

大阪電気通信大学
総合情報学部 デジタルゲーム学科
2024 年度 卒業論文

LLM を用いた文化的な背景を持つ文章 や単語の翻訳に関する研究

HW21A091 中田淳貴
デジタルアーキテクチャ研究室
指導教員：上善 恒雄

2025 年 1 月 29 日

概要

翻訳ソフト (Google 翻訳や DeepL など) は日常的な翻訳で高い精度を持つ一方で、スラングや方言、混成語といった特定の文化的背景や独自表現を含む文章では精度が低下することが知られている。ただし、昨今では大規模言語モデルの活用が進み、既存の翻訳ソフトではなく大規模言語モデルを用いることで文章の翻訳を行うようになってきている。本研究では、大規模言語モデルとして一般ユーザーに広く利用される ChatGPT の翻訳結果を既存の日本語訳と比較し、スラングやリドルなどといった文化的背景を反映した表現の翻訳可能性を評価した。実験結果より、以下のことが明らかになった

(1) 一般的ではない語彙や複雑な文法構造に対して翻訳精度が著しく低下する (2) 既存の知識に基づいた訳が優先され、元の意図が歪む場合がある (3) 未知語に対してカタカナ表記に依存する傾向がある

これらから、文化的要素を考慮したトークン化技術や辞書の拡充や未知語へのカタカナ語表記に対応するためのプロンプトでの追加の指示などが必要であることが考えられる。

目次

1	はじめに	5
1.1	研究背景	5
1.2	研究目的	5
1.3	研究方法	5
1.4	本論文の構成	5
2	関連研究	6
2.1	大規模言語モデルの文化的理解と課題	6
2.2	オンラインコミュニティ上における言語の急速な発展に伴う変化への大規模言語モデルの理解度とその改善手法	6
2.3	単語中の文字の順序を入れ替えた場合の大規模言語モデルの復元能力	6
3	研究対象の定義	6
3.1	スラング	7
3.2	アナグラム	7
3.3	リドル	7
3.4	混成語	7
3.5	パロディ	7
3.6	方言	8
3.7	古語	8
4	実験	8
4.1	ChatGPT を用いる際の問題点	8
4.2	スラングに関する実験	8
4.3	アナグラムに関する実験	9
4.4	リドルに関する実験	9
4.5	混成語に関する実験	9
4.6	パロディに関する実験	9
4.7	方言に関する実験	10
4.8	日本語評価における BLUE の課題	10
4.9	LLM とトークナイザーの関係	10
5	結果	10
6	考察	16

6.1	スラングに関する実験	16
6.2	アナグラムに関する実験	17
6.3	リドルに関する実験	18
6.4	混成語に関する実験	19
6.5	パロディに関する実験	19
6.6	方言に関する実験	20
6.7	複数の実験をまとめた考察	20
7	今後の課題	21

1 はじめに

1.1 研究背景

以前より Google 翻訳や DeepL など, さまざまなウェブ翻訳ツールが登場している. これらの翻訳ツールの翻訳精度は一般に用いる目的であれば十分高いと言えるが, 翻訳可能なものは基本的に標準的な文章や単語に限られており, あまり一般には用いられない文化的な背景を持つ文章や単語などに対しては十分に元の翻訳精度が発揮できていない. ただし, 昨今では ChatGPT などの大規模言語モデルを用いて文章の翻訳を行うといった手段も一般的になっている. そこで, そのような文化的な背景を持つ文章や単語に対して大規模言語モデルを用いて翻訳を行うことで, 翻訳精度を向上させることができるのではないかと考えた.

また大規模言語モデルでも完全な翻訳は行えないため, どのような内容に対して翻訳性能が低下するのかを探りその改良点を検討した.

1.2 研究目的

文献 [4] では既存手法では大規模言語モデルはスラングを理解するのが得意ではないとあるが, 実際に文章中で用いられる一般的ではない語彙はスラングだけにとどまらない.

例えば, 古文やことわざなどからの引用であったり, リドルや言葉遊びなども一般的ではない表現を用いることがある. スラングだけではなくそのような場合での大規模言語モデルの翻訳を調査する.

1.3 研究方法

本研究では, ChatGPT を用いて俗語の翻訳を行い, その翻訳を評価する. 本研究では文章全体の翻訳精度を評価するのではなく, あくまで一部の語句が大元の単語の意図が伝わる程度に翻訳されているかを評価する. さまざまな俗語を用いてどのような場合が翻訳精度が低下するのかを調査する.

1.4 本論文の構成

本章では, 本研究の背景と目的, 方法をそれぞれ述べた. 2 章では関連研究について述べる. 3 章ではこの論文中における文化的な背景を持つ文章や単語とは一体どのようなものを指しているのかを述べる. 4 章では実験の内容について述べる. 5 章では実験の結果について示す. 6 章では結果の考察を述べる. 7 章では今後の課題について述べる.

2 関連研究

2.1 大規模言語モデルの文化的理解と課題

研究 [5] では, あまり研究されていなかった大規模言語モデルの文化的な常識の理解力についての調査が行われている. 実験は中国, インド, イラン, ケニア, アメリカの国々の文化について Vicuna や Falcon, LLAMA2, GPT-4 を対象に行われた. 内容は文化的背景を必要とする常識, 文化からの国の予測, これらの文化的な情報を述べた後の一般常識や一般常識から文化を推測することなどが含まれている. その結果, 大規模言語モデルは文化固有の常識について大きなパフォーマンスの違いがあることや, 一般常識が文化的な文脈によって影響を受けること, プロンプトに入力した言語によって文化に関する知識でパフォーマンスに影響が出ることなどが述べられている.

2.2 オンラインコミュニティ上における言語の急速な発展に伴う変化への大規模言語モデルの理解度とその改善手法

研究 [4] では, スラングなどの言葉の移り変わりの激しい分野では大規模言語モデルがついていけないという問題に対して FOCUS というアプローチと SLANG というベンチマークが提案されている. アプローチ手法は因果推論を通して新しいフレーズや文脈の理解力を向上させるというものである. Mistral-7B や GPT-4, Claude での実験により, このアプローチを行った場合にスラングやインターネットミームの理解力が既存手法よりも向上することが示されている.

2.3 単語中の文字の順序を入れ替えた場合の大規模言語モデルの復元能力

研究 [1] では単語ごとに文字の順序が入れ替えられた英文であっても大規模言語モデルがその文章を理解できるかという実験が行われた. 実験は GPT-4 や GPT-3.5-turbo, text-davinci-003, Falcon-180b, Llama-2-70b を対象に文字の順序を入れ替えた文章の復元とその質問内容について回答するというものである. その結果, 多くの大規模言語モデルはそれらの文章を適切に元の文章に修正できることが示された. また, 特に GPT-4 はタイポグリセミアに近い結果を発揮し, より難易度の高い文章についても適切に修正できることが示された. これはこれらの改変された文章はトークナイザーによる適切なトークン化を妨害するにも関わらず, 非常に高い復元力を発揮するという直感的に反する結果である.

3 研究対象の定義

一般的な単語とは異なり, 各国の文化や歴史に根差した言葉や表現を指す. 一般的な文章と同等の文章構造を持つものの, 意味として一般的でない結果になることがある. そのため, 翻訳ツールによっては翻訳精度が低下することがある. この研究では以下のような内容を文化的な背景を持つ文

章や単語として扱う。

3.1 スラング

主に特定の集団内でのみ通用する言葉を指す。ここでは主にインターネットスラングと呼ばれるスラングを扱う。また、スラングはその性質上他の国に同様の意味を表す単語や文章が存在するとは限らない。今回の実験では英語圏でのスラングを用いたため、日本語に翻訳する際に対応する語彙が存在しない可能性がある。

例: アスキーアート, lol

3.2 アナグラム

既存の文字列を入れ替えて新しい意味を持つ文字列を作成することを指す。本来のアナグラムとは少し異なるが、本研究内では主にタイポグリセミアと呼ばれる現象を主として扱う。タイポグリセミアとは文章の最初と最後以外の文字を入れ替えても人間がその文章をある程度適切に読み取れる現象のことである。

例: 無理なマーク (生クリーム), flutterby (butterfly)

3.3 リドル

主に日本語では謎かけなどの和訳が当てられる。ここでは主に言葉遊びを主題としたリドルを扱う。

例: パンはパンでも食べられないパンは?, What is the center of gravity?

3.4 混成語

日本語では主にかばん語としても知られる。複数の単語の一部をそれぞれ組み合わせて作成された新しい単語のことを指す。ポケモン (ポケット + モンスター) のように、既存のある程度一般的な語彙にも含まれることがある。

例: ゴジラ (ゴリラ + クジラ), Oxbridge (Oxford + Cambridge)

3.5 パロディ

何か特定の元となる文字列などが存在し、それを元にしたリ、組み込むことで新しい文字列を作成することを指す。場合によっては元となった文章をそのまま全て引用している場合も考えられる。

例: 夜は短し歩けよ乙女 (いのち短し 恋せよ乙女), 情報畑でつかまえて (ライ麦畑でつかまえて)

3.6 方言

共通語・標準語とは異なった形で、特定の地方だけで使われる語を指す。一般に日本では日本国内における関西弁などが考えられるが、ここでは英語における Chinglish などの混成言語も含む。

例: ぼっこ (北海道), よか (博多)

3.7 古語

古代では用いられていたが、現代では一般的ではない語を指す。

例: あさまし (意外である), thou (you)

4 実験

本研究では翻訳用の大規模言語モデルとして ChatGPT の gpt-4o を用いた。理由としては先行研究でも系列の GPT-4 が用いられていることが多く、英語と日本語を十分に学習しているモデルだからである。また実際にユーザーが翻訳を行うことも考えると、数ある大規模言語モデルの中でもユーザーが多く一般的に認知されており、プログラム知識などを特に必要とせずとも簡単に翻訳の行える ChatGPT が今回の実験の対象として適当であると考えた。

4.1 ChatGPT を用いる際の問題点

GPT-4o は一般ユーザー向けにはサービスが制限されているため、一定時間ごとの文章量には限界が存在する。その制限を超えると GPT-4o から GPT-4 に自動的に切り替わってしまうため、使用する際に GPT-4o が使用できているかを確認する。また、ChatGPT にはメモリ機能とカスタム指示機能が存在する。これらは各チャット間で共通の知識を記憶させたり、プロンプトに入力を行う際に ChatGPT をパーソナライズする機能である。これらは今回の実験において影響を与える可能性が考えられるために設定から無効化しておく。同様に入力するプロンプトによって結果が大幅に変化することが考えられるために指示の内容は「以下の文章を日本語に翻訳してください」で統一する。

4.2 スラングに関する実験

先行研究 [4] で用いられているデータセット^{*1}のデータを一部抽出し、ChatGPT に文章を入力する。スラングの日本語訳と ChatGPT の出力結果の該当するスラング部分を比較し、どれだけ元の文章の意図が伝わっているかを確認する。

本実験での日本語訳はデータセットに英文で記載されているスラングの説明を Google 翻訳にかけ

^{*1} https://github.com/Meirtz/FocusOnSlang-Toolbox/blob/main/data/words_raw_Feb_2024.json

たものを参照する.

4.3 アナグラムに関する実験

ChatGPT に適当な 10 単語程度の英文とその日本語訳を用意させる. その後, その英文をもとに単語単位で文字の順序を入れ替えた文章を作り, 翻訳させる文章とし, ChatGPT に文章を入力する. ChatGPT の出力結果と日本語訳を比較し, どれだけ元の文章の意図が伝わっているかを確認する. 本実験での日本語訳は ChatGPT が英文と同時に生成した日本語訳を参照する.

4.4 リドルに関する実験

DMM 英会話の提供するサービス^{*2}から英語圏でのリドルをいくつか用意し, ChatGPT に文章を入力する. 元のリドルの意図と ChatGPT の翻訳結果を比較し, リドルとして日本語でも適切に解くことができるかどうかを確認する.

本実験での日本語訳は上記サービスに英文とともに併記してある日本語訳を参照する.

4.5 混成語に関する実験

Web 上で公開されている Through the Looking-Glass, and What Alice Found There^{*3*4}のうち, 混成語が使用されている箇所を抽出し, ChatGPT に文章を入力する. ChatGPT の出力結果と日本語訳を比較し, どれだけ元の文章の意図が伝わっているかを確認する.

本実験での日本語訳は新訳として出版されている鏡の国のアリス [6] を参照する.

4.6 パロディに関する実験

日本でも提供されているゲームである Minecraft の実績の名前のうち, パロディなどが使用されているものを一部抽出し, ChatGPT に文章を入力する. ChatGPT の出力結果と日本語訳を比較し, どれだけ元の文章の意図が伝わっているかを確認する.

本実験で日本語訳は Minecraft のゲーム内における日本語版の翻訳をそのまま活用する. Minecraft は 2 種類のエディションが存在し, それぞれのエディションにおける翻訳が一部異なるが, 今回の実験では一般によく知られている Java エディションの翻訳を参照する.

^{*2} <https://eikaiwa.dmm.com/blog/learning-english/tips/riddle/>

^{*3} 邦題: 鏡の国のアリス

^{*4} <https://www.gutenberg.org/cache/epub/12/pg12-images.html>

4.7 方言に関する実験

DMM 英会話の提供するサービス^{*5}などを元に方言や Chinglish をいくつか抽出し, ChatGPT に文章を入力する. 元の方言の意味と ChatGPT の翻訳結果を比較し, 適切に本来の意味が伝わっているかを確認する.

本実験での方言の日本語訳は上記サービスに英文とともに併記してある日本語訳を参照する. Chinglish の翻訳は^{*6}や^{*7}を参考に自ら翻訳を行った.

4.8 日本語評価における BLUE の課題

BLUE スコアは機械翻訳の評価指標の一つである. ただし, 元が英文用の指標であるため, 日本語に対しては適切な評価ができない. そのため, 日本語の評価を行うためには, 一度形態素解析を行う必要がある. ただし, 日本語への翻訳は特に類義語や意識などの変化が考えられるために, BLUE スコア以外にも人間による評価などの評価指標を設ける必要がある.

4.9 LLM とトークナイザーの関係

本研究では ChatGPT を用いて翻訳を行うが, ChatGPT などの大規模言語モデルは直接その言語そのものを理解しているのではなく, 実際に入力や出力された文章と大規模言語モデルの間にトークナイザーが存在する. トークナイザーは実際に入力された文章をルールに従ってトークンに分割し数値に変換, その後テンソルに変換する.

英文などを単純に単語ごとに分解すると語彙のサイズが肥大化し, メモリや計算時間が増加する. また, 文字単位での分割の場合には語彙サイズは大きくはならないもののトークンごとの意味を学習することが難しくなる. そのために “annoyingly” などの珍しい単語を “annoying” と “ly” などに分割し, それぞれの意味の合成として連結するなどの工夫によって語彙サイズを抑えるなどの工夫がある. またこれにより, 未知の単語に対しても既知のトークンに分解することによって理解することができる.

また単純な分割や変換だけではなく, 大規模言語モデルに応じた特殊なトークンや複数の文章の長さを一定に変換するためのトークン, トークンが実際の文章に関係するかの情報, 文章の区切りの情報などを追加する.[3][2]

5 結果

それぞれの実験結果についていくつかの例を示す.

^{*5} <https://eikaiwa.dmm.com/blog/culture-and-community/southern-american-english/>

^{*6} <https://www.urbandictionary.com/define.php?term=no%20zuo%20no%20die>

^{*7} <https://languagelog.ldc.upenn.edu/nll/?p=9686>

表1 スラングに関する実験の結果の例

英文	スラング	スラングの日本語説明	ChatGPT 日本語訳
“what are you doing this summer?” -having a cmbyn summer	a cmbyn summer	夏に悲しくて失恋し、それを楽しむのが大好きな人々のための美しい美学。	「この夏、何をする予定？」 - 「『君の名前で僕を呼んで (Call Me by Your Name)』みたいな夏を過ごしてるよ」
I may not have finished it, but I at least got a Germ-X Percent done.	a Germ-X Percent	99.99% という速くて楽しい方法です。Germ-X の主張からの造語です。多くの一般的な有害な細菌やバクテリアの 99.99% をわずか 15 秒で除去するのに効果的です。	まだ全部終わらせてないけど、少なくとも「ジェル式消毒液」くらいの割合は終わった気がする。
They couldn't handle a little teasing	a little teasing	背の低い友達をからかうとき	彼らはちょっとからかっただけでも耐えられなかった。
Omg!! Is that Aarunya Omg yes! It is her I wish I was beautiful as her!!!	Aarunya	彼女は最高の友達で、とても素晴らしい双子座の女性です。肌は色白です。	「オーマイガーッ！！あれってアールニャだよね！？」「オーマイガーッ！！そうだね、彼女だ！！」「私も彼女みたいに綺麗になりたい！！」

表1 はスラングに関する実験で使用した英文, スラング, スラングの説明の日本語訳, ChatGPT 日本語訳の例を示している. 「a little teasing」や「Aarunya」などはあまり違和感を感じられな
いが, 「a cmbyn summer」や「a Germ-X Percent」はやや日本語として破綻している. 4 つとも
に ChatGPT の日本語訳は元の意図が伝わっているとは言い難い.

表 2 アナグラムに関する実験の結果の例

英文	日本語訳	変換後	ChatGPT 日本語訳
I enjoy walking along the beach at sunset.	私は夕日が沈むビーチを歩くのが好きです。	I ejony wlinkag anlog the baceh at ssenut.	私は夕日が沈むころに海辺を歩くのが好きです。
She always carries a small notebook to jot down her thoughts.	彼女はいつも考えをメモするための小さなノートを持ち歩いています。	She aylwas crieras a salm1 neobtook to jot dwon her thtg-hous.	彼女は考えをメモするために、いつも小さなノートを持ち歩いています。
The train was delayed due to heavy snow in the area.	列車は地域の大雪の影響で遅れました。	The tairn was daeyled due to hveay sonw in the aera.	列車はその地域の大雪のため遅れました。
He spends hours studying the stars with his telescope.	彼は望遠鏡で星を観察するのに何時間も費やします。	He snpdes hrous suditnyg the srtas wtih his teeclpsoe.	彼は望遠鏡を使って星を研究するのに何時間も費やします。
The cake was so delicious that everyone asked for seconds.	ケーキがとても美味しくて、みんなおかわりを求めました。	The ckae was so di-iouecls taht eovnyree akesd for scondes.	ケーキがとても美味しかったので、みんなおかわりを求めました。
We woke up early to watch the sunrise from the mountain-top.	山頂から日の出を見るために私たちは早起しました。	We wkoe up erlay to wcath the sisnrue form the maoiuton-ntp.	私たちは山頂から日の出を見るために早起しました。

表 2 はアナグラムに関する実験で使用した英文, 日本語訳, 単語ごとに文字の順序を入れ替えた文章, ChatGPT 日本語訳の例を示している。基本的には日本語訳と ChatGPT による翻訳の違いはなく, 一部に見られる違いも同義語などの細かな違いのため, 十分に元の文章の意図が伝わっていると考えられる。

表 3 リドルに関する実験の結果の例

英文	日本語訳	ChatGPT 日本語訳
If you say it you break it. What is it?	名前を言うと壊れるものはなんだ？	言えは壊れてしまうものは何でしょう？
What letter of the alphabet has the most water?	アルファベットの中で最も水を含んでいる文字はなんだ？	アルファベットの中で最も水を持っている文字はどれでしょう？
What flowers grow under your nose?	鼻の下に育つお花はなんでしょう？	鼻の下で育つ花は何でしょう？
What can you catch but not throw?	キャッチすることはできるけど、投げることができないものはなんだ？	捕まえることはできるけど投げることはできないものは何でしょう？
What is a snake' s favorite subject?	蛇が一番好きな科目はなんだ？	ヘビの大好きな教科は何でしょう？
What starts with T, ends with T, and is full of T?	Tからはじまり T で終わり、中が T でいっぱいなのはなんだ？	T で始まり、T で終わり、中身が T でいっぱいのは何でしょう？

表 3 はリドルに関する実験で使用した英文, 日本語訳, ChatGPT による日本語訳の例を示している。日本語訳, ChatGPT による日本語訳共に細かな差異はあるものの概ね内容は同一であると推測される。ただし, 「If you say it you break it. What is it? 」のように日本語に翻訳しても適切にリドルとして成り立っているものから「What is a snake' s favorite subject?」のように日本語のまま考えていても答えが出ないものなど, 大元のリドルによってそのまま活用できるかに関しては違いが感じられる。

表 4 混成語に関する実験の結果の例

英文	日本語訳	ChatGPT 日本語訳
'Twas brillig, and the slithy toves Did gyre and gimble in the wabe; All mimsy were the borogoves, And the mome raths outgrabe. “That’ s enough to begin with,” Humpty Dumpty interrupted: “there are plenty of hard words there. ‘Brillig’ means four o’ clock in the afternoon—the time when you begin broiling things for dinner.” “That’ ll do very well,” said Alice: “and ‘slithy’ ?” “Well, ‘slithy’ means ‘lithe and slimy.’ ‘Lithe’ is the same as ‘active.’ You see it’ s like a portmanteau —there are two meanings packed up into one word.”	そはやきに時 めなやかな 濤蔀ら、にもずをじゃいり、 錐めく。 皆みじろい、檻樓濠蕪ら、い えかはな拉子ら、わめしゃ めく。 「まずはそこまででじゅぶん だ」ハンプティ・ダンプティ が止めました。「難しい言葉 がたくさんある。『やきに時』 とは午後四時のこと——夕飯 のために色々焼き物や煮物を 始める時だ。」 「なるほど、よくわかります」 とアリス。「では『めなやか な』は？」 「うむ、『めなやかな』は、『し なやか』——『しなやか』は 『活発に動いている』という 意味だが——かつ『ぬるぬる している』ということだ。つ まり、旅行カバンみたいなも んでね、二つの意味がひとつ の言葉におさまっているわけ だ。」	「あたかもブリリグだった、 そして滑らかでぬめぬめした トープたちが草むらの中で回 りくねり、穴をあけた。 全てのボロゴーフたちはひど くみすぼらしく、そしてモー ム・ラスたちは奇妙な音を立 てた。」 「そこまでで十分だ」とハン プティ・ダンプティが話に割 り込んできた。「そこには難 しい言葉がたっぷりあるよ。 『ブリリグ』ってのは午後 4 時のことで、その時間に夕食 の準備を始めるんだ。」 「なるほど、それはわかりや すいね」とアリス。「じゃあ、 『スライシー』って？」 「『スライシー』は『柔らかく てぬめぬめしている』という 意味だよ。『柔らかい』とい うのは『活発な』という意味 と同じだ。つまりこれは『か ばん語』なんだ——2 つの意 味が 1 つの言葉に詰め込まれ ているわけだ。」

表 4 は混成語に関する実験で使用した英文, 日本語訳, ChatGPT 日本語訳を一部抜粋したものを示している。この表では日本語訳と ChatGPT による翻訳に大きな違いが見られる。特に詩の部分である

'Twas brillig, and the slithy toves Did gyre and gimble in the wabe; All mimsy were the borogoves, And the mome raths outgrabe.

とその中に含まれる混成語自体の翻訳には大きな違いが見られる。これは日本語訳として参照した

『鏡の国のアリス』の翻訳スタイルがなるべく元の文章の意図を保持するように翻訳されているためであると考えられる。

表5 パロディに関する実験の結果の例

英文	日本語訳	ChatGPT 日本語訳
Those Were the Days	兵どもが夢の跡	あの頃はよかった
Withering Heights	荒が丘	ウィザリングの頂き
Crafting a New Look	おニューの衣装	新しい見た目をクラフトする
Isn't It Iron Pick	鉄のツルハシで決まり	鉄のツルハシ、最高でしょ？
We Need to Go Deeper	さらなる深みへ	もっと深く潜らなくちゃ
Spooky Scary Skeleton	不気味で怖いスケルトン	不気味で恐ろしい骸骨
The Next Generation	ザ・ネクストジェネレーション	次世代
Not Today, Thank You	今日はやめておきます	今日じゃないよ、ありがとう
Ol' Betsy	おてんば	古き良きベッツィ
Adventuring Time	冒険の時間	冒険の時間
Very Very Frightening	とてもとても恐ろしい	とてもとても怖い
Minecraft: Trial(s) Edition	Minecraft: トライアル版	マインクラフト：お試し版
Bee Our Guest	秘密の晩餐会	蜂をゲストに
Bukkit Bukkit	おけおけ	バケツ、バケツ
Whatever Floats Your Goat!	あなたのヤギたい様に！	ヤギが浮かべばそれでいい！
Total Beelocation	綿蜜に引越し	完全なるミツバチの移動

表5はパロディに関する実験で使用した英文、日本語訳、ChatGPTによる日本語訳の例を示している。「Spooky Scary Skeleton」や「Adventuring Time」などはほとんどそのまま類義語である。「Those Were the Days」などは一見すると日本語訳と ChatGPT の日本語で異なる結果に見えるが、「兵どもが夢の跡」の意味としてコトバンク*8では

あるできごとのあった現場のようすが、すっかり変わってしまったことのたとえ

となっているため、基本的には類義語であると推測される。ただし、「Ol' Betsy」や「Bee Our Guest」などは日本語役と ChatGPT による翻訳でかなりの違いが見られる。

*8 <https://kotobank.jp/word/つわものどもが夢の跡-2236731>

表 6 方言に関する実験の結果の例

英文	日本語訳	ChatGPT 日本語訳
good good Study, day day up	よく勉強し、毎日進歩しなさい。	よくよく勉強して、日に日に進歩しよう。
No zuo no die	自業自得	自分から面倒を起こさなければ問題は起きない。
bubba	兄弟	お兄ちゃん
cattywampus	斜めっている、調子が悪い	斜めになった、歪んだ、不揃いな。
y' all	みんな	みなさん、あなたたち

表 6 は方言に関する実験で使用した英文, 日本語訳, ChatGPT 日本語訳の例を示している. 表の上 2 つは中国語英語で用いられる言葉で, 下 3 つはアメリカの南部の方言である. 共に標準語的な英語とは異なる表記ではあるものの, ChatGPT の翻訳は概ね日本語訳と同義語であると感じられる.

6 考察

6.1 スラングに関する実験

スラングに関する実験ではデータセットを利用した結果, ChatGPT の利用規定に引っかかるという問題が発生した. これは実験当初想定していなかった問題で, スラングあるいは例文が ChatGPT にとっては不適切な内容を含んでいたためと推測される. そのため, 今回の実験では問題となりそうな文に関しては実験を行わないこととした. ただし, 実際には翻訳のために利用しようとして同じような問題が発生した場合に, その箇所だけは翻訳を行わないなどといった運用は不可能であるために, そういった内容に関しても適切な翻訳を行う大規模言語モデルの利用の検討をするべきである. また, 結果に関しては特に aadhik や aaliyahism などの造語的な単語であるスラングに関してはカタカナ語への翻訳が多く見られた. これにより, ChatGPT は特に英語から日本語への翻訳に関してはその単語に関する知識を持っていない場合には英語の綴りから推測される発音をもとに適当なカタカナ語への変換を行なっている可能性が考えられる. つまり, これらのスラングに関しては ChatGPT 側の知識が存在しないために本来のスラングとしての意図を適切に伝えることは困難である. またいくつかの ChatGPT の翻訳に関して疑問が発生したため, 対応する単語を Google にて検索を行った. 例えば「ジェル式消毒液」の場合は原文が「Germ-X」でありこれを Google で検索すると図 1 のような商品が複数ヒットした.

おそらくはこの商品が「Germ-X」であり, 図 1 の右下の画像からこれが消毒用アルコールのような商品ではないかと推測される. データセットにはこのスラングは「Germ-X」をもとに作られた造語であるとあるためにおそらくはこの商品名がスラングの元であると推察される. つまり

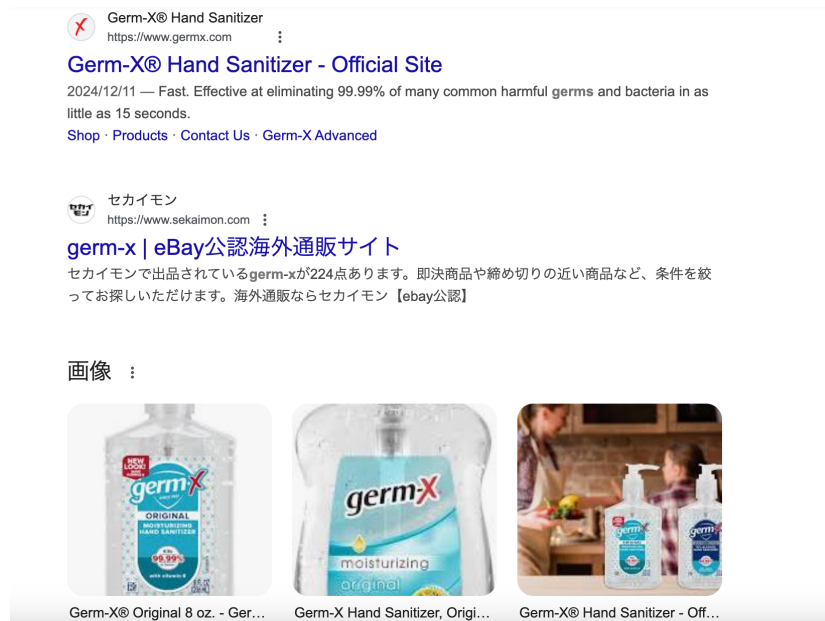


図1 GoogleでGerm-Xと検索を行った結果

「Germ-X」を「ジェル式消毒液」と翻訳しているということは「Germ-X」自体の学習は行っていたものの、それを元にしたスラングに関して学習が不足しているなどが理由として考えられる。このような場合、スラングの元のニュアンスを正確に伝えきれていないどころか人間でも十分に間違えうるものの、文章の意味が通らないレベルには間違った翻訳を行なってしまっている。

6.2 アナグラムに関する実験

この実験は関連研究 [1] で言われているようにアナグラムのような構造になっている文章に関して特に高い対応力があるように感じられた。また関連研究での実験に用いられたのは GPT-4 であり ChatGPT (GPT-4o) ではないが、少なくともこの実験の範疇ではそのバージョンの違いによる差は感じられなかった。つまりこれらの実験から、ChatGPT にとっては誤字脱字あるいはアナグラムなどの発生した未知の単語であってもおそらくは文字の内容などから大元の単語を推測することが可能であると推測される。

これは大規模言語モデルが単語単位ではなくトークナイザーで分割されたトークンに対しての処理を行っていることも関係するだろうが、文章を先頭から逐次理解していくのではなく文章を読み取る際に全体あるいは複数先のトークンもまとめて読み取っていると仮定される。

また図2は OpenAI が公開している GPT-4o で使用されているトークナイザーでのトークン化の例である。この図のように ChatGPT などで使用されているトークナイザーであっても日本語などのトークナイザーによる語彙が網羅されていない場合には、「宰」が「\xe5\xae」と「\xb0」に分割されたように一つの単語が複数のトークンに分割されることがある。ただし、それらの場合においても大規模言語モデルは問題なくその文を読み取ることが可能であるため、同様のことがこの

単語ごとに文字の順序を入れ替えた場合も起こっていると言える。



```
走れメロス 是太宰治の短編小説です。  
['走', 'れ', 'メ', 'ロ', 'ス', 'は', '太', '宰', '治', 'の', '短', '編', '小', '説', 'で', 'す', '。']
```

図2 トークナイザーに網羅されていない日本語を分割させ、utf-8 でエンコードした例

これらのことから少なくとも単語単位であればトークナイザーによる適切なトークン化が行われるかにかかわらず、大規模言語モデルは複数のトークンを組み合わせることでその元の語彙を推測することが可能である。これは一般的な文章においては軽い誤字脱字等を無視してくれると推測されるために便利かもしれないが、今回の実験のような場合では歴史的な誤字、意図した同一発音への単語の置き換え等を意図した文章であってもそのような場合を考慮できずに翻訳を行ってしまう可能性が考えられる。

ただし、今回の考察により発生した文中の単語を一部同音異義語に置き換えても元の文章を復元することが可能なのかどうかに関しては今回の内容にも関わってくるために今後研究する必要があると感じられる。

6.3 リドルに関する実験

リドルに関する実験では全体的に特殊な変化は見られなかった。そのため、日本語にも対応する語彙ベースがある場合には直接翻訳を行っても問題は感じられなかったが、英語特有の内容を聞かれている場合にはそのままの翻訳では問題があるように感じられた。例えば、アルファベットの中で最も水を含んでいる文字はなんだ？と聞かれ、適切な回答を求めるには 1. これが英語のリドルであること、2. Sea という英単語を知っていること、3. Sea が C と同音であることを知っていることが前提となっている。単純に和訳されたリドルが英語の知識を要求するかは個々の事例においてバラバラだが、英単語としては比較的簡単であると考えることができる。キャッチすることはできるけど、投げることができないものはなんだ？の場合には前提として風邪をひくという日本語を英語では catch a cold ということを知っているかが重要になってくる。このように英語におけるリドルは元の文化圏の言語に依存することが多いため、単純な和訳だけでは日本語のなぞなぞとしては一切成立せず、これが英語から日本語に翻訳されたものであるという前提のもとに英語でリドルを得必要性が発生する。このような場合には、日本語でもそのままなぞなぞが解けるように改変すると元の意図を残したままの適切な翻訳が行えるが、指示自体は一貫して翻訳してくださいのみを入力しているためかリドルとして成り立つかではなく、ただの英文としての翻訳を行っていると考えられる。実際に、リドル自体には英文には不思議な表記などを行うところはなく、基本的には自然な表記であるため、ChatGPT もやはりそのまま自然に翻訳を行ってしまうのではないかと推測される。

6.4 混成語に関する実験

今回の実験では混成語はカタカナ語発音の結合ではない訳が用いられている鏡の国のアリスを採用している。実験結果としては直接カタカナ語に翻訳した結果が多かったが、「slithy」のように文章上原文では混成語、その後に混成語とその意味の説明の構造になっているところを意味の説明、混成語と意味の説明のように翻訳されてしまった箇所が存在した。これによりその訳を見ただけでは突然混成語が出てきたように見える可能性がある。ただし、原文の1回目の混成語登場時点ではその混成語の意味は説明されていないため、ChatGPTは完全に先頭から確率的に文章の翻訳を行っているだけでなく、あとから文章後半で登場した単語を文章前半から探し出して置き換え直した可能性も考えられる。

ChatGPTのサービス自体には特定の内容の禁止などの条件が課せられており、そういった内容は表示されないように規制されているため、そのように文章全体を再評価している可能性は十分に考えられる。また今回の実験の内容には含んでいなかったが、カタカナ語ベースでの混成語、例えば ブレイクファースト と ランチ の組み合わせである ブランチ などは今回の実験結果から翻訳が行えると推測できるかもしれない。

6.5 パロディに関する実験

基本的には日本語訳と ChatGPT の翻訳の内容自体には基本的に大きな違いが見られない。だが、特に日本語訳にはいくつか特徴的な翻訳がされている。この理由は ChatGPT はこの文章が表示されるタイミングを把握していないが、日本語訳の翻訳を行う際にはそのタイミングが考慮されていることであると言える。例えば「Those Were the Days」という英文はそのまま翻訳を行うと ChatGPT の翻訳通り「あの頃は良かった」*⁹になる。実際のゲーム上ではボロボロになった砦を発見することでこの文章が表示される。そのために、元はボロボロになっていない状態の砦があったのであろうことを考慮し、この翻訳が採用されているのであると推測される。他にも例えば秘密の晩餐会という文章の場合、この文章の表示タイミングは蜂に関係するものである。そして英文は「Bee Our Guest」となっているため、この Bee の部分が誤字などではなく意図的に蜂のことを指していると考えられる。そのため「Be Our Guest」の訳である「ひとりぼっちの晩餐会」の訳を改変し、「密」を「蜜」に置き換えていると推測される。

ただ、今回の翻訳時にはそのような背景までは文章だけでは読みとることが難しく、結果として単純な文章の翻訳になってしまったと考えられる。また、他と明確に違って見える翻訳に「Ol'betsy」の訳が存在する。この言葉で単純に Google で翻訳を行っても「オールベッツィー」となり、つまりは他に適切な日本語訳が見つけれなかったのであろうことが推察できる。ただし、この英文の日本語訳を検討する際にはこの英文が映画『ロビン・フッド』に登場するキャラクターに因んでいることが示唆されている。そのために日本語訳が映画のそのシーンからおてんばと翻訳されていると

*⁹ <https://ejje.weblio.jp/content/those+were+the+days>

推測される。

これらのことから ChatGPT は文章自体を翻訳する際の意味自体は基本的には適切ではあるものの、特に短文の場合にはその文章の由来を推察することが困難な場合があり、そういった場合には由来などを理解せずに直接的な翻訳を行ってることが考えられる。また「Bukkit Bukkit」や「Bee Our Guest」などの翻訳からみて、単純に翻訳だけを行わせる場合には言葉遊び的な側面は重視されないように感じられる。

6.6 方言に関する実験

方言に関する実験では ChatGPT は適切な翻訳を行なったが、これは単に ChatGPT が学習済みの語彙の中に含まれていたという要因が強いと考えられる。そのため今回利用したある程度有名な方言よりも、よりマイナーな方言に関しては ChatGPT は適切な翻訳を行えない可能性が十分にある。

6.7 複数の実験をまとめた考察

スラングや混成語などの結果をもとに ChatGPT は特に学習を行えていない新出語句に関しては知識が存在しないため、意味ベースの翻訳ではなく発音ベースでカタカナ語への翻訳が行われる可能性が考えられる。そのため、そのような発音ベースのカタカナ語の登場をキーとして大規模言語モデルのその語彙の学習の有無を判断し、翻訳を行う際の間違った翻訳であるかどうかの判断を行うことが可能かもしれない。

また [7] では言語変種という概念においてはこの実験でいう方言を

地域ごとに異なる地域方言 (regional dialect)

スラングを

社会の異なるグループによって使われる、社会方言 (social dialect)

と捉えることができる。そのような場合、リドルやアナグラムのような文章や単語とは異なり、どちらも特定の単語が特定の言語から変化したものだと考えることができるので、ほぼ同一のものだと考えることができる。そのため、スラングと方言では同じような翻訳パターンが発生してもおかしくない。今回の実験ではスラングの翻訳は失敗に終わったが、方言の翻訳は成功していたことに関しては実験したデータによる違いが大きいと考えられる。方言はデータを集める際に検索などを行なった一般的な用語を用いたが、スラングは既存のデータセットを用いたために、大規模言語モデルへの学習の差がある程度激しいことが推測される。また、一般に地域方言と異なり研究 [4] で述べられている通り、スラングなどの社会方言はインターネットの普及によって変化が激しく、大規模言語モデルの学習タイミングによってはその単語や文章を学習していない可能性が高い。

7 今後の課題

一般大規模言語モデルによる文章の生成などを検証する際には BLUE などの定量的な指標を用いることが用いられてきたが, 今回の実験のリドルや混成語などのより単なる日本語訳だけではなく, 適切な類義語への翻訳が求められる場合には特に BLUE では正確に評価を行うことが難しいと考えられる. そのため, 今後は複数のモデルを用いて 2 つの文章の意味が一致しているかどうかを評価するなどのよりファジーな評価を行うことが必要である. ただ, モデルによる評価の場合, BLUE よりも定量的な評価を行うことは難しいと推測されるため, 新たな問題が発生する可能性が高い.

また今回は一般性の観点や先行研究に基づき, ChatGPT での実験を行ったが ChatGPT は海外で開発されたモデルであるため, 英語などの言語と比べると ChatGPT は日本語ネイティブではないと言えるが. 今回は特に日本語への翻訳へ焦点を当てた内容であるために生成する日本語の精度は大規模言語モデルの日本語の学習量に大きく影響を受ける. そのため, より日本語での学習を行い日本語表現を得意とするモデルでの実験を行う必要性がある. ただし, 日本で作成されている大規模言語モデルは日本語への理解度は ChatGPT などより高いはずだが, 肝心の翻訳元の言語における文化的背景なども考慮する必要性があることを考えると日本語だけではなく, 他の言語についてもより深い学習が求められるため, より適切な大規模言語モデルを模索する必要がある. また, 今回は英語から日本語への翻訳のみを行ったが, 昨今では原神, アークナイツなどの中国が開発したゲームも日本で多く見かける. そのため, 英語から日本語だけではなく中国語から日本語などの場合について同様に実験を行う必要性がある.

参考文献

- [1] Q. Cao, T. Kojima, Y. Matsuo, and Y. Iwasawa. Unnatural error correction: Gpt-4 can almost perfectly handle unnatural scrambled text, November 2023.
- [2] HuggingFace. Summary of the tokenizers. https://huggingface.co/docs/transformers/ja/tokenizer_summary, 2025-01-27.
- [3] HuggingFace. Tokenizer. <https://huggingface.co/docs/transformers/ja/preprocessing>, 2025-01-27.
- [4] L. Mei, S. Liu, Y. Wang, B. Bi, and X. Cheng. Slang: New concept comprehension of large language models, October 2024.
- [5] S. Shen, L. Logeswaran, M. Lee, H. Lee, S. Poria, and R. Mihalcea. Understanding the capabilities and limitations of large language models for cultural commonsense, May 2024.
- [6] 河合 祥一郎 (翻訳). 鏡の国のアリス. 角川文庫, 2010. 原タイトル: Through the Looking-Glass, and What Alice Found There 原著者名: Lewis Carroll 原出版社: Macmillan 原著出版年: 1871.

- [7] 鈴木恵子. 英語の多様性 –変種をどうとらえるか–. <https://www.shizuoka-eiwa.ac.jp/se-info/wp-content/uploads/2023/09/F-6> 紀要 – 第 8 号 – 英語の多様性 – 変種をどうとらえるか – (鈴木恵子).pdf ,2025-01-24.