We find that a **neural network** part-of-speech tagger implicitly **learns** to model **syntactic change**.

Detecting Syntactic Change Using a Neural Part-of-Speech Tagger

William Merrill[1,2]*, **Gigi Felice Stark**[1]*, **Robert Frank**[1]
[1] Department of Linguistics, Yale University
[2] Allen Institute for Artificial Intelligence
** Equal contribution*

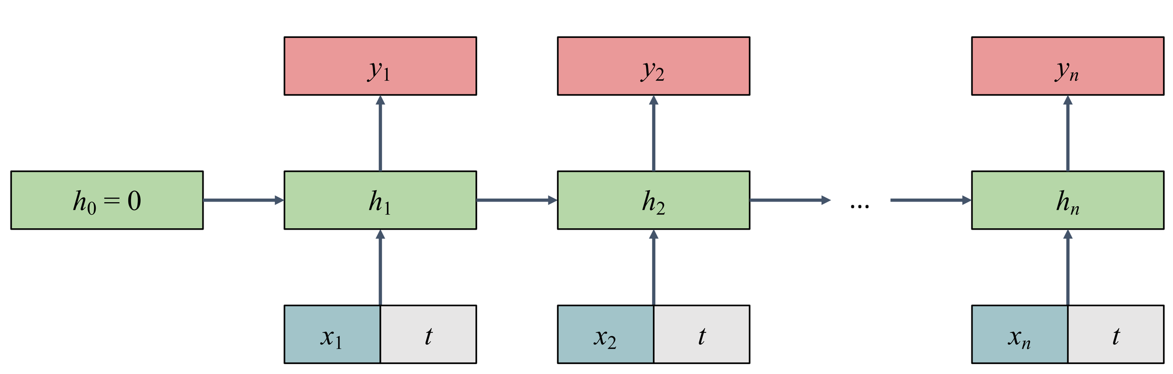
RESEARCH QUESTIONS

Does the network learn a temporal progression?

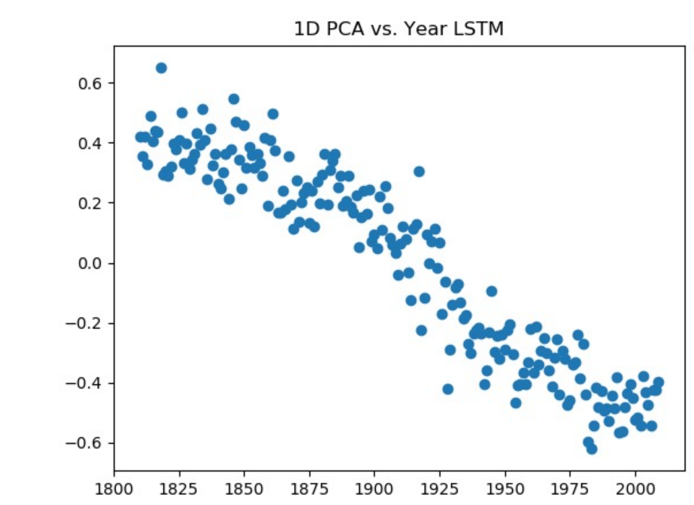
- **PCA** analysis of year embeddings
- **Perplexity** analysis to date new sentences

Does the temporal progression encompass syntax in addition to word frequency?

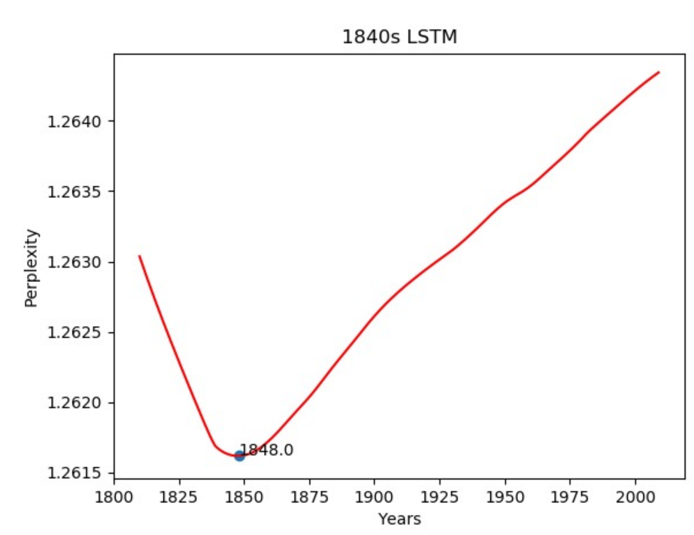
- **Feedforward** baseline comparisons



LSTM tagger architecture



First principal component of year embeddings **correlates strongly** with time (R^2 = 0.89)



1848 is **predicted composition year** for 1840s sentences

DISCUSSION

- Diachronic knowledge learned by network must encompass syntactic–not just lexical–change
- Neural networks are a viable method for further computational work regarding syntactic change

TAGGING PERFORMANCE

	FF	LSTM
Year	82.6	**95.5**
No Year	77.8	95.3

PCA ANALYSIS

- LSTM R^2 = **0.89**
- Feedforward R^2 = 0.68

PERPLEXITY ANALYSIS

	Baseline	FF	LSTM
Decade	50.0	26.6	**12.5**
Year	50.0	37.5	**21.9**

Average prediction error when dating sentences using decade and year-level buckets

EXAMPLE DATED SENTENCE

It is of great consequence, that we adorn the religion we profess, and that our light shine more and more that we grow in grace as we advance in years, and that we do not resemble the changing wind or the inconstant wave.

Actual Year: 1817
Predicted Year (LSTM): 1817

Prediction Error (FF): 86
Prediction Error (LSTM): 0

Yale