



What about lexicon features?

The character LSTM captures the same information

Heavily engineered lexicons used to be critical to good performance, but neural models typically don't use them

Word features from the Berkeley Parser (Petrov and Klein 2007) can be predicted with over 99.7% accuracy from the character LSTM representation

What's Going on in Neural Constituency Parsers? An Analysis

David Gaddy, Mitchell Stern, and Dan Klein University of California, Berkeley

How do they perform so well without it?

What word representations do we need?

A character LSTM is sufficient

Word Only

- Word and Tag
- Character LSTM Only
- Character LSTM and Wo
- Character LSTM, Word,

	91.44
	92.09
	92.24
ord	92.22
, and Tag	92.24

Do LSTMs introduce useful inductive bias compared to feedforward networks?

Yes!

We compare a truncated LSTM with feedforward architectures that are given the same inputs

The LSTM outperformed the best feedforward by 6.5 F1



BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

