Deep RNNs can Learn Syntax

- **Method**: Predicting syntactic constituents from different layers of a recurrent neural network (RNN)
- **Experimental Variations**: RNNs pre-trained on different forms of supervision; predicting different levels of syntactic constituency
- **Results**: Networks encode soft notions of hierarchical syntax, with deeper layers encoding more about higher-level constituents

**Methodology**

- Extract word-level representations from each hidden layer in pre-trained RNN
- Pre-trained models share same architecture: deep bidirectional LSTM network with four hidden layers
- Train a feedforward classifier to run syntactic auxiliary prediction tasks on these representations

**Experimental Setup**

- **Forms of supervision**: dependency parsing, SRL, MT, and LM
- **Prediction tasks**: POS tagging, and parent, grandparent, and great-grandparent constituent prediction
- **Baselines**: compare against a per-word majority baseline

**Results**

- RNNs can induce syntax
  - Models outperform baselines on every task.
  - Indicates models encode syntax, even when not explicitly trained on syntax (i.e., SRL, MT, and LM)
- Deeper layers encode higher-level syntax: For most models, the more abstract syntax tasks peak on deeper layers than the shallower tasks.
- Language models learn some syntax
  - LM induces syntax despite "unsupervised" training signal
  - Results on language modeling in line with previous work

**Conclusions**

- The internal representations learned by deep NLP models induce syntax without explicit supervision.
- Results also suggest that these deep RNNs induce a soft hierarchy over the syntax they encode, using the different layers of the network.
- These results provide some insight as to why deep RNNs perform well on NLP tasks without annotated linguistic features.