

非テキスト情報を用いたフローグラフ解析

○友利 涼, 森 信介

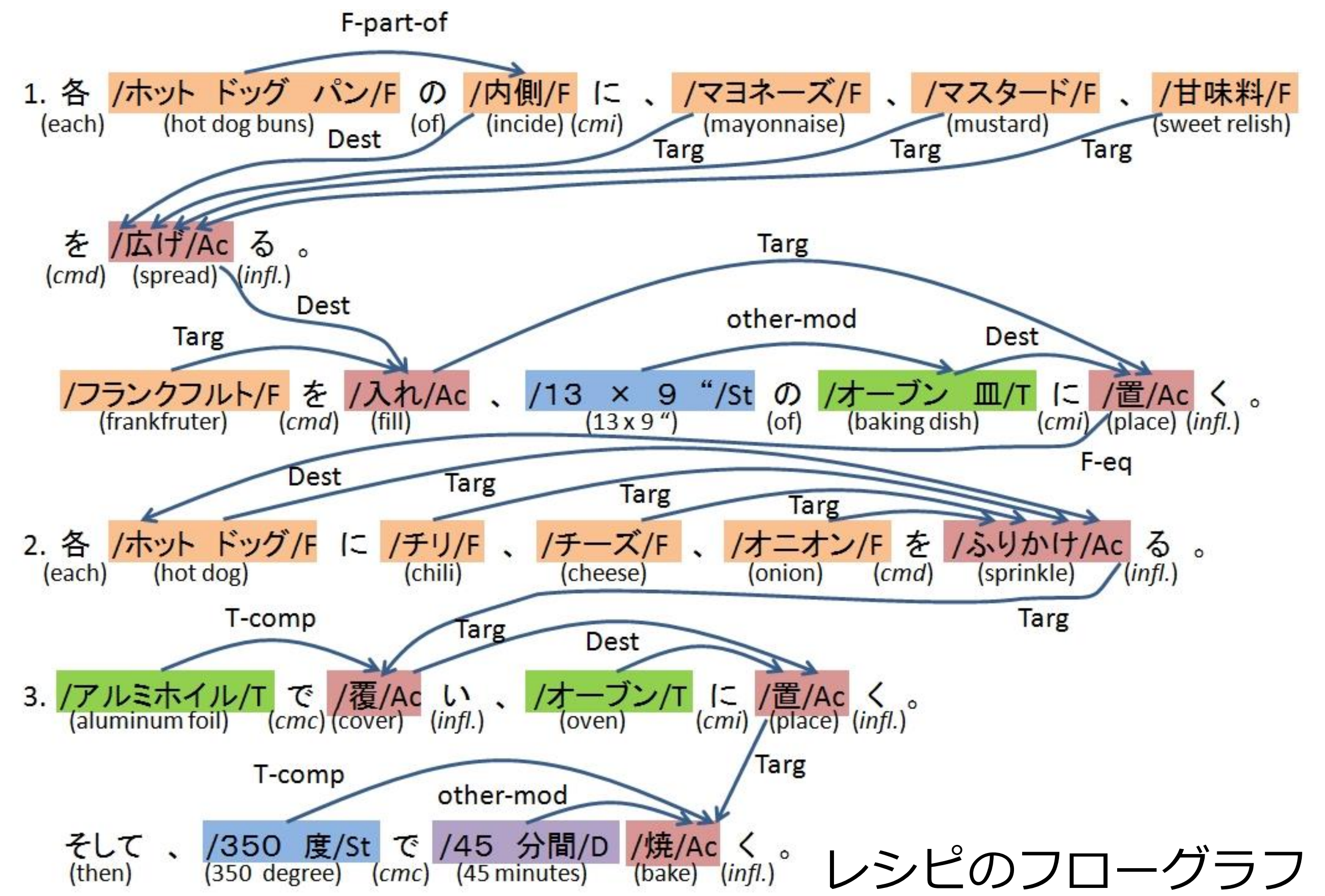
京都大学

概要

- 計算機がテキストだけから手順(フロー)を理解することには限界があるのでは?
- レシピには理解補助として説明画像が付随

貢献

- フローグラフ解析モデルをニューラル化
- テキストと説明画像を用いて、同時事前学習



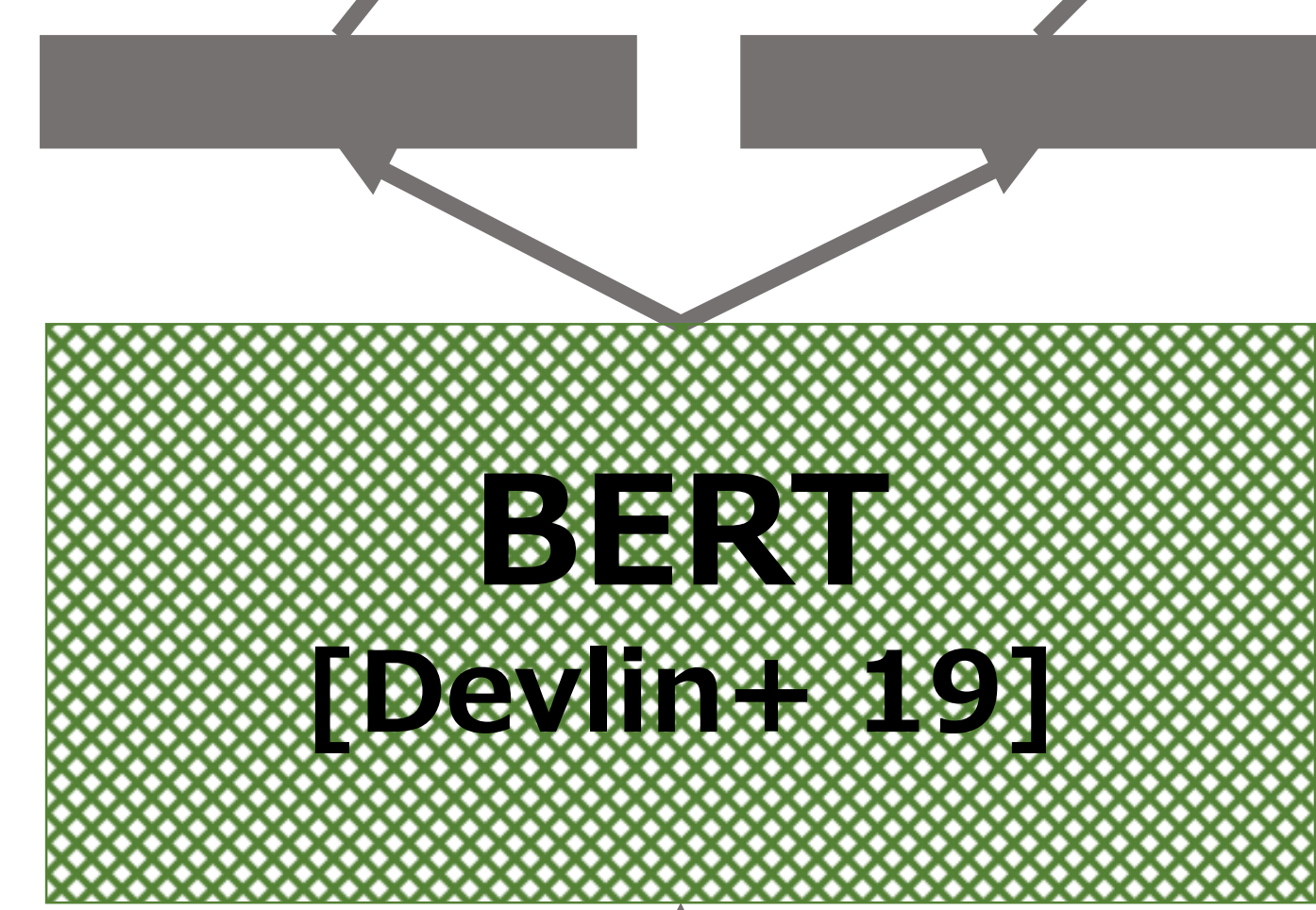
提案手法

事前学習

$$L_{image} = \max(0, d_p - d_n + \alpha) + \max(0, d_p - \beta)$$

テキスト穴埋め問題

油を薄く [延ば] し、豚肉や [玉ねぎ] を...



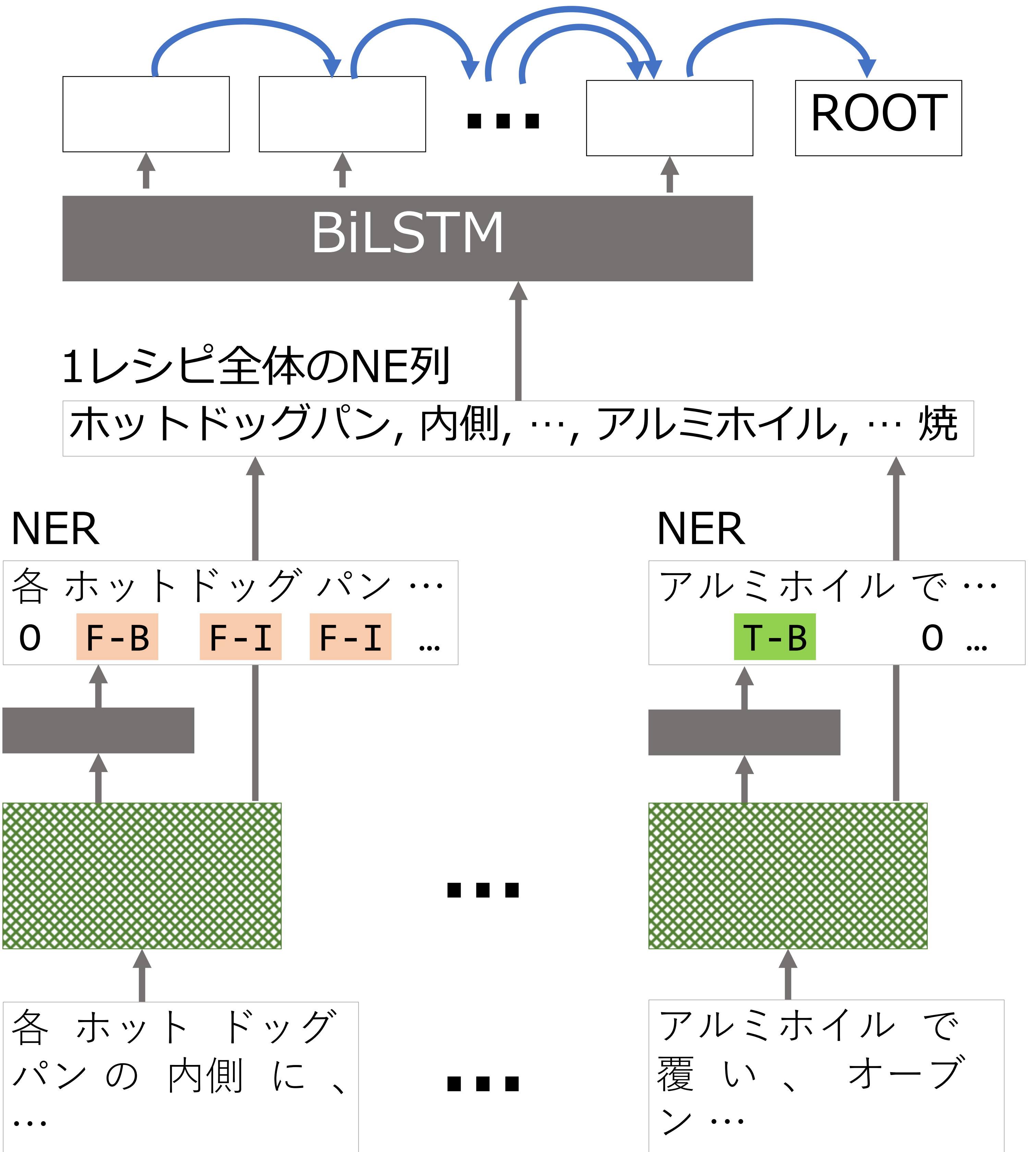
画像エンコーダ (PNASNet [Liu+ 18])



● 説明画像

● ネガティブサンプル

フローグラフ解析

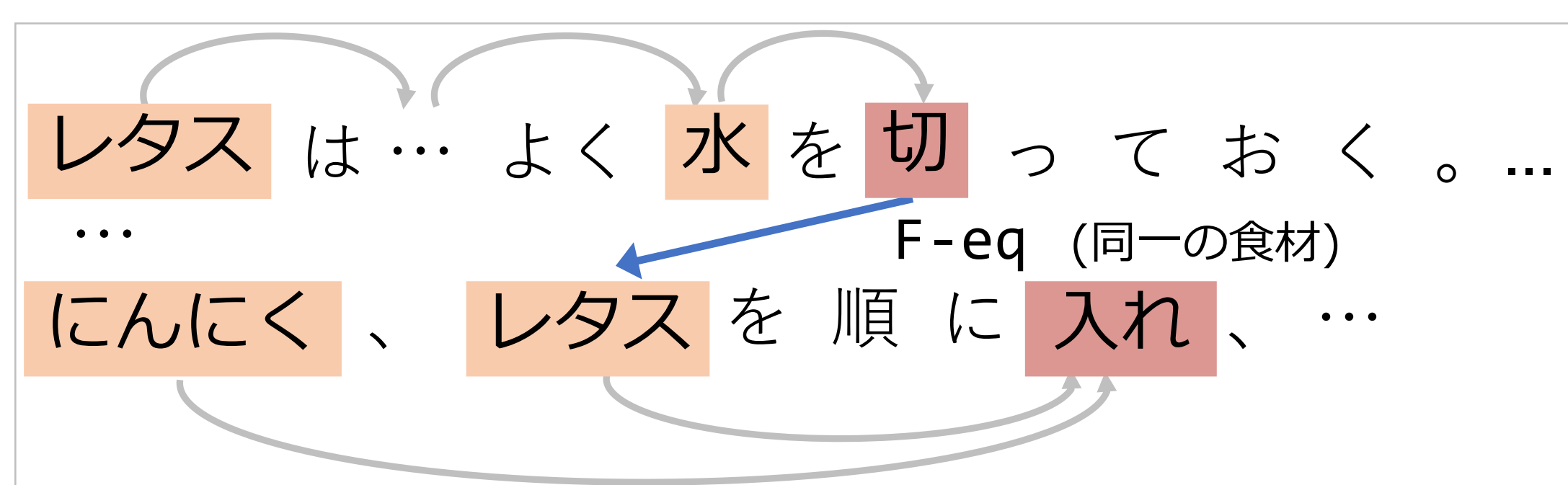
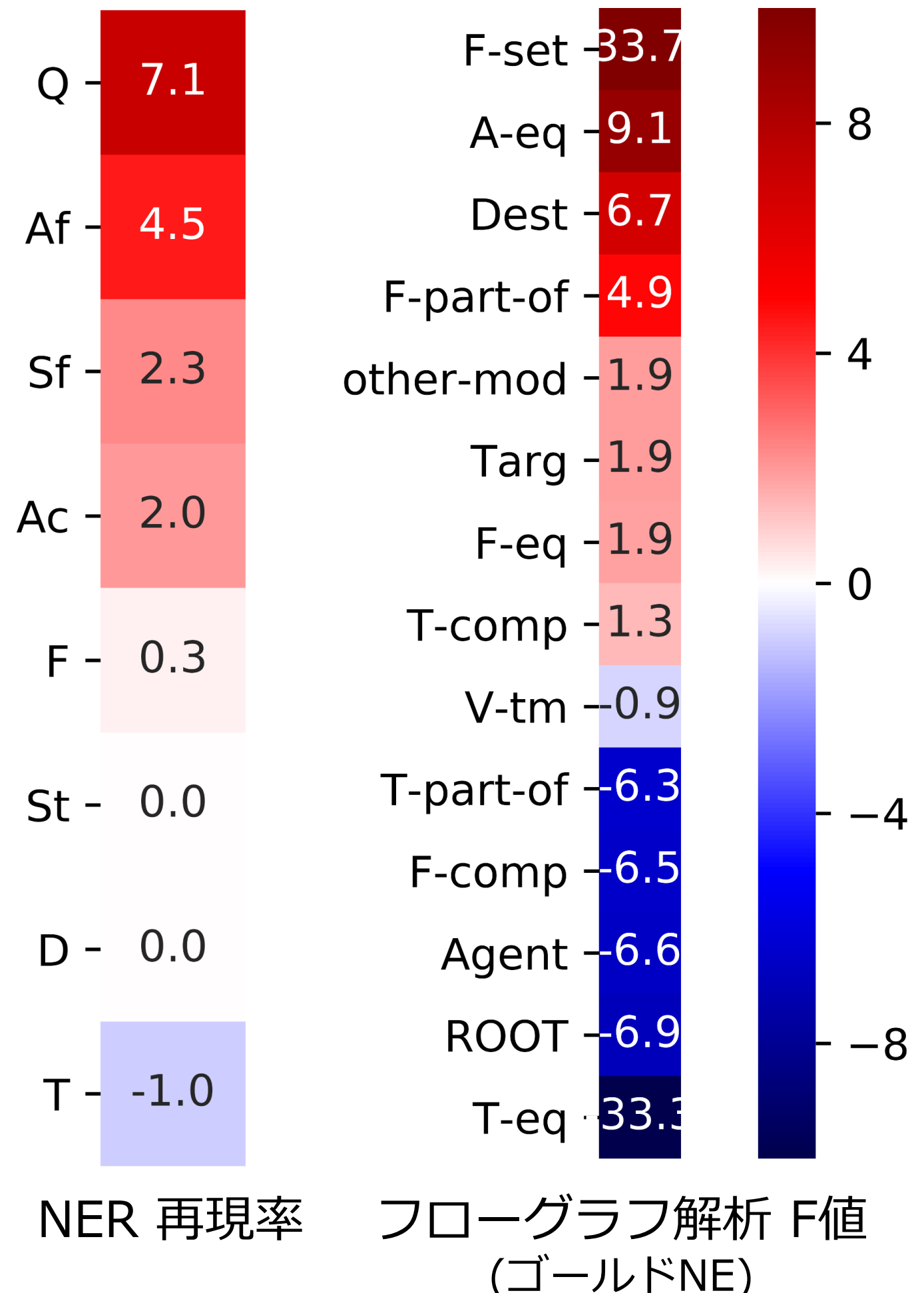


実験

実験設定

- 事前学習時
 - データセット
 - Wikipedia : 約1800万文
 - レシピ : 約60万文, 約45万画像
 - トークン : 超短単位をサブワード化
 - 語彙サイズ : 32,000
- ファインチューニング時
 - データセット
 - 学習, 開発, テスト : 245, 32, 30 (レシピ数)
 - NERと同時学習
 - ファインチューニング・解析時に画像は不使用

	NER F値	フローグラフ (ラベルなし) F値	フローグラフ F値
[Maeta+ 15] + ゴールドNE	-	78.8	69.8
BERT + ゴールドNE (画像なし事前学習)	-	79.8	72.3
提案手法 + ゴールドNE	-	80.5	74.0
BERT (画像なし事前学習)	88.4	68.9	64.8
提案手法	89.6	70.5	66.3



提案手法により、正解に変化した例

画像あり/画像なし事前学習の精度差
赤が画像あり事前学習の方が精度が高い