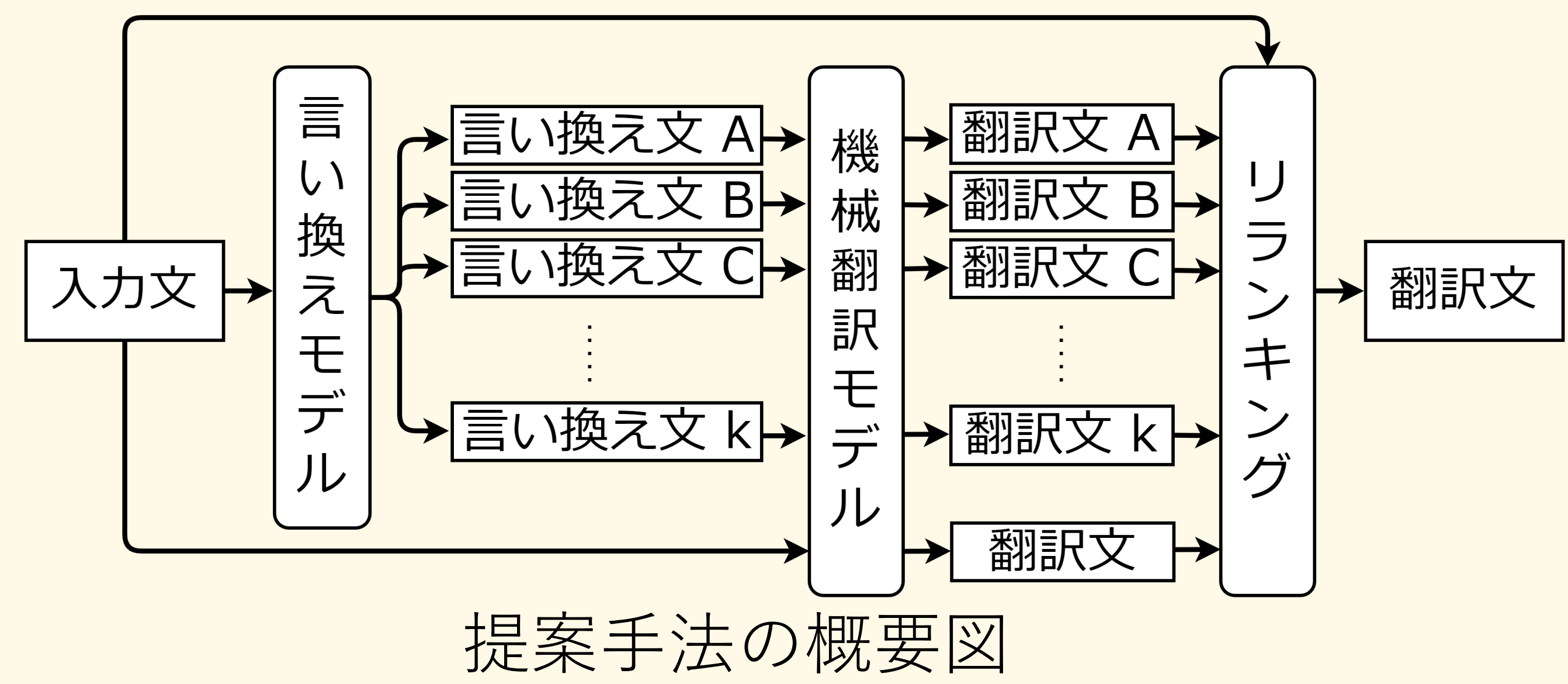


1. 背景

訓練済み翻訳器には得意な表現と不得意な表現がある
(要因の一つ：ドメインのギャップ)

入力文	Google翻訳
私は愛媛大学を受けます。	I will attend Ehime University.
私は愛媛大学を受験します。	I will take the Ehime University entrance exam.

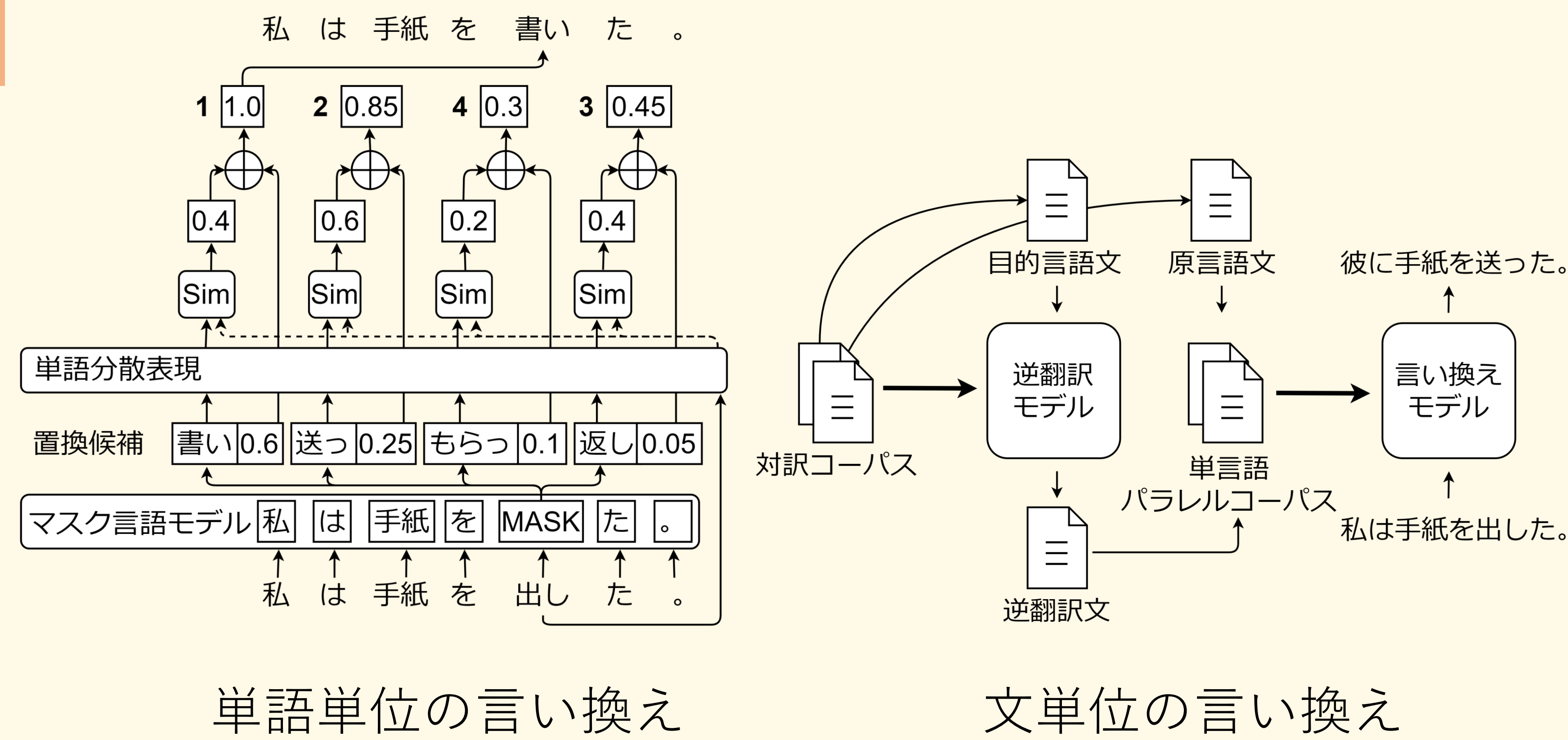


2. 提案手法：入力文のドメインと機械翻訳の訓練ドメインのギャップを埋める

- 言い換え生成：任意の機械翻訳モデルがより適切な翻訳候補を生成できるように、多様な言い換えを得る
- リランキング：入力文を考慮して翻訳文候補をリランキングし、翻訳品質の高い候補文を出力する

2.1 言い換え生成

- 単語単位と文単位の言い換えをk個生成
- 単語単位の言い換え生成
BERTなどのマスク言語モデルとfastTextなどの単語分散表現を用いて局所的な言い換え表現を得る
- 文単位の言い換え生成
Transformerなどの系列変換モデルを訓練して大域的な言い換え表現を得る



2.2 リランキング

- 教師なし・教師ありの2種類
- 教師なしリランキング
入力文と翻訳候補の間で双方向にforced decodingを行い、対数尤度の平均値を用いる
- 教師ありリランキング
複数の素性の線形和を用いる
(重みはK-best Batch MIRAで最適化)

素性	説明
順方向翻訳	順方向の機械翻訳モデルによる 入力文から翻訳候補へのforced decode確率
逆方向翻訳	逆方向の機械翻訳モデルによる 翻訳候補から入力文へのforced decode確率
言語モデル	対訳コーパスの目的言語側で訓練した言語モデルによる 翻訳候補の言語モデル確率
文長	入力文と翻訳候補の文長の差分およびその絶対値
言い換え (単語)	マスク言語モデルの単語穴埋め確率と 単語分散表現の余弦類似度の和
言い換え (文)	言い換えモデルによる入力文から 言い換え文へのforced decode確率

教師ありリランキングに用いた素性

3. 実験結果：日→英機械翻訳のBLEU

- データ
訓練用：JParaCrawl (ver 3.0) 1,000万文対
評価用：ASPEC (学術論文) 1,812文対
評価用：WMT20 (ニュース) 993文対
評価用：JParaCrawl 2,000文対

➤ 機械翻訳の自動評価結果

言い換え	リランキング	ASPEC			WMT20			JParaCrawl		
		k=5	k=10	k=20	k=5	k=10	k=20	k=5	k=10	k=20
-	-	19.9	20.0	20.3	19.7	19.7	19.8	35.3	35.2	35.3
-	教師なし	20.3*	20.4*	20.6*	19.8	20.2*	19.9	35.3	35.4	35.6*
単語単位	教師なし	20.6*	20.7*	21.1*	19.9	20.1*	20.2*	35.5*	35.5*	35.6*
文単位	教師なし	20.7*	20.8*	21.0*	19.9	20.0	20.1	35.4	35.3	35.4
-	教師あり	20.2*	20.2	20.5	19.6	19.6	19.9	35.9*	36.2*	36.6*
単語単位	教師あり	20.7*	20.8*	20.8*	20.2*	20.1*	20.4*	35.7*	35.6*	35.8*
文単位	教師あり	20.6*	20.4*	20.3	20.1*	20.0	20.1	35.7*	35.7*	35.7*

(kは言い換えの数, 太字は段ごとの最高値, *は最上段との統計的有意差 (p<0.05))

➤ 言い換え生成の人手評価結果

	同義性	流暢性
単語単位	74%	78%
文単位	86%	85%

- 各ドメインの評価用データから20文を無作為に抽出
- 各文についてk=5の単語単位および文単位の言い換え
- 各言い換え文を同義性と流暢性の2つの観点でそれぞれ良いか悪いかの2値で評価

➤ 分かったこと

- 未知ドメインにおいて効果的
- 既知ドメインにおいては教師ありリランキングが有用
- 高品質な言い換えが必ずしも翻訳品質には寄与しない

➤ 今後の予定

- ブラックボックスの機械翻訳にも適用できるようにリランキングの素性を再検討
- 言い換え生成器の工夫