

Transformerを用いた古文→現代文のニューラル機械翻訳



中城裕之 / 武蔵野大学データサイエンス学部1年 / 佐々木・ウィラット・ソムゼミ

研究概要

深層学習モデル「Transformer」を用いた機械翻訳システムを構築し、古文から現代文へのニューラル機械翻訳を行う。
学習データには源氏物語約1万5000文対を使用する。

活用

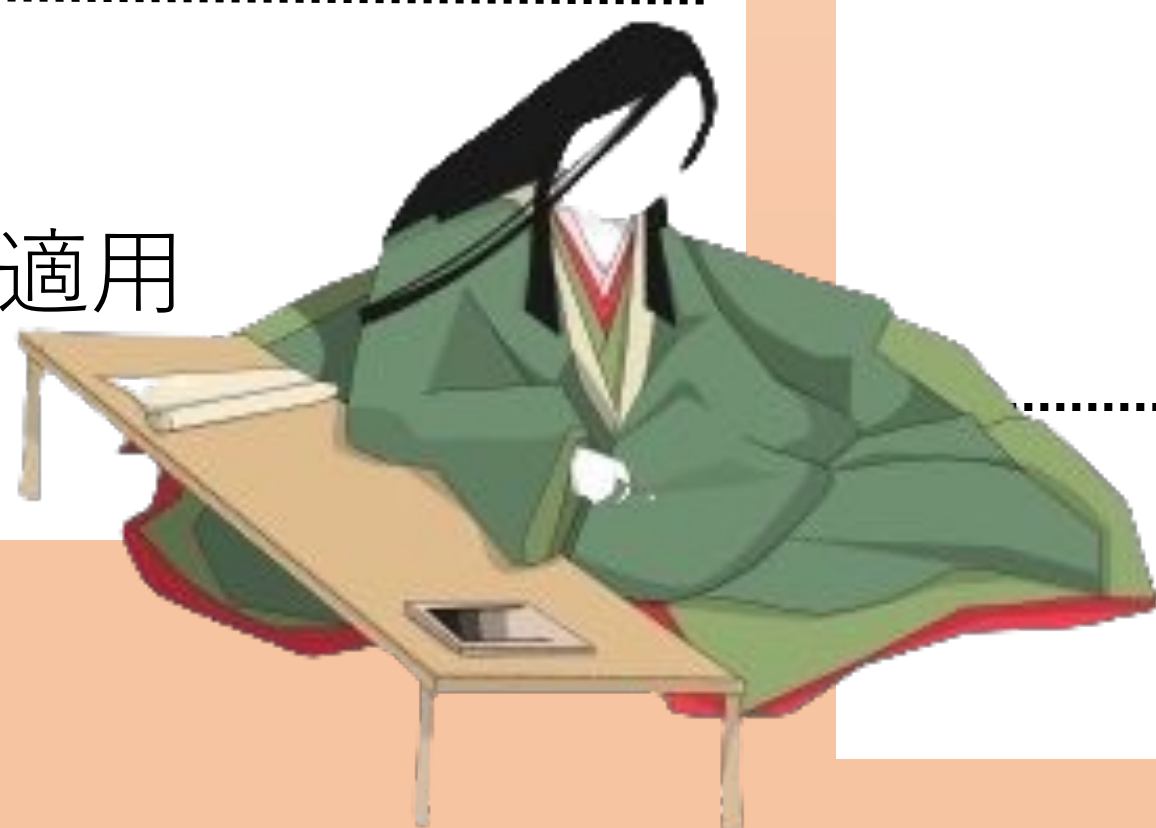
- 歴史研究向けポストエディットサービス
- 一般層向け翻訳サービス

背景

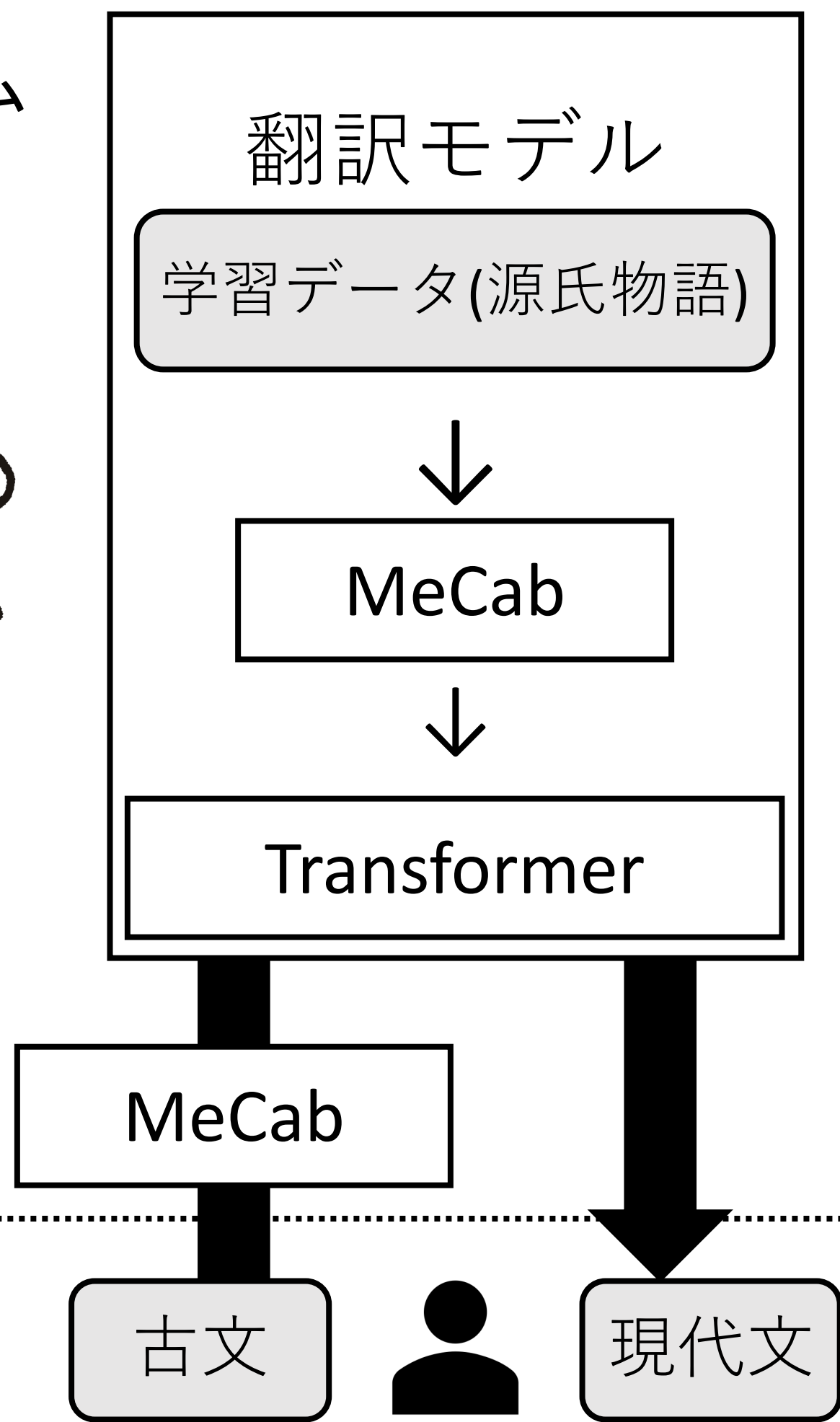
古文の機械翻訳研究は外国語より大幅に遅れている。古文翻訳システムを広く普及させるためには、この遅れを取り戻す必要がある。

目的

- Transformerの古文翻訳への適用
- NMTモデルの精度向上



システム構成図



成功例(紫式部日記)

原文	御帳の東面は、内裏の女房参り集ひてさぶらふ
人手翻訳	御帳台の東面の間には、主上付きの女房たちが参集して伺候する
翻訳結果	御帳台の東面は、宮中の女房を参集して伺候する

失敗例(平家物語)

原文	この一門にあらざらむ人は、みな人非人なるべしとぞ宣ひける
人手翻訳	この一門でない人は、みな、人でないとおっしゃった
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

まとめ

- 長文に特に弱い
- 時代が違う文章に対応できなかった



今後の課題

- 通時性を持たせる
- ↓
- 新規アルゴリズムの実装
 - データ量を増やす

