

IBIS 2017 企画セッション

# 自然言語処理への 機械学習の応用



IBIS2017

第 20 回 情報論的学習理論  
ワークショップ

The 20th Information-Based Learning Science Workshop

# 深層学習と 自然言語処理

東京大学大学院  
情報理工学系研究科  
鶴岡 慶雅



IBIS2017

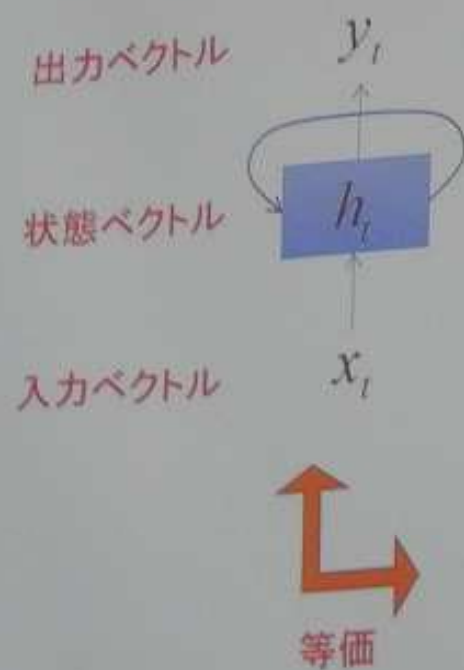
第20回 情報論的  
ワークショップ

The 20th Information Theoretical Workshop



# リカレントニューラルネットワーク (Recurrent Neural Network, RNN)

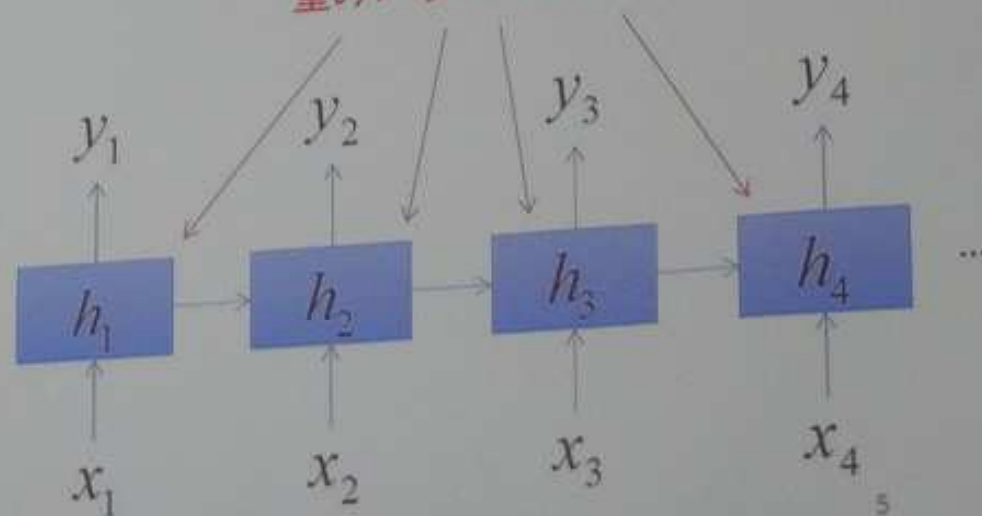
- 任意の長さの**系列**を扱うことができる



$$h_t = \text{sigmoid}(W^{\text{hx}} x_t + W^{\text{hh}} h_{t-1})$$

$$y_t = W^{\text{yh}} h_t$$

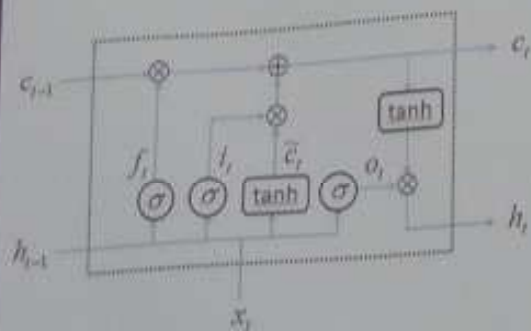
重みパラメータを共有





# LSTM (Long Short-Term Memory)

- 単純な RNN の問題点
  - 勾配消失問題
  - 長距離の依存関係をとらえられない
- Long Short-Term Memory (LSTM)



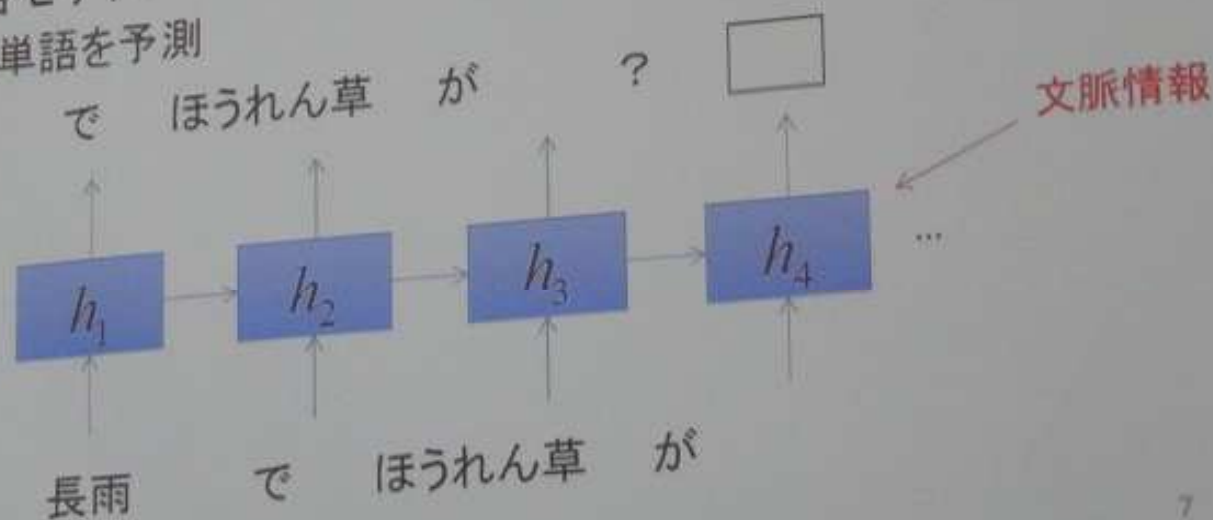
$$\begin{aligned} i_t &= \sigma(W^{(i)}x_t + U^{(i)}h_{t-1} + b^{(i)}) \\ f_t &= \sigma(W^{(f)}x_t + U^{(f)}h_{t-1} + b^{(f)}) \\ o_t &= \sigma(W^{(o)}x_t + U^{(o)}h_{t-1} + b^{(o)}) \\ \tilde{c}_t &= \tanh(W^{(c)}x_t + U^{(c)}h_{t-1} + b^{(c)}) \\ c_t &= i_t \odot \tilde{c}_t + f_t \odot c_{t-1} \\ h_t &= o_t \odot \tanh(c_t) \end{aligned}$$



# RNNと自然言語処理

- 自然言語処理では文字や単語の**系列**を扱う
  - 言語モデル、品詞タグ付け、固有表現認識、機械翻訳、etc.

- 例) 言語モデル
  - 次の単語を予測



# 品詞タグ付け

- 文中の各単語に品詞情報を付与

Paul Krugman , a professor at Princeton University, was  
awarded the Nobel Prize in Economics on Monday.

- 品詞タグ

NN: 名詞

NNP: 固有名詞

DT: 限定詞

:

IN: 前置詞

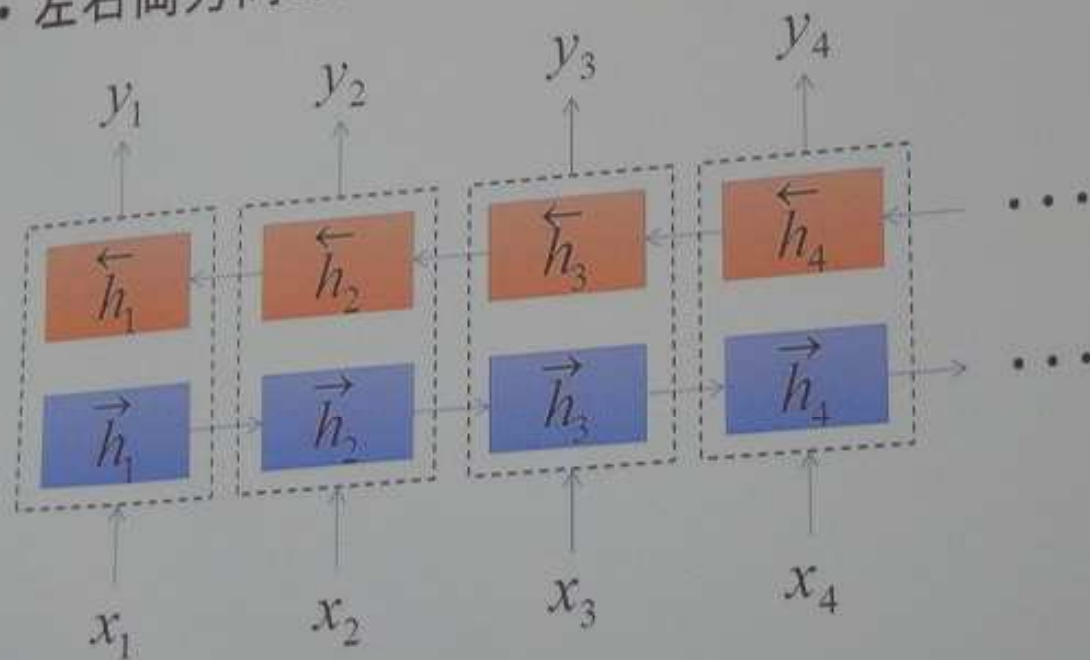
VBD: 動詞(過去形)

VCN: 動詞(過去分詞)

:

# Bidirectional LSTM (BiLSTM)

- 2つのRNN(順方向と逆方向)
- 左右両方向の文脈情報を捉えられる

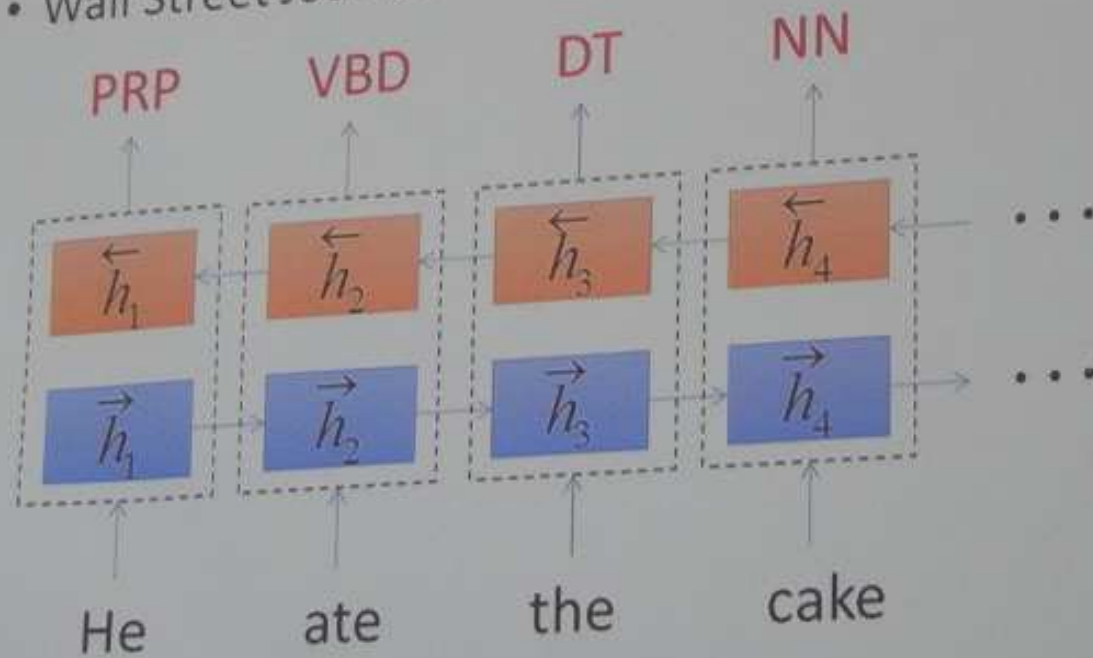




# BiLSTMによる品詞タグ付け

- 学習データ

- Wall Street Journal コーパス: 約40,000文



→ CRF 層を追加 (Huang et al., 2015; Ma and Hovy, 2016)

## チャンキング (shallow parsing)

He reckons the current account deficit will narrow to  
NP VP NP VP PP  
only #1.8 billion in September.  
NP PP NP

- 文をフラットな句に分割
- 再帰的な分割は行わない

# チャンキング (shallow parsing)

He reckons the current account deficit will narrow to  
 $B_{NP}$   $B_{VP}$   $B_{NP}$   $I_{NP}$   $I_{NP}$   $I_{NP}$   $B_{VP}$   $I_{VP}$   $B_{PP}$   
only #1.8 billion in September .  
 $B_{NP}$   $I_{NP}$   $I_{NP}$   $I_{NP}$   $B_{PP}$   $B_{NP}$   $O$

- 各単語に対するタグ付けの問題に変換できる
  - B: チャンクの先頭
  - I: チャンクの中(先頭以外)
  - O: チャンクの外

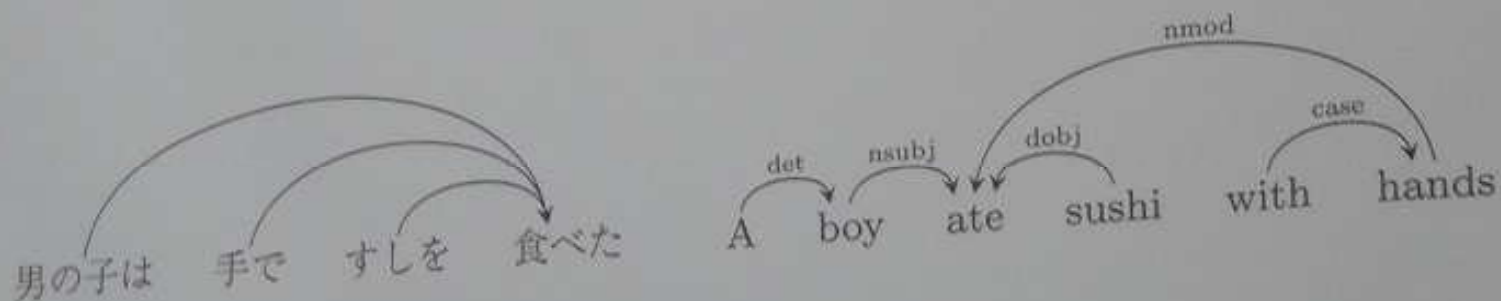
## 固有表現認識 (named entity recognition)

The peri-kappa B site mediates human immunodeficiency  
DNA virus type 2 enhancer activation in monocytes ...  
cell\_type

- 文中の固有表現を認識
- チャンキングと同様、系列ラベリングの問題として処理できる

# 構文解析

- 依存構造 (dependency structure)
  - 文中の単語間の関係をグラフで表す
    - 係り元 (dependent) から係り先 (head, 主辞) へのエッジ
    - (矢印を逆向きに描くことも多い)
  - 係り受け構造とも呼ばれる





# Shift-Reduce法による依存造解析

- Shift-Reduce 法
  - バッファ (buffer)
    - 解析前の単語列を格納
  - スタック (stack)
    - 解析後の依存構造を格納
  - アクション (action)
    - Shift
      - バッファの先頭の単語をスタックに移動
    - Reduce
      - スタックトップの2つの単語の間にエッジを生成

# Shift-Reduce 法

I saw a dog with eyebrows



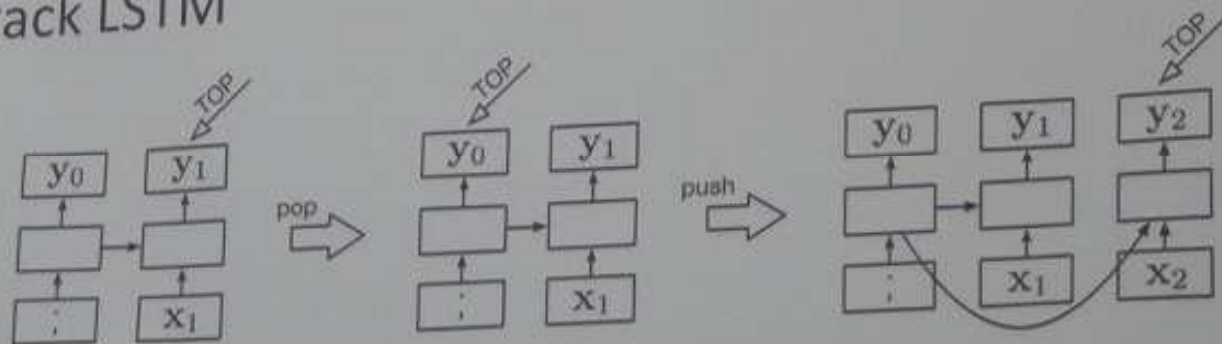
OPERATION	STACK	BUFFER
Shift ReduceL ReduceR	saw dog	

# Shift-Reduce 法

- 学習モデル

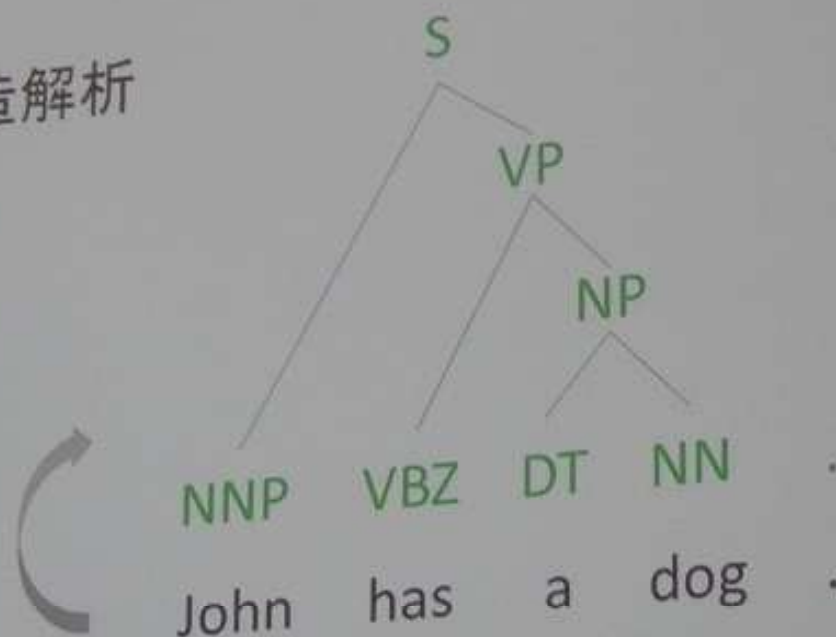
- Stack LSTM (Dyer et al., 2015)
- FFNN + ビーム探索 + Early Update (Andor et al., 2016)
- BiLSTM (Kiperwasser & Goldberg, 2016)

- Stack LSTM



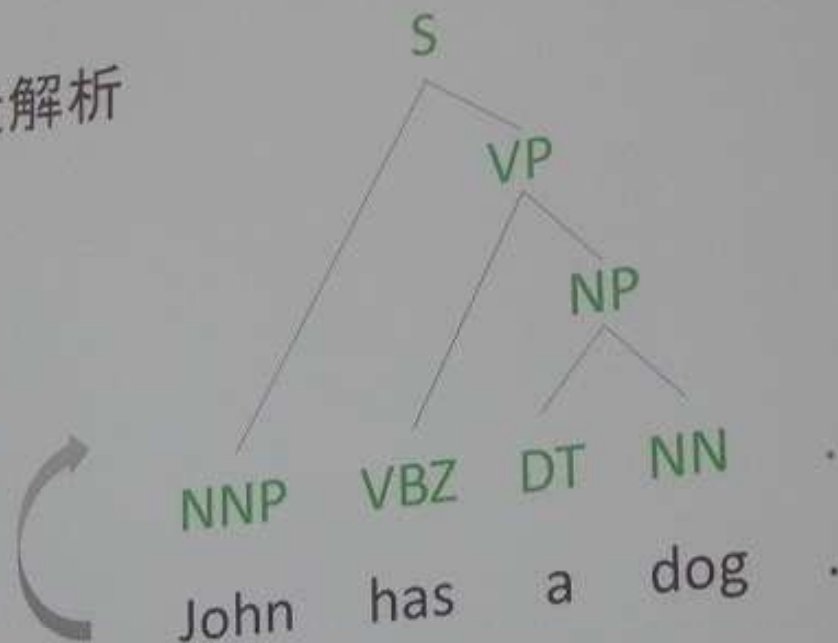
# 構文解析

- 句構造解析



# 構文解析

- 句構造解析



$(S (NP NNP)_{NP} (VP VBZ (NP DT NN)_{NP})_{VP} \cdot)_{S}$

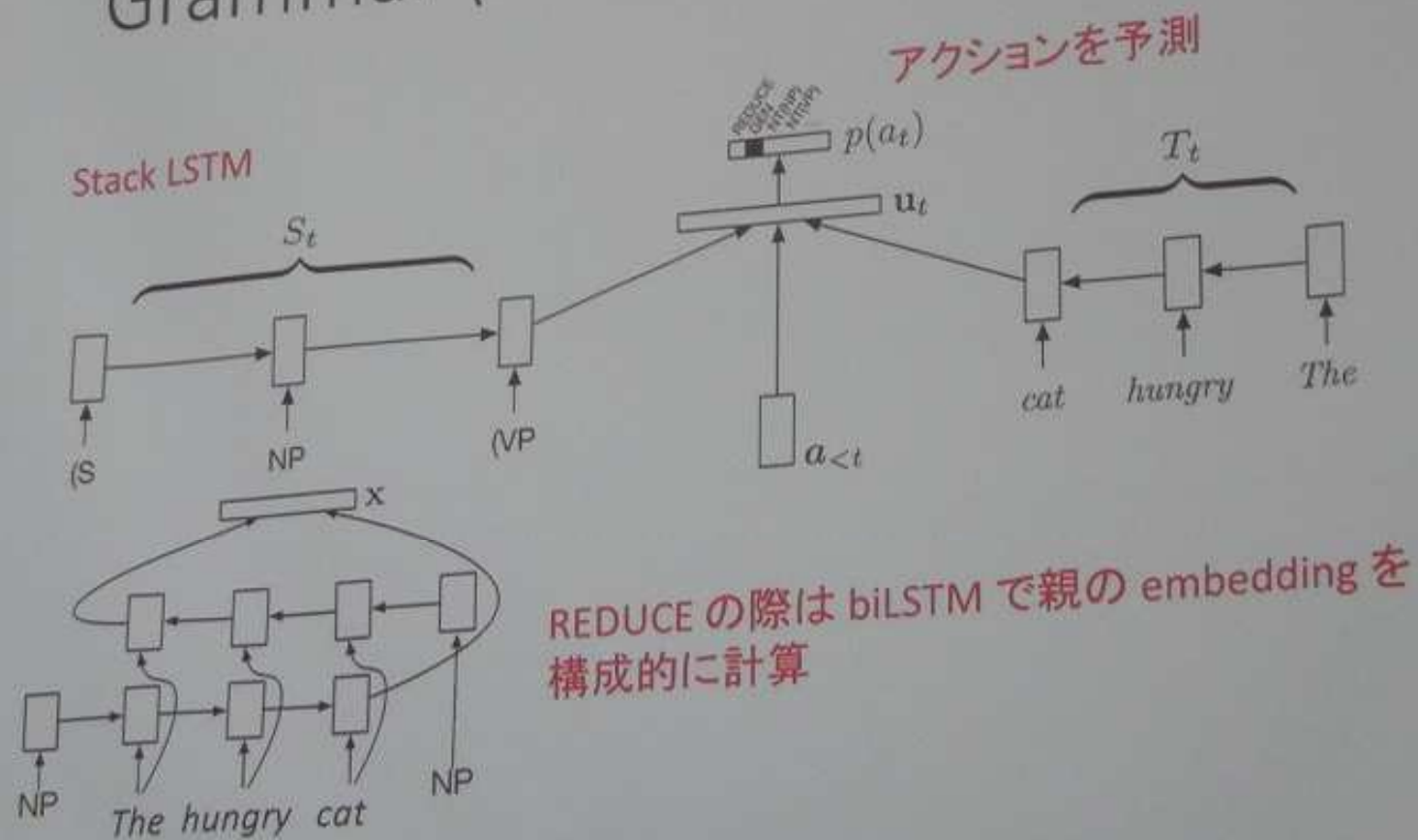


# Recurrent Neural Network Grammar (RNNG) (Dyer et al., 2016)

入力文: *The hungry cat meows.*

	Stack	Buffer	Action
0		<i>The   hungry   cat   meows   .</i>	NT( <i>S</i> )
1	( <i>S</i>	<i>The   hungry   cat   meows   .</i>	NT( <i>NP</i> )
2	( <i>S</i>   ( <i>NP</i>	<i>The   hungry   cat   meows   .</i>	SHIFT
3	( <i>S</i>   ( <i>NP</i>   <i>The</i>	<i>The   hungry   cat   meows   .</i>	SHIFT
4	( <i>S</i>   ( <i>NP</i>   <i>The hungry</i>	<i>hungry   cat   meows   .</i>	SHIFT
5	( <i>S</i>   ( <i>NP</i>   <i>The hungry   cat</i>	<i>cat   meows   .</i>	REDUCE
6	( <i>S</i>   ( <i>NP</i>   <i>The hungry   cat</i> )	<i>meows   .</i>	NT( <i>VP</i> )
7	( <i>S</i>   ( <i>NP</i>   <i>The hungry   cat</i> ) ( <i>VP</i>	<i>meows   .</i>	SHIFT
8	( <i>S</i>   ( <i>NP</i>   <i>The hungry   cat</i> ) ( <i>VP meows</i>	<i>meows   .</i>	REDUCE
9	( <i>S</i>   ( <i>NP</i>   <i>The hungry   cat</i> ) ( <i>VP meows</i> )	<i>meows   .</i>	SHIFT
10	( <i>S</i>   ( <i>NP</i>   <i>The hungry   cat</i> ) ( <i>VP meows</i> )   .	<i>meows   .</i>	REDUCE
11	( <i>S</i>   ( <i>NP</i>   <i>The hungry   cat</i> ) ( <i>VP meows</i> )   .)	<i>meows   .</i>	

# Recurrent Neural Network Grammar (RNNG) (Dyer et al., 2016)



# 構文解析の精度

- 依存構造解析

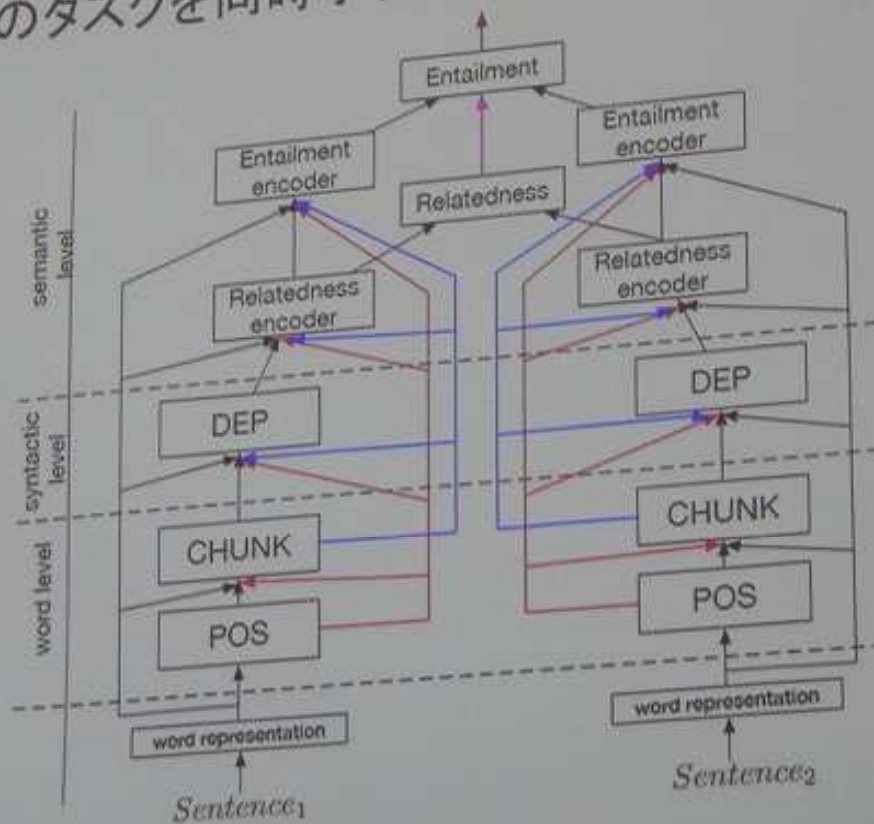
	UAS	LAS
パーザー		
Turbo parser (Martins et al., 2013)	92.9	90.6
SyntaxNet (Andor et al., 2016)	94.6	92.8
Deep Biaffine Attention (Dozat and Manning, 2016)	95.4	93.8
Recurrent Neural Network Grammar (Dyer et al. 2016)	95.8	94.6

- 句構造解析

	F1 score
パーザー	
Symbol-refined CFG (Petrov and Klein, 2007)	90.1
Bayesian Symbol-refined TSG (Shindo et al. 2012)	91.1
Recurrent Neural Network Grammar (Dyer et al. 2016)	93.6

# マルチタスク学習

- Joint Many-Task Model (Hashimoto et al., 2016)
- 5つのタスクを同時学習





# ニューラル機械翻訳

- エンコーダ・デコーダモデル (Sutskever et al., 2014)

- Encoder RNN
  - 翻訳元の文を読み込み、実数値ベクトルに変換
- Decoder RNN
  - 実数値ベクトルから翻訳先言語の文を生成





# アテンション (Bahdanau et al., 2015)

- ・ 翻訳先の各単語を選択する際に翻訳元の文のどこに注目するのかを動的に決定

翻訳元の各単語の隠れ状態の加重平均

$$c_i = \sum_{j=1}^{T_x} \alpha_{ij} h_j$$

重み(すべて足すと1)

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T_x} \exp(e_{ik})}$$

$$e_{ij} = \text{FeedForwardNN}(s_{i-1}, h_j)$$

