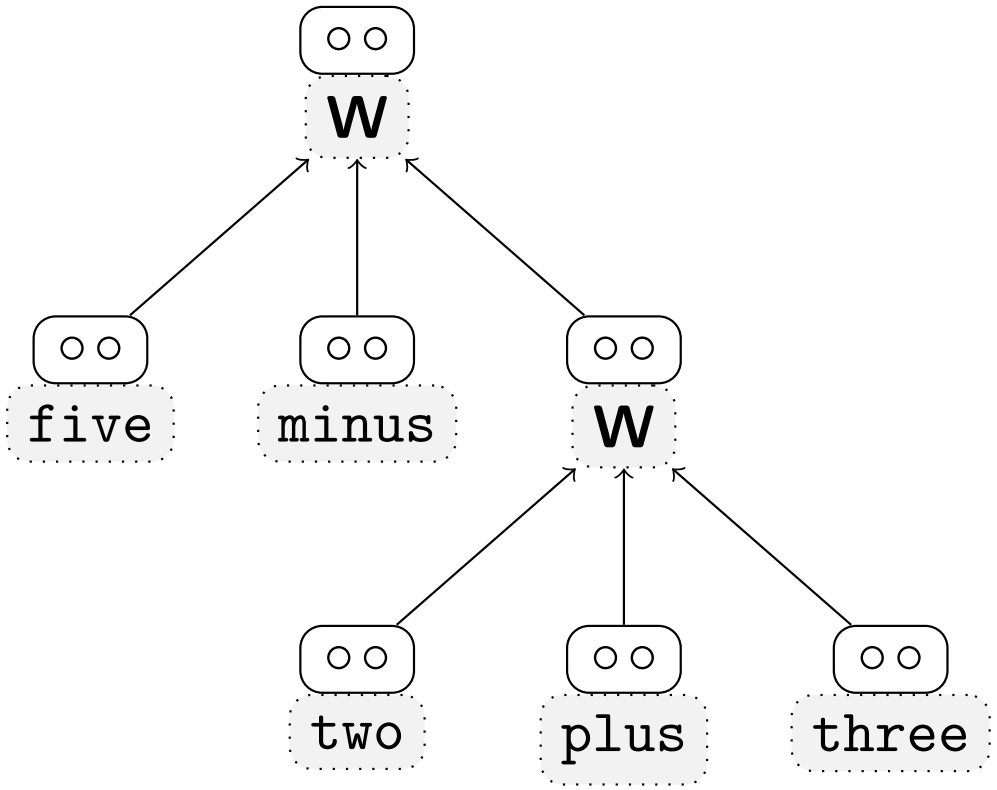
Recursive neural networks (TreeRNN)



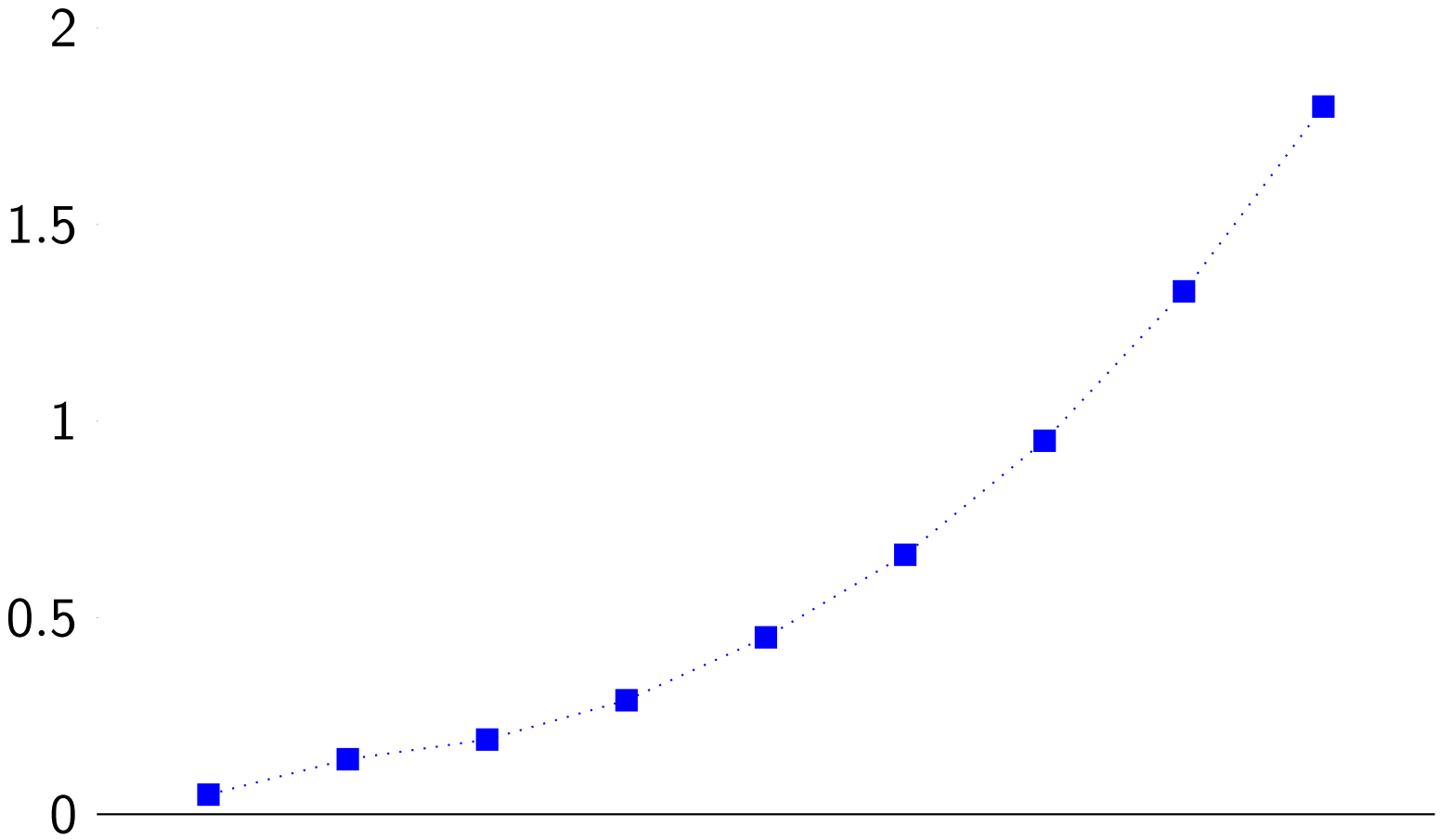
Socher et al. (2011, 2010)

Recursive Neural Network (TreeRNN)

TreeRNNs for arithmetics

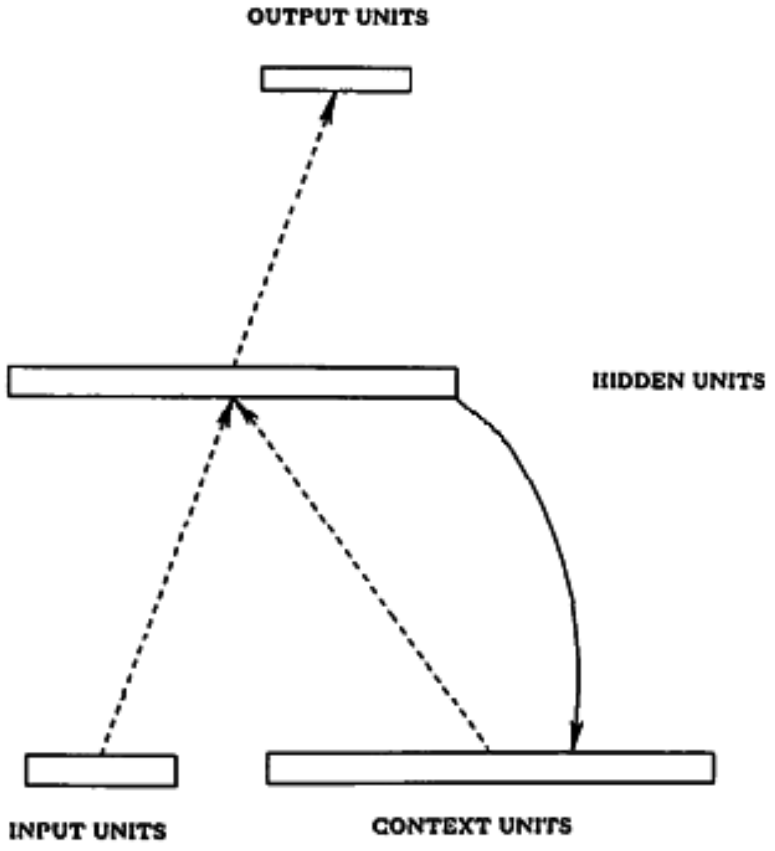


Results



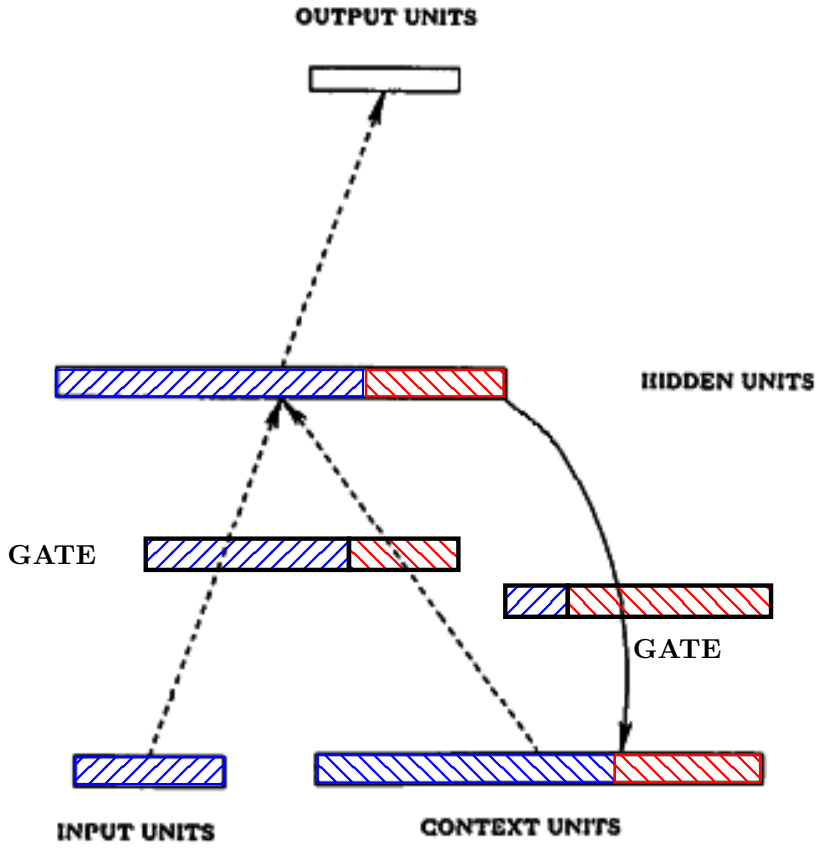
Recurrent neural networks

A simple recurrent neural network in wich activations are copied from the hidden layer to the context layer (Elman, 1990)

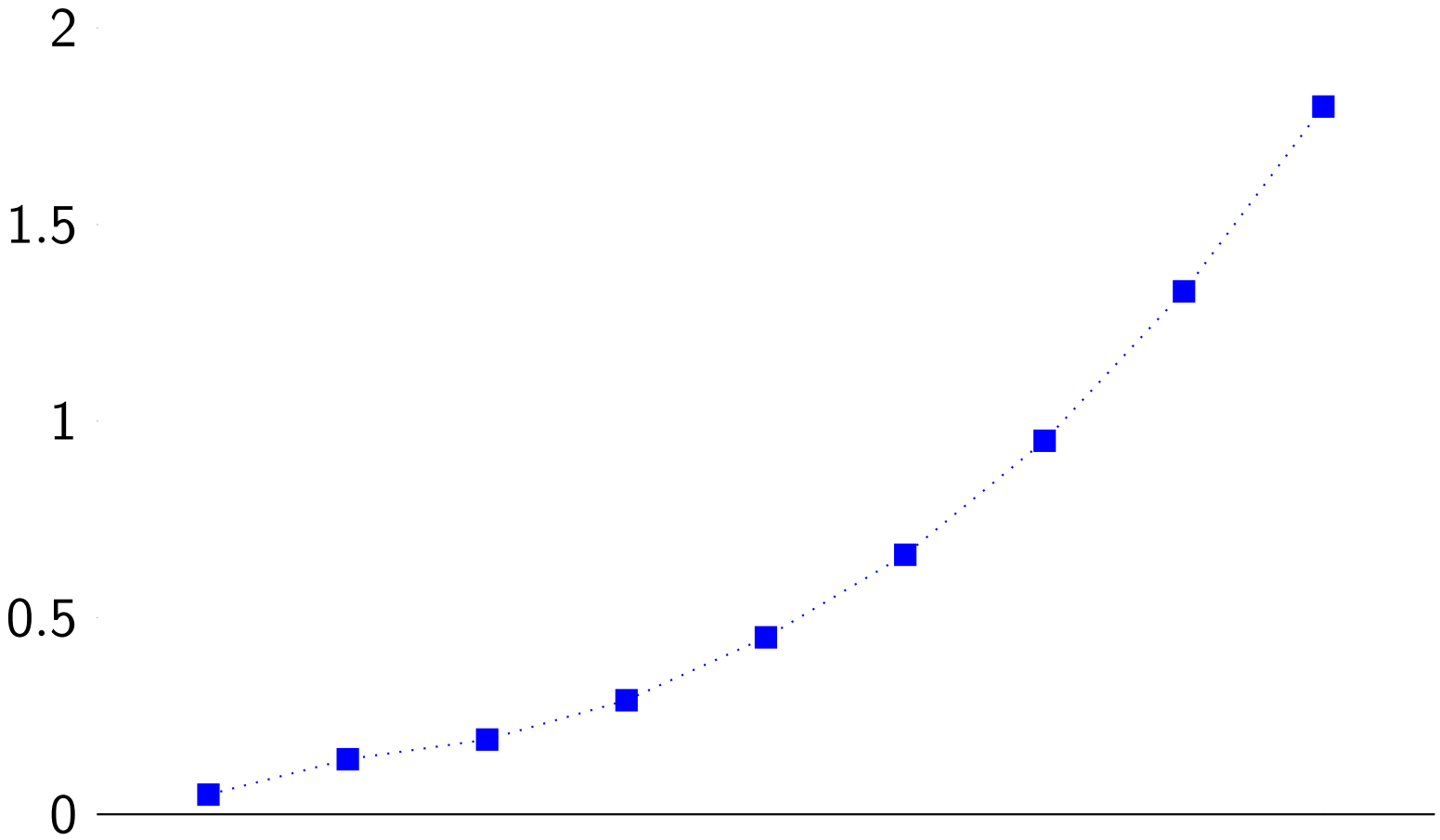


Gated recurrent neural networks

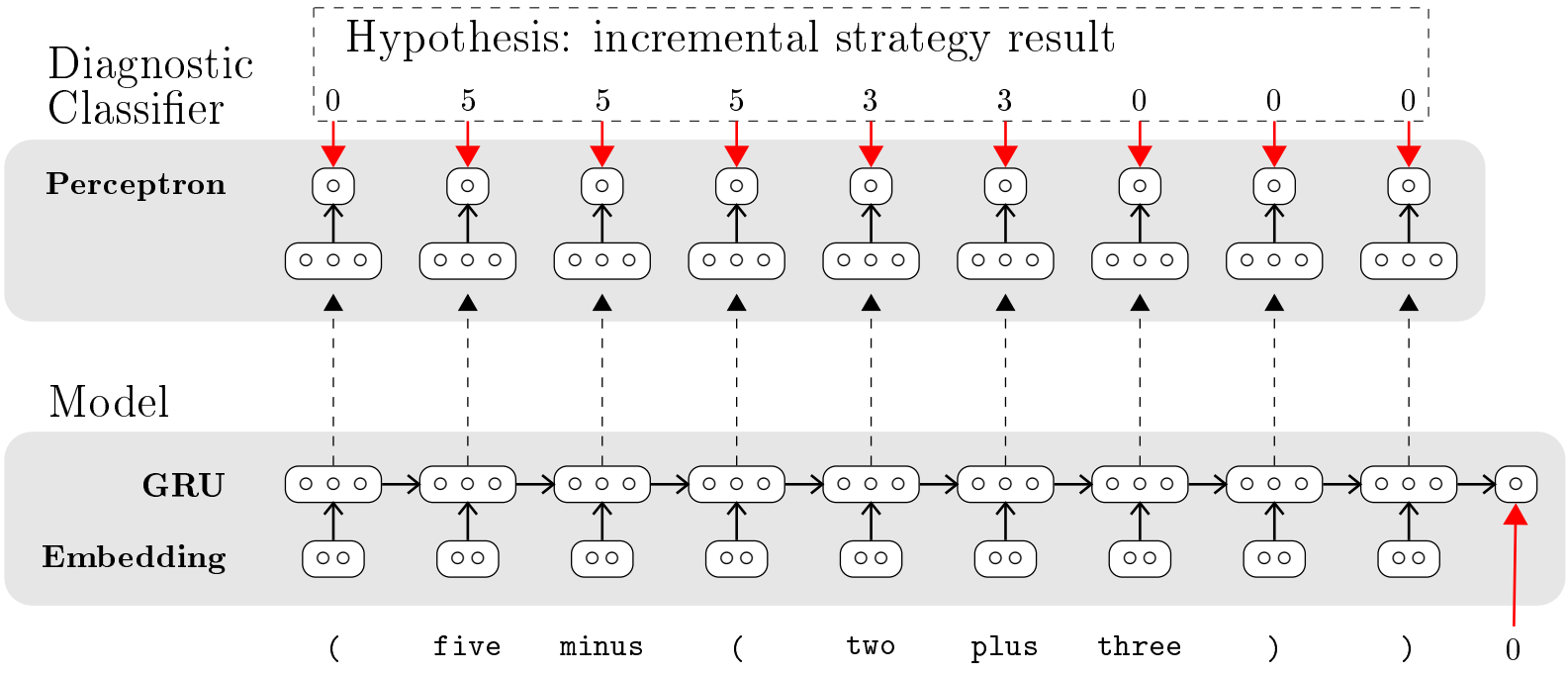
A gated recurrent neural network (Cho et al., 2014; Chung et al., 2015)



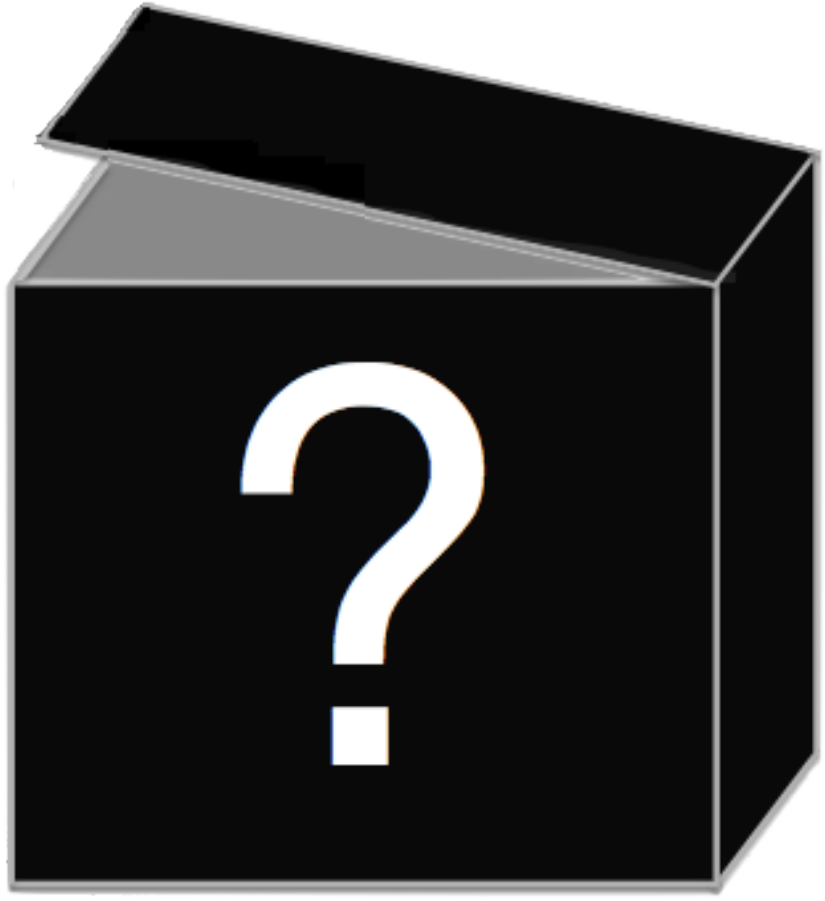
Results



What does the network do?



Conclusion



References

Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*, 2014.

Junyoung Chung, Caglar Gulcehre, Kyunghyun Cho, and Yoshua Bengio. Gated feedback recurrent neural networks. *arXiv preprint arXiv:1502.02367*, 2015.

Jeffrey L Elman. Finding structure in time. *Cognitive science*, 14(2):179–211, 1990.

Richard Socher, Christopher D Manning, and Andrew Y Ng. Learning continuous phrase representations and syntactic parsing with recursive neural networks. In *Proceedings of the NIPS-2010 Deep Learning and Unsupervised Feature Learning Workshop*, pages 1–9, 2010.

Richard Socher, Jeffrey Pennington, Eric H Huang, Andrew Y Ng, and Christopher D Manning. Semi-supervised recursive autoencoders for predicting sentiment distributions. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 151–161. Association for Computational Linguistics, 2011.