



# 深層学習時代の 計算言語学

持橋大地

統計数理研究所

NLP2023ワークショップ2

# オーガナイザー

○持橋大地 (統数研)

○窪田悠介 (国語研)

永田亮 (甲南大)

小木曾智信 (国語研)

大谷直輝 (東外大)

川崎義史 (東大)

小町守 (都立大)

高村大也 (産総研)

内田諭 (九大)

大関洋平 (東大)

(○は世話役)

# 言語処理学会と言語学

- 言語学者の参加が少ない時代が続いた
- その中でも、NLPの中で言語学をしようという試みは継続 → 永田さんイントロ
- なぜ少なかったのか？

# 以前の状況

NLPのモデルの複雑度 << 言語学が期待する  
複雑度

- という時代が長く続いた
- 例：nグラムモデル 「そんなものは言語学ではない！」
- 一方で、近藤先生の和歌の研究のような先駆的な試み (印度哲学でも同様の試みがあった)

# 現在の状況

NLPのモデルの複雑度 >> 言語学で扱っている  
複雑度

- 深層学習で、NLPのモデルが一気に複雑に！
- GPT-4は、もはや言語学が扱ってきた言語現象をほとんど再現しているように見える
- GPT-4はいったい何をしているのか  
→ あれ？人間を理解することと一緒に??

# 重要な違い

- LLM (GPT-4のような) は、摩訶不思議な言語能力を持っているが、その内部は完全に機械的  
← 人間の理解とは異なる
- そこで、
  - LLMの振る舞いを機械的(数学的)に解析すること
  - 人間の言語処理と比べること (含 脳科学による人間の解析)が重要になるのではないか

# 言語学冬の時代から、春の時代へ

- 目の前に、解析対象となる面白いLLMが  
転がっている  
→ 言語学の力を借りて、理解を深めよう
- 言語学的研究の試み：  
本WSでの発表のほかに、NLP2023では
  - D12-1: 窪田さんの寄生スコープの研究
  - B10-2: 近藤先生の言語地理学の研究などがあります (他にも多数)