

# 品質推定と大規模言語モデルによる 機械翻訳対象の自動編集

---

惟高日向<sup>1</sup>, 藤田篤<sup>2</sup>, 梶原智之<sup>1</sup>

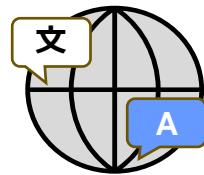
<sup>1</sup> 愛媛大学大学院理工学研究科, <sup>2</sup> 情報通信研究機構

# 背景

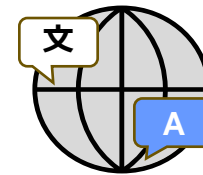
## 機械翻訳の実用化

- 機械翻訳の品質向上
- 商用・無料システムの普及（みんなの自動翻訳@KI, TexTra®など）

ブラックボックス化されているため、  
システムの中身を独自に調整することは難しい



Webページの翻訳  
(個人レベルの利用)



後編集

文書の翻訳  
(産業翻訳における利用<sup>[1]</sup>)

[1] Midori TATSUMI (2012) An Invitation to the Translation Studies in Japan

# 背景

機械翻訳システムを効果的に活用するための手法の一つとして**前編集**がある

→ 事前に翻訳しやすい表現に書き換えることで翻訳の誤りを低減する手法

(書換前) 私は愛媛大学を**受けます**。

(書換後) 私は愛媛大学を**受験します**。



翻訳器



**×** I **will attend** Ehime University.

**○** I **will take the entrance exam** for Ehime University.

表現によっては**適切に翻訳ができない**

書き換えによって**適切に翻訳ができる**

**後編集**や複数の翻訳文の**リランキング**と組み合わせることでさらに翻訳品質の向上が期待できる

# 先行研究：機械翻訳の前編集

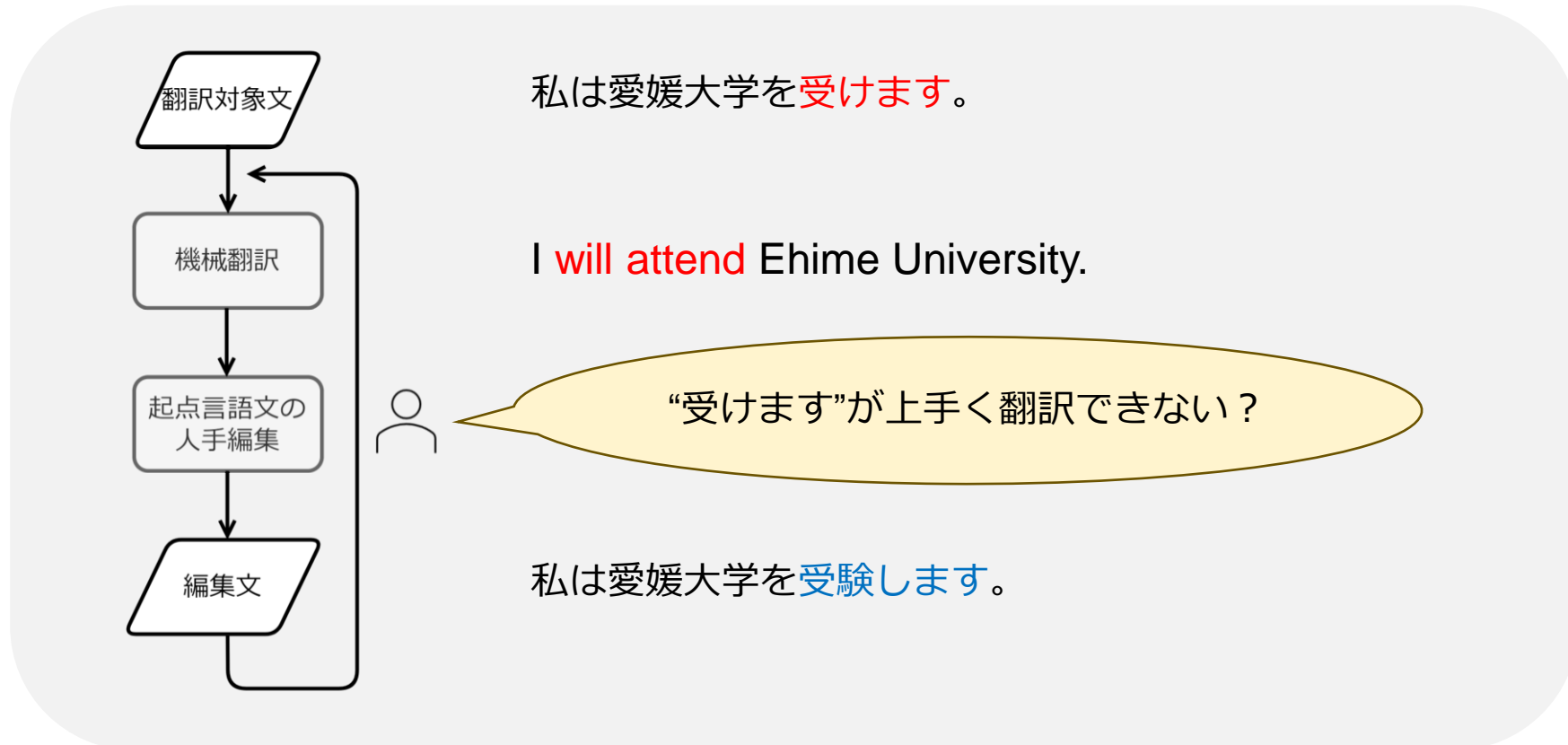
---

機械翻訳の前編集では以下の2つが研究されている

- 手動前編集
- 自動前編集

# 先行研究：手動前編集

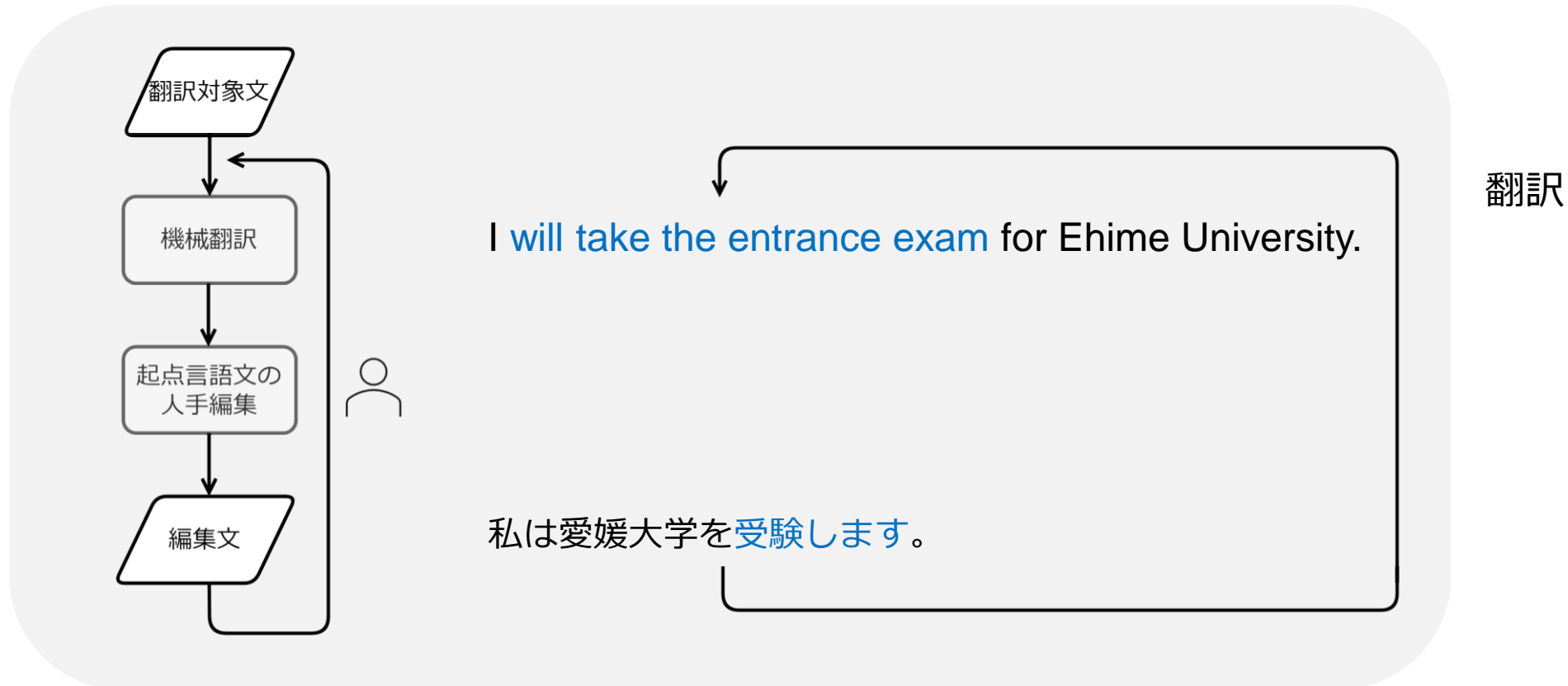
翻訳文を**参照しつつ**、**人手**で起点言語文を前編集<sup>[2,3]</sup>



# 先行研究：手動前編集

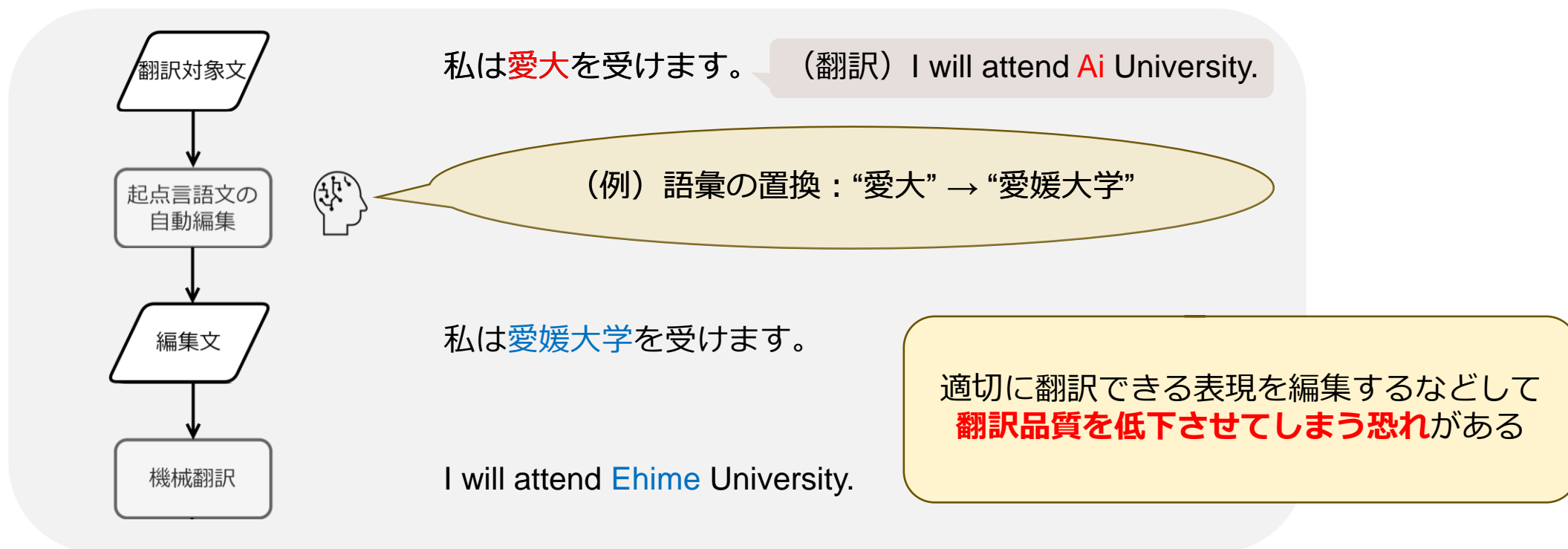
翻訳文を**参照しつつ**、**人手**で起点言語文を前編集<sup>[2,3]</sup>

言い換えを含む様々な編集操作により翻訳品質が向上



# 先行研究：自動前編集

翻訳文を**参照しないで**、**自動**で起点言語文を前編集<sup>[4-16]</sup>



[4] Shirai et al. (1993) Effects of Automatic Rewriting of Source Language within a Japanese to English MT System

[5] 金淵培, 江原暉将 (1994) 日英機械翻訳のための日本語長文自動短文分割と主語の補完

[6] 山口昌也ら (1998) 前編集結果を利用した前編集自動化規則の獲得

[7] Shirai et al. (1998) Automatic Rewriting Method for Internal Expressions in Japanese to English MT and Its Effects

[8] 吉見毅彦, 佐田いち子 (2000) 英字新聞記事見出し翻訳の自動前編集による改良

[9] 吉見毅彦ら (2000) 頑健な英日機械翻訳システム実現のための原文自動前編集

[10] Sun et al. (2010) A Novel Statistical Pre-Processing Model for Rule-Based Machine Translation System

[11] 南條浩輝ら (2012) 機械翻訳の品質向上のための対訳コーパスからの統計的前編集システムの自動構築

[12] Mirkin et al. (2013) SORT: An Interactive Source-Rewriting Tool for Improved Translation

[13] Stajner, S and Popovic, M (2016) Can Text Simplification Help Machine Translation?

[14] Stajner, S and Popovic, M (2018) Improving Machine Translation of English Relative Clauses with Automatic Text Simplification

[15] Mehta, S et al. (2020) Simplify-Then-Translate: Automatic Preprocessing for Black-Box Translation

[16] Koretaka et al. (2023) Mitigating Domain Mismatch in Machine Translation via Paraphrasing

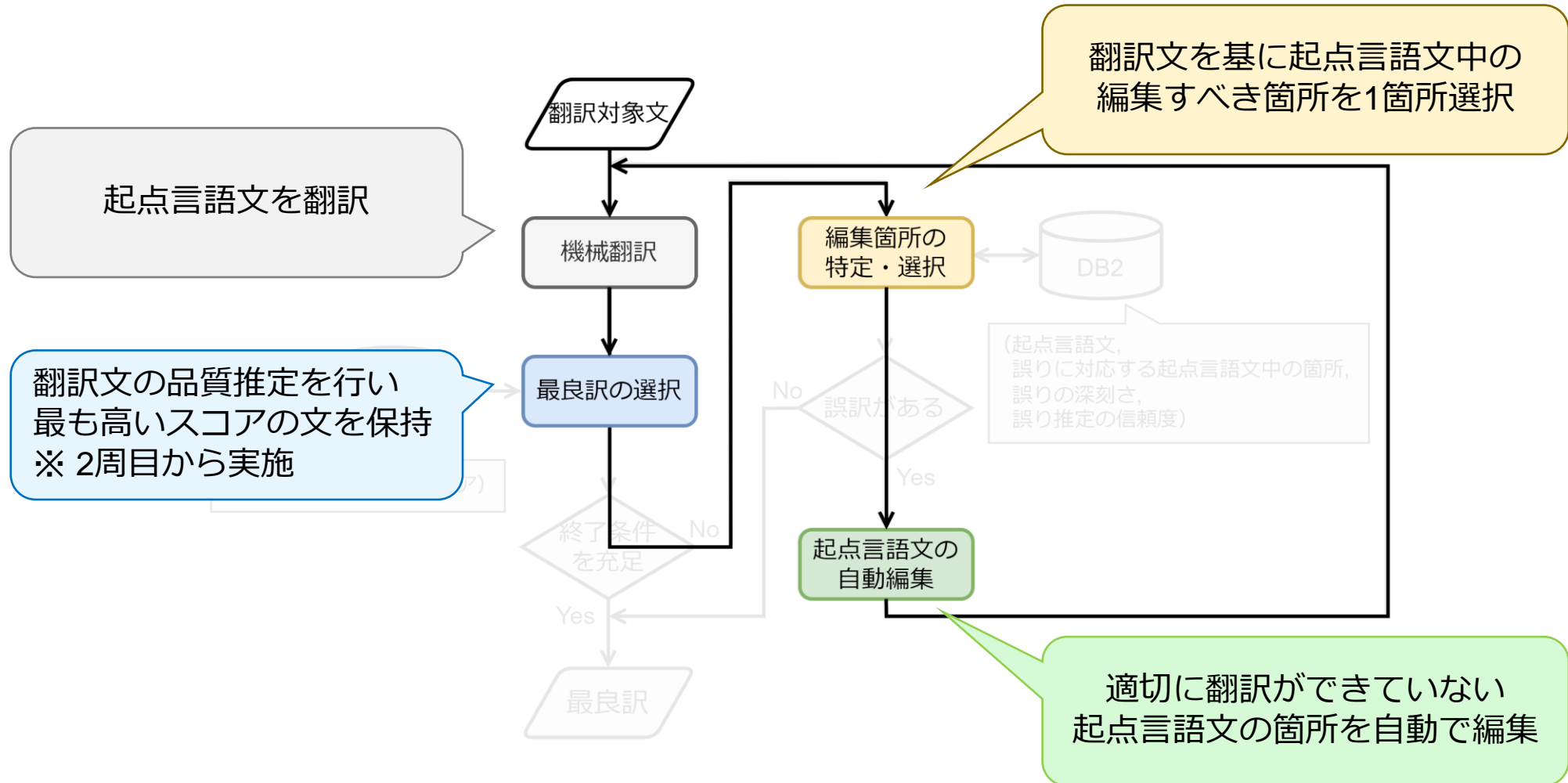
# 提案手法



# 概要

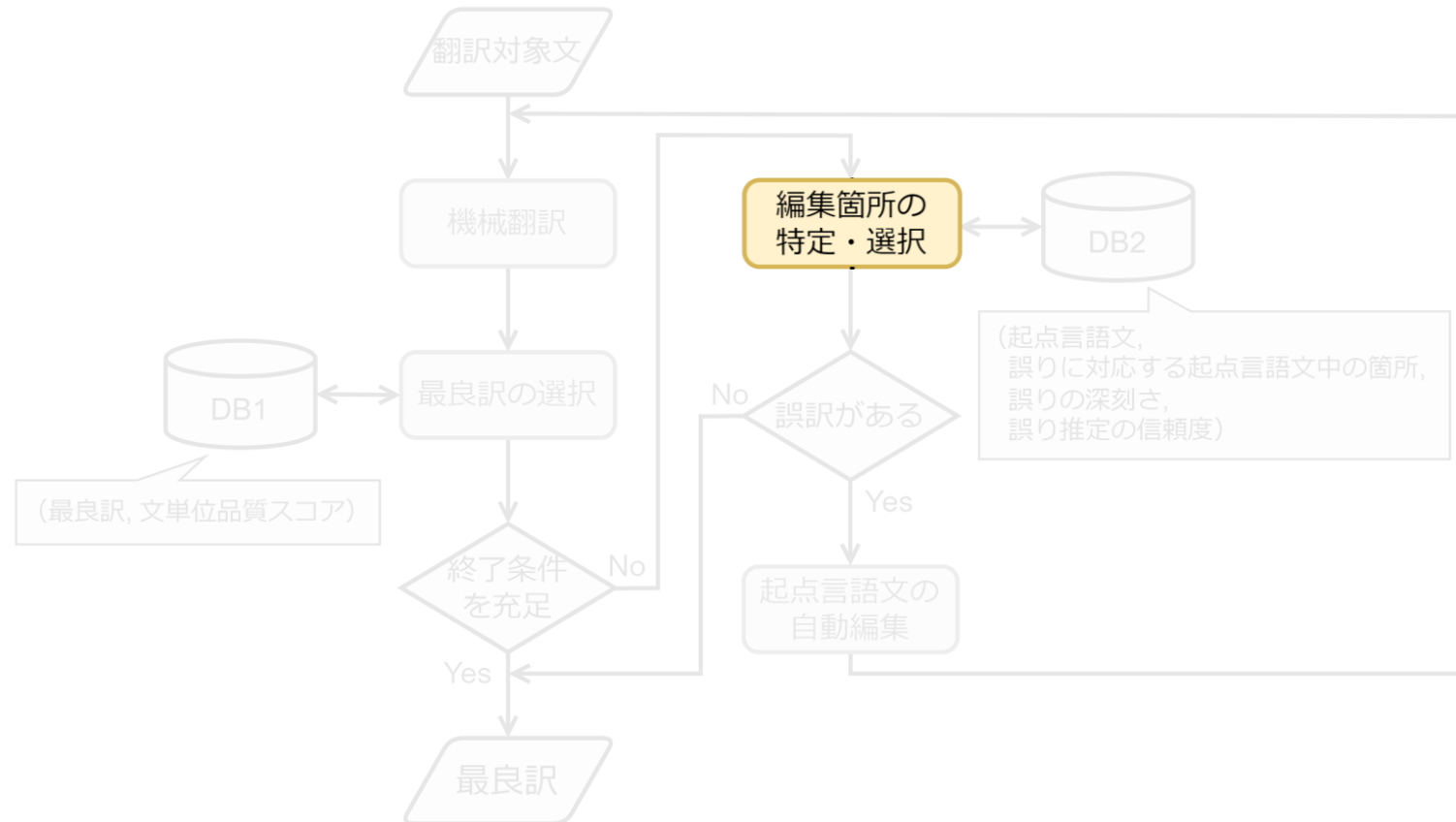
※ 複数の編集箇所が想定されるが、同時編集は困難と考え  
最も深刻な箇所から順に1つずつ繰り返し編集することで対処する

「機械翻訳」「編集箇所の特定・選択」「自動編集」「最良訳の選択」を繰り返し実行



# 編集箇所の特定・選択

編集箇所を特定し、最も深刻な誤りに対応するものを1箇所選択



# 編集箇所の特特定・選択

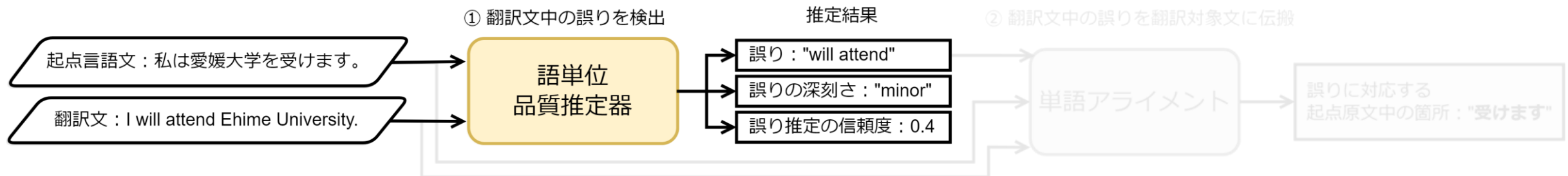
---

語単位の品質推定器を用いて翻訳文中の誤りを特定

翻訳文中の誤りを人手で注釈付けしたデータを用いて訓練されているため、  
これらの誤りの高精度な検出が期待できる

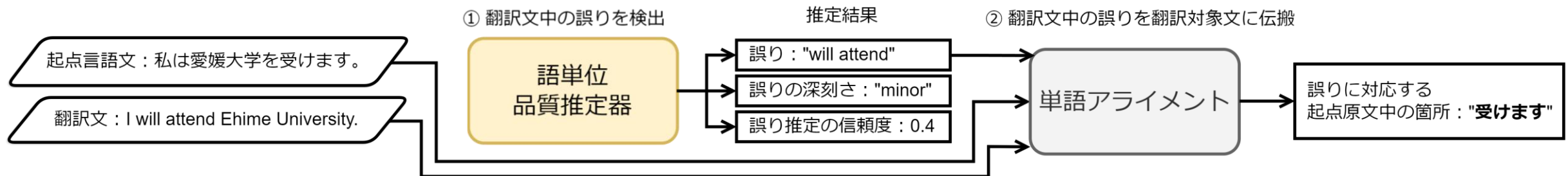
# 編集箇所の特定・選択

① 起点言語文と翻訳文を**語単位の品質推定器**に入力し、翻訳文中の誤りを検出



# 編集箇所の特定・選択

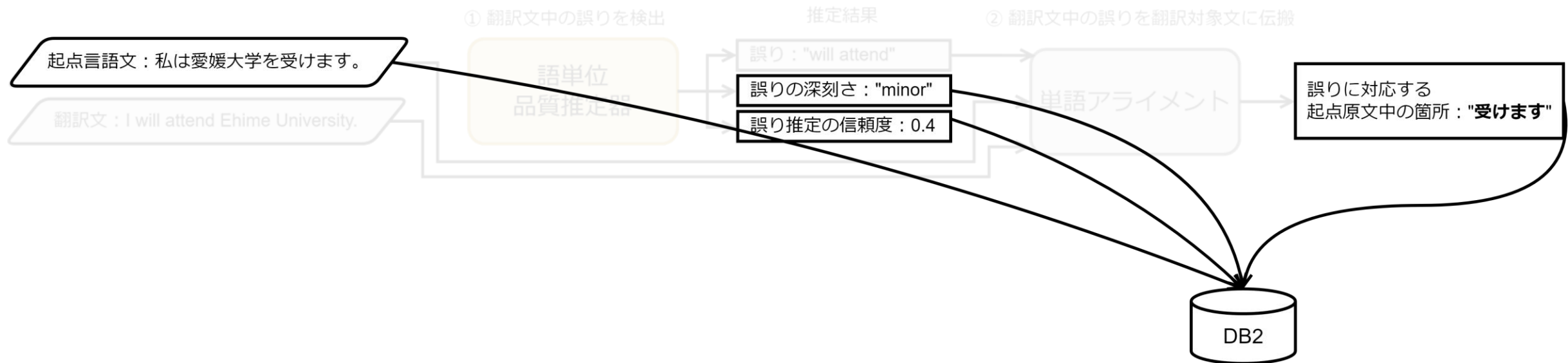
② 起点言語文、翻訳文、推定結果から**単語アライメント**を用いて翻訳文中の誤りを伝搬



得られた起点言語文中の箇所のうち最も深刻な誤りに対応するものを編集箇所として選択

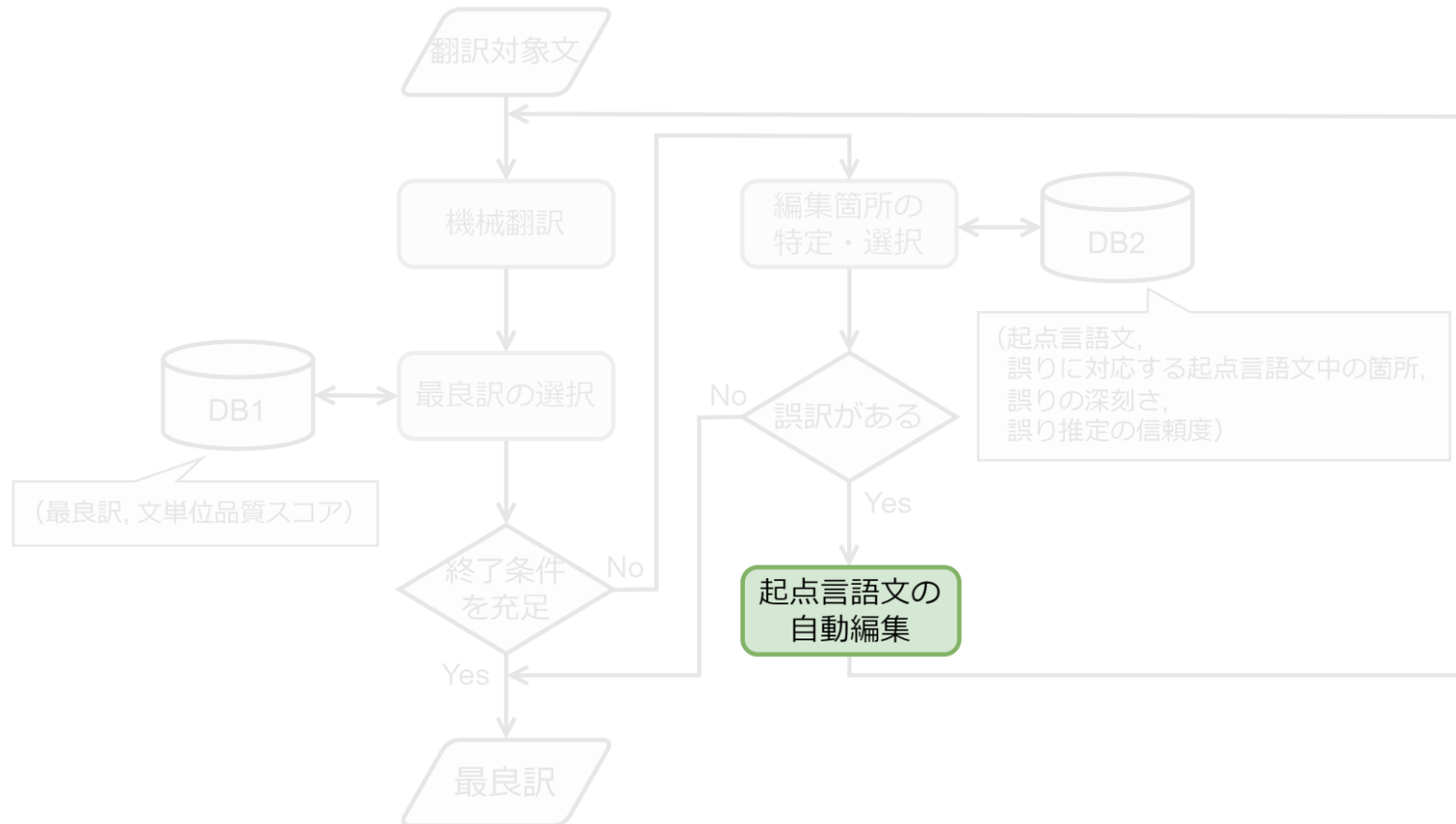
# 編集箇所の特特定・選択

より高品質な訳文の探索のためにデータベースに保持（後のスライドで説明）



# 起点言語文の自動編集

起点言語文中の指定された箇所を異なる表現に編集する



# 起点言語文の自動編集

---

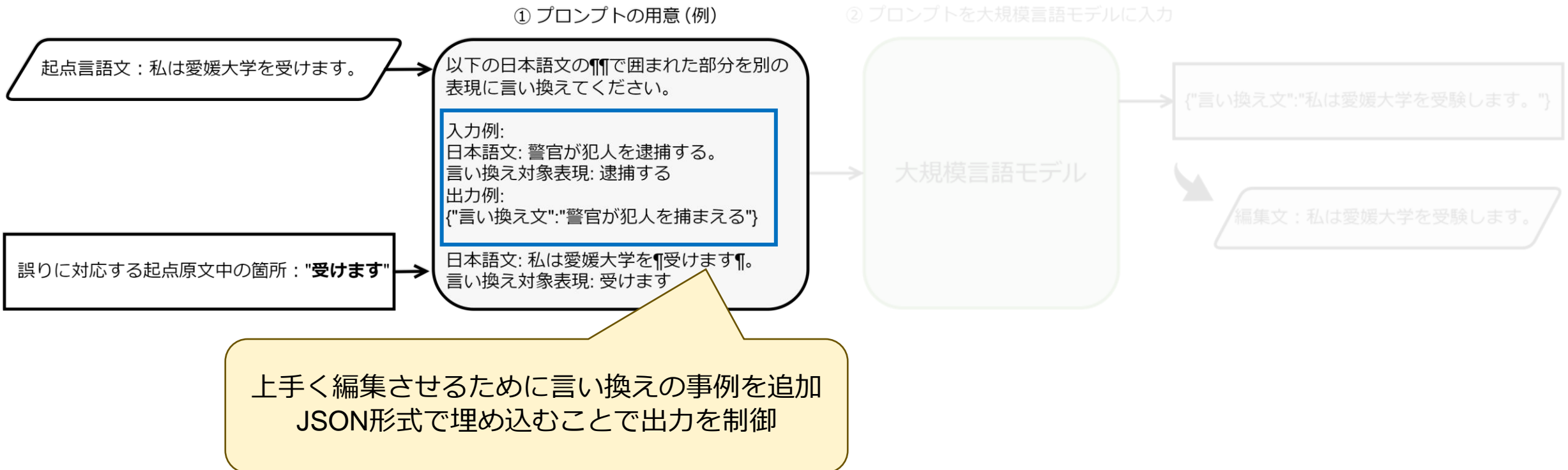
大規模言語モデルを用いて起点言語文を自動で編集

多様な表現を学習済みであり、多様な編集処理の実現が期待できる



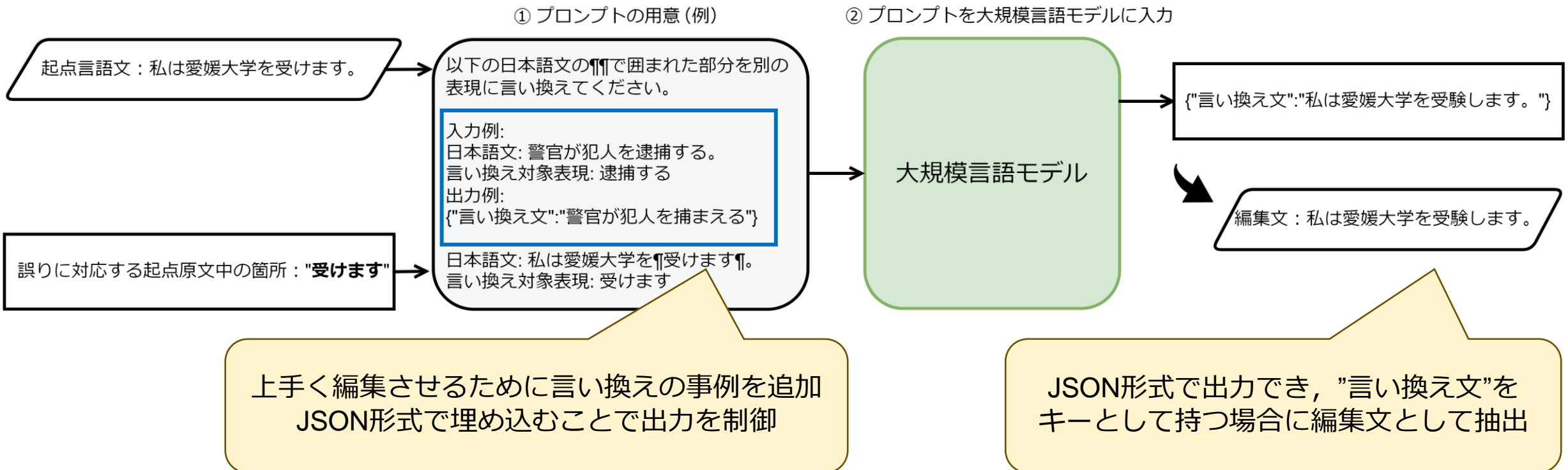
# 起点言語文の自動編集

## ① 起点言語文と誤りに対応する起点言語文の箇所を基にプロンプトを作成



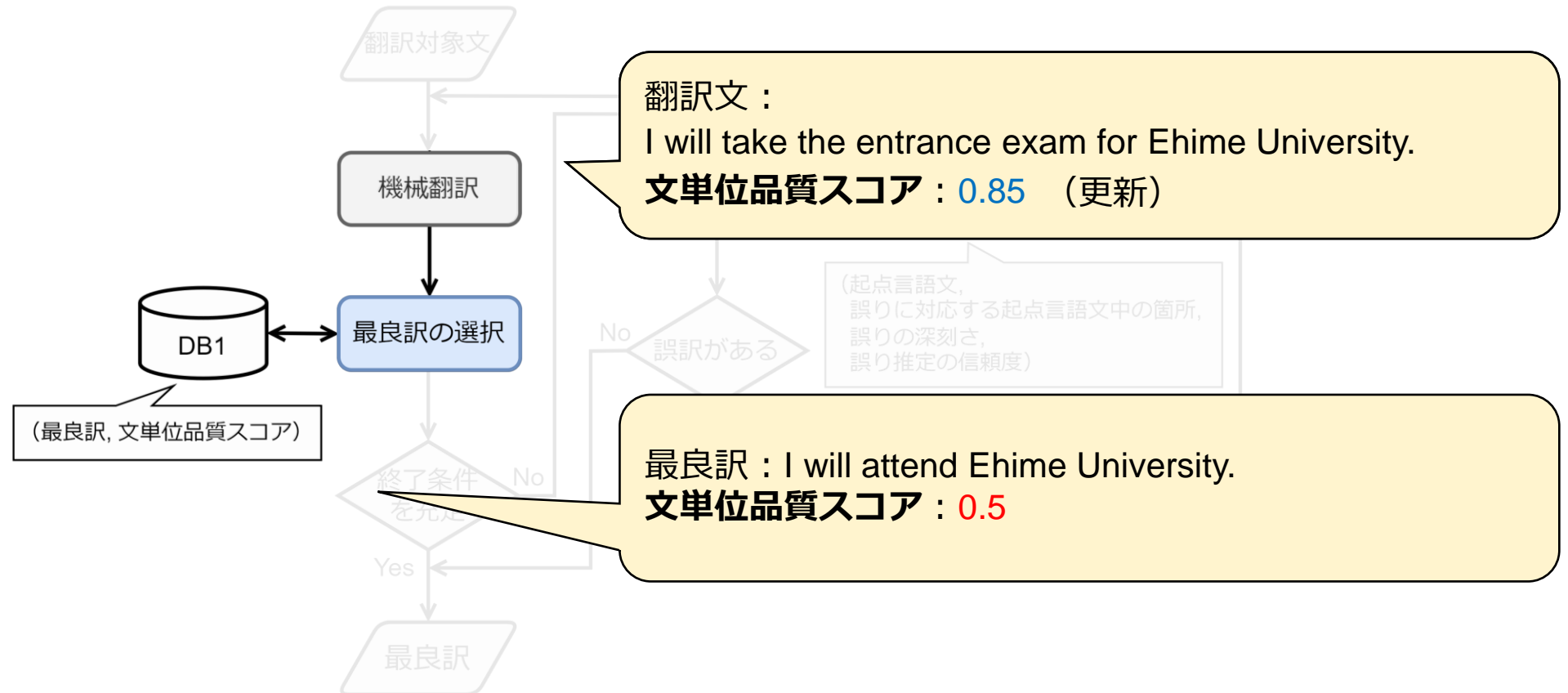
# 起点言語文の自動編集

## ① 起点言語文と誤りに対応する起点言語文の箇所を基にプロンプトを作成



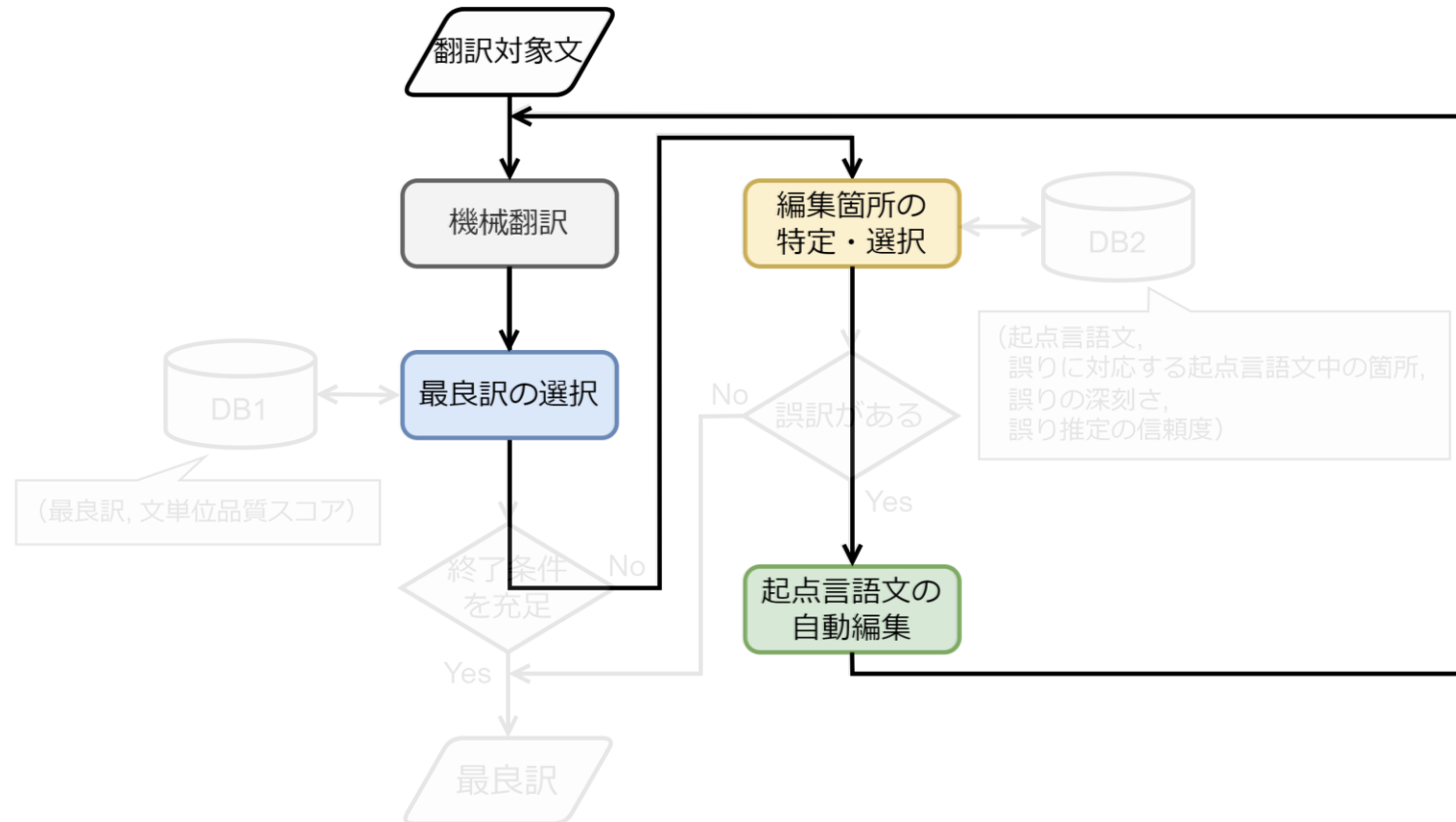
# 最良訳の選択

編集後の起点言語文を改めて機械翻訳し、文単位の品質スコアを基に最良訳を選択



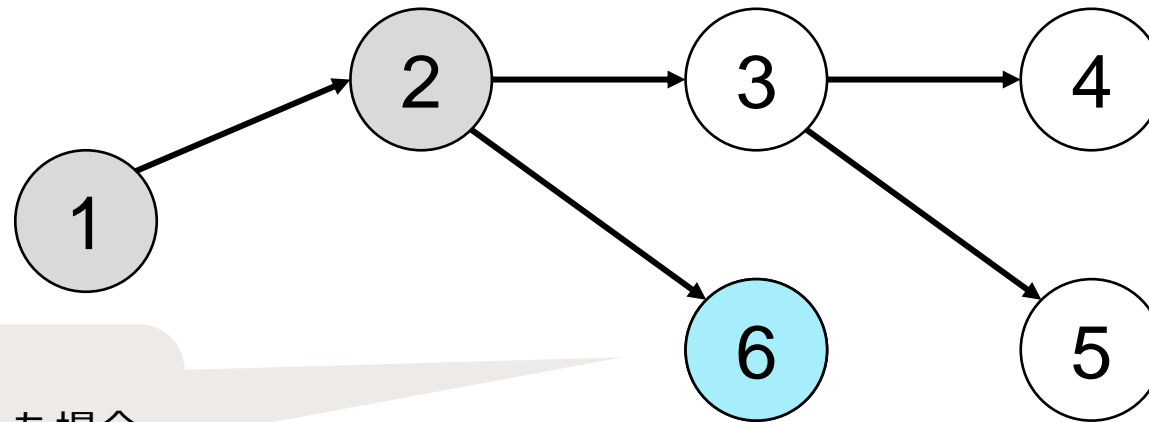
# より高品質な訳文の探索

編集を繰り返しても**翻訳品質が単調に改善するとは限らない** [2] → **深さ優先で最良訳を探索**



# より高品質な訳文の探索

- **翻訳品質が向上**：語単位の品質推定器の出力を手掛かりとして貪欲に編集を繰り返す
- **翻訳品質が低下**：次に深刻な誤りや編集過程で直前に得た起点言語文にバックトラックする



探索する対象がなくなった場合は現状の最良訳を出力

木構造で示した編集履歴のパスの例

※ 同じ箇所に対する異なる編集はしていない

# 評価実験

日英・日中・英日の機械翻訳タスクで評価

# 実験設定

## • 機械翻訳器 :

- NMT : TexTra  
NLLB-200  
JParaCrawl (small, base, big)
- LLM : Llama (BM25, vector)  
Llama-Swallow (BM25, vector)

## • 評価用データ :

- 日英 : ASPEC, WMT23, MTNT19, KFTT
- 日中 : ASPEC
- 英日 : ALT / WMT23 / MTNT19

## • 評価指標 : COMET

## • 提案手法の設定 :

ループ上限 : 5回, 品質推定器 : XCOMET, 編集器 : Llama, Llama-Swallow

## • 比較手法 :

- 比較手法1 : 惟高ら<sup>[16]</sup>の語単位の言い換えとリランキングに基づく手法
- 比較手法2 : 比較手法1のリランキング部分を文単位の品質推定器に置換

翻訳文を参照しない既存の自動前編集手法

編集部分のみを比較するため

# 機械翻訳器におけるCOMETのマクロ平均

## 評価用データ全体における結果

	TexTra	NLLB-200	JParaCrawl (small)	JParaCrawl (base)	JParaCrawl (big)	Llama (BM25)	Llama (vector)	Llama-Swallow (BM25)	Llama-Swallow (vector)
編集なし	83.47	77.43	77.97	78.63	80.03	82.31	82.38	82.61	82.80
比較手法1	83.55 (2)	77.43 (4)	77.90 (4)	78.47 (4)	79.66 (1)	82.55 (3)	82.56 (4)	83.14 (5)	83.14 (3)
比較手法2	83.66 (4)	78.88 (8)	78.30 (4)	78.83 (5)	80.04 (3)	83.04 (7)	83.17 (7)	83.40 (7)	83.58 (7)
提案手法	83.82 (5)	78.47 (7)	79.34 (7)	80.07 (7)	81.06 (6)	82.71 (6)	82.68 (5)	82.81 (5)	83.04 (5)

27/69で最も高いCOMETスコアを示した

## 誤りを含む部分集合に対する結果 (61.6% ~ 96.7%)

46/69で最も高いCOMETスコアを示した

	TexTra	NLLB-200	JParaCrawl (small)	JParaCrawl (base)	JParaCrawl (big)	Llama (BM25)	Llama (vector)	Llama-Swallow (BM25)	Llama-Swallow (vector)
編集なし	82.40	77.48	76.90	77.46	78.88	81.99	82.01	82.55	82.54
比較手法1	82.55 (2)	77.17 (1)	76.86 (4)	77.32 (4)	78.52 (1)	81.88 (1)	81.80 (1)	82.45 (2)	82.32 (0)
比較手法2	82.61 (3)	78.38 (8)	77.24 (4)	77.70 (5)	78.85 (3)	82.34 (4)	82.40 (5)	82.66 (3)	82.78 (3)
提案手法	82.90 (5)	78.87 (7)	78.52 (7)	79.18 (7)	80.11 (6)	82.52 (6)	82.41 (5)	82.84 (4)	82.91 (5)

※ ( ) 中の数値は、編集なしと比較して統計的に有意な結果を示した評価用データの数



# 提案手法による品質の上り幅

機械翻訳器・評価用データごとの提案手法の効用の傾向を確認  
(誤り部分集合に対しての結果)

全体的な傾向として**負の相関**があった

→ 編集を行わない場合の**翻訳品質が低いほど**

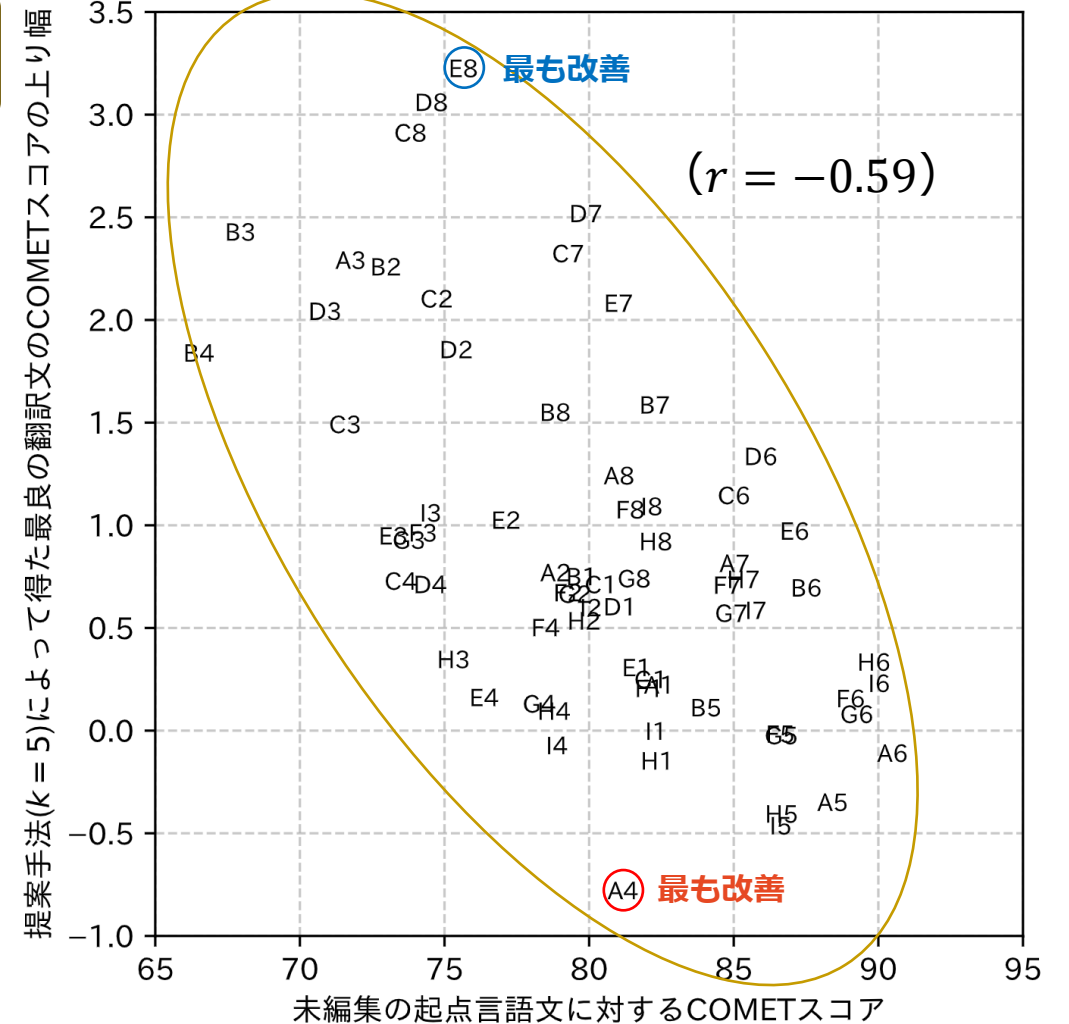
提案手法による**品質の改善が大きい**

モデルの識別記号

評価用データの識別番号

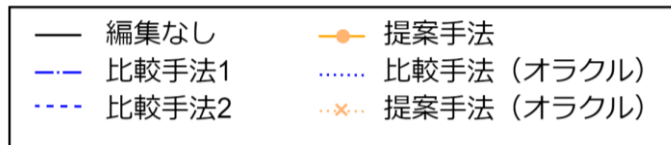
A : TexTra  
B : NLLB-200  
C : JParaCrawl (small)  
D : JParaCrawl (base)  
E : JParaCrawl (big)  
F : Llama (BM25)  
G : Llama (vector)  
H : Llama-Swallow (BM25)  
I : Llama-Swallow (vector)

1 : ASPEC (日英)  
2 : WMT23 (日英)  
3 : MTNT19 (日英)  
4 : KFTT (日英)  
5 : ASPEC (日中)  
6 : ALT (英日)  
7 : WMT23 (英日)  
8 : MTNT19 (英日)



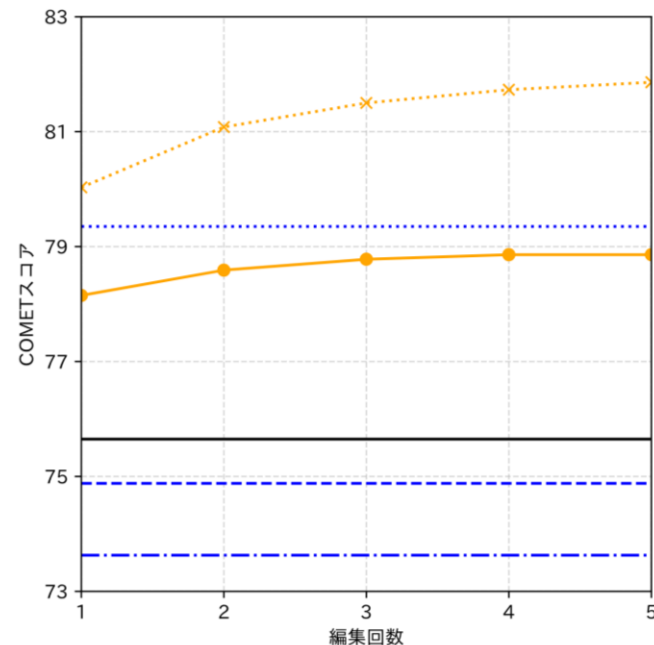
# 編集回数に応じた品質の変化

- 編集回数を重ねるにつれてMTNT19は翻訳品質が向上、KFTTは翻訳品質が劣化
- オラクル訳は両設定でCOMETスコアが向上したが、それらの選択はできていない  
→ **文単位の品質推定器**には改善の余地がある

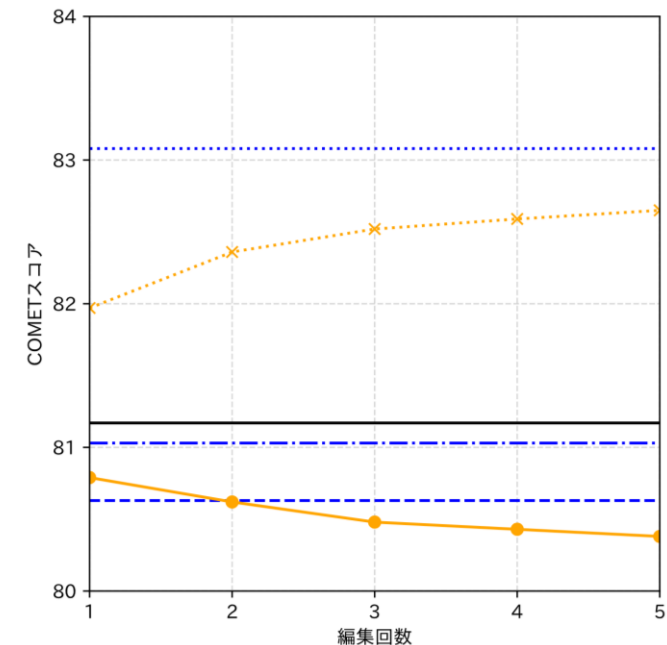


その時点までに取りうる最良の翻訳文 (オラクル訳) のCOMETスコア

※ 予稿における図の凡例が誤っていたため修正  
「比較手法1」と「比較手法2」を入れ替え



JParaCrawl (big) を用いた場合の英日MTNT19 (最も改善した設定)



TexTraを用いた場合の日英KFTT (最も劣化した設定)

# 分析

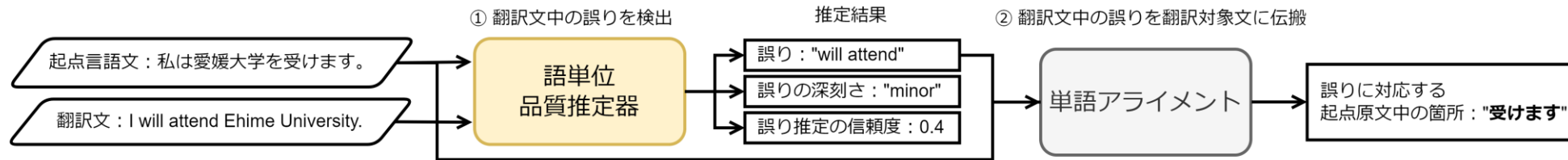
編集箇所の特的手法と編集器の是非を分析

データ：事前実験で用いた日英方向の誤りを含む100文  
翻訳器：JParaCrawl (big)

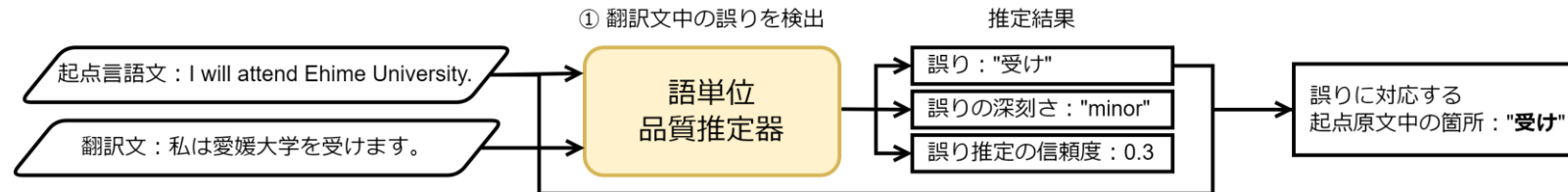
# 編集箇所の特定手法に関する分析

以下の3つの手法を比較

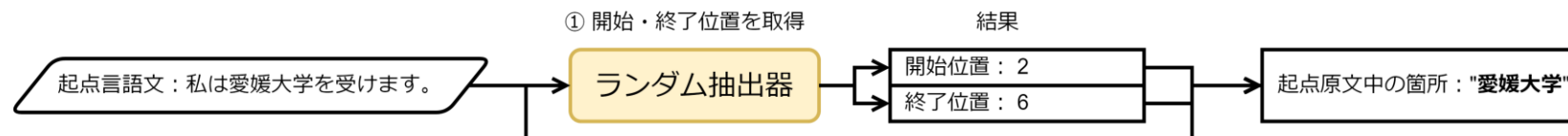
- **順入力**：「語単位の品質推定」と「単語アライメント」の2段階で特定する手法



- **逆入力**：語単位の品質推定器であるXCOMETに入力する際に逆入力することで直接特定する手法



- **ランダム**：編集すべき箇所の開始位置と終了位置をランダムに決定する手法



# 編集箇所の特定手法に関する分析

- ランダムでもCOMETスコアは向上するが、品質推定器を用いた2つの手法の方が高い  
→ 品質推定器を活用することは有用
- 順入力とは他の手法よりもオラクルのCOMETスコアが高く、編集回数によらずスコアが高い

	順入力	逆入力	ランダム
編集なし	71.82	71.82	71.82
$k = 1$	73.28*	72.33	72.75*
$k = 2$	73.65*	72.60	72.89*
$k = 3$	73.48*	72.69*	72.72
$k = 4$	73.39*	72.73*	72.72
$k = 5$	73.31*	73.04*	72.72
オラクル ( $k = 5$ )	75.42	74.43	73.76

\* : 編集なしと比較して統計的に有意 (以下同様)

# 自動編集に用いるモデルに関する分析

- 編集対象である起点言語に適応済みであり大規模なモデルが翻訳品質を大きく改善
- 現状のLLMおよびその使い方では、人手レベルでの編集が実現できていなかった

第一著者が編集	人手	Llama-Swallow 70B	Llama-Swallow 8B	Llama 70B	Llama 8B
編集なし	71.82	71.82	71.82	71.82	71.82
$k = 1$	73.58*	73.28*	72.39	72.90*	71.82
$k = 2$	74.20*	73.65*	72.45	73.04*	71.82
$k = 3$	74.50*	73.48*	72.67	72.95*	71.82
$k = 4$	74.40*	73.39*	72.64	73.05*	71.82
$k = 5$	74.22*	73.31*	72.66	72.99*	71.82
オラクル ( $k = 5$ )	77.61	75.42	74.78	75.50	71.82

翻訳品質：人手による編集 > LS-70B > L-70B > LS-8B > L-8B

# まとめ

- **背景**：機械翻訳の品質を向上させる手法として前編集（の自動化）がある
- **課題**：従来手法は適切に翻訳できる表現を編集するなどして品質を低下させる恐れがある
- **手法**：翻訳文を参照し、実際に生じている誤りを回避するように起点言語文を編集
- **結果**：既存手法よりも翻訳品質を改善

