

ニューラル自然言語処理

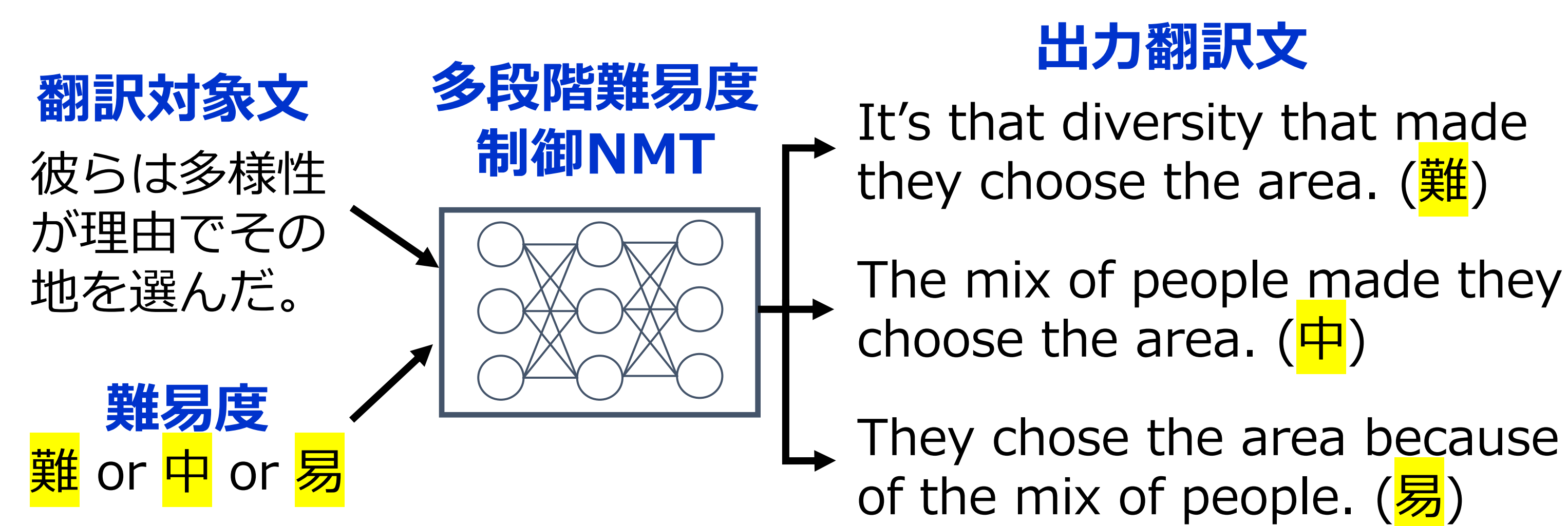
～人間のコミュニケーション支援や知的活動・業務支援に向けて～

概要

- ニューラルネットワークに基づく自然言語処理技術(ことばを計算機で処理・解析する技術)に関する⑦つの取り組みについて。

① 難易度制御機械翻訳

- 出力翻訳文の難易度を多段階で指定可能な機械翻訳



→ ユーザーの読解レベルにあわせた翻訳が可能

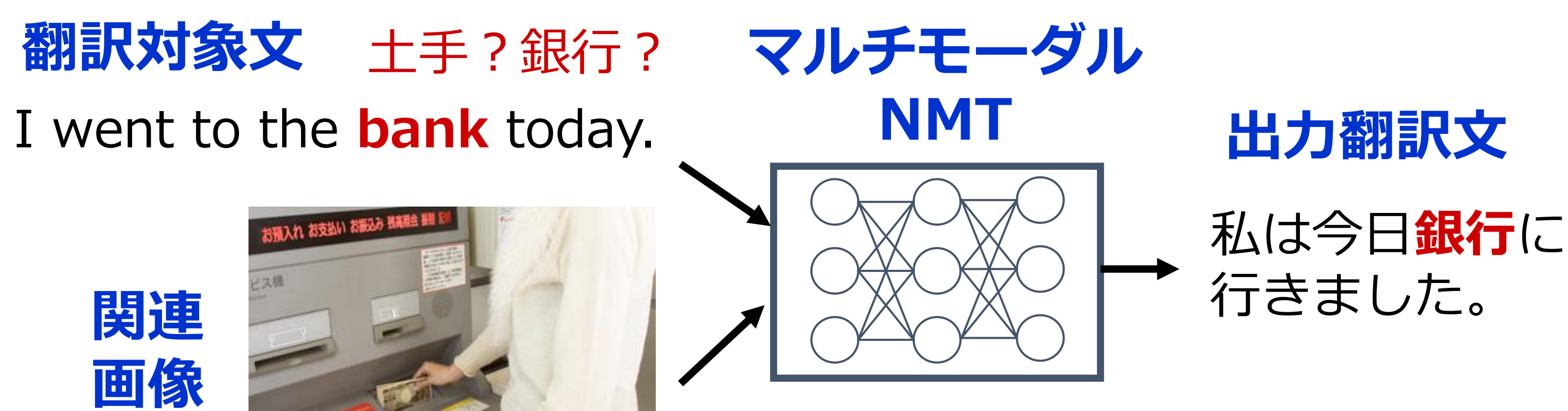
日英多段階難易度制御機械翻訳を実現^[1]

- ベンチマークデータ(教師データ, 評価データ)を構築
- 難易度制御機械翻訳向け学習方法を考案



② マルチモーダル機械翻訳

- 翻訳に関連する画像も入力とする機械翻訳



→ 関連画像を翻訳時の曖昧性解消などに活用し翻訳性能向上

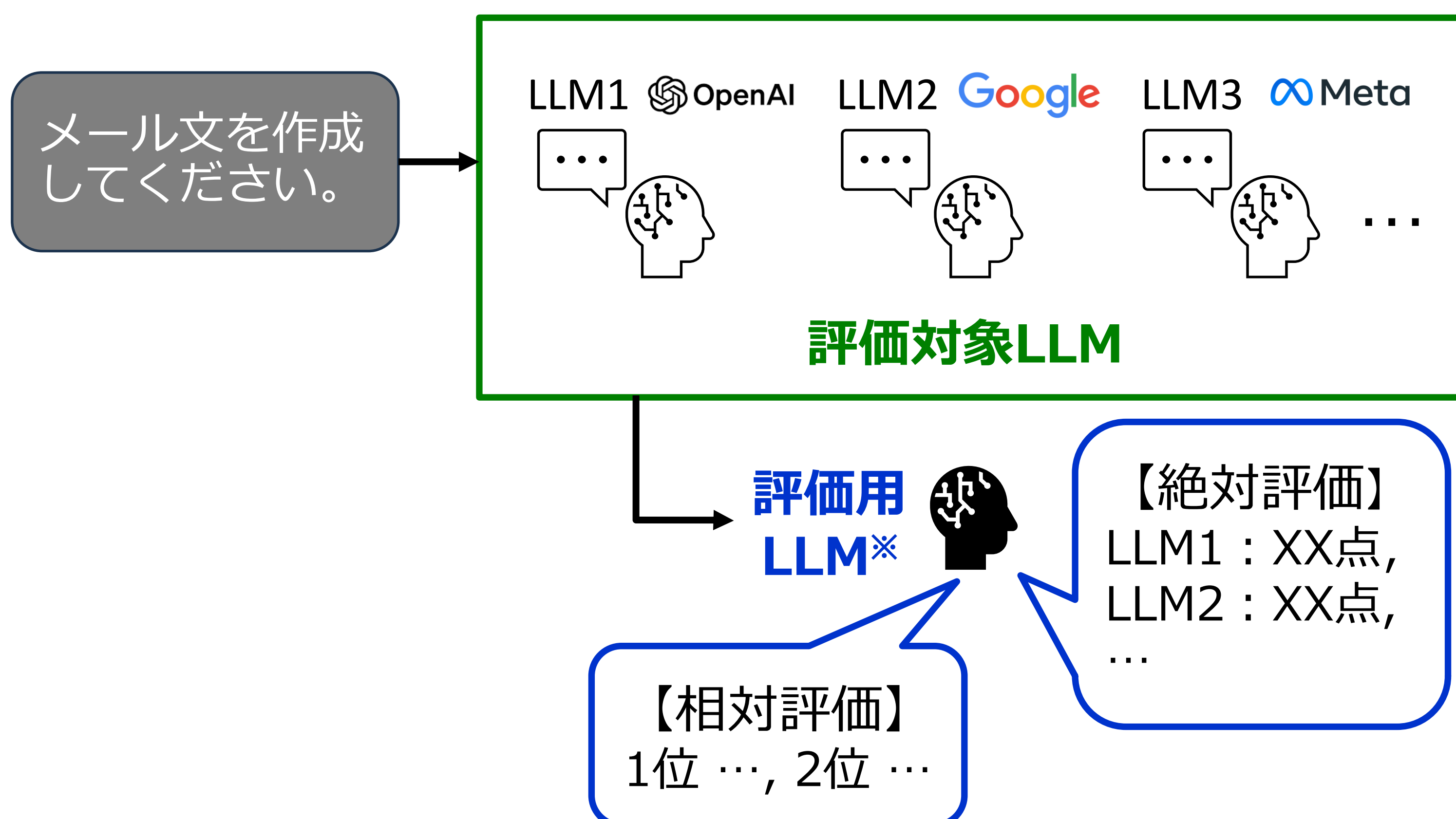
入力画像を翻訳に適した画像に変換して使う
マルチモーダルNMTを考案し、翻訳性能改善^[2]

- Ambiguous COCOデータ: BLEU(%) 27.91 → 28.30



③ 大規模言語モデル(LLM)の評価

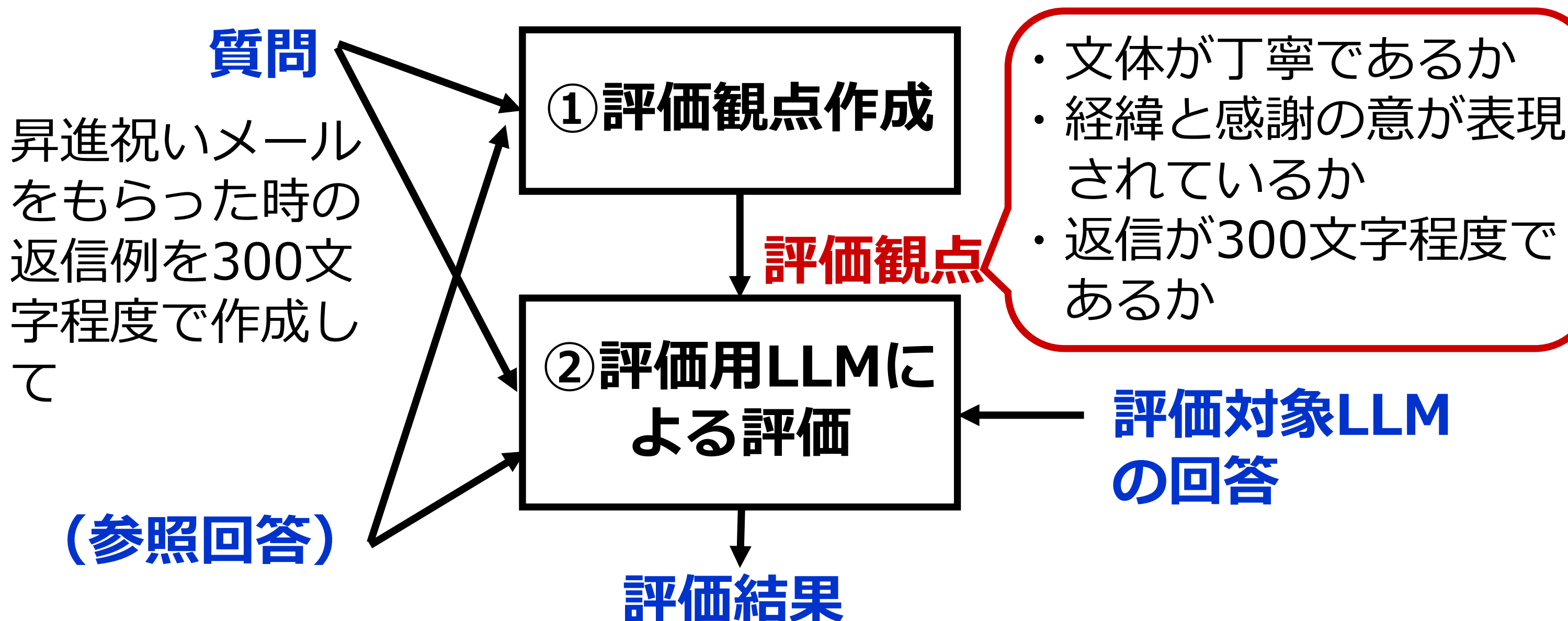
- 高性能なLLM*を用いてLLMの性能を自動評価



→ 乱立するLLMを比較可能

自動作成した評価観点をを用いるLLMによるLLMの評価手法を考案し、ビジネスタスク(文章作成, 校正, 要約)で有効性を確認^[3]

- 従来手法の問題点: 同じ視点で評価がされていない
- 実験結果:
 - ✓ 絶対評価: 人手評価との二乗平均平方根誤差 1.06 → 1.03
 - ✓ 相対評価: 人手評価との一致率(%) 61.0 → 62.8



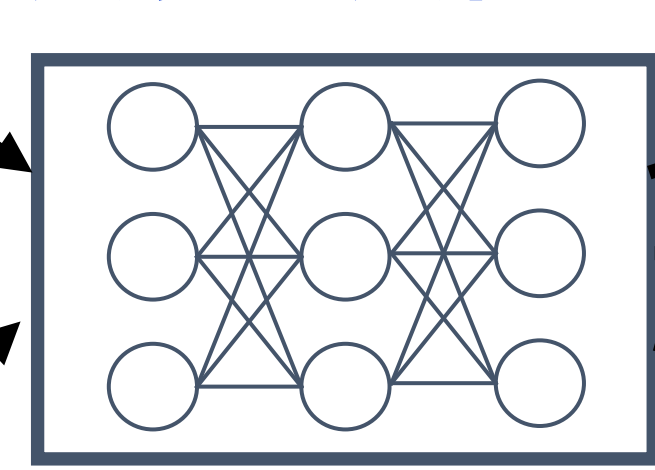
④ スタンス検出

- ニュース記事やSNS上での投稿に対して書き手のスタンスを自動特定

分析対象テキスト

通学時間もかからない
しオンライン授業って
効率的だね

スタンス検出



スタンス

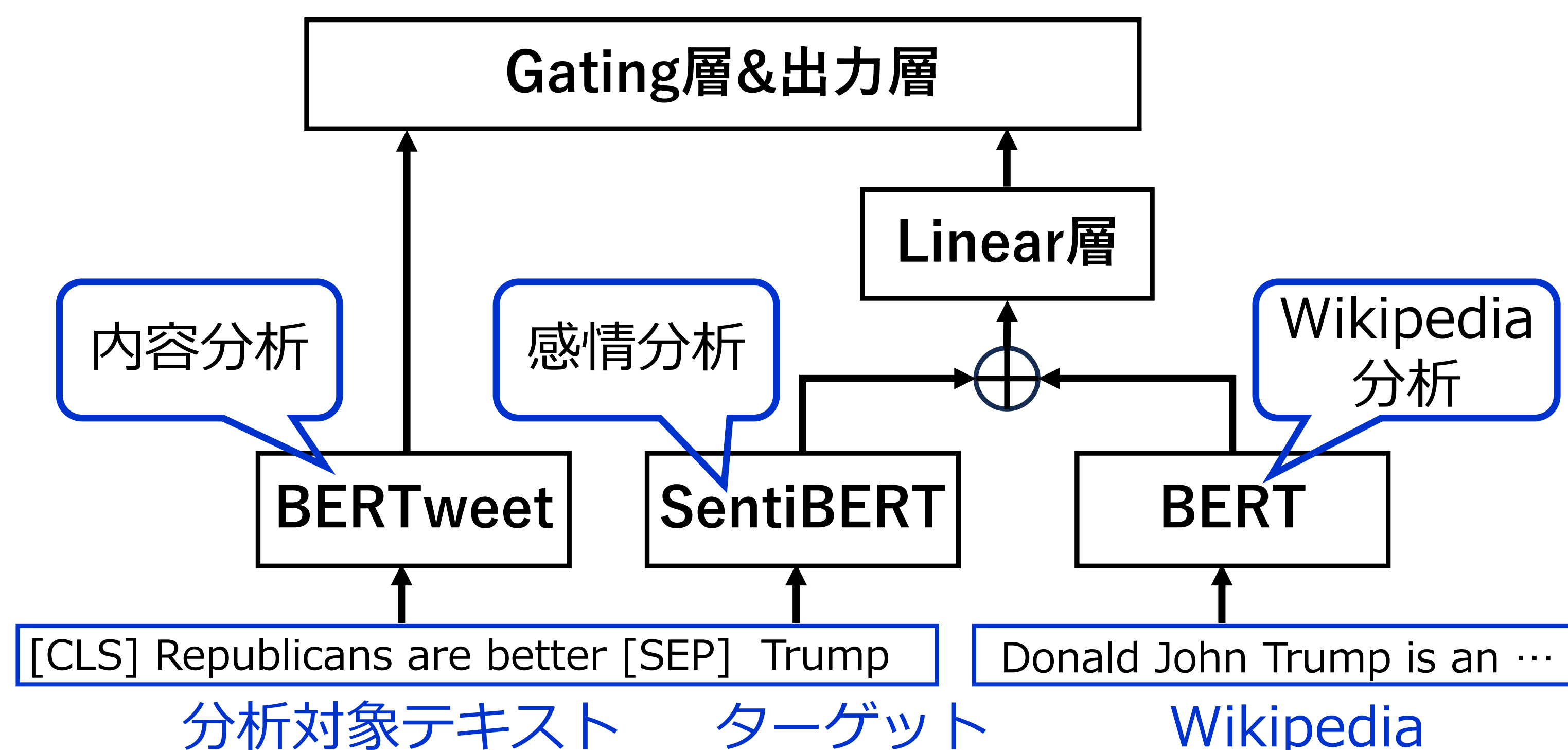


ターゲット
オンライン授業

→ 意見分析, 情報の妄信の回避などに役立つ

Wikipedia情報と感情情報を考慮するスタンス検出モデルを考案し, 検出精度改善^[4]

- P-stanceデータ: F値(%) 82.4 → 83.7



⑤ 特許の文書表現

- 特許文書をベクトルで表現

特許の文書構造を考慮した表現手法を考案^[5]

- 段落のつながりを判定できるようなベクトルで表現
- 各段落の見出しを推定できるようなベクトルで表現

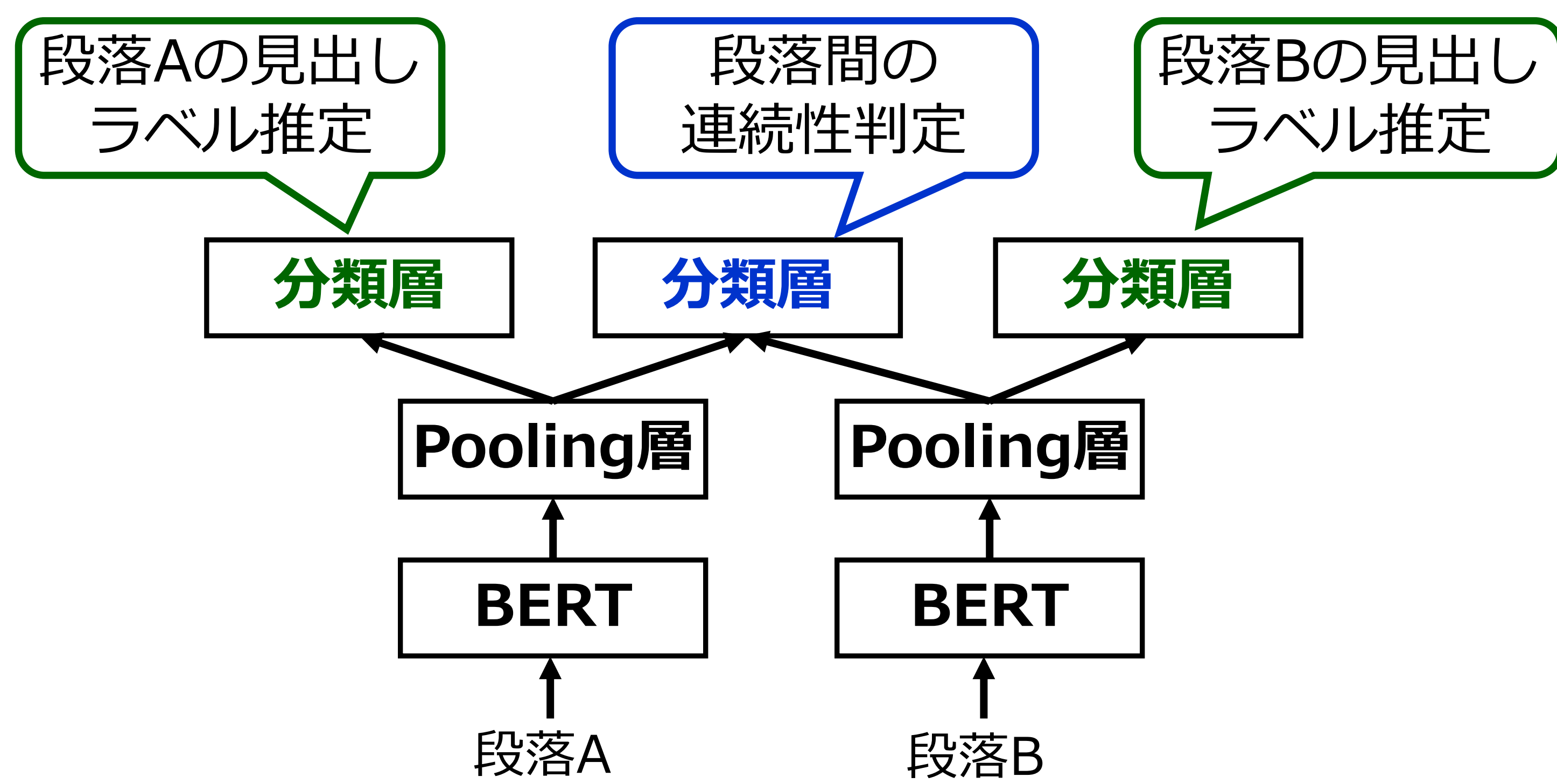
【発明の詳細な説明】

【技術分野】 【0001】 段落1
【背景技術】 【0002】 段落2
【発明の概要】 【0003】 段落3

段落単位でまとめ
られている

各段落は見出しで
整理されている

特許文書構造例



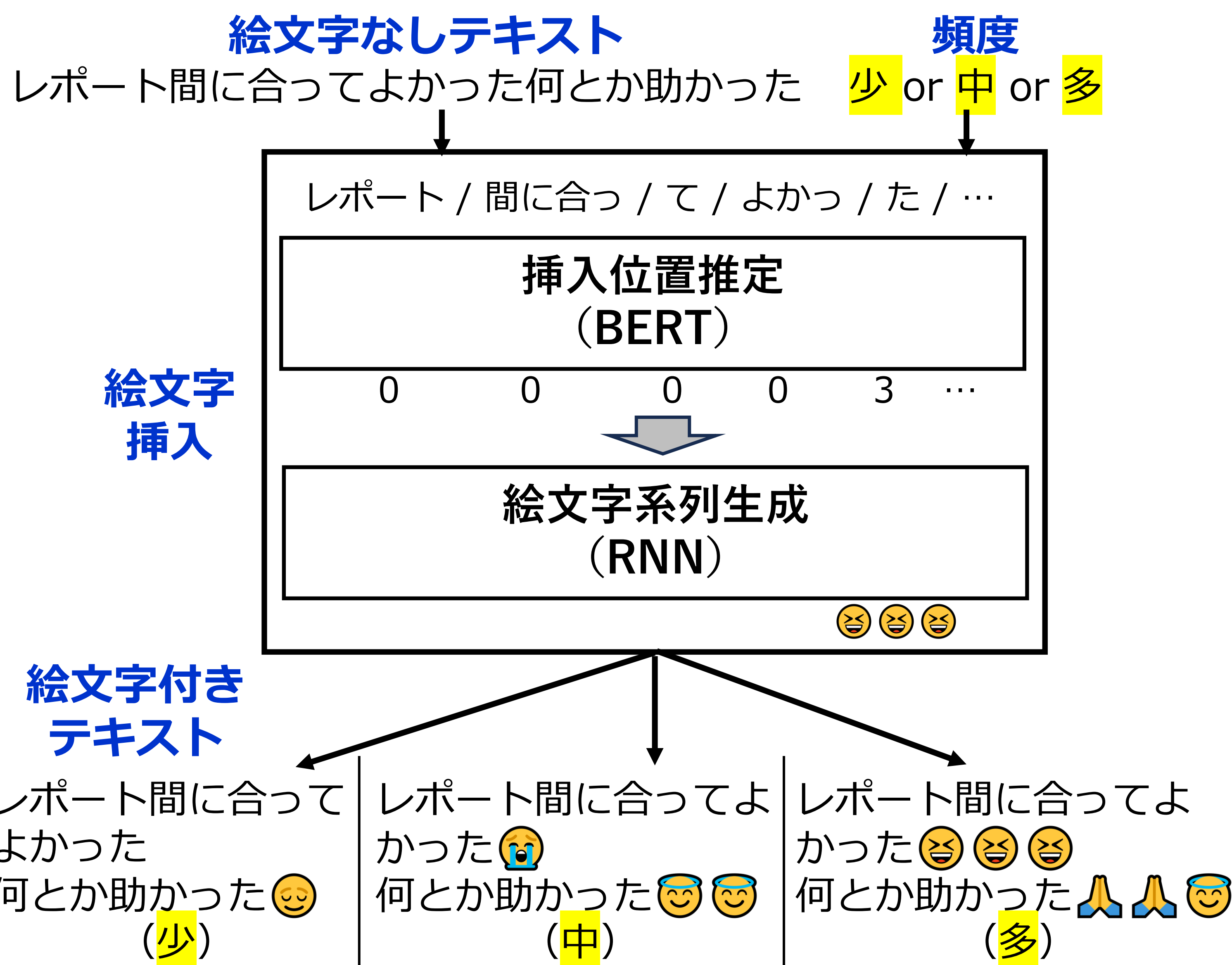
→ 特許文書のクラスタリング性能改善(F値(%) 47→54)

⑥ 絵文字自動挿入

- 絵文字を含まないテキストに絵文字を自動的に挿入

指定した割合で絵文字を自動挿入する手法を考案し, 人が行った絵文字挿入と同等以上であることを確認^[6]

- 従来: 1文に1つの絵文字のみ挿入
- 我々: 複数の箇所に複数の絵文字を挿入可能

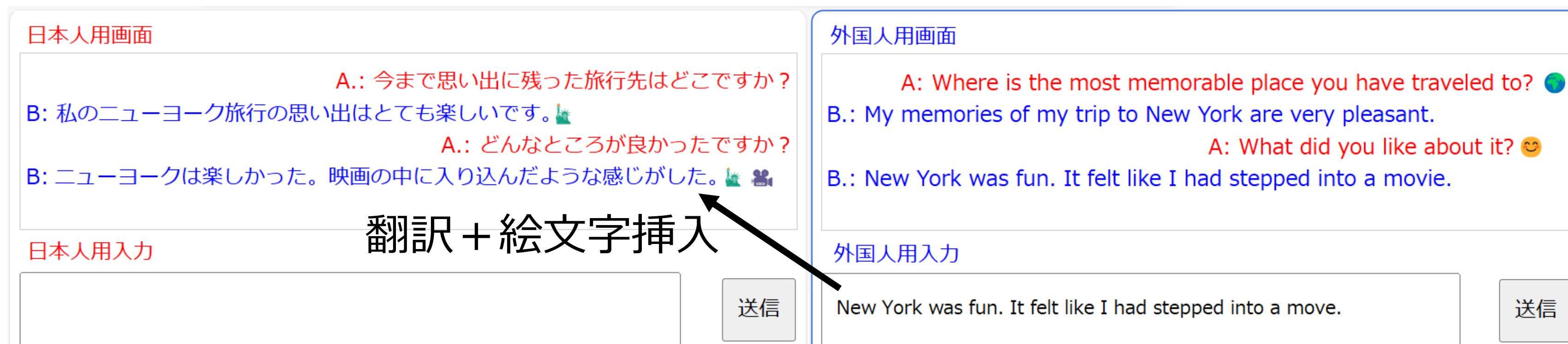


→ 感情のこもったコミュニケーションが可能

⑦ 機械翻訳+絵文字自動挿入システム

絵文字を自動挿入した翻訳文で外国人とコミュニケーションするシステムを開発

→ 外国人との豊かなコミュニケーションが可能



- 参考文献**
- [1] 谷ら, “日英多段階難易度制御機械翻訳: 評価データの作成および複数参照文に基づく学習の提案”, 自然言語処理, Vol.31, No.2, 2024.
 - [2] Yuasa et al., “Multimodal Neural Machine Translation Using Synthetic Images Transformed by Latent Diffusion Model”, ACL SRW, 2023.
 - [3] 森田ら, “自動作成した評価観点をを用いたLLMによるLLMの参照回答ベース評価”, 情報処理学会第260回自然言語処理研究会, 2024.
 - [4] Matsumotoら, “Combining Knowledge from Wikipedia and Sentiment Information for Stance Detection”, 情報処理学会第86回全国大会, 2024.
 - [5] 湯浅ら, “特許文書構造を利用したBERTによる事前学習”, 情報処理学会第84回全国大会, 2022.
 - [6] 富田ら, “絵文字の量を制御可能な絵文字自動挿入”, 言語処理学会第30回年次大会, 2024.