事前学習モデルを用いた近代文語文の現代語機械翻訳

喜友名 朝視顕 [1]　平澤 寅庄 [1]　小町 守 [1]　小木曽 智信 [2]
[1] 東京都立大学　[2] 国立国語研究所
kiyuna-tomoshige@ed.tmu.ac.jp,hirasawa-tosho@ed.tmu.ac.jp,
komachi@tmu.ac.jp,togiso@ninjal.ac.jp

1　はじめに

本研究では、近代文語文の現代語機械翻訳タスクを提案する。これは、近代（明治期から大正期）の文語から現代日本語への翻訳タスクである。本タスクは、同じ言語間の通時的な違いに着目して言語表現を言い換えるタスクの 1 つである。本タスクの特徴として、原言語と目的言語は同じ日本語であり、共通の語彙が多いことが挙げられる。英語では、William Shakespeare の初期近代英語文を現代英語文に変換する試み [1] が行われている。

本タスクでは、タスクの特徴を考慮し、新たに提案する評価尺度を使用する。ニューラル機械翻訳（neural machine translation: NMT）の出力は流暢であるが、そのまま公開することは難しい。そこで本研究では、ポストエディットのコストを考慮した翻訳性能の評価尺度として、原文との BLEU および参照訳との BLEU の調和平均を評価値とする調和平均 BLEU を提案し、既存の評価尺度よりも人手評価値との相関が高いことを示す。

また、本タスクの課題として、学習に使用できる対訳コーパスが低リソースであることが挙げられる。この問題に対処する方法の 1 つに、事前学習モデルの利用がある。Lample ら [2] は、Transformer [3] ベースの言語横断的な言語モデル cross-lingual language model (XLM) を学習する方法として、各単言語コーパスを用いた教師なし手法（masked language modeling: MLM）と対訳コーパスを用いた教師あり手法（translation language modeling: TLM）を提案した。そして、事前学習済みの XLM を使用して、教師あり NMT と教師なし NMT で実験を行い、両者共に翻訳性能が向上することを示した。

本研究でも、タスクの特徴を考慮し、MLM で学習した XLM を用いた NMT を行った。単言語コーパスを使用した教師なし NMT 手法の場合、訳抜けや誤訳はほとんど生じなかった。しかし、多くの単語は原文のままであるような出力に止まった。

本研究における貢献は、次の 3 点である。

(1) 近代文語文の現代語機械翻訳タスクを提案する。
(2) 原言語と目的言語に共通の語彙が多い場合の、ポストエディットのコストを考慮した翻訳性能の評価尺度として、原文も考慮する調和平均 BLEU を提案する。
(3) 古文の翻訳で初めて事前学習モデルを用いた NMT を行う。

2　近代文語文の現代語機械翻訳

近代文語文から現代語への機械翻訳タスクを提案する。本タスクの特徴として、原言語と目的言語は同じ日本語であり、その差は通時的に生じたもので、共通の語彙が多いことが挙げられる。

本タスクのデータセットには、『学問のすゝめ』とその現代語訳[1)]から抽出した段落単位の対訳コーパスを使用する。

2.1　学習

訓練・開発データには『学問のすゝめ』の二編以降を用いる。段落単位の対訳コーパスから、1,513 文対から成る文単位の対訳コーパスを作成する。作成方法は付録 A を参照。

訓練・開発データを 5.1 の方法で単語分割を行ったときの各言語別の単語トークン数と単語タイプ数を表 1 に示す。近代文語の単語タイプのうち 44.28% が現代語の単語タイプと等しく、現代語の単語タイプのうち 49.36% が近代文語の単語タイプと等しい。

2.2　テスト

評価データには『学問のすすめ』（福澤諭吉、1872）の初編を用いる。16 段落から成る段落単位の対訳

1）福澤諭吉著・齋藤孝訳『現代語訳 学問のすすめ』ちくま文庫（2009）

表 1　訓練・開発データの単語トークン数とタイプ数。

言語	単語トークン数	単語タイプ数
近代文語	55,684	5,354
現代語	60,925	4,803
近代文語 + 現代語	—	7,786

コーパスであり、近代文語文は 75 文、その現代語訳が 121 文である。文ごとにモデルに入力し、75 文の翻訳結果を得た後、それらを連結して 16 個の段落に整形し、段落単位のテストを行う。

3　調和平均 BLEU（hm-BLEU）

原言語と目的言語に共通の語彙が多いときを考える。仮に適切な翻訳ができなかった場合、文意と無関係の単語を出力するより、原文の単語をそのまま出力した方が、翻訳として妥当であろう。その場合、原文も考慮して評価するのが自然である。しかし、機械翻訳の評価尺度として一般に用いられている BLEU [4] は、翻訳結果を参照訳との n-gram の一致率によって翻訳性能を評価し、原文は考慮しない。そこで、原文と参照訳の両方を考慮する BLEU として、原文との BLEU と、参照訳との BLEU に対し、単純にそれらの調和平均を評価値とする、調和平均 BLEU（hm-BLEU）を提案する。

hm-BLEU は、原文との n-gram の一致率を考慮する評価尺度である。原文との n-gram の一致率が高いほど、原文の情報を多く保存していると言える。原文の情報を保存していれば、訳抜けや文意に無関係な単語の出現に気付きやすくなるだろう。

4　事前学習モデルを用いた NMT

言語モデルの事前学習と、事前学習済み言語モデルで初期化した翻訳モデルの fine-tuning の 2 つに大別される。

言語モデルの事前学習　原言語と目的言語に共通の語彙が多いため、単言語の言語モデルよりも言語横断的な言語モデルを用いることで、高い翻訳性能が期待できる。そこで、言語モデルには XLM [2] を用いる。言語モデルの事前学習には、BERT (bidirectional encoder representations from Transformers) [5] と似た MLM を用いる。XLM の MLM は BERT の MLM と異なり、文対ではなく任意の数の文（最大トークン数で切り捨てる）で学習を行う。近代文語と現代語の各単言語コーパスを用いて、MLM により共通の言語モデルを作成する。

翻訳モデルの fine-tuning　事前学習済みの言語モデルで翻訳モデルの encoder 全体と decoder の一部を初期化し、教師なし手法または教師あり手法により学習を行う。各手法は、Lample and Conneau (2019) [2] の翻訳手法に従う。教師なし手法は、Lample et al. (2018) [6] の denoising と逆翻訳を用いる手法を拡張した手法で、各単言語コーパスを用いて学習を行う。教師あり手法は、Ramachandran et al. (2016) [7] の手法を multilingual NMT [8] に拡張した手法で、対訳コーパスを用いて系列変換の目的関数に従い学習を行う。どちらも、単一のモデルで、近代文語から現代語の翻訳だけでなく、現代語から近代文語の翻訳も行うため、近代文語の単語を多く残すような現代語への翻訳が期待できる。

5　実験

5.1　データセット

対訳データセットに加えて、言語モデルの事前学習に必要な単言語データセットを用意した。それぞれ、MeCab [9] を用いて形態素解析を行った。形態素解析の辞書には、近代文語は近代文語 UniDic [10][2]、現代語は現代書き言葉 UniDic [11][3] を用いた。また、1 文に含まれる短単位の数を 100 語に制限した。

対訳データセット　2.1 の訓練・開発データのうち、9 割を訓練データ（1,362 文対）、1 割を開発データ（151 文対）として使用した。翻訳モデルの fine-tuning にのみ用いる。

単言語データセット　訓練データは近代文語と現代語の各単言語コーパス（詳細は付録 B を参照）を用いる（それぞれ 226,793 文）。開発データには『文明論之概略』（福澤諭吉、1875）とその現代語訳[4] から付録 A の方法で作成した文単位の対訳コーパスを使用した（3,306 文対）。言語モデルの学習と翻訳モデルの fine-tuning の両方に用いる。

5.2　実験方法

実験には Lample ら [2] の実装[5] を利用し、6 つの手法で翻訳モデルの学習を行った。特に言及がない

2）https://unidic.ninjal.ac.jp/download_all#unidic_kindai（UniDic-kindai_1603）
3）https://unidic.ninjal.ac.jp/download#unidic_bccwj（unidic-cwj-2.3.0）
4）福澤諭吉著・齋藤孝訳『現代語訳 文明論之概略』ちくま文庫（2013）
5）https://github.com/facebookresearch/XLM

表 2　各評価尺度および人手評価による各翻訳手法の評価値。各評価尺度の値は、評価値を 100 倍したものである。

手法	LM	1 回目の fine-tuning		2 回目の fine-tuning		BLEU				GLEU	SARI	人手評価
		手法	データ	手法	データ	s	r	sr	hm			平均
原文						100.00	13.47	100.00	23.74	0.39	10.99	5.0
参照訳						13.68	100.00	100.00	24.07	100.00	89.16	–
1	XLM	教師あり	対訳	—	—	3.28	7.15	9.80	4.50	6.59	34.23	3.5
2	XLM	教師なし	対訳	—	—	2.97	6.25	8.21	4.03	5.34	32.30	1.5
3	XLM	教師なし	単言語	—	—	57.50	16.42	62.41	25.55	0.33	41.81	6.5
4	XLM	教師なし	単言語	教師あり	対訳	19.43	17.17	29.35	18.23	9.05	43.25	5.0
5	XLM	教師なし	単言語	教師なし	対訳	3.16	5.87	8.41	4.11	5.16	33.09	1.0
6	BERT	教師なし	単言語	—	—	56.69	15.85	60.98	24.77	0.64	42.64	7.0

限り、ハイパーパラメータは既定値を使用した。

5.2.1　言語モデルの事前学習と翻訳モデルの初期化

2 つの方法で翻訳モデルの初期化を行った。

言語横断的な初期化　単言語データセットを用いて学習した XLM で、翻訳モデルの encoder 全体と decoder の一部を初期化する（XLM）。

言語横断的でない初期化　近代文語と現代語の各単言語コーパスを用いて近代文語と現代語の言語モデルをそれぞれ学習し、翻訳モデルの encoder を近代文語の言語モデルで、decoder を現代語の言語モデルで初期化する（BERT [6]）。開発データは単言語データセットと同様である。

5.2.2　翻訳モデルの fine-tuning

言語横断的な初期化を行った翻訳モデルに対し、対訳データセットを用いて教師あり手法および教師なし手法（手法 1、手法 2）、単言語データセットを用いて教師なし手法（手法 3）により fine-tuning を行った。単言語データセットを用いて fine-tuning を行ったモデルに対しては、対訳データセットを用いて教師あり手法および教師なし手法により 2 回目の fine-tuning を行った（手法 4、手法 5）。

また、言語横断的でない初期化を行った翻訳モデルに対し、単言語データセットを用いて教師なし手法により fine-tuning を行った（手法 6）。

5.3　評価方法

各翻訳手法の性能を比較するために、6 つの評価尺度で評価した：(1) 原文との BLEU [4]（s-BLEU）、(2) 参照訳との BLEU（r-BLEU）、(3) 原文と参照訳をマルチリファレンスとして用いる BLEU（sr-BLEU）、(4) 3 で提案した hm-BLEU、(5) 文法誤り訂正で使用される原文と参照訳の両方を用いる GLEU [12]、(6) テキスト平易化で使用される原文と参照訳の両方を用いる SARI [13]。これら 6 つの評価尺度は値が大きいほど評価が高くなり、最大値は 1 である。

5.4　実験結果

各評価尺度による各翻訳手法の評価値を表 2 に示す。低リソースの対訳データセットを用いた手法 1 と手法 2 は、s-BLEU と r-BLEU のいずれも低い。単言語データセットを用いた手法 3 と手法 6 は、s-BLEU が高く、r-BLEU は原文よりも高い。手法 3 の後に教師あり手法により fine-tuning を行う手法 4 は、s-BLEU は手法 3 より低いが、r-BLEU は手法 3 より高い。

また、r-BLEU で評価すると手法 4 が最も高いが、hm-BLEU で評価すると手法 3 が最も高い。これは、hm-BLEU が s-BLEU を考慮した結果である。

6　メタ評価

5.3 で挙げた 6 つの評価尺度がどの程度、ポストエディットのコストを考慮した翻訳性能の評価できているかを調べるために、評価尺度の評価（メタ評価）を行った。まず、評価データから 2 つの段落を無作為に選び、各翻訳手法について 2.2 と同じ方法で翻訳を行い、システム単位の人手評価を行った。そして、各評価尺度の評価値（原文の評価値を含む）と人手評価値とのスピアマンの順位相関係数で評価した。

6.1　人手評価

国立国語研究所で『日本語歴史コーパス』の構築に従事する 2 人の評価者（A、B）によるシステム単位の人手評価を行った。評価の観点は、クラウドソーシングを使用して、公開できる品質の現代文に

6）　next sentence prediction タスクは行わない。

表 3　各評価尺度の評価値と人手評価値とのスピアマンの順位相関係数。

評価尺度	A	B	平均
s-BLEU	0.85	0.45	0.77
r-BLEU	0.74	**0.81**	0.81
sr-BLEU	0.85	0.45	0.77
hm-BLEU	**0.96**	0.65	**0.92**
GLEU	-0.70	0.09	-0.45
SARI	0.37	**0.81**	0.54

するための時間・人手コストを考慮した翻訳の精度とした。各評価値は、1 から 10 の離散値で、原文を 5 とし、最も精度が良い場合を 10 とした。原文および各翻訳手法の出力に対する人手評価の平均値を表 2 の右段に示す。各評価者の評価値は付録 D を参照。2 人の評価者間のスピアマンの順位相関係数は 0.72 であった。

6.2　メタ評価の結果

表 3 にメタ評価の結果を示す。人手評価の平均値との相関は、hm-BLEU が最も高い。これは hm-BLEU が、ポストエディットのコストを考慮した翻訳性能の評価尺度として優れていることを示している。また、評価者 A は hm-BLEU との相関が最も高いが、評価者 B は s-BLEU との相関が高い。

7　考察

事前学習モデルを用いた NMT　表 2 より、単言語データセットを用いた教師なし手法（手法 3、手法 6）は、s-BLEU を高く保ちつつ、r-BLEU は原文より高いことがわかる。これは、付録 C の出力例のように、多くの単語は原文のままで、一部の単語のみを翻訳しているためである。また、言語横断的な初期化（手法 3）と言語横断的でない初期化（手法 6）を比較するとあまり差がないことがわかる。これらの結果は、単言語データセットを用いた教師なし手法が、原文の構造をできるだけ保存するような翻訳を可能にしていることを示唆している。

また、事前言語モデルが翻訳性能に与える効果を調べるために、事前学習モデルで初期化しない手法との比較や、各手法の単語分散表現を使って通時的な単語の対応を調べる必要があるだろう。

調和平均 BLEU（hm-BLEU）　ポストエディットのコストは、原文との対応を調べるコストと、対応の取れた部分同士の翻訳が不適当な場合に正しい翻訳に訂正するコストの 2 つに大別できると考えられる。前者は s-BLEU で、後者は r-BLEU で評価できると仮定すると、表 3 より、評価者 A は s-BLEU と r-BLEU の両方で相関が高く、2 つのコストを同程度考慮して評価したと推測できる。一方、評価者 B は s-BLEU との相関が低く、原文との対応よりも、出力の翻訳精度や、正しい翻訳に訂正するコストを重視したと推測できる。

ポストエディットのコストを考慮した翻訳性能の評価尺度を考える際に、本研究では単純に s-BLEU と r-BLEU の調和平均を評価値とする評価尺度を提案したが、正しい翻訳に訂正するコストを重視するような重み付き調和平均も検討する必要があるだろう。

8　関連研究

星野ら [14] は古代から近世までの古文とその現代語訳からなる段落単位の対訳コーパスから、文分割のためのルールベースのスコア関数を用いて、文単位の対訳コーパスを抽出する手法を提案した。抽出した対訳コーパスを用いて統計的機械翻訳を行い、提案手法による文分割の有用性を示した。

また、Takaku ら [15] は単語分散表現を現代文の単言語コーパスで学習した後、古文の単言語コーパスを近代から古代に遡るように通時適応した。得られた単語分散表現で翻訳モデルの encoder の単語埋め込み層の重みを初期化し、星野らが抽出した対訳コーパスを用いて、古文の翻訳で初めて教師あり手法の NMT を行い、時代固有の単語と多様な語彙を訳出できるモデルを作成した。

9　おわりに

本研究では、近代文語文の現代語機械翻訳タスクを提案し、事前学習モデルを用いた手法で実験を行った。その結果、単言語コーパスを用いた手法により、原文の構造をなるべく保存するような翻訳が可能であることがわかった。また、ポストエディットのコストを考慮した翻訳性能の評価尺度として、原文も考慮する調和平均 BLEU を提案し、人手評価値との相関が最も高いことを示した。

今後は、原文の情報をできる限り保ちつつ、より多くの単語を正しく翻訳できるような手法を検討したい。

謝辞　本研究は国立国語研究所の共同研究プロジェクト「通時コーパスの構築と日本語史研究の新展開」及び所長裁量経費の助成を受けたものです。

参考文献

[1] Wei Xu, Alan Ritter, Bill Dolan, Ralph Grishman, and Colin Cherry. Paraphrasing for style. In *Proceedings of COLING*, pp. 2899–2914, 2012.

[2] Alexis Conneau and Guillaume Lample. Cross-lingual language model pretraining. In *NeurIPS*, Vol. 32, pp. 7059–7069, 2019.

[3] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NIPS*, pp. 5998–6008. 2017.

[4] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of ACL*, pp. 311–318, 2002.

[5] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL*, pp. 4171–4186, 2019.

[6] Guillaume Lample, Myle Ott, Alexis Conneau, Ludovic Denoyer, and Marc'Aurelio Ranzato. Phrase-based & neural unsupervised machine translation. In *Proceedings of EMNLP*, pp. 5039–5049, 2018.

[7] Prajit Ramachandran, Peter Liu, and Quoc Le. Unsupervised pretraining for sequence to sequence learning. In *Proceedings of EMNLP*, pp. 383–391, 2017.

[8] Melvin Johnson, Mike Schuster, Quoc V. Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, Macduff Hughes, and Jeffrey Dean. Google's multilingual neural machine translation system: Enabling zero-shot translation. *TACL*, Vol. 5, pp. 339–351, 2017.

[9] Taku Kudo, Kaoru Yamamoto, and Yuji Matsumoto. Applying conditional random fields to Japanese morphological analysis. In *Proceedings of EMNLP*, pp. 230–237, 2004.

[10] 小木曽智信, 小町守, 松本裕治. 歴史的日本語資料を対象とした形態素解析. 自然言語処理, Vol. 20, No. 5, pp. 727–748, 2013.

[11] 伝康晴, 小木曽智信, 小椋秀樹, 山田篤, 峯松信明, 内元清貴, 小磯花絵. コーパス日本語学のための言語資源：形態素解析用電子化辞書の開発とその応用. 日本語科学, Vol. 22, pp. 101–123, 2007.

[12] Andrew Mutton, Mark Dras, Stephen Wan, and Robert Dale. GLEU: Automatic evaluation of sentence-level fluency. In *Proceedings of ACL*, pp. 344–351, 2007.

[13] Wei Xu, Courtney Napoles, Ellie Pavlick, Quanze Chen, and Chris Callison-Burch. Optimizing statistical machine translation for text simplification. *TACL*, Vol. 4, pp. 401–415, 2016.

[14] 星野翔, 宮尾祐介, 大橋駿介, 相澤彰子, 横野光. 対照コーパスを用いた古文の現代語機械翻訳. 言語処理学会第 20 回年次大会, pp. 816–819, 2014.

[15] Masashi Takaku, Tosho Hirasawa, Mamoru Komachi, and Kanako Komiya. Neural machine translation from historical Japanese to contemporary Japanese using diachronically domain-adapted word embeddings. In *PACLIC*, 2020.

[16] 国立国語研究所 (2019)『日本語歴史コーパス 明治・大正編 I 雑誌』(短単位データ 1.2) https://pj.ninjal.ac.jp/corpus_center/chj/meiji_taisho.html#zasshi.

[17] 国立国語研究所 (2018)『日本語歴史コーパス 明治・大正編 II 教科書』(短単位データ 1.0) https://pj.ninjal.ac.jp/corpus_center/chj/meiji_taisho.html#kyokasho.

表 4　各翻訳手法の出力例。

手法	出力
原文	学問をするには分限を知ること肝要なり。
参照訳	学問をするには、なすべきことを知ることが大事である。
1	学問をするには、人間の見識品格は、このような学問をする人である。
2	学問をするならおおいに学問をするべきである。
3	学問をするには分限を知ることが肝要です。
4	学問をするには、分限を知ることが大切である。
5	学問をするには、分限を知ることが肝要だ。
6	学問をする者はがんばらなくてはならない。
原文	譬えば、いろは四十七文字を習い、手紙の文言、帳合いの仕方、算盤の稽古、天秤の取扱い等を心得、なおまた進んで学ぶべき箇条ははなはだ多し。
参照訳	たとえば、いろは四十七文字を習って、手紙の言葉や帳簿の付け方、そろばんの稽古や天秤の取り扱い方などを身につけることをはじめとして、学ぶべきことは非常に多い。
1	たとえば、文章にすればたいして意味がないようなものでも、口で言葉にすれば、理解もしやすく、人の心を動かすものがあるのだ。
2	たとえば、乞食を禁止するという法律は、もちろん公明正大なものだけれども、それぞれの個人が乞食に物を与えようとする心情はとがめなくてよろしい。
3	譬えば、いろは四十七文字を習い、手紙の文言、帳合いの仕方、算盤の稽古、天秤の取扱いなどを心得、なおまた進んで学ぶべき箇条ははなはだ多い。
4	たとえば、一人の新聞を見てみれば、その趣旨を示し、本を書き、新聞をの上、「政治」の法則を心得、それでも高いレベルの勉強をして、最後には自分でも「成功」ということは、たいへん多い。
5	萬えば、危機、危機、危機、危機、柱、いや、爪の仕方、算盤の稽古、天秤の習慣などを心得、なおまた進んで学ぶべき箇条ははなはだ多い。
6	たとえば、いま、若い学生が、酒や異性にもおぼれず、きちんと生活してちゃんと勉強すれば、家族や目上の人間にも叱られることなく、誇らしげになることに思えるけれども、これはただほかのだらしない学生に比べてのことにすぎない。

表 5　人手評価による各手法の評価値。

手法	BLUE			人手評価		
	s	r	hm	A	B	平均
原文	100.00	13.47	23.74	5	5	5.0
参照訳	13.68	100.00	24.07	–	–	–
1	3.28	7.15	4.50	1	6	3.5
2	2.97	6.25	4.03	1	2	1.5
3	57.50	16.42	25.55	7	6	6.5
4	19.43	17.17	18.23	3	7	5.0
5	3.16	5.87	4.11	1	1	1.0
6	56.69	15.85	24.77	6	8	7.0

A　文単位の対訳の作成方法

1 つの原文にいくつかの現代文を割り当てることで文単位の対訳を作成する。原文 S に対し、窓幅 10 の範囲でスコアが最小となる現代文を H_S とする。スコアは、句読点を除いた単語列の組に対する、「単語が一致すると距離が 2.5 だけ短くなる」という操作を追加した編集距離である。原文 A, B に対して貪欲に対応付けた現代文をそれぞれ H_A, H_B とし、H_A と H_B の間の現代語訳を、原文 A に割り当てる強さ α を定める。対応の取れない現代語訳が最も少なくなる α を $[0.5, 2.0)$ の範囲で二分探索により求める。探索が停止した時点での、対応の取れた原文と現代語訳の対を文単位の対訳として用いる。

B　単言語コーパス

単言語データセットの訓練データに使用した近代文語と現代語の各単言語コーパスは以下のとおりである。

近代文語の単言語コーパス　『日本語歴史コーパス　明治・大正編 I 雑誌』[16] に収録されている『明六雑誌』、『東洋学芸雑誌』、『国民之友』、『太陽』、『女学雑誌』、『女学世界』、『婦人倶楽部』の文語記事と、『日本語歴史コーパス　明治・大正編 II 教科書』[17] に収録されている国語教科書の文語部分を用いた。語種が漢語または固有名の場合は、表層形の代わりに語彙素を用いた。

現代語の単言語コーパス　『現代日本語書き言葉均衡コーパス』[7] の「出版・書籍」を用いた。

C　各翻訳手法の出力例

表 4 に各翻訳手法の出力例を示す。手法 1、手法 2 はどちらも、文意と無関係の単語を出力している。手法 3 は文末の単語以外、原文の単語をほとんどそのまま出力している。手法 4 は上段では「肝要」を「大切」と訳しているが、下段では文意と無関係の単語を出力している。手法 5 は上段では手法 3 と似ているが、下段では同じ単語の繰り返しが見られる。手法 6 は上段では「分限」の訳抜け、下段では文意と無関係の単語が見られる。

D　人手評価の結果

評価者 A、B による各翻訳手法の評価値を表 5 に示す。評価者によって、評価が異なる箇所がいくつか見られた。特に、手法 1 と手法 4 に対して、評価者 A は 5 未満、評価者 B は 5 以上と評価している。

7）https://pj.ninjal.ac.jp/corpus_center/bccwj/