

- | 計量的な
- | 語の意味分析から
- | 視点と主観は
- | 捉えられるのか

言語処理学会 第29回年次大会

「深層学習時代の計算言語学」

森下 裕三

桃山学院大学

本研究は JSPS 科研費 20K13069 の助成を受けている

INTRODUCTION

00 | 認知言語学とコーパス言語学

| 沖縄に来る前に深層学習について「事前学習」

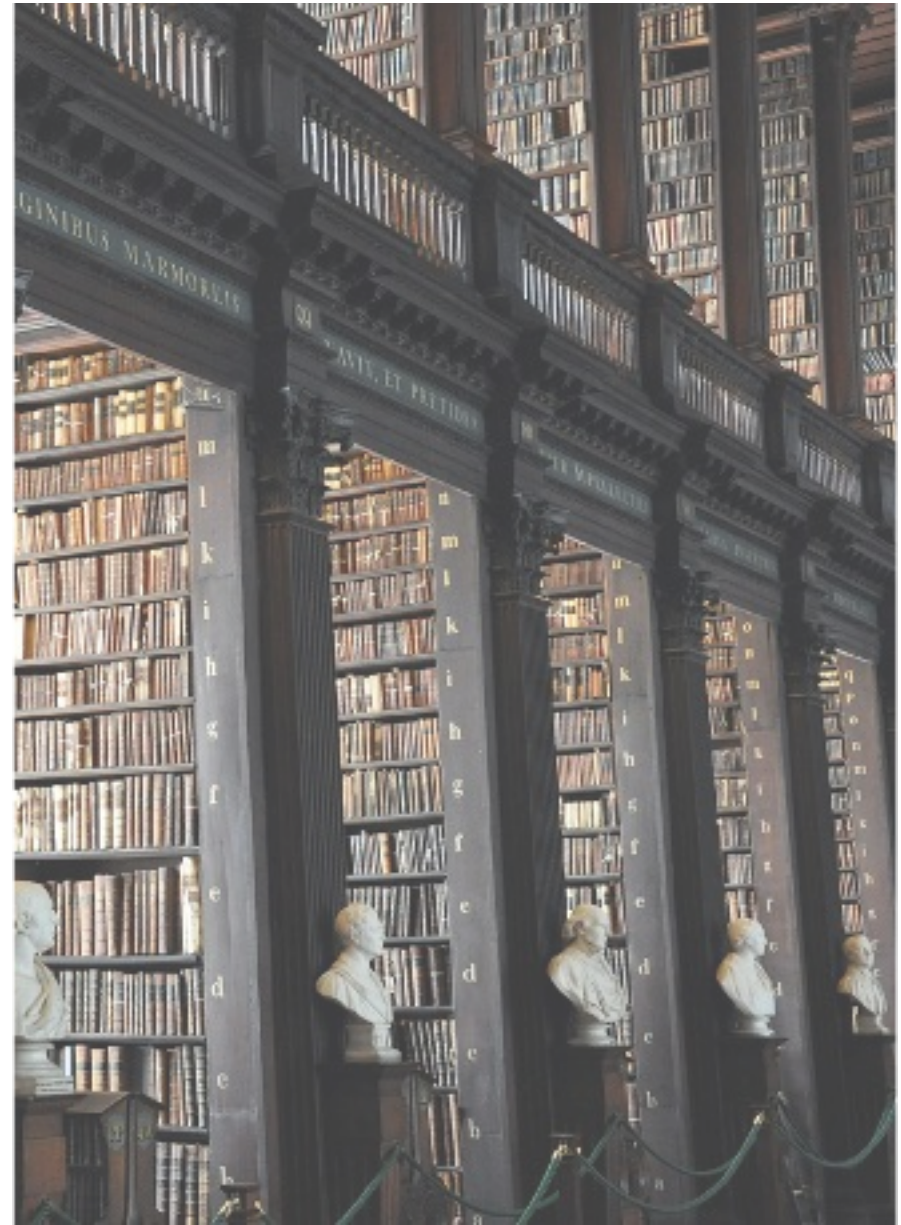
| **NLP2023** で **fine-tuning**

CONTENTS

- 01 | backgrounds and aims
研究の背景と目的
- 02 | Methodologies and results
実験手法と実験結果
- 03 | linguistics and NLP
言語学と自然言語処理

Backgrounds and aims

01 | 研究の背景と目的



研究の背景

語の意味 | ことばがどう使われているのかを考える

Wittgenstein (1953)

| 共起する語を知らなければならない

Firth (1957)

Sinclair (1991) ▶ カウントベース

構造主義 | 分布仮説に基づく意味の理論

| 似た文脈に生起する語は似た意味を持つ

Harris (1954) ▶ 推論ベース

研究の目的

視 点 | 語の背後にある視点を客観化したい

▷ **frame semantics** e.g. Fillmore (1980)

類義関係 | ***land* vs. *ground***

研究の目的

視 点 | 語の背後にある視点を客観化したい

▷ **Deictic Path Verbs.** e.g. Matsumoto+ (2017)

| **come** and **go**

NLP の実験方法を実装

分散表現 | カウントベース

| 推論ベース [BERT]

word prediction ▷ attention

カウントベース

corpus | **Penn Tree Bank (pre-processed)**

window size |

w_{t-4}	w_{t-3}	w_{t-2}	w_{t-1}	w_t	w_{t+1}	w_{t+2}	w_{t+3}	w_{t+4}
-----------	-----------	-----------	-----------	-------	-----------	-----------	-----------	-----------

procedure | [...] the balloon to the **ground** <unk> all the air [...]
[...] much of their lost **ground**

▷ 共起行列作成

$$\text{PMI}(x, y) = \log_2 \frac{P(x, y)}{P(x) P(y)}$$

特異値分解 (Singular Value Decomposition: SVD)

カウントベース

[query]

land

acres: 0.609

desert: 0.580

building: 0.558

las: 0.548

[...]

tours: 0.461

カウントベース

[query]	ground
	loose: 0.406
	broke: 0.391
	break: 0.3851
	egg: 0.3693
	[...]
	losing: 0.343

カウントベース

[query]	come
	out: 0.403
	bother: 0.381
	go: 0.381
	las: 0.548
	[...]
	came: 0.348

カウントベース

[query]

go

forth: 0.498

trap: 0.447

maybe: 0.430

sleep: 0.4134

[...]

come: 0.3815

推論ベース [BERT]

BERT

| **pre-trained model**

prediction

| We can live off the **land**.

| ▷ We can live off the **[MASK]**.

'land' | weights: 0.1609

'water' | weights: 0.0528

'money' | weights: 0.0471

'ground' | weights: 0.0261

推論ベース [BERT]

BERT

pre-trained model

prediction

Much of the surrounding **land** was covered in dung.

▷ Much of the surrounding **[MASK]** was covered in dung.

'area' | weights: 0.3450

'ground' | weights: 0.1479

'land' | weights: 0.1382

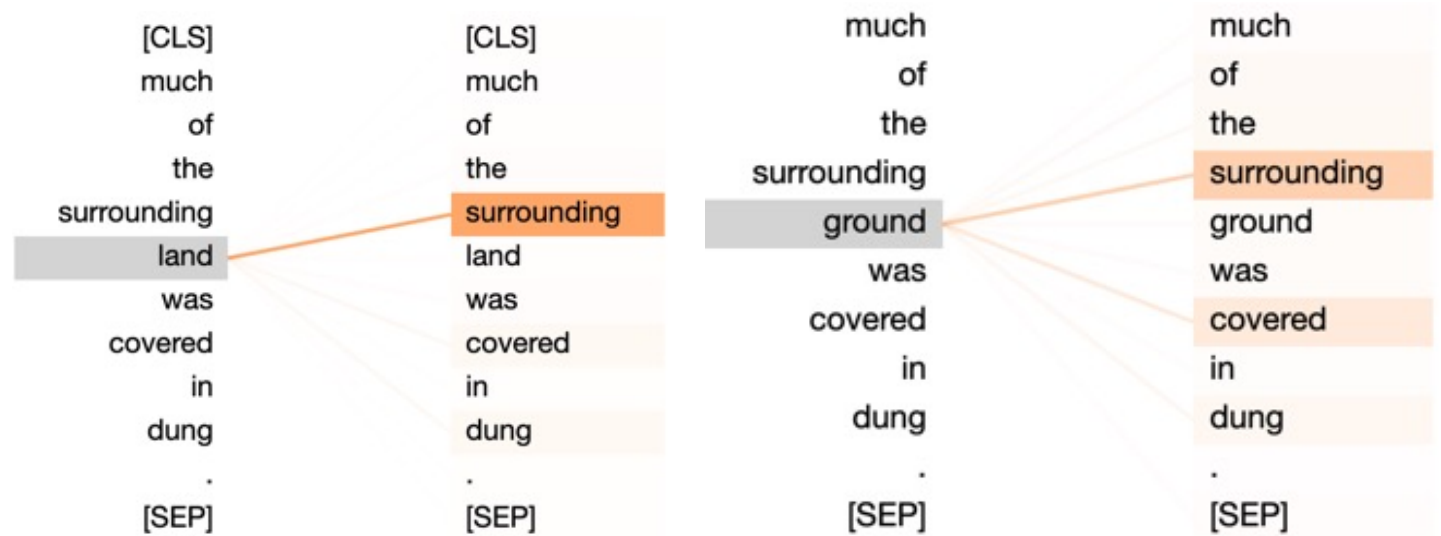
推論ベース [BERT]

BERT

pre-trained model

exchange

Much of the surrounding {land/ground} was covered in dung.



推論ベース [BERT]

BERT

| **pre-trained model**

prediction

| Money should **come** out of every department.

| ▷ Money should **[MASK]** out of every department.

'come' | weights: 0.3367

'flow' | weights: 0.1622

'be' | weights: 0.1131

'go' | weights: 0.1125

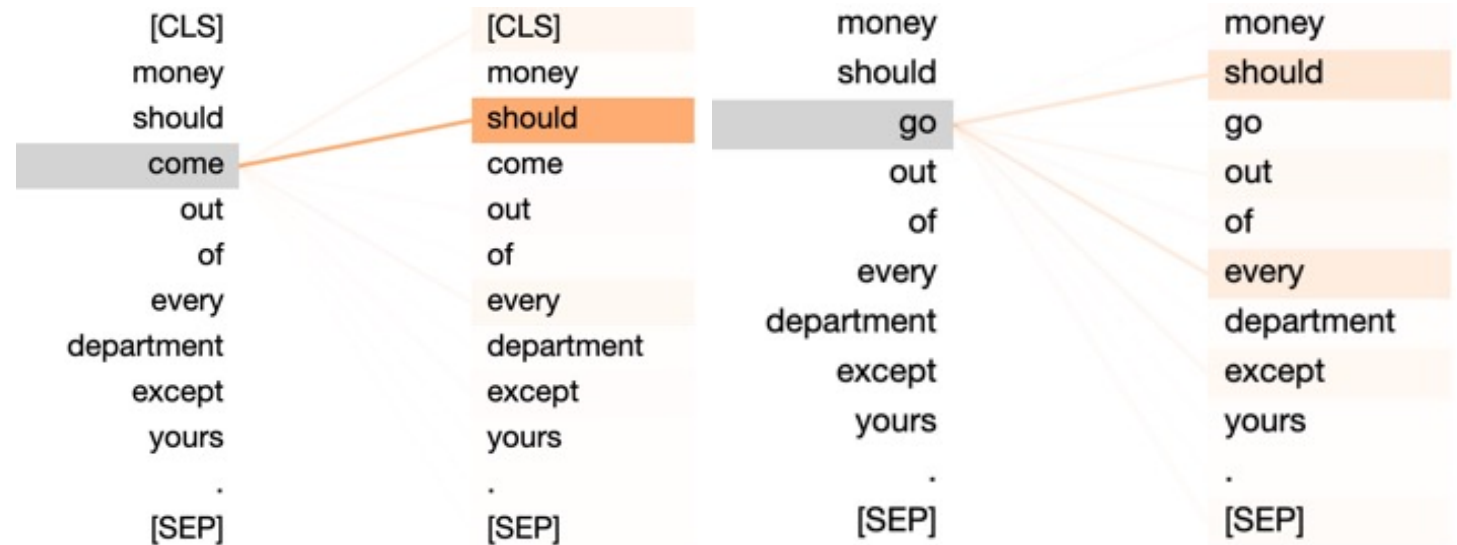
推論ベース [BERT]

BERT

pre-trained model

exchange

Money should {**come/go**} out of every department.



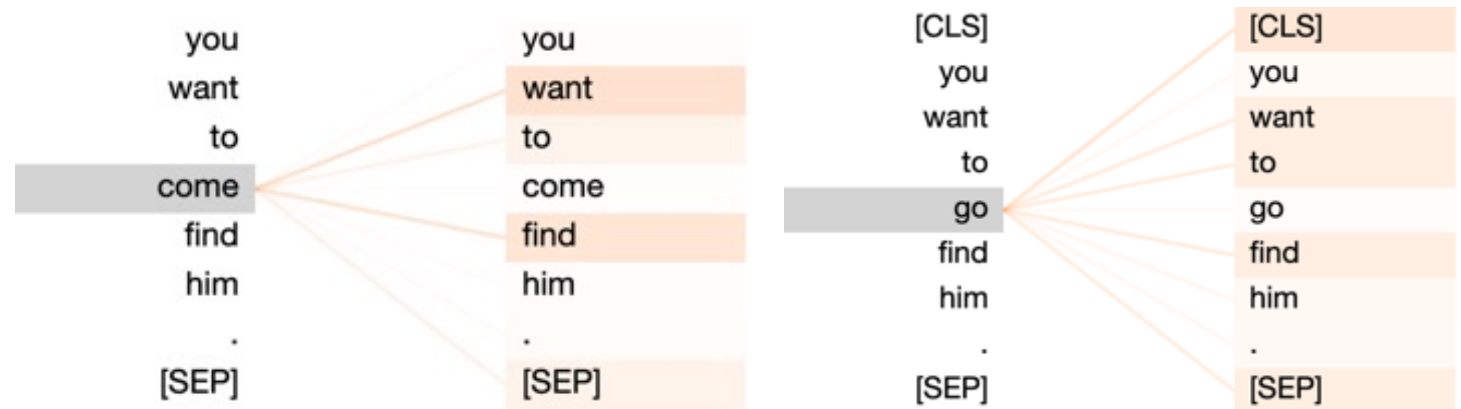
推論ベース [BERT]

BERT

pre-trained model

exchange

You want to {come/go} find him.



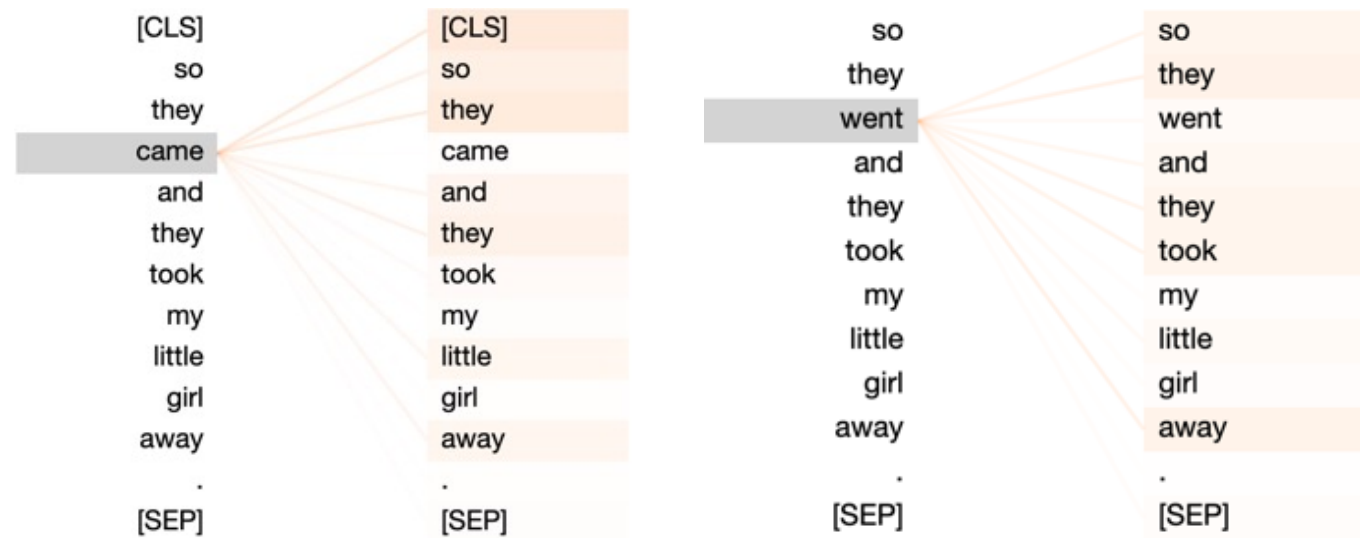
推論ベース [BERT]

BERT

pre-trained model

exchange

They **{came/went}** and they took my little girl away.



考察

問題 | なぜ **come** は **go** より助動詞や主語の影響大

先行研究 | 従来の研究では動詞の後続要素の方が重要

Fillmore (1971)

He came to me.

I'm coming to the restaurant.

本研究 | 従来の研究とは異なる結果

cf. Matsumoto+

今後の展望

言語観察 | 大規模コーパス [e.g. BNC, COCA] の用例

モデル化 | 例：ロジスティック回帰で **come/go** を分類

今 後 | 仮説検証のみならず探索的な **NLP** の活用