

未来創造PJ

古文-現代文の機械翻訳

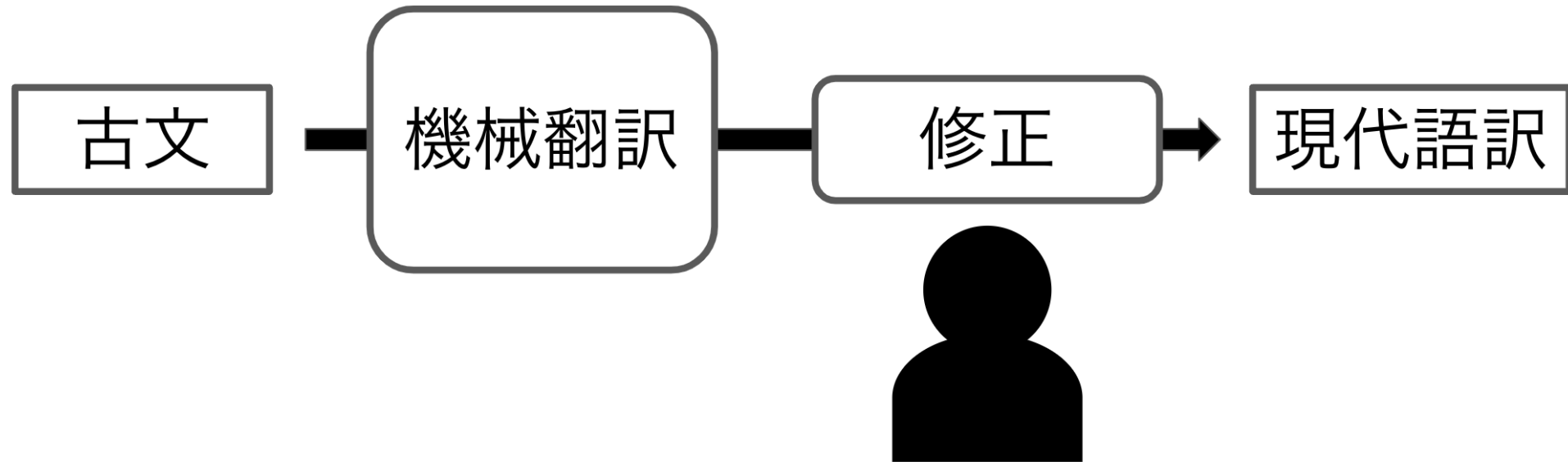
2222027 中城裕之

テーマ：古文-現代文の機械翻訳

- 古文-現代文のニューラル機械翻訳モデルを構築する
- 精度を向上させ、**実用化**を目指す

社会的意義

- 歴史研究者向けポストエディットサービス
研究者の翻訳作業を機械翻訳で効率化する



社会的意義

- 一般層向け古文翻訳サービス

大河ドラマなどの流行により、歴史ファンが増加

一般層に古文を読むのは難しい

“生の歴史”である古文に触れることをサポートする

1年後期の成果

- Transformerを用いてニューラル機械翻訳システムを構築

学習データは源氏物語1万5千文

推論データは紫式部日記、平家物語など

翻訳結果

この一門にはならない人は、皆非難されるのであった	この一門でない人は、みな、人でないとおっしゃった
秋の様子がお入りになって、お部屋の境遇は、どうしてもどうしてもよく美しい	秋の風情が現れ立ってくるにつれて、土御門邸の様子は、何とも言い表わしようもないほどに趣がある
無理を手紙の鐘の音、一緒にいつもの響き合っている	祇園精舎の鐘の音には、諸行無常すなわちこの世のすべての現象は絶えず変化していくものだという響きがある
今は昔、どうしてもどうしているものであった	今となっては昔のことだが、竹取の翁というものがいた
このような道路は、どうしてもどうしてお帰りになる	こんな道を、どうして行かれるのですか

課題

- データがない

古文の現代語訳には著作権が発生するため、ネット上で集めるには限度がある

- データを所持している研究機関はあるが、こちらに実績がないので共同研究に応じてもらえない

LLMによる 古文-現代文の機械翻訳

概要

- GPT等の大規模言語モデルは、プロンプトを調整することでその能力を更に引き出すことができる
- 古文翻訳に最適なプロンプトを探索する
- 古文翻訳データをプロンプトとして与え、翻訳力を向上させる