

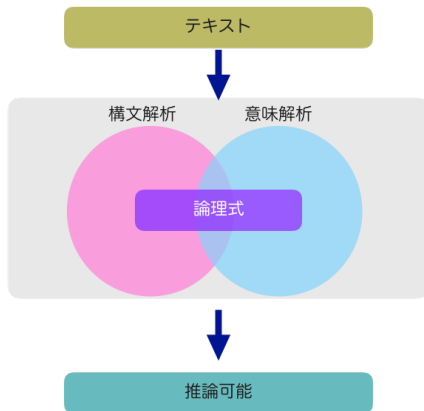
# RNN 系列変換モデルを用いた高階論理からの 文生成

馬目 華奈<sup>1</sup> 戸次 大介<sup>1</sup>

<sup>1</sup> 戸次研究室

January 29, 2018

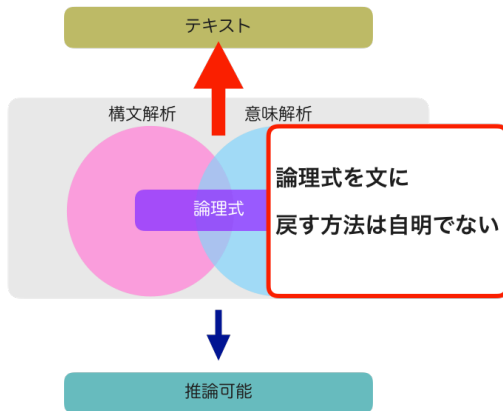
# 研究背景



<応用>

- ・含意関係認識
- ・文間の類似度

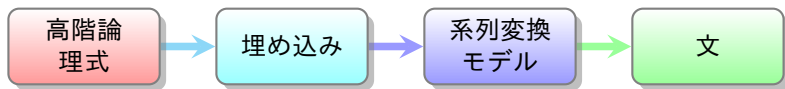
# 研究背景



<応用>

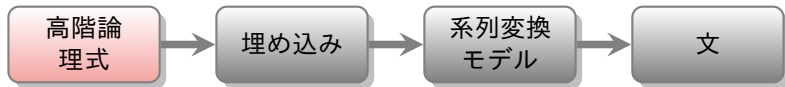
- ・含意関係認識
- ・文間の類似度

# 研究概要



- ニューラルネットによる系列変換モデルを用いて高階論理式から文を生成する手法を提案.
- 埋め込みの際, 4つの手法  
(記号, トークン, 木構造, グラフ) による比較実験を行う.

## 関連研究 1



- ニューラルネットによる系列変換モデルを用いて高階論理式から文を生成する手法を提案.
- 埋め込みの際, 4つの手法  
(記号, トークン, 木構造, グラフ) による比較実験を行う.

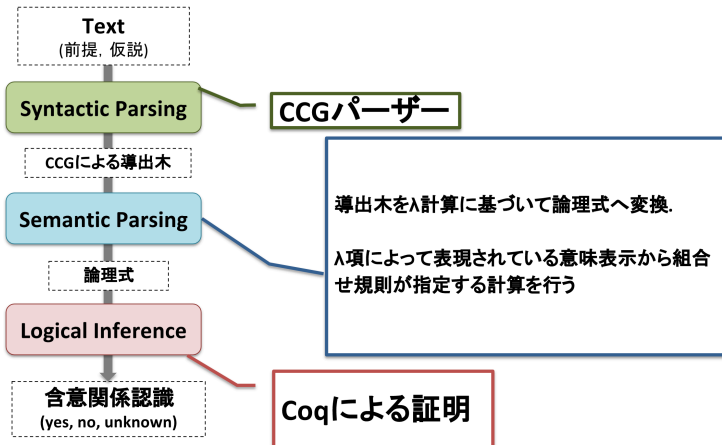
# 関連研究：CCG に基づく論理式による文の意味表現

## cgc2lambda

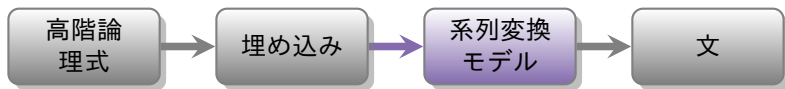
<https://github.com/mynlp/cgc2lambda>

Mineshima+ [EMNLP2015]

Martínez-Gómez+ [ACL2016]



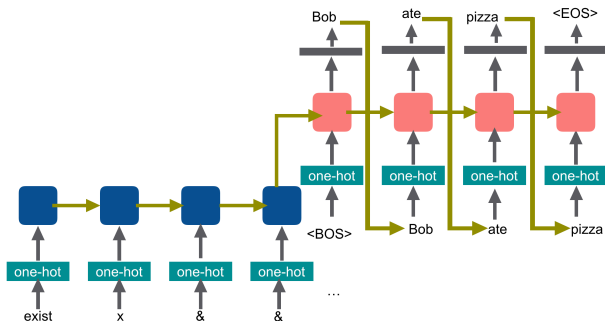
## 関連研究 2



- ニューラルネットによる系列変換モデルを用いて高階論理式から文を生成する手法を提案.
- 埋め込みの際, 4つの手法 (記号, トークン, 木構造, グラフ) による比較実験を行う.

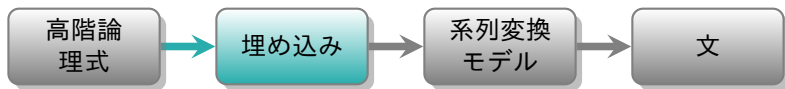
## 関連研究：系列変換モデル

- 入出力がシーケンスとなるニューラルネットのモデル
- エンコータ：入力列を再帰型 NN により隠れ状態ベクトルに変換
- デコーダ：隠れ状態ベクトルを初期値とし、隠れ状態と自身のこれまでの出力結果を基に次のトークンを生成





# 提案手法



- ニューラルネットによる系列変換モデルを用いて高階論理式から文を生成する手法を提案.
- 埋め込みの際, 4つの手法  
(記号, トークン, 木構造, グラフ) による比較実験を行う.

## 提案手法：論理式埋め込み 1

記号ごとに区切る

[e,x,i,s,t,s, ,x, ,, ,(,(x, ,=,...]

トークンごとに区切る

[exists,x,and,=,x,Bob,,exists,z1,and,]

# 提案手法：論理式埋め込み 2

## 記号ごとに区切る

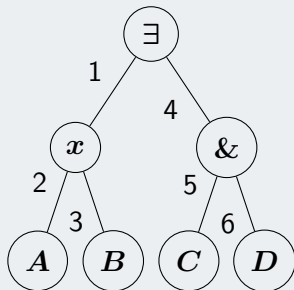
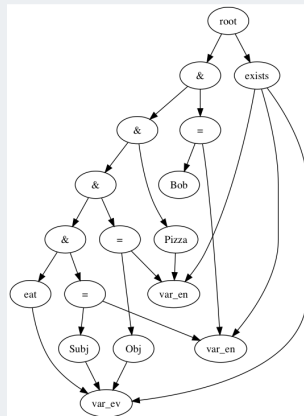


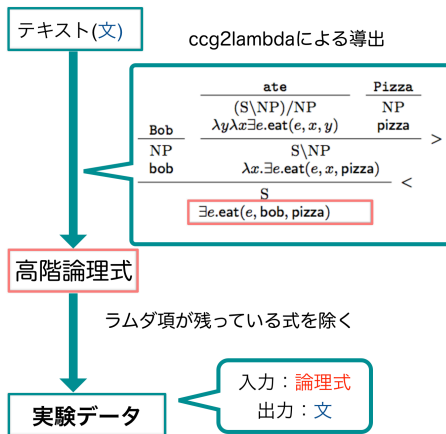
Figure: 木構造の DFS

## グラフ



# 提案手法：データセット

- SNLI を用い論理式と文のペアを作成
- 60 単語以内の文例を対象 train:18087/dev:3617/test:1500



## 実験：実験設定

- 系列変換モデルによる文生成  
入力：論理式，出力：文
- 記号ベースの埋め込みベクトルをエンコーダ 70 次元，デコーダ 78 次元，  
その他の埋め込みベクトルを 256 次元

	文字	単語	木構造	グラフ
入力語彙数	70	5,118	5,107	4,991
出力語彙数	78	7,214	7,214	7,214
入力列最長	2,097	699	451	259
出力列最長	270	55	53	53

## 実験：評価方法

### BLEU による評価

$$score = BP \exp \left( \sum_{i=1}^N \frac{1}{N} \log P_n \right)$$

$$BP = \begin{cases} 1 & (c \geq r) \\ \exp \left( 1 - \frac{r}{c} \right) & (c < r) \end{cases}$$

$$P_n = \frac{\sum_{i=0} \text{出力文 } i \text{ 中と解答文 } i \text{ 中で一致した } n\text{-gram 数}}{\sum_{i=0} \text{出力文 } i \text{ 中の全 } n\text{-gram 数}}$$

## 実験：実験結果

### BLEU 評価

指標	文字	トークン	木構造	グラフ
BLEU	36.6	39.7	41.8	44.7

“Two kids are playing tag.” の例

- トークン : Two dogs are playing a game.
- 木構造 : Two kids are playing tennis.
- グラフ : Two kids are playing together.

## まとめ

- 系列変換モデルを用いて  
高階論理式から文を生成する手法を提案した.
- 含意関係認識用データセットを用いて  
提案手法の評価を行った結果, 論理式をグラフ化し埋め込む  
ことで  
精度向上がみられた.



## 今後の課題

- 他の意味表現からの文生成との比較や他のデータセットによる評価を行う.
- アテンション付き系列変換モデルやコピー機構を用いるなどモデルの改良に取り組む.

## 参考文献