

LLM が機械翻訳を捉えた桎梏から脱したのか — 翻訳創造性について —

Huacheng Song¹ Haoying Yang¹ Hongzhi Xu¹

¹ 上海外国語大学 コーパス研究院

hinayang331@gmail.com

概要

翻訳の文化的・技術的転回に伴って、翻訳創造性の重要さが高まる一方、それが自動翻訳モデルに一つの難題だと認識され、訳文における創造性の評価研究がますます必要となる。本研究では、訳者、Google および GPT4 による日本文学の中国語訳を対象に、各翻訳方式の創造性をめぐって細粒度的かつ量的評価を行った。その結果、LLM の翻訳創造性上の性能がニューラル機械翻訳より優れたが、この桎梏から完全に逃したというわけでもなく、翻訳者に比べれば成長の余地が多にある。

1 はじめに

現在のニューラル機械翻訳 (NMT: Neural Machine Translation) が直面した最大な問題点は抽象性、柔軟性および多様性という言葉遣いの特徴を持った文学の翻訳と言えよう [1, 2]。近年、ChatGPT を始め、大規模言語モデル (LLM: Large Language Models) が登場し、自動翻訳を含む数多くのテキスト生成タスクに高精度の性能を発揮したことから、世間で大きな期待が寄せられている [3, 4]。ところが、NMT と LLM による自動翻訳の翻訳創造性を覗き込む研究はまだ少なく、ましてや日本語訳を対象とした研究も管見の限り行われていない。こうした現状はモデルによる訳出精度の向上に支障を生じるだろう。

本研究は、Bayer-Hohenwarter [5] と Guerberof-Arenas ら [6, 7] の研究を踏まえ、受容可能性 (Acceptability) と新奇性 (Novelty) という 2 つの側面から、訳者ならびに Google、GPT4 による日本文学の中国語訳上の翻訳創造性を考察した。まず、原文から 8 種類の潜在的創造性単位 (UCP: Unit of Creativity Potential) を抽出し、翻訳品質評価枠組である DQF-MQM [8] および Bayer-Hohenwarter [5] と Pedersen [9, 10] が構築した翻訳ストラテジー分類体系を採用し、訳文の

受容可能性スコア (AS: Acceptability Score) と新奇性スコア (NS: Novelty Score) を算出した上で、創造性スコア (CS: Creativity Score) を計算する。さらに、4 つの翻訳自動評価指標を利用し、UCP 翻訳の全体的・局所的性能を観察する。

結果として、UCP の創造的翻訳は今も自動翻訳の問題点であることが得られた。LLM は一般の NMT より性能が上がったが、人間並みのレベルにはまだ達していないことに、本研究は以下の課題から具体的に論述する。1) UCP 翻訳は 3 つの翻訳方式にどのような支障を及ぼすか? 2) 3 つの翻訳方式による各種類の UCP 訳文にどのような性能が見られるか? 3) LLM は翻訳創造性の課題を克服したか?

2 関連研究

「創造性」は、抽象的な心理学の概念で、適切かつ正確な定義と測量は極めて難しい [11]。この概念を借りて、翻訳分野では創造的翻訳 (Creative Translation) という概念が生まれた。翻訳目的の実現を前提に原文と変わった表現を用いる翻訳方略を指し、受容可能性と新奇性という 2 つの性質を持っている [12, 13, 5, 11]。しかし、今のところ、翻訳創造性に関する量的研究はまだ始まったばかりで、運用可能な量的理論体系の構築が中心になっている [14, 5, 11]。コーパスで専門訳者と翻訳学習者の翻訳創造性を比較し、翻訳教育に示唆を与える実証研究もいくつか進められている [5, 15, 16]。どちらも本研究の理論基礎の構築に役立っている。

自動翻訳の領域では、NMT に創造性があるかという課題に取り組んだのは Guerberof-Arenas ら [6, 7] のみだ。彼らは、翻訳創造性評価指標の核心である創造的シフト (Creative Shift) [5] をめぐって、NMT が SF 小説翻訳上の性能を分析した。結果として、文学テキストを用いて訓練されたモデルでも、期待された創造性に至らないのみならず、それに基づ

きポストエディットを行った専門記者の創造性さえ低下させたことに、翻訳創造性の研究がNMTの更なる発展に重大な意味を持つことが裏付けられた。日本でのNMTパフォーマンスを細粒度的に考察する研究は、多くが語彙・文法レベルに着目した[17, 18, 19, 20]。一方、LLM領域の最新研究は、自動翻訳の全体的性能を向上させることが主流である[3, 4, 21]。要するに、NMTでもLLMでも文学翻訳の創造性への注目が十分に集まっていない。

3 研究デザイン

本研究は2つの段階からなっている。1) 原文と訳文の収集とアノテーション、2) 訳文の品質と翻訳創造性の自動評価と人手評価。

3.1 デキスト収集とアノテーション

デキスト収集 本研究では、1) 代表的な現代文学作品であること、2) 創造的翻訳が必要であること、3) モデルの訓練データに含まれていないこと [7] を基準に、日本人作家村田喜代子の受賞された短編小説「百のトイレ」、「蟹女」、「真夜中の自転車」、「白い山」¹⁾を原文として選んだ。原文デキストは合計3114文、93265文字で、平均文長は30文字である。比較分析の対象とする中国語訳文は3つの翻訳方式：人間翻訳の訳文は2023年に出版された中国語版²⁾、NMT訳文はNMTシステムGoogle翻訳³⁾、LLM訳文はOpenAI GPT4⁴⁾から入手する。

テキストアノテーション Bayer-Hohenwarte [5]によると、翻訳創造性を測るには、その第一歩が原文のUCP表現を抽出することである。UCP表現とは、特殊あるいは新奇な表現と形式を使ってある意味や目的を伝達・実現することである。本研究は、日本語で最も使用されている8種類のUCP表現⁵⁾を対象に、原文に現れたUCP表現を4名の日本語母語話者に抽出・分類してもらった。クロスチェックした結果、908文で1192のUCP表現が見つかった。

本文での翻訳創造性は、Guerberof-Arenasら[6, 7]の研究と同様に、「受容可能性」と「新奇性」から考察が展開される。そこで、UCP表現の訳文をアノテーションする際、前者はDFQ-MQMI[8]に基づいたエラーの識別とエラーの重大度(Severity)の判

断、後者は具体的な翻訳方略・手法の判断を2人の中国人日本語翻訳者に頼んだ。それぞれの詳しい注釈キームは表1と表2が示されている。

表1 受容可能性の注釈スキーム

側面	種類	重大度	スコア
受容可能性	不可能	深刻	-15
		重度	-5
		軽度	-1
	可能	中立	0
		称賛	1

表2 新奇性の注釈スキーム

側面	種類	翻訳方略	翻訳手法
新奇性がない		保持	保持
		直接訳 定訳	直接訳 定訳
新奇性		抽象化	上位語 パラフレーズ
		具体化	拡大化 補償
新奇性がある		調整	文化的・歴史的要因 コンテキスト的要因 創造的省略

特に、受容可能性、新奇性と翻訳創造性の関係は2つの仮説を基盤とする。

- **受容可能性は新奇性の前提**である。適切かつ正確(中立あるいは称賛の場合)であるUCP訳文のみ、新奇性スコアの算出を行う。
- **新奇性は翻訳創造性の核心**である。新奇性がない訳文は正しくても翻訳創造性がないとする。

3.2 指標とスコア

アノテーションから得たデータを基に、本文は1) 4つの翻訳自動評価指標BLEU[22]、COMET[23]、COMETKIWI[24]、XCOMET-QE[25]、2) 受容可能性、新奇性および翻訳創造性を定量化した人手評価体系を通じて、記者ならびにGoogle翻訳、GPT4の翻訳創造性について全体から局所までの量的比較分析を実施した。

人手評価は3つの公式によって計算される。AS(公式1)はUCP訳文の語彙・文法上の正確さと意味上の適切さ、NS(公式2)はUCP訳文における創造性シフトの比率を測る。 N_{UCPs} 、 N_{Kudos} 、 $N_{Creative\ Shifts}$ はそれぞれUCP、表1の称賛、表2の「新奇性のある」カテゴリーの中に含まれた翻訳方略の数で、 $ErrorPoints$ は表1の「受容不可能」に含まれたエラー重大度の総点である。CS(公式3)はASとNSの加重調和平均値で、本研究は β を1とする。3つのスコアは区間[0,1]内にあり、翻訳性能との関係は正の相関にある。

1) <https://a.co/d/6Ds7zjv>

2) <https://a.co/d/2CCBPba>

3) <https://translate.google.com>

4) <https://openai.com/gpt-4>

5) 詳しい定義と例文は付録の表A

表3 8種類のUCP訳文の自動評価平均スコア

番号	カテゴリー	BLEU		COMET		COMETKIWI			XCOMET-QE			
		人間	Google	GPT4	人間	Google	GPT4	人間	Google	GPT4	人間	Google
A	メタファー	20.56	26.11		0.70	0.74	0.37	0.38	0.40	0.80	0.82	0.84
B	イディオム・慣用語・ことわざ	22.45	28.72		0.69	0.73	0.39	0.41	0.40	0.79	0.81	0.78
C	掛け言葉・語呂合わせ・パロディ	5.60	5.53		0.54	0.53	0.52*	0.49	0.43	0.81*	0.65	0.63
D	オノマトペ	16.77	21.95		0.62	0.65	0.36*	0.35	0.36	0.73*	0.63	0.68
E	口だけた言い方・業界用語・隠語	22.42	29.80		0.69	0.72	0.47	0.51	0.49	0.87	0.91	0.89
F	文化・歴史用語	31.15	32.30		0.72	0.76	0.35	0.40	0.38	0.78	0.84	0.84
G	外来語・新語・流行語	55.01	55.15		0.83	0.84	0.53	0.55	0.55	0.87	0.90	0.90
H	固有名詞	38.81	48.00		0.78	0.82	0.44	0.46	0.45	0.77	0.81	0.84

$$\text{Acceptability Score(AS)} = \left(\frac{N_{\text{Kudos}} + \sum \text{Error Points}}{N_{\text{UCPs}}} + 15 \right) / 16 \quad (1)$$

$$\text{Novelty Score(NS)} = \frac{N_{\text{Creative Shifts}}}{N_{\text{UCPs}}} \quad (2)$$

$$\text{Creativity Score(CS)} = \frac{(1 + \beta^2) \times \text{Acceptability Score} \times \text{Novelty Score}}{\beta^2 \times \text{Acceptability Score} + \text{Novelty Score}} \quad (3)$$

4 結果分析

4.1 自動評価

3つの翻訳方式から生成されたUCP表現を含む文全体の訳文(以下 Sent)と、UCP表現だけに対応する訳文(以下 UCP 訳文)に対する4つの自動評価指標による採点は表4にまとめた。BLEUとCOMETは人間翻訳を参照訳とし、NMT訳文と参照訳との一致性を測るため、人間翻訳への点付けはしない。一方、COMETKIWIとXCOMET-QEは、原文と訳文間の意味的類似度を測るメカニズムとなっている。また、BLEU⁶⁾スコアだけが区間[0,100]の実数で、残りの3つ⁷⁾は[0,1]の間の実数で表現される。

表4 UCPを含む文の翻訳とUCP訳文の自動評価スコア

翻訳方式	BLEU		COMET		COMETKIWI		XCOMET-QE	
	Sent	UCP	Sent	UCP	Sent	UCP	Sent	UCP
人間翻訳					0.52	0.41	0.65	0.80
Google 翻訳	18.11	15.79	0.77	0.72	0.55	0.43	0.71	0.79
GPT4 翻訳	22.67	22.20	0.80	0.76	0.57	0.44	0.73	0.82

表4の示すとおり、GPT4は文レベルにおいても句レベルにおいてもより高い性能を発揮できた。BLEUとCOMETのスコアからみては、GPT4はGoogle翻訳より人間翻訳と一致する一方、COMETKIWIとXCOMET-QEのスコアからみては、原文に潜んだ意味をより正確に掘り出せる。そして、各指標のSentスコアとUCPスコアを比較し、XCOMET-QEを除く3つの指標は前者より後者のほ

6) <https://github.com/mjpost/sacrebleu>

7) <https://github.com/Unbabel/COMET>

うが低いことから、UCP翻訳に必要な創造性がどの翻訳方式にも支障をもたらしたことが考えられる。

表3は各種のUCP訳文に対する自動評価指標の平均得点を統計した。GoogleとGPT4の性能について、C類UCP訳文がBLEUとCOMETからもっとも低い評価をもらったのに対し、COMETKIWIとXCOMET-QEのスコアはD類UCPが一番低い。わりと性能高く訳出できたUCPはG類UCPで、翻訳方式と評価指標にかかわらず0.5以上の得点を得た。

異なった翻訳方式のスコアを比較分析すると、表4との共通点と相違点がいくつか見つかった。共通点として、GPT4の性能が全体的にGoogle翻訳より優れていることがあげられる。相違点として、C類UCPとD類UCP訳文に対するCOMETKIWIとXCOMET-QEのスコアは、人間翻訳のほうが高いこと、そして、Google翻訳は数種類のUCPに対して、GPT4と太刀打ちできる情報抽出能力を備えていることがある。

自動評価指標の採点は、GPT4によるUCP翻訳がより高い品質を持っていることを示した。ところが、COMETKIWIとXCOMET-QEが人間翻訳への点付けが一番低いという常識外れな現象も見られ、主流となった自動評価指標は単語の埋め込みと意味特徴によって、参考訳への依存から解放されたが、まだ十分にはならないことがわかった。そこで、本研究は信頼性の高い人手評価に重きを置いた。

4.2 人手評価

表6では、訳者、Google翻訳、GPT4によるUCP訳文のAS、NSおよびCSが表されている。どのスコアでも人間 > Google > GPT4という順序が観察される。CSについて、GPT4とGoogle翻訳は人間翻訳に比べてそれぞれ0.13と0.2の差が見られるが、自動翻訳同士の差は0.07しかない。要するに、翻訳創造性の各方面において、GPT4がGoogle翻訳より優れたといえども、人並みのレベルまではかなり

表5 8種類のUCP訳文のAcceptability Score、Novelty ScoreおよびCreativity Score

番号	カテゴリー	Creativity Score			Acceptability Score			Novelty Score		
		人間	Google	GPT4	人間	Google	GPT4	人間	Google	GPT4
A	メタファー	0.80	0.54	0.55	0.94	0.86	0.90	0.70	0.39	0.40
B	イディオム・慣用語・ことわざ	0.86	0.60	0.75	0.93	0.81	0.90	0.80	0.48	0.64
C	掛け言葉・語呂合わせ・パロディ	0.49	0.00	0.48	0.94	0.71	0.85	0.33	0.00	0.33
D	オノマトペ	0.56	0.41	0.47	0.93	0.86	0.89	0.40	0.27	0.32
E	口くだけた言い方・業界用語・隠語	0.48	0.30	0.41	0.95	0.86	0.85	0.32	0.18	0.27
F	文化・歴史用語	0.68	0.48	0.53	0.92	0.85	0.88	0.54	0.34	0.38
G	外来語・新語・流行語	0.37	0.18	0.22	0.93	0.90	0.92	0.23	0.10	0.13
H	固有名詞	0.24	0.06	0.19	0.94	0.85	0.89	0.14	0.03	0.10

の距離があるので、創造性という桎梏を完全に脱したというわけではない。ASにおける3つの翻訳方式間の差は0.1に及ばいが、新奇性におけるGoogle翻訳とGPT4の得点はそれぞれ人間翻訳の50%と70%となっており、翻訳方式間の創造性の差異は新奇性によるものとみなせるだろう。

表6 UCP訳文のAcceptability Score、Novelty ScoreおよびCreativity Score

翻訳方式	Creativity Score	Acceptability Score	Novelty Score
人間翻訳	0.58	0.93	0.43
Google 翻訳	0.38	0.86	0.24
GPT4 翻訳	0.45	0.90	0.30

UCPの種類を考えながら、翻訳方式間の得点を比較した。表5によると、各種のUCP訳文のどの得点でも全体と一致した順序が現れた。人間翻訳は必ずしも正しいとは言えないが、ASが0.92と0.95の間で分布している。それに比べて、GPT4もGoogle翻訳も0.85以上の点数を得たため、受容可能性上の性能において今の自動翻訳にはかなりの競争力があると考えられる。Google翻訳による全種類のUCP訳文はNS値が人間翻訳の半分以下で、新奇性のない翻訳方略が多用されている。GPT4はC類UCPをはじめ、5種のUCP訳文のNSが人間翻訳と0.1弱の差が観察されたが、残りの3種(A類、B類、F類)訳文のパフォーマンスには改善の余地がかなり大きい。

UCP全体において、B類UCPのCSが一番高く、どの翻訳方式にとってもこれを創造的に訳す難易度が一番低いことがわかった。図1によっては、B類UCPを訳す際、訳者のみならず、自動翻訳も直接訳、抽象化または調整の翻訳方略を活用できる。これは、原文に出たB類UCPの大半は、身体部位に基づいたものであることが原因だろう。このような表現が頻ぱんに使われている間、それに対応する

中国語訳も定着した。そして、訓練データにも使用されたため、Google翻訳とGPT4に学習された。その反対に、G類UCPとH類UCP訳文の創造性は一段と低く、借用・直接訳・定訳という新奇性のない翻訳方略の使用が多くみられる。主な原因は：1) 定訳が多いこと、2) 一部のH類UCPは漢字からなっており、中国語との意味的連合性が強いことにある。A類UCPの訳文においては、翻訳方式間のCSに著しい差がある一方、人間翻訳のNSは自動翻訳の70%を超えている。我々の認知構造に深く根差しているA類UCPは、文化・社会・コンテクストと深く関わっているが、B類UCPなどの言語文化に定着されたUCPとは違って、個人的な性質を持っており、形式的にも意味的にも柔軟性と多様性が顕著である。専門訳者は起点言語と目標言語の文化を熟知しているため、認知的・文化的負荷の高いA類UCPを翻訳するとき、抽象化・具象化・調整などの方略を使いこなせる。ところが、現時点でのNMTとLLMにとって、認知的に原文の意味合いを理解するのは今も一つ大きな課題と言える。

5 おわりに

本研究は、Guerberof-Arenasら[7]が提唱した翻訳創造性スコアの公式を改善した上で、日本文学に出たUCP表現の中国語訳を対象に、量的アプローチでGPT4を代表とするLLMの性能を考察した。結果的には、LLMの翻訳創造性および2つの側面「受容可能性」と「新奇性」における性能はGoogle翻訳を代表とするNMTより増しているが、専門訳者のレベルに到達するまで見過ごせない距離があると言える。ゆえに、認知的・文化的要素を考慮に入れることで、LLMの文学翻訳における性能向上が期待される。また、翻訳自動評価指標による翻訳創造性への評価も今後の課題として改善する必要がある。

参考文献

- [1] Cynthia Beatrice Costa and Igor A Lourenço da Silva. On the translation of literature as a human activity par excellence: ethical implications for literary machine translation. **Aletria: Revista de Estudos de Literatura**, Vol. 30, No. 4, pp. 225–248, 2020.
- [2] Katherine Thai, Marzena Karpinska, Kalpesh Krishna, Bill Ray, Moira Inghilleri, John Wieting, and Mohit Iyyer. Exploring document-level literary machine translation with parallel paragraphs from world literature. In **Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing**, pp. 9882–9902, 2022.
- [3] Wenhao Zhu, Hongyi Liu, Qingxiu Dong, Jingjing Xu, Lingpeng Kong, Jiajun Chen, Lei Li, and Shujian Huang. Multilingual machine translation with large language models: Empirical results and analysis. **arXiv preprint arXiv:2304.04675**, 2023.
- [4] Biao Zhang, Barry Haddow, and Alexandra Birch. Prompting large language model for machine translation: A case study. **arXiv preprint arXiv:2301.07069**, 2023.
- [5] Gerrit Bayer-Hohenwarter. **Translational creativity: how to measure the unmeasurable**, Vol. 37. Samfundslitteratur Copenhagen, 2009.
- [6] Ana Guerberof-Arenas and Antonio Toral. The impact of post-editing and machine translation on creativity and reading experience. **Translation Spaces**, Vol. 9, No. 2, pp. 255–282, 2020.
- [7] Ana Guerberof-Arenas and Antonio Toral. Creativity in translation: Machine translation as a constraint for literary texts. **Translation Spaces**, Vol. 11, No. 2, pp. 184–212, 2022.
- [8] Arle Lommel and Alan K Melby. Tutorial: Mqm-dqf: A good marriage (translation quality for the 21st century). In **Proceedings of the 13th Conference of the Association for Machine Translation in the Americas (Volume 2: User Track)**, 2018.
- [9] Peter Sibson. Subtitling norms for television: an exploration focussing on extralinguistic cultural references, 2012.
- [10] J. Pedersen. **Subtitling Norms for Television: An Exploration Focussing on Extralinguistic Cultural References**. Benjamins translation library. John Benjamins Publishing Company, 2011.
- [11] Gerrit Bayer-Hohenwarter and Paul Kussmaul. Translation, creativity and cognition. In **The Routledge handbook of translation and cognition**, pp. 310–325. Routledge Abingdon and New York, 2020.
- [12] Ana Rojo. The role of creativity. **The handbook of translation and cognition**, pp. 350–368, 2017.
- [13] Paul Kussmaul. Types of creative translating. **Benjamins translation library**, Vol. 39, pp. 117–126, 2000.
- [14] Paul Kussmaul. Creativity in the translation process: Empirical approaches. In **Translation Studies: The state of the art**, pp. 91–101. Brill, 1991.
- [15] Gerrit Bayer-Hohenwarter. **Comparing translational creativity scores of students and professionals: flexible problem-solving and/or fluent routine behaviour**, Vol. 83. Samfundslitteratur Copenhagen, 2010.
- [16] Gerrit Bayer-Hohenwarter. “creative shifts” as a means of measuring and promoting translational creativity. **Meta**, Vol. 56, No. 3, pp. 663–692, 2011.
- [17] 毅彦吉見. 人間による翻訳文と機械翻訳文における係り受け距離の比較分析. **Information : an international Interdisciplinary journal**, Vol. 17, No. 6, pp. 2155–2168, 06 2014.
- [18] TAKEHIKO YOSHIMI. Comparative analysis of familiarity rating of verbs in human-translated and machine-translated sentences. **Journal of Natural Language Processing**, Vol. 11, No. 2, pp. 101–113, 2004.
- [19] 毅彦吉見. 人間による翻訳文と機械翻訳文における係り受け構造の統計的性質. 計量国語学 = Mathematical linguistics : 計量国語学会機関誌, Vol. 24, No. 5, pp. 218–227, 2004.
- [20] 毅彦吉見. 人間による翻訳文と機械翻訳文における名詞の馴染み度の計量的比較分析. **Information : an international Interdisciplinary journal**, Vol. 16, No. 7, pp. 5329–5344, 07 2013.
- [21] Chenyang Lyu, Jitao Xu, and Longyue Wang. New trends in machine translation using large language models.
- [22] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In **Proceedings of the 40th annual meeting of the Association for Computational Linguistics**, pp. 311–318, 2002.
- [23] Ricardo Rei, José GC De Souza, Duarte Alves, Chrysoula Zerva, Ana C Farinha, Taisiya Glushkova, Alon Lavie, Luisa Coheur, and André FT Martins. Comet-22: Unbabel-ist 2022 submission for the metrics shared task. In **Proceedings of the Seventh Conference on Machine Translation (WMT)**, pp. 578–585, 2022.
- [24] Ricardo Rei, Marcos Treviso, Nuno M Guerreiro, Chrysoula Zerva, Ana C Farinha, Christine Maroti, José GC De Souza, Taisiya Glushkova, Duarte Alves, Luísa Coheur, et al. Cometkiwi: Ist-unbabel 2022 submission for the quality estimation shared task. In **Proceedings of the Seventh Conference on Machine Translation (WMT)**, pp. 634–645, 2022.
- [25] Nuno M Guerreiro, Ricardo Rei, Daan van Stigt, Luisa Coheur, Pierre Colombo, and André FT Martins. xcomet: Transparent machine translation evaluation through fine-grained error detection. **arXiv preprint arXiv:2310.10482**, 2023.

A 8種類のUCPについて

番号	カテゴリー	特徴	【例文】
A	メタファー	ある物事の様子や性質を言い表す場合に、ほかのより具体的な物事になぞらえて表現したり、ほかの物事を言い表す際に用いられる特徴的な語彙を借りて表現したりすること。比喩表現の一種であるが、「まるで〜」「〜のような」などの明確な比喩表現を使わない。	従姉は【悩みの種】をひとつ持ってきた。
B	イディオム・慣用語・ことわざ	イディオムにおける単語の配列が慣習的となっており、字面から意味を推測できないことが多い。	仕事上の悩みにはわたしの【首がかかっている】。
C	掛け言葉・語呂合わせ・パロディ	言葉の持つ音の響きやリズムを運用することで、聞き手や読み手に同音異義語を連想させ、面白さを楽しませる表現。字面上の意味と裏の意味が異なっている。	【損して徳とれ】(損して得とれ)
D	オノマトペ	物事の声を模した言葉である擬音語、事物の様態を言語音によって象徴的に表す言葉である擬態語を総称したものの。	あたりには【ポツポツ】と夏の草が噴き出ている。
E	口くだけた言い方・業界用語・隠語	格式ばらずに、より平易で日常的な言葉を用いた表現。多くの場合、俗っぽい感じや親しみやすい雰囲気醸成。そして、ここでの業界用語も、正式名称ではない略語、隠語のようなもの。	ヤクザ:【サツ】だ!!
F	文化・歴史用語	独自の文化現象や歴史典籍と深く関わっており、その文化に特有な物事を指したり、あるいは特定の歴史意味を持っている表現のこと。	二葉亭のお父さんは長い間江戸の【御家人】化していた。
G	外来語・新語・流行語	新旧文化・独自文化と外来文化の相互交流において、新たに生まれ、新しい概念を取り入れた語。主に外来語、新語・流行語。	彼氏が怒られる姿を見て【蛙化現象】した…
H	固有名詞	人名・地名・事件名・物事名。ある類に属する個物に与えられた名称を表わす語。人名、地名、国名のほか、書名、会社名、学校名、年号、商号、商品名などの呼び名。	【スターバックス】、ぼっち・ぎ・ろっく!

B 人手評価について

各訳文のアノテーションの結果

翻訳方式	UCP数	Creative Shift	Error Point	Kudos
人間翻訳	1192	508	-140	41
Google 翻訳	1192	289	-1483	3
GPT4 翻訳	1192	357	-809	12

8種のUCP訳文のアノテーションの結果

番号	カテゴリー	数	Creative Shift			Error Point			Kudos		
			人間	Google	GPT4	人間	Google	GPT4	人間	Google	GPT4
A	メタファー	132	92	52	53	-6	-156	-90	14	1	4
B	イディオム・慣用語・ことわざ	143	115	68	92	-12	-283	-93	4	0	3
C	掛け言葉・語呂合わせ・パロディ	3	1	0	1	0	-11	-5	0	0	1
D	オノマトペ	227	90	62	73	-21	-280	-184	10	0	2
E	口くだけた言い方・業界用語・隠語	22	7	4	6	0	-26	-31	3	0	0
F	文化・歴史用語	206	112	70	78	-49	-301	-189	5	2	1
G	外来語・新語・流行語	286	67	28	36	-47	-173	-80	4	0	1
H	固有名詞	173	24	5	18	-5	-253	-137	1	0	0

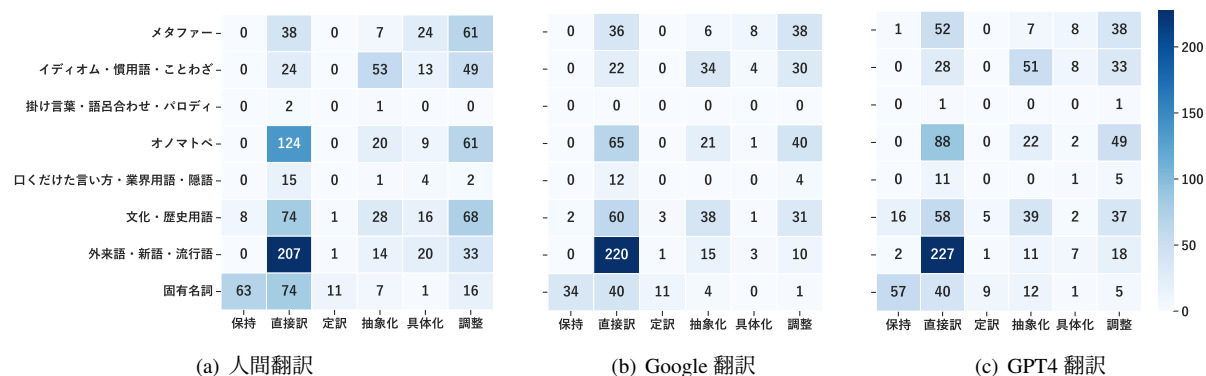


図1 8種類のUCPに対する翻訳方略の分布