

# マスク言語モデルを利用した データ拡張に基づく 日本語文内ゼロ照応解析

---

今野颯人<sup>1</sup>,

松林優一郎<sup>1, 2</sup>, 清野舜<sup>2, 1</sup>,  
大内啓樹<sup>2, 1</sup>, 高橋諒<sup>1, 2</sup>, 乾健太郎<sup>1, 2</sup>

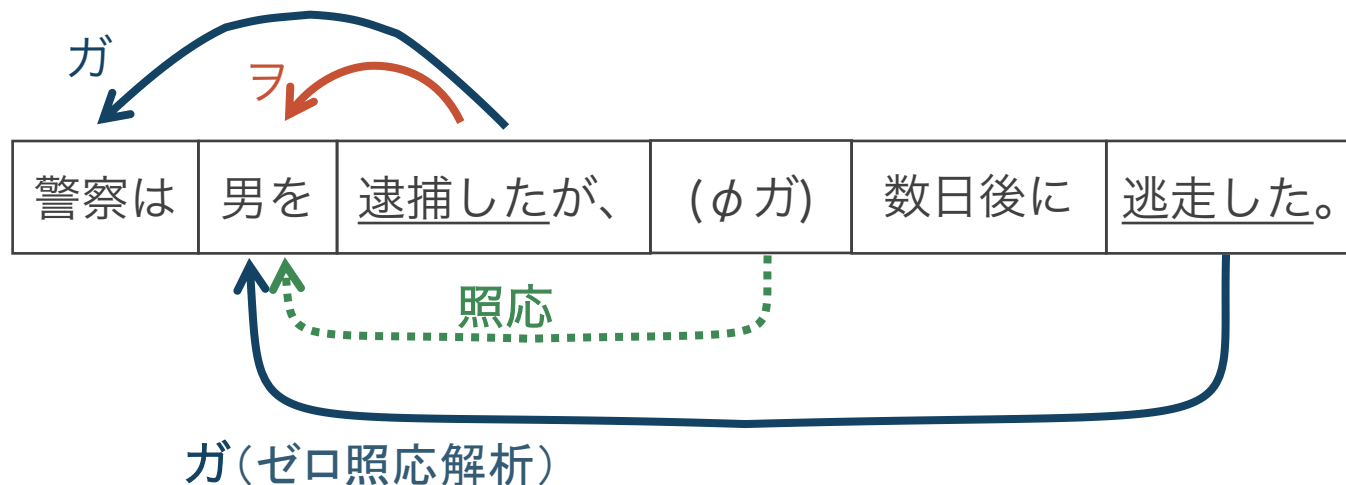
<sup>1</sup> 東北大学 <sup>2</sup> 理化学研究所

2019/3/19 言語処理学会第26回年次大会

# ゼロ照応解析

## ■ 文中の述語の省略された項を特定するタスク

- 例：「逃走した」の主格（ガ格）である「男」が省略



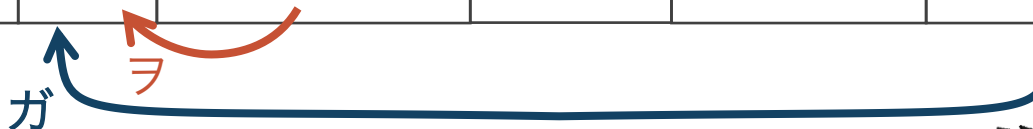
- 省略された項を特定し構造化された意味関係を獲得する
- 文の理解に重要

# 日本語ゼロ照応解析における問題点

## ■ 日本語ゼロ照応解析：解析が困難

解析対象を文内に限っても解析精度は $F_1$ 値で**58%**程度<sup>[1]</sup>

警察は	男を	逮捕したが、	( $\phi$ ガ)	数日後に	逃走した。
-----	----	--------	-------------	------	-------



逮捕される



逃走する



- 複雑な構文構造、統語的手がかりが少ない
- **訓練事例数が少ない**
  - 最も規模の大きなコーパスでも事例数は**2万程度**<sup>[2]</sup>
  - 人手によるアノテーションは高コスト

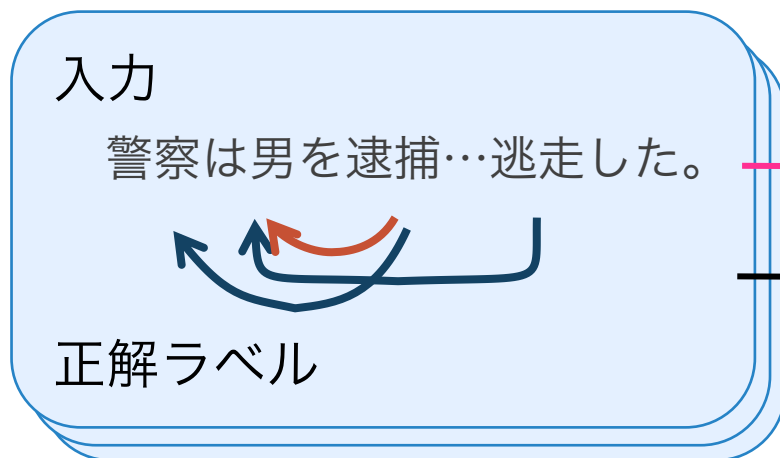
[1] Matsubayashi & Inui '18

[2] NAIST Text Corpus 文内ゼロ

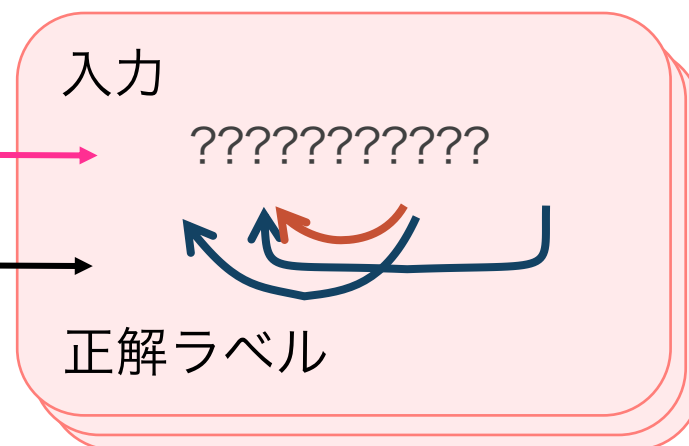
## 本研究の概要

- 目的：日本語ゼロ照応解析の精度向上
- 手段：データ拡張 (Data Augmentation)
  - データ不足問題の解消を試みる
  - 訓練データからの拡張を行う

アノテーション付き訓練データ



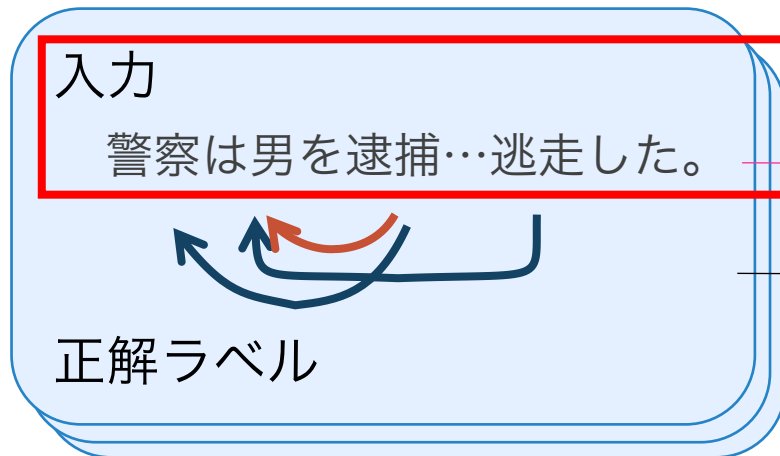
拡張された訓練データ



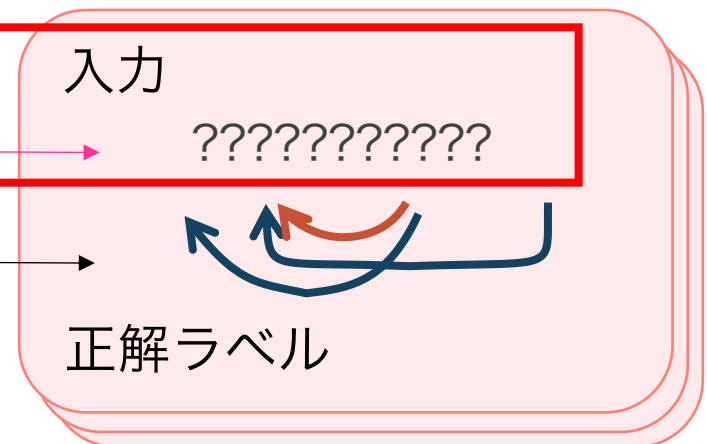
# 本研究の概要

- 目的：日本語ゼロ照応解析の精度向上
- 手段：データ拡張 (Data Augmentation)
  - データ不足問題の解消を試みる
  - 訓練データからの拡張を行う

アノテーション付き訓練データ



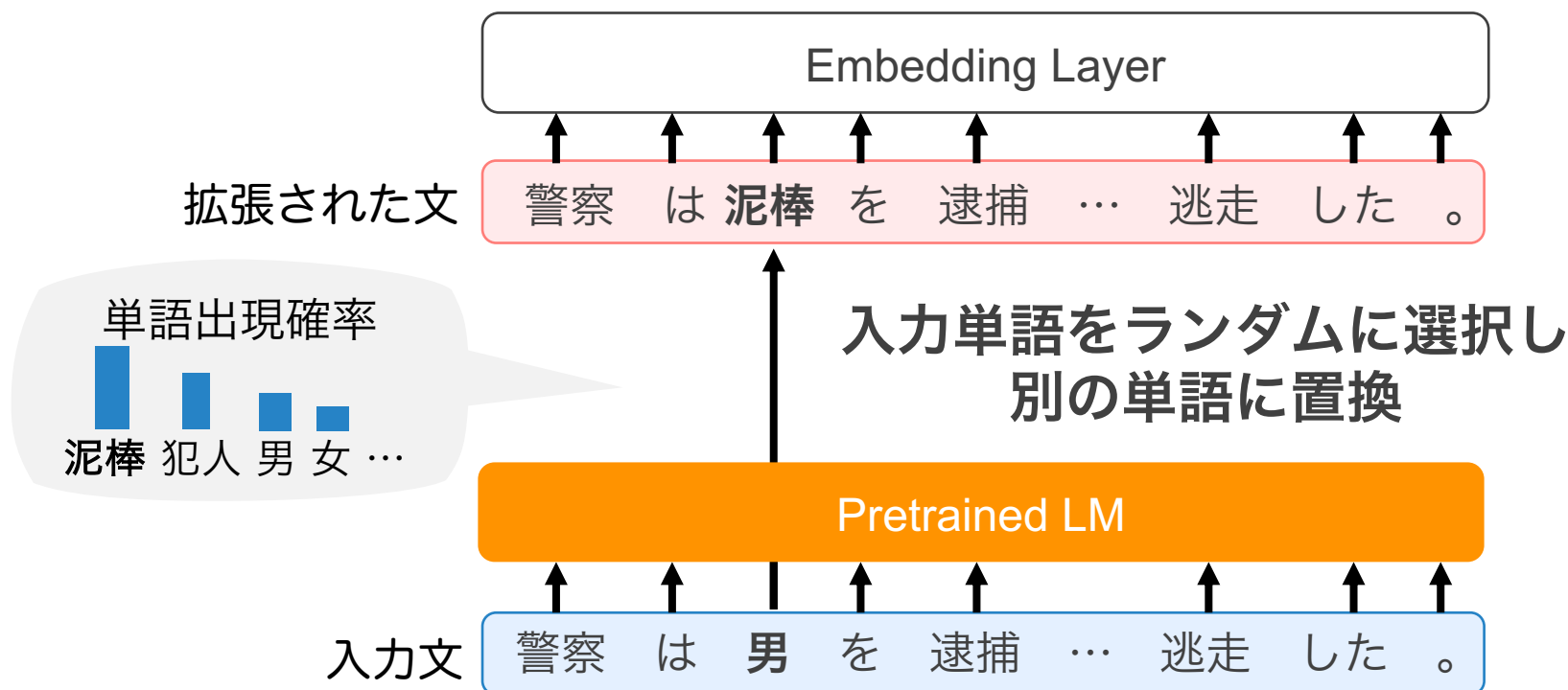
拡張された訓練データ



拡張

# 既存のデータ拡張手法： contextual data augmentation (CDA)

- 文脈を考慮して訓練データ中の単語を別の単語に置換



- MTや感情分析においてよい性能<sup>[3][4][5]</sup>

[3] Kobayashi'18

[4] Wu+'19

[5] Gao+'19

# CDAを日本語ゼロ照応解析に適用したい

## ■ 問題 1 : 言語モデルの使い方が異なる

提案手法① : マスクに基づくデータ拡張

- 二つの異なる言語モデルの使用方法を統合

## ■ 問題 2 : 任意の単語が置換対象となる

提案手法② : 言語情報を利用したマスク戦略

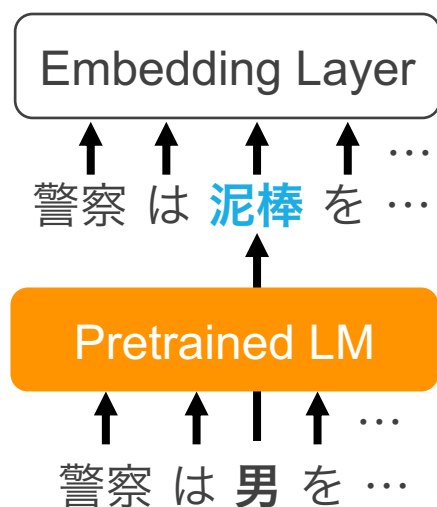
- マスクする箇所を制御

# 問題 1 : 言語モデルの使い方が異なる

## ■ 二つの言語モデル (LM) の使用方法

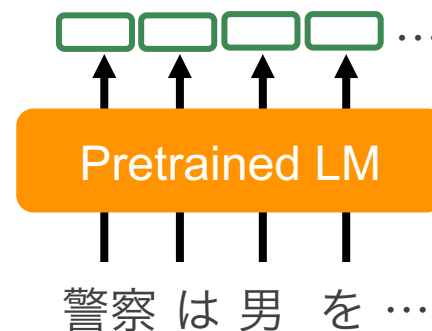
### 1. LM-for-replacement

- LMで単語置換
- CDAにおける使用方法



### 2. LM-as-feature

- LMの最終隠れ層が入力素性
- 分類問題 (SRL, NERなど) において一般的な使い方
- 精度が大幅に向上



## ■ 二つの方法をうまく統合したい



# CDAを日本語ゼロ照応解析に適用したい

## ■ 問題 1 : 言語モデルの使い方が異なる

提案手法① : マスクに基づくデータ拡張

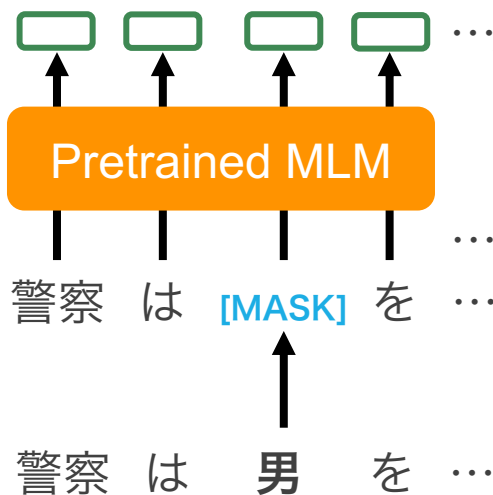
- 二つの異なる言語モデルの使用方法を統合

## ■ 問題 2 : 任意の単語が置換対象となる

提案手法② : 言語情報を利用したマスク戦略

- マスクする箇所を制御

# 提案手法①：マスクに基づくデータ拡張



MLMの最終隠れ層を  
入力素性とする

マスク言語モデル (MLM) の  
マスクトークンで置換

- MLMは[MASK]に文脈を考慮し単語を埋めるよう学習
- [MASK]の意味表現は文脈上適切な単語表現であると期待
- LM-for-replacementとLM-as-featureを統合

# CDAを日本語ゼロ照応解析に適用したい

## ■ 問題 1 : 言語モデルの使い方が異なる

提案手法① : マスクに基づくデータ拡張

- 二つの異なる言語モデルの使用方法を統合

## ■ 問題 2 : 任意の単語が置換対象となる

提案手法② : 言語情報を利用したマスク戦略

- マスクする箇所を制御

## 問題 2 : 任意の単語が置換対象となる

- 文の統語的・意味的な構造が変化する恐れあり
- 意味構造を表す正解ラベルとの整合性がとれない

例 1) 警察が男を逮捕したが、数日後に逃走した。



警察が男を **殺した** が、数日後に逃走した。



男が逃走



警察が逃走

例 2) 男が、逃走したことを知った。



男が**再び**逃走したことを知った。



男が知る



私（一人称）が知る

# CDAを日本語ゼロ照応解析に適用したい

## ■ 問題 1 : 言語モデルの使い方が異なる

提案手法① : マスクに基づくデータ拡張

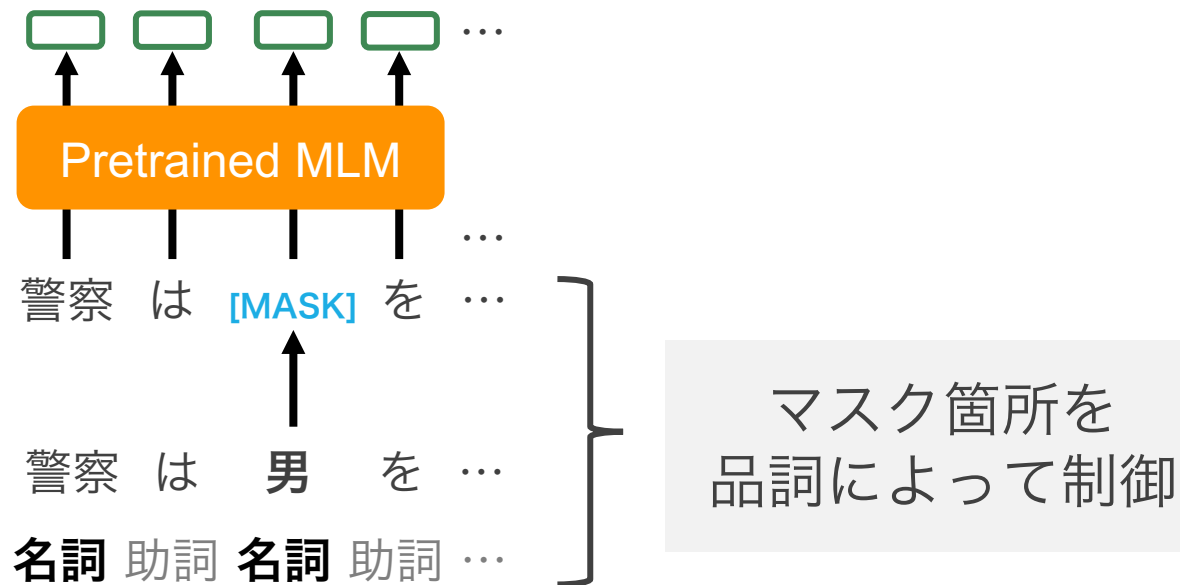
- 二つの異なる言語モデルの使用方法を統合

## ■ 問題 2 : 任意の単語が置換対象となる

提案手法② : 言語情報を利用したマスク戦略

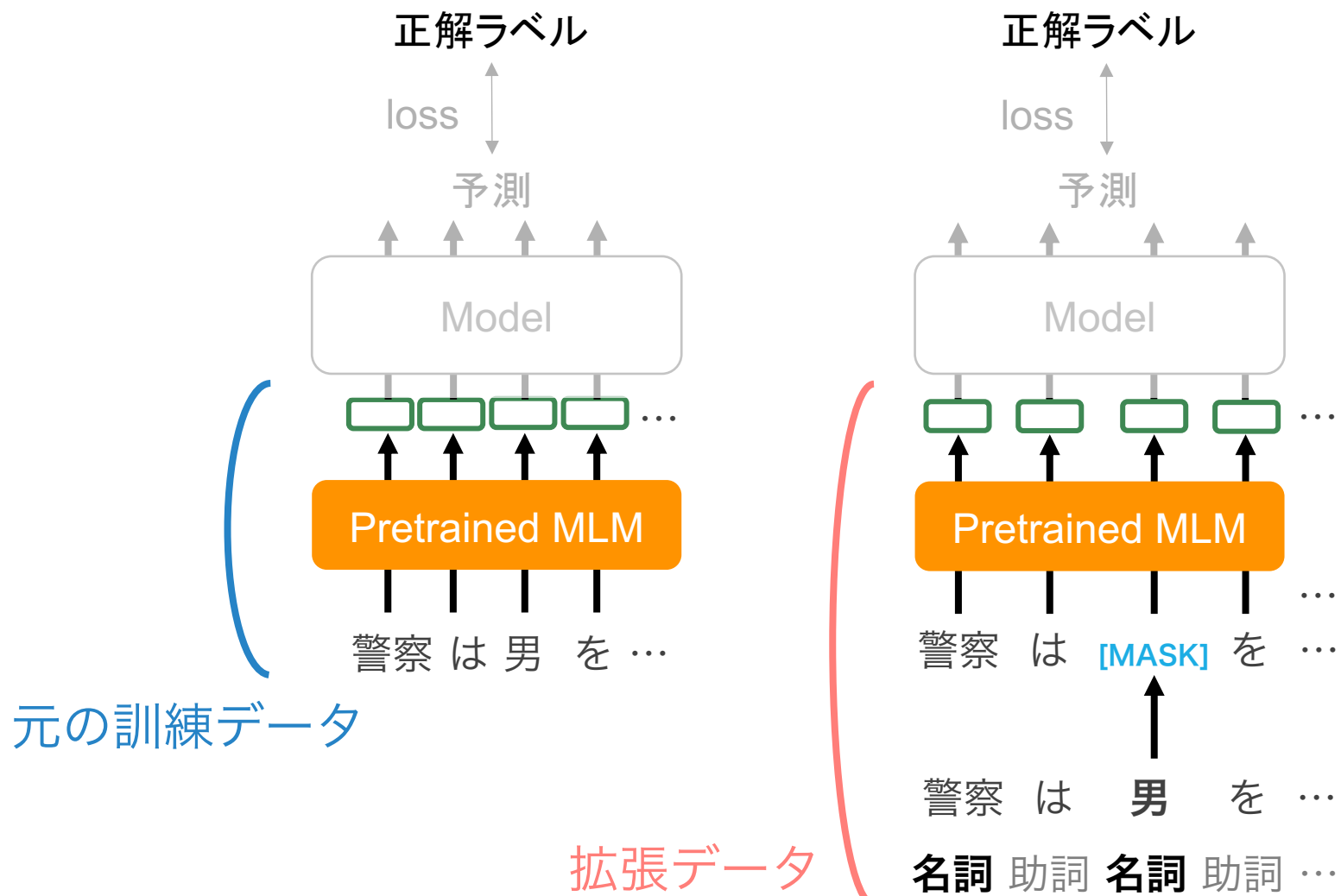
- マスクする箇所を制御

## 提案手法②：言語情報を利用したマスク戦略



- マスク箇所を品詞によって制御
  - 入力素性と正解ラベルとの構造的な整合性を保つ

# 提案手法のまとめ



# 実験

## ■ 1. 提案手法の効果の検証

以下の設定の組み合わせ全てでモデルを訓練

- 品詞タグの種類：{名詞, 動詞, 助詞, 記号, 全単語}
- マスク確率 $\alpha$ ：{0.1, 0.3, 0.5, 0.7, 0.9, 1.0}

## ■ 2. 他手法との比較

- 既存研究：
  - Matsubayashi & Inui '18
  - Omori & Komachi '19
- Baselineから性能の向上が見込める他手法
  - ZeroDrop
  - Replace



# 実験①：提案手法の効果の検証

## ■ 最適なマスク確率でのvalidation setにおける比較

マスク対象	ALL $F_1$	SD	文内ゼロ $F_1$
Baseline	87.43	$\pm 0.14$	64.08
全単語	87.64	$\pm 0.09$	64.89
名詞	87.53	$\pm 0.09$	64.62
動詞	87.35	$\pm 0.09$	64.15
助詞	87.43	$\pm 0.19$	64.31
記号	87.29	$\pm 0.16$	64.12
名詞以外	87.44	$\pm 0.16$	64.34
動詞以外	87.67	$\pm 0.11$	<b>65.02</b>
助詞以外	87.44	$\pm 0.15$	64.23
記号以外	87.59	$\pm 0.19$	64.66

提案①：

マスクに基づくデータ拡張  
が有効であることを示唆

動詞以外をマスク

→ 文内ゼロ  $F_1$  が最も高い

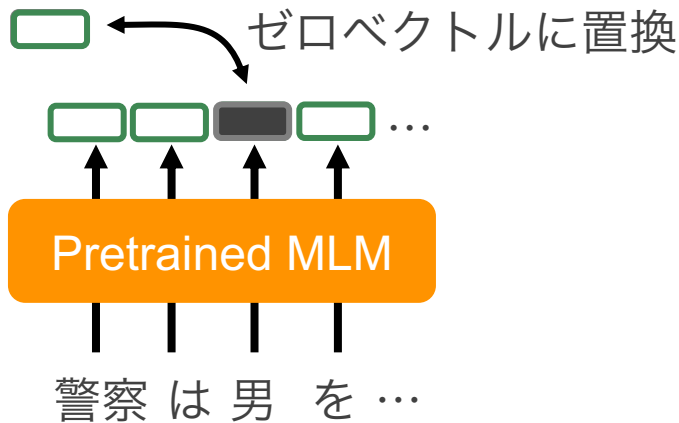
提案②：

言語情報を利用したマスク戦略  
が有効であることを示唆

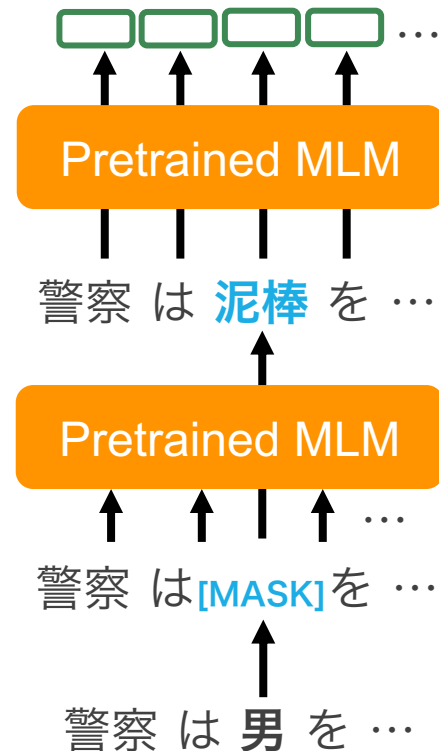
# 実験②：他手法との比較

- Baselineから性能の向上が見込める他手法との比較

## ZeroDrop



## Replace



- 既存研究との比較

- Matsubayashi & Inui '18
- Omori & Komachi '19

## 実験②：他手法との比較

### ■ test setにおける他手法との比較

Method	ALL	SD	文内ゼロ F <sub>1</sub>
Matsubayashi&Inui	83.94	±0.12	55.55
Omori&Komachi	83.82	±0.10	53.50
Baseline	86.85	±0.11	63.89
Replace	86.84	±0.19	63.87
ZeroDrop	86.94	±0.14	64.23
Masking	86.98	±0.13	64.15
ZeroDrop+Masking	87.14	±0.11	64.86

BERTにより

Baselineが既存研究を大幅に上回っている

Replaceは  
Baselineと  
ほぼ同等の性能

Replace：単純にCDAを適用することは  
LM-as-featureのモデルには効果が見込めない

## 実験②：他手法との比較

### ■ test setにおける他手法との比較

Method	ALL	SD	文内ゼロ $F_1$
Matsubayashi&Inui	83.94	$\pm 0.12$	55.55
Omori&Komachi	83.82	$\pm 0.10$	53.50
Baseline	86.85	$\pm 0.11$	63.89
Replace	86.84	$\pm 0.19$	63.87
ZeroDrop	86.94	$\pm 0.14$	64.23
Masking	86.98	$\pm 0.13$	64.15
ZeroDrop+Masking	<b>87.14</b>	$\pm 0.11$	<b>64.86</b>

ZeroDropとMaskingは  
Baselineから性能が向上

ZeroDropとMaskingを組み合わせることにより最高精度を達成

# 考察

- ZeroDropとMaskingを組み合わせた手法が最高精度を達成
  - これら二つの方法が相補的な関係にあることを示唆
- ZeroDropがノイズを除去している
  - マスクトークンが意味のない反復的な助詞や句読点で埋められていることがある
  - 訓練中にノイズとして機能するこれらを除去

										NOM		PRED
$X$	内閣	改造	を	通常	国会	召集	前	に	やる	考え	は	ない
$X'$	内閣	[M]	[M]	[M]	[M]	[M]	[M]	に	やる	考え	[M]	ない
$X''$	内閣	は	は	,	,	,	ため	に	やる	考え	は	ない

# 結論

- 問題提起：日本語ゼロ照応解析における**データ不足**
- 提案手法：CDAを日本語ゼロ照応解析に適用
  - (1)マスクに基づくデータ拡張
  - (2)言語情報を利用したマスク戦略
- 実験：
  - (1) 提案手法の効果の検証
    - **動詞以外をマスク**する設定が最も良かった
  - (2)他手法との比較
    - **MaskingとZeroDropを組み合わせたデータ拡張**で最高精度を達成
- 現在就職活動中です