

令和 6 年度茨城大学工学部情報工学科  
卒業研究論文  
外部知識を利用する LLM  
による制約付き MT

所属 情報工学科

著者 石野稜馬 (21T4010R)

指導教員 新納浩幸教授

令和 7 年 2 月 4 日 (火)

令和 6 年度茨城大学工学部情報工学科 卒業研究論文

## 外部知識を利用する LLM による制約付き MT

著者

石野稜馬 (21T4010R)

指導教員

新納浩幸教授

### 論文要旨

近年、他言語の人とのコミュニケーションや文章に触れる機会が増加している傾向にある。それに伴い、翻訳機能を必要とする状況も増加している。通訳がその場になくとも、所持している携帯電話などの電子端末を用いた機械翻訳により円滑に交流することが可能である。現在では、Web 上から利用することが可能な機械翻訳のシステムが多種多様に存在する。日々発展し精度が向上している機械翻訳だが、その一方で、科学論文などの専門用語の翻訳や、特定の固有名詞などの翻訳では、用語の専門性の高さや、住んでいる地域の差異などによって想定している翻訳とは違う意味として出力されてしまうという課題がある。そのため制約語彙を含む文の翻訳を正確に行えることが求められている。本論文では、制約語彙を含む文の翻訳の精度向上を目的とし、これまで自然言語処理の機械翻訳分野において提案されていた手法である、統計的機械翻訳を用いたシステムに制約語彙を参照させ翻訳を行う手法とは代わり、LLM に外部知識として制約語彙リストを参照させ、制約付き機械翻訳を行う手法を提案する。本研究では、Langchain の Agents 機能を用いて、原文を入力した際に、Agents が LLM に制約語彙リストと原文を渡して出力を行うというプログラムを作成した。「制約語彙リストを参照させて機械翻訳を行う」「制約語彙リストを参照させずに機械翻訳を行う」「プロンプト内で直接制約語彙の変換を指示して機械翻訳を行う」といった 3 種類の条件で、同じ入力をした際に出力結果にどのような差が生じるかという実験を行った。実験の結果、制約語彙リストを参照させる方が翻訳の精度が向上した。LLM に翻訳を任せているため、LLM の質やプロンプトの質によっても翻訳の精度に差が出ることも判明した。

# 目次

<b>第 1 章</b>	<b>序論</b>	<b>5</b>
1.1	背景 . . . . .	5
1.2	目的 . . . . .	5
<b>第 2 章</b>	<b>事前知識</b>	<b>6</b>
2.1	日本語を含む言語対における制約付き翻訳タスクの近況 . . . . .	6
2.2	準備 . . . . .	10
<b>第 3 章</b>	<b>提案手法</b>	<b>13</b>
<b>第 4 章</b>	<b>実装方法</b>	<b>14</b>
4.1	プログラムの方向性 . . . . .	14
4.2	使用技術 . . . . .	14
4.3	プログラム概要 . . . . .	17
<b>第 5 章</b>	<b>実験</b>	<b>19</b>
5.1	研究方法 . . . . .	19
5.2	実験結果 . . . . .	20
<b>第 6 章</b>	<b>考察</b>	<b>22</b>
6.1	実験前推測 . . . . .	22
6.2	考察 . . . . .	22
<b>第 7 章</b>	<b>結論</b>	<b>24</b>
	<b>参考文献</b>	<b>26</b>

目次	4
付録	27
A    プログラム . . . . .	27

# 第1章

## 序論

### 1.1 背景

近年、他国の人とのコミュニケーションや、普段使用していない言語の文章を読むといった、翻訳機能を必要とする場面が日常的に増加している傾向にある。この傾向に伴い、翻訳も様々な形で発展している。他国の人とのコミュニケーションでは、通訳という選択肢もあるが、近年では、自身の所有している携帯電話などの電子端末を用いた機械翻訳 (Machine Translation、以下 MT と略す) による円滑なコミュニケーションを行う場面の方が日常的である。MT とは、人が翻訳を行うのではなく、その名の通りコンピュータが人工知能 (Artificial Intelligence、以下 AI と略す) を利用し、自動で行う翻訳の名称である。日々改良・発展され精度が向上している MT だが、科学論文などの専門用語の翻訳では、想定している翻訳とは違う意味で出力されてしまうことがある。また同様に、特定の固有名詞の翻訳では、住んでいる地域などによって翻訳に違いが生じ、想定している翻訳とは違う語彙で出力されてしまうことがある。これらの語彙は制約語彙と呼ばれ、MT の分野の問題のひとつとなっており、制約語彙を含む文の翻訳を正確に行えることが求められる。

### 1.2 目的

本研究では、外部知識として制約語彙のリストを用意し、このリストと制約語彙を含む原文を大規模言語モデル (Large Language Models、以下 LLM と略す) に渡すことで、翻訳の精度を向上させることが目的である。

## 第2章

# 事前知識

### 2.1 日本語を含む言語対における制約付き翻訳タスクの近況

この関連研究は、日本語を含む言語対における制約付き翻訳タスクの近況をまとめられたものである。本研究と同じように、専門用語や特定の固有名詞といった制約語彙を適切に翻訳したいという背景がある。こちらの関連研究では、2021年の機械翻訳共有タスクで提出されたシステムでも、「用語」に関するエラーが一定数出現したことが述べられている。この関連研究で言及されていることは3つある。

#### 2.1.1 制約付き翻訳共有タスクの設立

1つ目は制約付き翻訳共有タスクの設立である。この共有タスクは ASPEC 科学論文ドメインデータを元に再設計された。ASPEC 科学論文ドメインデータを用いた理由として、科学論文を参考にしているため専門用語が多く使用されているからである。タスク内容は、与えられた制約語彙に従い、文を翻訳するというものである。共有タスクを設計するのは、様々なモデルを、同じデータセット・同じ条件下で比較可能である統一的な評価基盤を用意する必要があるからである。これにより、現時点で最も優れているモデルが何であるかを評価することができ、また最も優れたモデルで解くことが出来ない課題が何かを明白にすることができる。また共有タスクを予め用意しておくことで、新規参入がしやすく技術促進に繋がるという狙いもある。共有タスクを設立する上で必要な準備は2つある。

1つ目は図 2.1 のように、元データに対する「制約語彙リスト」を作成する必要がある。

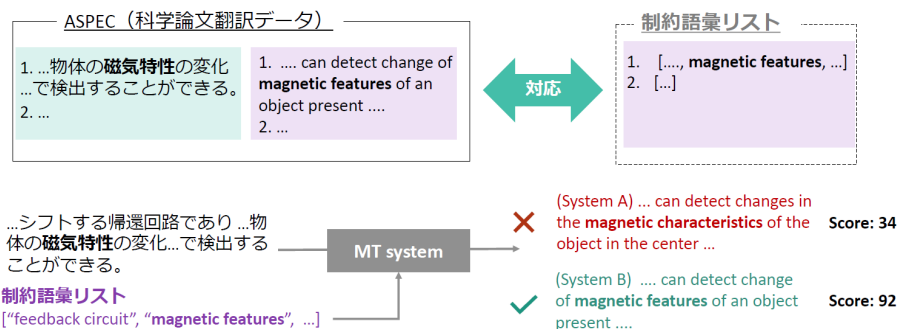


図 2.1: 制約語彙リストの作成

2つ目は評価指標の策定を行う必要がある。

評価のために、図 2.2 のように“制約語彙”の正解を文単位でアノテーションする必要がある。

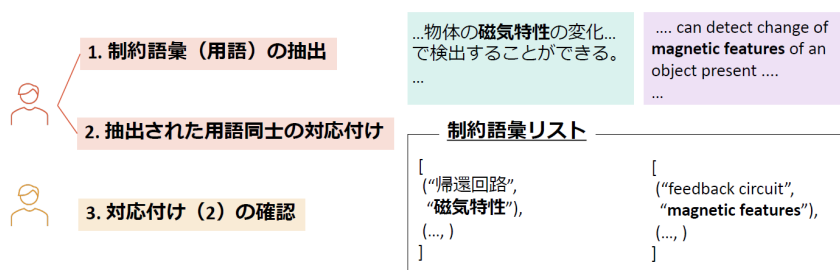


図 2.2: 2021 年のアノテーションの方法

2021 年に行われた際、全工程を人手作業で行っていたため、高コストであった。これを踏まえて 2022 年には図 2.3 のように制約語彙の抽出を自動抽出に変更してアノテーションが行われた。

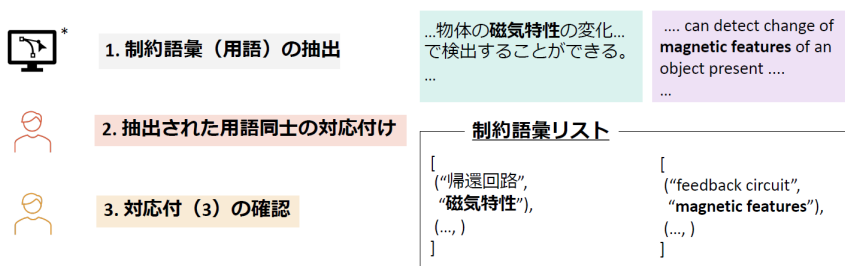


図 2.3: 2022 年のアノテーションの方法

しかし、自動対応付けはうまくいかず、依然として高コストのままであった。続いて、「翻訳の質」と「用語の一貫性」の両方を考慮した上で自動評価指標の策定を行った。図

2.4のように、制約語彙が含まれていない例をフィルタリングし、参照訳との overlap を計測し BLEU スコアで評価を行った。しかし、自動翻訳だけでは完全には翻訳の質を測ることは困難であるため、人手評価も行った。その方法は、評価者に、元文とシステム出力文を提示して点数を付与して評価するという形である。

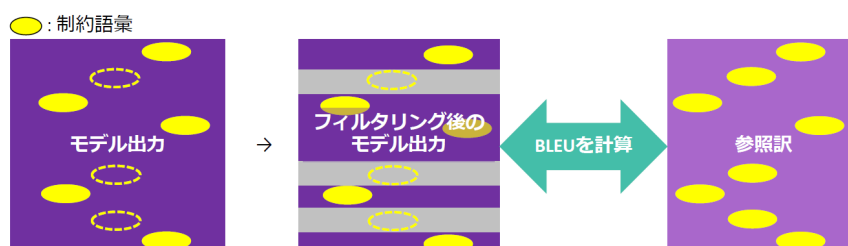


図 2.4: 自動評価指標の策定

### 2.1.2 制約付き翻訳モデルの紹介

この関連研究では、2つの制約付き翻訳モデルを紹介する。1つ目は「緩い」制約モデルである。まずモデル学習時に、以下の形式でデータを学習させる。

- input 元文 +  $\langle sep \rangle$  + 制約語彙 1 +  $\langle sep \rangle$  + 制約語彙 2...
- output 翻訳文

学習を通じて、入力に付加された制約語彙を含むように仕向ける。明示的なルール・制約のための機構を組み込んではいないため、必ずしも制約語彙を含むとは限らないという注意点があるモデルである。

2つ目は「厳しい」制約モデルである。デコーディング時、出力探索中に制約語彙に相当するトークン列を強制的に出力させるため、モデルの翻訳の質が下がることがあるモデルである。

この関連研究では、これら2つの制約モデルを応用したシステムを2つ提出している。1つ目は「緩い」制約モデルと「厳しい」制約モデルを組み合わせたものであり、事前に「緩い」制約モデルである程度制約語彙が探索中に出現するモデルにし、「厳しい」制約モデルによる翻訳の質の低下を抑えるシステムである。2つ目は「緩い」制約モデルに「後編集」を組み合わせたモデルであり、制約語彙を含まない可能性のある「緩い」制約モデルに対し、制約語彙を必ず含むように「後編集」モデルを導入したシステムであ

る [1] [2] [3]。

### 2.1.3 共有タスクの結果と今後の課題

この関連研究では、図 2.5 の通り結果が得られた。

年	提出システム	En-Ja		Ja-En	
		自動評価 (Exact+BLEU)	人手評価 (src-based DA)	自動評価 (Exact+BLEU)	人手評価 (src-based DA)
2021	<b>Hard&amp;Soft const (ensemble) (NTT)</b>	<b>57.2</b>	<b>77.5</b>	<b>44.1</b>	<b>75.6</b>
	Soft const (NHK)	33.9	74.1	37.5	73.9
	Hard&Soft const (NICT)	28.8	73.6	31.8	72.1
	Hard const (TMU)	-	-	22.6	50.2
	(human ref.)	-	73.4	-	74.1
2022	Soft const+Post-edit (ensemble) (TMU)	52.7	76.4	40.8	74.1
	Soft const+Post-edit (TMU)	50.5	76.6	38.1	72.0
	Soft const (TMU)	37.6	74.9	23.0	73.3
	(human ref.)	-	76.6	-	74.7

図 2.5: 共有タスクの結果

最もスコアが高かったモデルは、「緩い」&「厳しい」制約モデル (Hard&Soft const (ensemble)(NTT)) であった。厳しい制約モデルであるため、制約語彙を 100% 出力可能であり、自動評価・人手評価ともに非常に高い質で翻訳を出力可能であった。

次にスコアが高かったモデルは、「緩い」制約 + 「後編集」モデル (Soft const+Post-edit(ensemble) (TMU)) であった。「後編集」の補完によって制約語彙を 100% 出力可能であり、一般的な「厳しい」制約モデルより早い (0.115 秒/文) と推測される。

自動評価	相関値
⇔ BLEU	0.795
⇔ Exact match	0.498
⇔ <b>Ours (Exact+BLEU)</b>	<b>0.836</b>

図 2.6: 自動評価のスコア

また、図 2.6 のような結果が得られたため、考案した自動評価 (Exact+BLEU) は人

手評価との相関が高いことが分かり、これにより「用語の一貫性」(Exact match)を考慮することの重要性が示される。

最高性能モデルであっても低い人手評価スコアの例を確認することが出来た。推測できる課題が大きく分けて、「制約語彙に含まれていない「用語」の欠落(データアノテーションの質)」「制約語彙に含まれない誤訳(翻訳モデルの質)」「アノテータのドメイン知識の不足(人手評価の質)」「ASPEC 参照訳の問題(元のデータの質)」の4つが挙げられている。

## 2.2 準備

本研究を行う上で必要となる前提知識は以下の通りである。

### 2.2.1 機械学習

機械学習とは、大量のデータに対して、どのようなパターンがあるのかを学習させ、機械が任意のタスクを処理出来るようにすることである。機械学習は様々な分野で利用されており、様々なデータからどのような傾向があるのか分析を行う「分析」、顔認識などの機能に用いている、「認識」、自動運転などの「制御」、画像生成や音声生成といった「生成」などで利用されている。学習の方法は大きく分けて「教師あり学習」「教師無し学習」「強化学習」などの手法がある。機械学習の手法の一つにニューラルネットワークというものがある。

ニューラルネットワークとは、機械学習の手法の一つであり、人間の脳内に存在する神経細胞(ニューロン)を数理モデル化したものである。入力層、隠れ層(中間層)、出力層から成っており、ニューロンを模したものに与えられたデータを基に新しい概念を学習することが可能となる。そして、その学習を繰り返すことによって最適なモデルを構築していくことを目的としている。ニューラルネットワークをより多層にし複雑化し、この構成過程全体のことを称したものをディープラーニングという。ディープラーニングの目的は誤差関数の値を最小化するような重みパラメータを見つけ出すことである。

## 2.2.2 LLM

LLM とは、膨大な「データ量」や「パラメータ数」「計算量」によって構築された言語モデルのことであり、様々なモデルが提案されている。大規模言語モデルはいくつか提案されており、そのうちの 1 つとして GPT-3 というモデルがある。

GPT-3 (Generative Pre-trained Transformer 3) は、OpenAI によって開発された大規模な言語モデルであり、1750 億個のパラメータを所持している。GPT-3 の基になっている言語モデルである GPT と同じく Transformer の構造を用いているモデルであり、事前に大量のテキストデータで学習され、様々な自然言語処理タスクに対応できる。GPT-3 はゼロショットとワンショットの設定で有望な結果を達成し、少数ショットの設定では最先端のモデルと競合、あるいはそれを超えることもあるという実績を持っている。本研究で使った LLM のモデルはこの GPT-3 の改良版である GPT-3.5turbo である。

LLM を用いることで、文章の生成や情報の抽出、プログラミング、翻訳など様々なことを行うことができる [4] [5] [6] [7]。

## 2.2.3 Langchain

LangChain は、LLM を用いたアプリケーションの開発を効率的に行うためのライブラリで、Python 言語用に提供されているライブラリである。ライブラリとはプログラミングをする際に高頻度で利用される機能を事前に作成し、呼び出すだけでその機能を利用できるようにしているものである。ライブラリには様々な種類のものが存在するが、本研究はその中の Langchain というライブラリを主軸にプログラムを作成している。

Langchain 中の「LLM」という機能は、名前の通り、言語モデルを呼び出すためのものである。OpenAI などの言語モデルを指定することができ、回答のブレ具合を意味する Temperature や、プロンプトなどのパラメーターを付与し、LLM の API へリクエストを送信するといった形をとっており、LLM の API をプログラムから便利に呼び出すための機能を提供してくれているモジュールである。Temperature の値や、プロンプトの内容によって回答文の生成に違いが生じる。

他には、「Agents」という機能があり、本研究のプログラムの要の機能である。プロンプトに基づいて何を必要とするのかを自動で判断し、用意されているツールを自分

で選択して使用し、タスクを処理してくれるモジュールである。[?]

#### 2.2.4 API

API とは、Application Programming Interface の略称である。アプリケーションやプログラム間でやり取りする際のルールや手段を提供してくれる。アプリが外部のサービスからデータを取得したり、操作をリクエストしたりする際に使用される。主な役割として、「データの提供」「データの取得」「通信の効率上昇」「セキュリティ強化」などが挙げられる。API を利用するには、API キーを予め取得する必要がある。本研究では、OpenAI の API を利用するため、OpenAI の API キーを取得する必要がある。

## 第 3 章

# 提案手法

これまで、自然言語処理の機械翻訳分野の研究において、統計的機械翻訳を用いたシステムに制約語彙リストを参照させ翻訳を行う手法は提案されていた。統計的機械翻訳とは、大量の対訳文データから統計モデルを作成し、確率の高い文を翻訳結果として出力するモデルである。しかし、統計的機械翻訳ではなく LLM に制約付き機械翻訳を行わせる手法は提案されていなかった。そのため本研究では、LLM に外部知識として制約語彙リストを参照させ、制約付き MT を行う手法を提案する。

## 第4章

# 実装方法

### 4.1 プログラムの方向性

プログラムの方向性

Langchain の Agents 機能と OpenAI の API を利用し、原文を入力すると Agents が原文と用意されている制約語彙リスト、LLM を活用し翻訳を行って結果を出力するというもの。

### 4.2 使用技術

・使用技術

使用したモデル：gpt-3.5-turbo。

使用したデータ：アジア学術論文抜粋コーパスデータ。[翻訳文、原文、制約リスト] のセット。今回は英語を日本語に翻訳することを想定した。

表 4.1: アジア学術論文抜粋コーパスデータの対訳リスト (抜粋)

ID	番号	日本語	英語
A-05A0698890	0	DERSソフトウェアを用いて「ふげん発電所」の線量率を詳細に計算できる。	Details of dose rate of "Fugen Power Plant" can be calculated by using DERS software.

A-05A0698890	1	線量率を計算する際の状況の変化, すなわちオブジェクトの変化をVRdoseソフトウェアがDERSに特別なマーカにより知らせる。	The changes of conditions for computation of the dose rate, namely, object changes, are inputted to DERS from VRdose software by using a special marker.
A-05A0698890	2	DERSはこれらの変化に対応して新たな線量率を計算できる。	Responding to these changes DERS can compute new dose rate.
A-05A0698890	3	このソフトウェアのR5バージョンの特徴, 利用マニュアルと設計文書をまとめた。	The characteristics of R5 version of this software, instruction manual, and design document were summarized.

入力：コーパスデータの原文。本研究では43個使用。

制約リストは、制約語彙を辞書にしたファイルを用いて人手作業で作成した。形式は変換したい語彙→変換したあとの語彙。今回は約130組の制約語彙をリストに記載。

表 4.2: 制約語彙の英語辞書ファイル (抜粋)

---

"Fugen Power Plant"

DERS software

dose rate

dose rate

VRdose software

object changes

DERS

DERS

---

dose rate

software

R5 version

表 4.3: 制約語彙の日本語辞書ファイル (抜粋)

---

線量率
DERSソフトウェア
「ふげん発電所」
VRdoseソフトウェア
線量率
オブジェクトの変化
DERS
線量率
DERS
R5バージョン
ソフトウェア
増幅器
コイル
位相シフト回路
感知
出力
位相シフト磁気センサーシステム

---

表 4.4: 作成した制約リスト (抜粋)

---

DERS software	→	DERSソフトウェア
dose rate	→	線量率
VRdose software	→	VRdoseソフトウェア
object changes	→	オブジェクトの変化

---

DERS	→	D E R S
software	→	ソフトウェア
R5 version	→	R 5バージョン

---

原文と制約リストを Agents に渡して翻訳を行ってもらおう。

出力：変換、翻訳が適用された文。

### 4.3 プログラム概要

本研究では、3つプログラムを作成したが、Langchain の Agents 機能と OpenAI の API を利用し、原文を入力すると Agents が原文と用意されている制約語彙リスト、LLM を活用し翻訳を行って結果を出力するという点は統一されている。それぞれのプログラムには、翻訳する際に異なる条件を設定する。

- 1つ目のプログラムは、制約語彙リストを Agents が参照して機械翻訳を行う。
  - プログラムの流れは、必要なモジュールを import し、続いてあらかじめ取得してある OpenAI の API キーを設定する。続いて制約リストを辞書として定義。このリストには、約 130 組の制約語彙を記述。続いて、文章内の単語やフレーズをこの辞書に従って変換する関数を用意。また、与えられた文章を日本語に翻訳する関数も用意。この翻訳用の関数は、事前に単語やフレーズを変換したのちに翻訳を行うように設定。続いてプロンプトを記述し、指示を与えます。最後に Agents を作成し、Agents を実行したときのコードを記述。
- 2つ目のプログラムは、制約語彙リストを Agents が参照せずに機械翻訳を行う。
  - プログラムの流れは1つ目のプログラムとほぼ同じであるが、こちらのプログラムには制約リストを参照させないため、辞書の定義しない。また、辞書を参照して翻訳前に制約語彙の変換をする必要がないため、プロンプトの指示も制約語彙の変換部分を省略。
- 3つ目のプログラムは、プロンプト内で直接単語ごとの変換を指示して機械翻訳を行う。
  - プログラムの流れは1つ目のプログラムとほぼ同じであるが、プロンプト内

で制約語彙の具体的な変換を指示するという変更点があるため、プロンプト内に追加で記述。辞書の定義や Agents の作成は 1 つ目のプログラムと同じであるため、変更はない。

## 第5章

# 実験

### 5.1 研究方法

本研究では用意したコーパスデータの原文43個を、3つ用意したそれぞれのプログラムに入力として渡し、それぞれの条件でどのような結果が出力されるのかを評価する。

実験結果は表 5.1 のようにまとめる。「リスト有」であるならば、制約語彙リストを Agents が参照して機械翻訳を行うプログラムに原文を入力した際の出力結果をまとめ、「リスト無」では、制約語彙リストを Agents が参照せずに機械翻訳を行うプログラムに原文を入力した際の出力結果をまとめ、「プロンプトで細かく指示した場合」では、プロンプト内で直接単語ごとの変換を指示して機械翻訳を行うプログラムに原文を入力した際の出力結果をまとめる。

表 5.1: 実験結果まとめ方例

○-○文目	
原文：	原文内容
リスト有：	出力結果
リスト無：	出力結果
プロンプトで細かく指示した場合：	出力結果

## 5.2 実験結果

Langchain の Agents 機能を用いて作成した Agents に、OpenAI の API を利用して LLM を使用させ、43 個の原文を翻訳させた。その際、制約語彙リストを Agents に参照させて出力させたもの、制約語彙リストを Agents に参照させずに出力させたもの、プロンプト内で直接単語ごとの変換を指示して出力させたものの 3 つの結果をまとめ、比較した。

結果の例をいくつか取り上げ表 5.2、表 5.3、表 5.4 にまとめた。

表 5.2: 実験結果 1

1-1 文目	
原文：	Details of dose rate of "Fugen Power Plant" can be calculated by using DERS software.
リスト有：	「ふげん発電所」の線量率の詳細は、DERS ソフトウェアを使用して計算することができます。
リスト無：	フーゲン発電所の線量率の詳細は、DERS ソフトウェアを使用して計算できます。
プロンプトで細かく指示した場合：	「ふげん発電所」の線量率の詳細は、DERS ソフトウェアを使用して計算することができます。

表 5.3: 実験結果 2

2-3 文目	
原文：	For the coils, here were used two coils with 6mm of inner diameter and 800 of coil number by inversely connecting.
リスト有：	コイルには、内径 6mm、コイル数 800 のコイルを 2 つ使用し、逆接続されました。
リスト無：	コイルには、内径 6mm のコイル 2 個が逆に接続された状態で 800 個使用されました。

プロンプトで細かく指示した場合： コイルには、内径 6mm のコイルを 2 つ使用し、コイル数は 800 で逆接続されました。

---

表 5.4: 実験結果 3

---

**6-1 文目**

---

原文：	The article introduces technological progresses to fabricate single crystal thin films of insulator, metal and ferroelectrics on a silicon substrate.
リスト有：	記事は、シリコン基板上に強誘電体、金属、絶縁体の単結晶膜を製造するための技術進歩を紹介しています。
リスト無：	この記事では、絶縁体、金属、強誘電体の単結晶薄膜をシリコン基板上に製造するための技術進歩について紹介しています。
プロンプトで細かく指示した場合：	その記事は、シリコン基板上に強誘電体、金属、および絶縁体の単結晶膜を製造するための技術進歩を紹介しています。

---

## 第6章

# 考察

### 6.1 実験前推測

1つ目のプログラム（制約語彙リストを Agents が参照して機械翻訳を行う。）の出力結果を結果1、2つ目のプログラムは（制約語彙リストを Agents が参照せずに機械翻訳を行う。）の出力結果を結果2、3つ目のプログラム（プロンプト内で直接単語ごとの変換を指示して機械翻訳を行う。）の出力結果を結果3とする。実験前の推測では、結果3の翻訳結果が最も精度が良く、次に結果1の翻訳結果、最後に結果2の翻訳結果という順番に翻訳の精度が落ちていくと推測した。

### 6.2 考察

実際の実験結果は、出力結果が43個のうち、前述した推測と合致した出力結果が31個となった。推測と合致した出力結果となった31個のデータ内、18個の出力結果が制約語彙リストを与えただけでも一字一句正しく変換を行うことが出来ていた。残る12個の出力結果では、推測とは異なる結果を残した。その内、翻訳自体を行わず、正しく変換・翻訳を行うことが出来なかったものは1個見受けられた。制約語彙リストを参照している場合と、プロンプトで直接制約語彙の変換を指示している場合とでは、プロンプトで指示している場合の方がより精度が良い傾向にあるが、まれに制約リストを渡して変換している場合の方が正確な場合も見受けられた。これは、利用した LLM の翻訳の質や、プロンプトでの指示の質が影響していると考えられる。

制約語彙は、語彙のつながりが長いものほど、一字一句正しく変換を行えていない場合

が存在した。

～ted などの過去分詞となる語彙は「～された」と変換されてしまうことがほとんどであり、意味としては正しくとも一字一句正しく変換することが出来ない場合が見受けられた。これは制約リストよりも、LLM が持つ翻訳機能による変換が強く影響していると考えられるため、プロンプトでの指示を強化する必要があると考えられる。

数字とアルファベットは全角でリストに記載されていても、半角で出力されることが多い。ただし、プロンプトで指示した場合は、全角で出力してくれる場合の方が多い傾向にあった。しかし半角で出力された例外も見受けられた。

three-dimensional shapes → 「3次元形状」などの数字部分が制約語彙リストを参照していない場合は漢数字になってしまう場合が多い傾向にあった。

phase → 位相と phase plate → 位相板という、ある制約語彙に別の制約語彙が内包されている場合、例でいうなら phase の部分と plate の部分でわけて先に phase を位相に変換してしまっているため、plate 部分を制約語彙の一部だと判別せず、plate → プレートと変換してしまっている場合がある。しかし変換をプロンプトで直接指示している場合は、正確に変換できていた。

これらのことから、外部知識を利用する LLM による制約付き MT において、プロンプトで直接制約語彙をどのように変換するかを指示することがもっとも翻訳の精度が高いと考えられる。次に、制約語彙リストを参照させる翻訳の精度が高いと考えられる。この翻訳の精度は、利用している LLM のモデルの質や、制約語彙リストの内容、制約語彙リストを参照させるタイミング、プロンプトによる指示の内容といった要因が影響していると考えられる。

## 第7章

# 結論

本研究では、Langchain の Agents 機能によって、Agents が OpenAI の API を利用することで、LLM に制約語彙リストと原文を渡して制約付き MT を行い、制約語彙リストを参照した際の翻訳精度にどのような影響があるのかを調査した。その結果、制約語彙リストを参照させることで、LLM に原文を翻訳させるよりも、制約付き MT の精度が向上することが認められた。また制約語彙リストを参照させるだけよりも、プロンプト内で制約語彙の変換を指示する方が翻訳の精度が向上することが認められた。LLM を利用しているため、LLM のモデルの質や、制約語彙リストの内容、制約語彙リストを参照させるタイミング、プロンプトによる指示の内容といった要因が、翻訳の精度に影響していることがわかった。

前述の通り、本研究ではプロンプト内で制約語彙の変換を指示する場合がもっとも翻訳精度が高い傾向にあることがわかった。この結果を受けて、プロンプト内では出来る限り簡潔に指示を出し、制約語彙リストの参照による翻訳精度の向上が当面の課題である。

# 謝辞

本論文の作成にあたり、多くの方々にご指導ご鞭撻を賜りました。

指導教官の茨城大学工学部情報工学科新納浩幸教授には終始多大なご指導を賜りました。ここに深謝の意を表します。

また、新納研究室の皆様には、本研究の遂行にあたり多大なご助言、ご協力頂きました。ここに誠意の意を表します。

## 参考文献

- [1] David Grangier Viresh Ratnakar Qijun Tan Markus Freitag, George Foster and Wolfgang Macherey. Experts, errors, and context: A large-scale study of human evaluation for machine translation. *Transactions of the Association for Computational Linguistics*, 2021.
- [2] Christian Federmann. Appraise evaluation framework for machine translation. *In Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, 2018.
- [3] Yong Wang Victor O.K. Li. Guanhua Chen, Yun Chen. Lexical-constraint-aware neural machine translation via data augmentation. 2020.
- [4] Tim Salimans Ilya Sutskever. Alec Radford, Karthik Narasimhan. Improving language understanding by generative pre-training. 2018.
- [5] 徹底解説！初代 gpt 論文「improving language understanding by generative pre-training」. 2025.02.04. <https://qiita.com/Toyamanokinsan/items/adff5e927fe26148c69c>.
- [6] et al Tom. Language models are few-shot learners. *Advances in neural information processing systems 33*, 2020.
- [7] 【論文毎日読んでみる 3】 gpt-3 の論文【”gpt-3: Language models are few-shot learners” (brown et al., 2020)】. 2025.02.04. <https://qiita.com/ryo18/items/d755ddc10fc93b397627>.

# 付録

## A プログラム

1つ目のソースコードを A.1 に示す.

ソースコード A.1: ronbun.py

---

```
1 import openai
2 from langchain_core.tools import tool
3 from langchain_core.prompts import ChatPromptTemplate
4 from langchain_openai.chat_models import ChatOpenAI
5 from langgraph.prebuilt import create_react_agent
6
7
8 openai.api_key = 'sk-proj-*****'
9 model = ChatOpenAI(model='gpt-3.5-turbo', openai_api_key=openai.
    api_key)
10
11 term_dict = {
12     'Fugen_Power_Plant': '「ふげん発電所」',
13     'DERS_software': 'DERSソフトウェア',
14     'dose_rate': '線量率',
15     'VRdose_software': 'VRdoseソフトウェア',
16     'object_changes': 'オブジェクトの変化',
17     'DERS': 'DERS',
18     'software': 'ソフトウェア',
19     'R5_version': 'R5バージョン',
20     'sensing': '感知',
21     'phase_shifts': '位相シフト回路',
22     'amplifiers': '増幅器',
23     'phase_shift_magnetic_sensor_system': '位相シフト磁気センサーシステム',
24     'output': '出力',
```

- 25 'coils': 'コイル',  
26 'input\_signal\_phase': '入力信号位相',  
27 'sensitivity': '感度',  
28 'magnetic\_features': '磁気特性',  
29 'resolution': '分解能',  
30 'resonance\_frequency': '共振周波数',  
31 'feedback\_circuit': '帰還回路',  
32 'inversely\_connecting': '逆接続',  
33 'inner\_diameter': '内径',  
34 'pure\_water': '純水',  
35 'magnetized\_device': '磁化デバイス',  
36 'flow\_rate': '流速',  
37 'volume': '流量',  
38 'detect': '検出',  
39 'SAW\_resonant\_frequency(SAW-RF)\_shifts': 'SAW共振周波数シフト',  
40 'strain': '歪み',  
41 'radio\_frequency\_IC\_tag(RFID)': '無線ICタグ(RFID)',  
42 'strain\_testing\_method': '歪み測定法',  
43 'surface\_acoustic\_waves(SAW)': '表面音響波(SAW)',  
44 'sensor\_network': 'センサーネットワーク',  
45 'high\_sensitivity\_strain\_sensor': '高感度歪みセンサー',  
46 'frequency\_resolution': '周波数分解能',  
47 'SAW-RF': 'SAW共振周波数',  
48 'detectable\_strain': '検出可能歪み',  
49 'quartz\_substrate': '水晶基板',  
50 'SAW\_sensor': 'SAWセンサー',  
51 'resonant\_frequency': '共振周波数',  
52 'micromachining': 'マイクロマシニング',  
53 'minimum\_detection\_strain': '最小検出歪み',  
54 'three-phase\_correlation\_image\_sensor(3PCIS)': '3相相関イメージセン  
サー(three-phase correlation image  
sensor: 3PCIS)',  
55 'real-time\_phase\_stamp\_range\_finder(RT-PSRF)': '実時間の位相スタン  
プ・レンジファインダー(phase-stamp range fi  
nder: PSRF)',  
56 '3PCIS': '3PCIS',  
57 'optical\_stripe\_range\_finder': '光ストライプ・レンジファインダー',  
58 'three-phase\_reference\_signal\_generator\_from\_RT-PSRF': '3相基準信号  
発生装置',  
59 'single\_scanning': '単一走査',  
60 'sheet\_beam': 'シートビーム',  
61 'phase': '位相',  
62 'pixels': 'ピクセル',

- 63 'reference\_signal': '基準信号',  
64 'optical\_stripes': '光ストライプ',  
65 'phase': '位相',  
66 'angle': '角度',  
67 'detected': '検出され',  
68 'converting': '変換する',  
69 'real\_time': '実時間',  
70 'three-dimensional\_shapes': '3次元形状',  
71 'projected\_sheet\_beam': '投影シートビーム',  
72 'object': '対象物',  
73 'time-stamp\_VLSI\_range\_finder': '時間スタンプVLSIレンジファイン  
ダー',  
74 'spatial\_ununiformity': '空間的不均一性',  
75 'PSRF': 'PSRF',  
76 'surface\_reflection': '表面反射度',  
77 'modulated\_PSRF': '変調型PSRF',  
78 'proposed\_PSRF': '提案PSRF',  
79 'hydrogen\_bond': '水素結合',  
80 'hydrogen\_bond\_networks': '水素結合ネットワーク',  
81 'water\_molecules': '水分子',  
82 'hydrogen\_bond\_move\_collectively': '水素結合',  
83 'intermittent\_movement': '間欠運動',  
84 'energy\_fluctuations': 'エネルギー揺らぎ',  
85 'pH': 'pH',  
86 'pH7': 'pH7',  
87 'insulator': '強誘電体',  
88 'silicon\_substrate': 'シリコン基板',  
89 'single\_crystal\_thin\_films': '単結晶膜',  
90 'ferroelectrics': '絶縁体',  
91 'silicon\_substrate': 'シリコン基板',  
92 'γ-alumina': 'γ-アルミナ',  
93 'functional\_thin\_films': '機能性薄膜',  
94 'CVD\_and\_MBE\_methods': 'CVD法とMBE法',  
95 'CVD\_methods': 'CVD法',  
96 'MBE\_methods': 'MBE法',  
97 'alumina\_film': 'アルミナ膜',  
98 'multilayer\_thin\_films': '多層薄膜',  
99 'various\_sensors': '各種のセンサ',  
100 'electrification\_problem': '帯電問題',  
101 'phase\_difference\_electron\_microscope': '位相差電子顕微鏡',  
102 'phase\_plate': '位相板',

- 103 'electrification\_problem': '帯電問題',  
104 'fine\_processing\_technology': '微細加工技術',  
105 'phase\_plate': '位相板',  
106 'FIB\_(focus\_ion\_beam\_equipment)': 'FIB (フォーカスイオンビーム装置)',  
107 'immanent\_fouling\_within\_an\_electron\_microscope': '電顕内在的汚れ',  
108 'new-model\_electron\_microscope\_system': '新型電子顕微鏡装置',  
109 'electrification\_problem': '帯電問題',  
110 'organic\_substance\_sample': '有機物試料',  
111 'sandwich\_carbon\_coating': 'サンドイッチ炭素コーティング',  
112 'thin\_film\_processing': '薄膜加工',  
113 'phase\_plate': '位相板',  
114 'electrification\_fouling': '帯電汚れ',  
115 'tip\_gate\_dielectrics': '先端ゲート誘電体',  
116 'Hafnium-related\_thin\_film\_oxides': 'ハフニウム関連薄膜酸化物',  
117 'field\_effect\_transistors\_(FET)': 'フィールド効果トランジスタ (FET)',  
118 'HfO2': 'HfO<sub>2</sub>',  
119 'SiNx': 'SiNx',  
120 'physical\_properties': '物理特性',  
121 'k\_nano-laminate\_materials': 'kナノラミネート材料',  
122 'deposition\_process': '沈着プロセス',  
123 'annealing\_process': 'アニーリングプロセス',  
124 'surface\_analysis\_method': '表面分析法',  
125 'composition\_profile': '組成プロフィール',  
126 'band\_offset': 'バンドオフセット',  
127 'XPS': 'XPS',  
128 'bandgap': 'バンドギャップ',  
129 'ESEELS\_method\_(elastic\_scattering\_electronic\_energy\_loss\_spectrum)': 'ESEELS法 (弾性散乱電子エネルギーロススペクトル)',  
130 'band\_offset': 'バンドオフセット',  
131 'bandgap': 'バンドギャップ',  
132 'Hf/(Hf+Si)\_ratio': 'Hf / (Hf + Si) 比',  
133 'nano-laminate': 'ナノラミネート',  
134 'nuclei': '原子核',  
135 'pairing\_density\_distribution': '対密度分布',  
136 'Hartree-Fock-Bogoliubov\_method': 'Hartree-Fock-Bogoliubov法',  
137 'nuclei': '原子核',  
138 'axial\_symmetric': '軸対称な',  
139 'quadrupole\_deformed': '四重極変形した',  
140 'density\_distribution': '密度分布',

```

141 'pairing_density_distribution': '対密度分布',
142 'surface_type_pairing_correlation': '表面型対相関',
143 'Angular_dependences': '角度依存性',
144 'radial_axis': '動径軸',
145 '102Zr': '1 0 2 Z r',
146 'pairing_density_distribution': '対密度分布',
147 'intermediately': '中程度に',
148 'magnetic_phase_diagram': '磁気状態図',
149 'single_crystals': '単結晶',
150 'Bi2Sr2CaCu2O8 δ + (- Bi2212)': 'B i 2 S r 2 C a C u 2 O 8 + δ ( B i
    - 2 2 1 2 )',
151 'Josephson_vortex_flow_resistance': 'J o s e p h s o n 渦糸フロー抵抗',
152 'over_doped': '過剰添加',
153 'doped': '過剰添加',
154 'upper_and_lower_phase_boundaries': '上と下の相境界',
155 'periodic_oscillation': '周期的振動',
156 'flow_resistance': 'フロー抵抗',
157 'three-dimensional_regular_Josephson_vortex_lattice_phase': '三次元
    の規則的 J o s e p h s o n 渦糸格子相',
158 'lower_boundary': '下界',
159 'doping': 'ドーピング',
160 'three-dimensional_orderly_vortex_phase': '三次元規則的渦糸相',
161 'anisotropy': '異方性',
162 'phase_boundary': '相境界',
163 'formation_of_Josephson_vortex_lattice': 'J o s e p h s o n 渦糸格子
    形成',
164 'Josephson_vortices': 'J o s e p h s o n 渦糸',
165 'two-dimensional_quasi-orderly_state': '二次元の準規則的状态',
166 'boundary': '境界',
167 'phase_transition': '相転移',
168 'Josephson_vortex_lattice_phase': 'J o s e p h s o n 渦糸格子状態',
169 }
170
171
172 def apply_term_dict(text, term_dict):
173     """文章内の単語やフレーズを用語辞書に従って変換する関数。辞書の順番で
        フレーズを置き換えるようにする。
174
175
176     """
177     for term, translation in term_dict.items():
178         text = text.replace(term, translation)

```

```
179     return text
180
181 # のモデルを指定してを呼び出す ChatGPTAPI
182 # デコレーターを関数に付ける Tool
183 @tool
184 def ask_gpt_with_terms(prompt1):
185     """与えられた文章を日本語に翻訳する関数。事前に用語辞書を適用して、特
        定の単語やフレーズを変換する。
186
187     """
188
189     translated_prompt = apply_term_dict(prompt1, term_dict)
190
191     # 以下に自分のキーを設定 API
192     openai.api_key = 'sk-proj-*****'
193
194     prompt = f"{translated_prompt}を日本語に翻訳してください。解説や追
        加の情報は不要です。"
195
196     res = openai.chat.completions.create(
197         model='gpt-3.5-turbo',
198         messages=[{'role': 'user', 'content': prompt}],
199         temperature=0,
200         top_p=1,
201         n=1
202     )
203     return res.choices[0].message.content
204
205 #プロンプトの定義
206 prompt2 = ChatPromptTemplate.from_messages(
207     [
208         (
209             'system',
210             """与えられた文章を日本語に翻訳してください。解説や追加の情報
                は不要です。翻訳以外の情報を提供しないでください。
211
212             """,
213         ),
214         ('placeholder', '{messages}'),
215     ]
216 )
217 )
```

```
218
219 #エージェントの作成
220 agent = create_react_agent(
221     model=ChatOpenAI(model='gpt-3.5-turbo'),
222     tools=[ask_gpt_with_terms],
223     state_modifier=prompt2
224 )
225
226 #でモデルとを紐付ける bind_toolsLLMtool
227 #chain = prompt2 | ChatOpenAI().bind_tools([ask_gpt])
228 #から関数の呼び出し結果を取得 LLM
229 #result = chain.invoke({'messages':['']})
230
231 # のテスト API
232 prompt1 = input('翻訳したい文章を入力してください。 \n')
233 #print(ask_gpt.invoke(prompt1 + を日本語にして'))
234
235 #エージェントの実行
236 result = agent.invoke({'messages':[prompt1]})
237 print(result['messages'][-1].content)
```

---

また、2つ目のソースコードを A.2 に示す。

---

ソースコード A.2: ronbun2.py

---

```
1 import openai
2 from langchain_core.tools import tool
3 from langchain_core.prompts import ChatPromptTemplate
4 from langchain_openai.chat_models import ChatOpenAI
5 from langgraph.prebuilt import create_react_agent
6
7 openai.api_key = 'sk-proj-*****'
8 model = ChatOpenAI(model='gpt-3.5-turbo')
9
10 # のモデルを指定してを呼び出す ChatGPTAPI
11 # デコレーターを関数に付ける Tool
12 @tool
13 def ask_gpt(prompt1):
14     """与えられた文章を日本語に翻訳する関数。翻訳だけを行ってください。
15
16     """
17     # 以下に自分のキーを設定 API
```

```
18     openai.api_key = 'sk-proj-*****'
19
20     prompt = f"{prompt1}を日本語に翻訳してください。解説や追加の情報は
        不要です。"
21
22     res = openai.chat.completions.create(
23         model='gpt-3.5-turbo',
24         messages=[{'role': 'user', 'content': prompt}]
25     )
26     return res.choices[0].message.content
27
28 #プロンプトの定義
29 prompt2 = ChatPromptTemplate.from_messages(
30     [
31         (
32             'system',
33             """与えられた文章を日本語に翻訳してください。解説や追加の情報は
                不要です。翻訳以外の情報を提供しないでください。
34
35
36             """,
37         ),
38         ('placeholder', '{messages}')
39     ]
40 )
41
42 #エージェントの作成
43 agent = create_react_agent(
44     model=ChatOpenAI(model='gpt-3.5-turbo'),
45     tools=[ask_gpt],
46     state_modifier=prompt2
47 )
48
49 #でモデルとを紐付ける bind_toolsLLMtool
50 #chain = prompt2 | ChatOpenAI().bind_tools([ask_gpt])
51 #から関数の呼び出し結果を取得 LLM
52 #result = chain.invoke({'messages':['']})
53
54 # のテスト API
55 prompt1 = input('翻訳したい文章を入力してください。 \n')
56 #print(ask_gpt.invoke(prompt1 + を日本語にして'))
57
```

```
58 #エージェントの実行
59 result = agent.invoke({'messages':[prompt1]})
60 print(result['messages'][-1].content)
```

---

また、3つ目のソースコードを A.3 に示す.

ソースコード A.3: ronbun3.py

---

```
1
2 import os
3 import openai
4 import openai
5 from langchain_core.tools import tool
6 from langchain_core.prompts import ChatPromptTemplate
7 from langchain_openai.chat_models import ChatOpenAI
8 from langgraph.prebuilt import create_react_agent
9
10 openai.api_key = 'sk-proj-*****'
11 model = ChatOpenAI(model='gpt-3.5-turbo', openai_api_key=openai.
    api_key)
12
13 term_dict = {
14     'Fugen_Power_Plant': '「ふげん発電所」',
15     'DERS_software': 'DERSソフトウェア',
16     'dose_rate': '線量率',
17     'VRdose_software': 'VRdoseソフトウェア',
18     'object_changes': 'オブジェクトの変化',
19     'DERS': 'DERS',
20     'software': 'ソフトウェア',
21     'R5_version': 'R5バージョン',
22     'sensing': '感知',
23     'phase_shifts': '位相シフト回路',
24     'amplifiers': '増幅器',
25     'phase_shift_magnetic_sensor_system': '位相シフト磁気センサーシステム',
26     'output': '出力',
27     'coils': 'コイル',
28     'input_signal_phase': '入力信号位相',
29     'sensitivity': '感度',
30     'magnetic_features': '磁気特性',
31     'resolution': '分解能',
32     'resonance_frequency': '共振周波数',
```

- 33 'feedback\_circuit': '帰還回路',  
34 'inversely\_connecting': '逆接続',  
35 'inner\_diameter': '内径',  
36 'pure\_water': '純水',  
37 'magnetized\_device': '磁化デバイス',  
38 'flow\_rate': '流速',  
39 'volume': '流量',  
40 'detect': '検出',  
41 'SAW\_resonant\_frequency\_(SAW-RF)\_shifts': 'SAW共振周波数シフト',  
42 'strain': '歪み',  
43 'radio\_frequency\_IC\_tag\_(RFID)': '無線ICタグ (RFID)',  
44 'strain\_testing\_method': '歪み測定法',  
45 'surface\_acoustic\_waves\_(SAW)': '表面音響波 (SAW)',  
46 'sensor\_network': 'センサーネットワーク',  
47 'high\_sensitivity\_strain\_sensor': '高感度歪みセンサー',  
48 'frequency\_resolution': '周波数分解能',  
49 'SAW-RF': 'SAW共振周波数',  
50 'detectable\_strain': '検出可能歪み',  
51 'quartz\_substrate': '水晶基板',  
52 'SAW\_sensor': 'SAWセンサー',  
53 'resonant\_frequency': '共振周波数',  
54 'micromachining': 'マイクロマシニング',  
55 'minimum\_detection\_strain': '最小検出歪み',  
56 'three-phase\_correlation\_image\_sensor\_(three-phase\_correlation\_image\_sensor:3PCIS)',  
57 'real-time\_phase\_stamp\_range\_finder\_(RT-PSRF)': '実時間の位相スタンプ・レンジファインダー (phase-stamp range finder:PSRF)',  
58 '3PCIS': '3PCIS',  
59 'optical\_stripe\_range\_finder': '光ストライプ・レンジファインダー',  
60 'three-phase\_reference\_signal\_generator\_from\_RT-PSRF': '3相基準信号発生装置',  
61 'single\_scanning': '単一走査',  
62 'sheet\_beam': 'シートビーム',  
63 'phase': '位相',  
64 'pixels': 'ピクセル',  
65 'reference\_signal': '基準信号',  
66 'optical\_stripes': '光ストライプ',  
67 'phase': '位相',  
68 'angle': '角度',  
69 'detected': '検出され',  
70 'converting': '変換する',

- 71 'real\_time': '実時間',  
72 'three-dimensional\_shapes': '3次元形状',  
73 'projected\_sheet\_beam': '投影シートビーム',  
74 'object': '対象物',  
75 'time-stamp\_VLSI\_range\_finder': '時間スタンプVLSIレンジファイン  
ダー',  
76 'spatial\_ununiformity': '空間的不均一性',  
77 'PSRF': 'PSRF',  
78 'surface\_reflection': '表面反射度',  
79 'modulated\_PSRF': '変調型PSRF',  
80 'proposed\_PSRF': '提案PSRF',  
81 'hydrogen\_bond': '水素結合',  
82 'hydrogen\_bond\_networks': '水素結合ネットワーク',  
83 'water\_molecules': '水分子',  
84 'hydrogen\_bond\_move\_collectively': '水素結合',  
85 'intermittent\_movement': '間欠運動',  
86 'energy\_fluctuations': 'エネルギー揺らぎ',  
87 'pH': 'pH',  
88 'pH7': 'pH7',  
89 'insulator': '強誘電体',  
90 'silicon\_substrate': 'シリコン基板',  
91 'single\_crystal\_thin\_films': '単結晶膜',  
92 'ferroelectrics': '絶縁体',  
93 'silicon\_substrate': 'シリコン基板',  
94 'γ-alumina': 'γ-アルミナ',  
95 'functional\_thin\_films': '機能性薄膜',  
96 'CVD\_and\_MBE\_methods': 'CVD法とMBE法',  
97 'CVD\_methods': 'CVD法',  
98 'MBE\_methods': 'MBE法',  
99 'alumina\_film': 'アルミナ膜',  
100 'multilayer\_thin\_films': '多層薄膜',  
101 'various\_sensors': '各種のセンサ',  
102 'electrification\_problem': '帯電問題',  
103 'phase\_difference\_electron\_microscope': '位相差電子顕微鏡',  
104 'phase\_plate': '位相板',  
105 'electrification\_problem': '帯電問題',  
106 'fine\_processing\_technology': '微細加工技術',  
107 'phase\_plate': '位相板',  
108 'FIB(focus\_ion\_beam\_equipment)': 'FIB (フォーカスイオンビーム装  
置)',  
109 'immanent\_fouling\_within\_an\_electron\_microscope': '電顕内在的汚れ',  
110 'new-model\_electron\_microscope\_system': '新型電子顕微鏡装置',

- 111 'electrification\_problem': '帯電問題',  
112 'organic\_substance\_sample': '有機物試料',  
113 'sandwich\_carbon\_coating': 'サンドイッチ炭素コーティング',  
114 'thin\_film\_processing': '薄膜加工',  
115 'phase\_plate': '位相板',  
116 'electrification\_fouling': '帯電汚れ',  
117 'tip\_gate\_dielectrics': '先端ゲート誘電体',  
118 'Hafnium-related\_thin\_film\_oxides': 'ハフニウム関連薄膜酸化物',  
119 'field\_effect\_transistors\_(FET)': 'フィールド効果トランジスタ (FET)',  
120 'HfO2': 'H f O 2',  
121 'SiNx': 'S i N x',  
122 'physical\_properties': '物理特性',  
123 'k\_nano-laminate\_materials': 'k ナノラミネート材料',  
124 'deposition\_process': '沈着プロセス',  
125 'annealing\_process': 'アニーリングプロセス',  
126 'surface\_analysis\_method': '表面分析法',  
127 'composition\_profile': '組成プロフィール',  
128 'band\_offset': 'バンドオフセット',  
129 'XPS': 'X P S',  
130 'bandgap': 'バンドギャップ',  
131 'ESEELS\_method\_(elastic\_scattering\_electronic\_energy\_loss\_spectrum)': 'ESEELS法 (弾性散乱電子エネルギーロススペクトル)',  
132 'band\_offset': 'バンドオフセット',  
133 'bandgap': 'バンドギャップ',  
134 'Hf/(Hf+Si)\_ratio': 'H f / (H f + S i) 比',  
135 'nano-laminate': 'ナノラミネート',  
136 'nuclei': '原子核',  
137 'pairing\_density\_distribution': '対密度分布',  
138 'Hartree-Fock-Bogoliubov\_method': 'H a r t r e e - F o c k - B o g o l i u b o v 法',  
139 'nuclei': '原子核',  
140 'axial\_symmetric': '軸対称な',  
141 'quadrupole\_deformed': '四重極変形した',  
142 'density\_distribution': '密度分布',  
143 'pairing\_density\_distribution': '対密度分布',  
144 'surface\_type\_pairing\_correlation': '表面型対相関',  
145 'Angular\_dependences': '角度依存性',  
146 'radial\_axis': '動径軸',  
147 '102Zr': '1 0 2 Z r',  
148 'pairing\_density\_distribution': '対密度分布',

```

149 'intermediately': '中程度に',
150 'magnetic_phase_diagram': '磁気状態図',
151 'single_crystals': '単結晶',
152 'Bi2Sr2CaCu2O8 δ +(- Bi2212)': 'Bi2Sr2CaCu2O8 + δ (Bi
    - 2212)',
153 'Josephson_vortex_flow_resistance': 'Josephson 渦糸フロー抵抗',
154 'over_doped': '過剰添加',
155 'doped': '過剰添加',
156 'upper_and_lower_phase_boundaries': '上と下の相境界',
157 'periodic_oscillation': '周期的振動',
158 'flow_resistance': 'フロー抵抗',
159 'three-dimensional_regular_Josephson_vortex_lattice_phase': '三次元
    の規則的 Josephson 渦糸格子相',
160 'lower_boundary': '下界',
161 'doping': 'ドーピング',
162 'three-dimensional_orderly_vortex_phase': '三次元規則的渦糸相',
163 'anisotropy': '異方性',
164 'phase_boundary': '相境界',
165 'formation_of_Josephson_vortex_lattice': 'Josephson 渦糸格子
    形成',
166 'Josephson_vortices': 'Josephson 渦糸',
167 'two-dimensional_quasi-orderly_state': '二次元の準規則的状态',
168 'boundary': '境界',
169 'phase_transition': '相転移',
170 'Josephson_vortex_lattice_phase': 'Josephson 渦糸格子状態',
171 }
172
173
174 def apply_term_dict(text, term_dict):
175     """文章内の単語やフレーズを用語辞書に従って変換する関数。辞書の順番で
        フレーズを置き換えるようにする。
176
177
178     """
179     for term, translation in term_dict.items():
180         text = text.replace(term, translation)
181     return text
182
183 # のモデルを指定してを呼び出す ChatGPTAPI
184 # デコレーターを関数に付ける Tool
185 @tool
186 def ask_gpt_with_terms(prompt1):

```

```
187     """与えられた文章を日本語に翻訳する関数。事前に用語辞書を適用して、特
188         定の単語やフレーズを変換する。
189     """
190
191     translated_prompt = apply_term_dict(prompt1, term_dict)
192
193     # 以下に自分のキーを設定 API
194     openai.api_key = 'sk-proj-*****'
195
196     term_dict_str = "\n".join([f'{term}:_{translation}' for term,
197                                translation in term_dict.items()])
198
199     prompt = (
200         "以下の文章を日本語に翻訳してください。解説や追加情報は不要です。
201         翻訳する際には、"
202         "用語辞書に記載されているすべての用語（以下参照）を必ず翻訳してく
203         ださい。"
204         "翻訳時に用語辞書を遵守し、指定された用語を正確に翻訳すること。"
205         "\n\n"
206         "用語辞書:\n"
207         "1._'Fugen_Power_Plant'\t→\t「ふげん発電所」'\n"
208         "2._'DERS_software'\t→\tD E R S ソフトウェア'\n"
209         "3._'dose_rate'\t→\t線量率'\n"
210         "4._'VRdose_software'\t→\tV R d o s e ソフトウェア'\n"
211         "5._'object_changes'\t→\tオブジェクトの変化'\n"
212         "6._'DERS'\t→\tD E R S'\n"
213         "7._'software'\t→\tソフトウェア'\n"
214         "8._'R5_version'\t→\tR 5 バージョン'\n"
215         "9._'sensing'\t→\t感知'\n"
216         "10._'phase_shifts'\t→\t位相シフト回路'\n"
217         "11._'amplifiers'\t→\t増幅器'\n"
218         "12._'phase_shift_magnetic_sensor_system'\t→\t位相シフト磁気セン
219         サーステム'\n"
220         "13._'output'\t→\t出力'\n"
221         "14._'coils'\t→\tコイル'\n"
222         "15._'input_signal_phase'\t→\t入力信号位相'\n"
223         "16._'sensitivity'\t→\t感度'\n"
224         "17._'magnetic_features'\t→\t磁気特性'\n"
225         "18._'resolution'\t→\t分解能'\n"
226         "19._'resonance_frequency'\t→\t共振周波数'\n"
227         "20._'feedback_circuit'\t→\t帰還回路'\n"
```

- 224 "21. 'inversely\_connecting' → '逆接続' '\n"
- 225 "22. 'inner\_diameter' → '内径' '\n"
- 226 "23. 'pure\_water' → '純水' '\n"
- 227 "24. 'magnetized\_device' → '磁化デバイス' '\n"
- 228 "25. 'flow\_rate' → '流速' '\n"
- 229 "26. 'volume' → '流量' '\n"
- 230 "27. 'detect' → '検出' '\n"
- 231 "28. 'SAW\_resonant\_frequency\_(SAW-RF)\_shifts' → 'SAW共振周波数シフト' '\n"
- 232 "29. 'strain' → '歪み' '\n"
- 233 "30. 'radio\_frequency\_IC\_tag\_(RFID)' → '無線ICタグ (RFID)' '\n"
- 234 "31. 'strain\_testing\_method' → '歪み測定法' '\n"
- 235 "32. 'surface\_acoustic\_waves\_(SAW)' → '表面音響波 (SAW)' '\n"
- 236 "33. 'sensor\_network' → 'センサーネットワーク' '\n"
- 237 "34. 'high\_sensitivity\_strain\_sensor' → '高感度歪みセンサー' '\n"
- 238 "35. 'frequency\_resolution' → '周波数分解能' '\n"
- 239 "36. 'SAW-RF' → 'SAW共振周波数' '\n"
- 240 "37. 'detectable\_strain' → '検出可能歪み' '\n"
- 241 "38. 'quartz\_substrate' → '水晶基板' '\n"
- 242 "39. 'SAW\_sensor' → 'SAWセンサー' '\n"
- 243 "40. 'resonant\_frequency' → '共振周波数' '\n"
- 244 "41. 'micromachining' → 'マイクロマシニング' '\n"
- 245 "42. 'minimum\_detection\_strain' → '最小検出歪み' '\n"
- 246 "43. 'three-phase\_correlation\_image\_sensor\_(3PCIS)' → '3相相関イメージセンサー (three-phase correlation image sensor: 3PCIS)' '\n"
- 247 "44. 'real-time\_phase\_stamp\_range\_finder\_(RT-PSRF)' → '実時間の位相スタンプ・レンジファインダー (phase-stamp range finder: PSRF)' '\n"
- 248 "45. '3PCIS' → '3PCIS' '\n"
- 249 "46. 'optical\_stripe\_range\_finder' → '光ストライプ・レンジファインダー' '\n"
- 250 "47. 'three-phase\_reference\_signal\_generator\_from\_RT-PSRF' → '3相基準信号発生装置' '\n"
- 251 "48. 'single\_scanning' → '単一走査' '\n"
- 252 "49. 'sheet\_beam' → 'シートビーム' '\n"
- 253 "50. 'phase' → '位相' '\n"
- 254 "51. 'pixels' → 'ピクセル' '\n"
- 255 "52. 'reference\_signal' → '基準信号' '\n"
- 256 "53. 'optical\_stripes' → '光ストライプ' '\n"
- 257 "54. 'phase' → '位相' '\n"
- 258 "55. 'angle' → '角度' '\n"
- 259 "56. 'detected' → '検出され' '\n"

- 260 "57. 'converting' → 変換する''\n"
- 261 "58. 'real\_time' → 実時間''\n"
- 262 "59. 'three-dimensional\_shapes' → 3次元形状''\n"
- 263 "60. 'projected\_sheet\_beam' → 投影シートビーム''\n"
- 264 "61. 'object' → 対象物''\n"
- 265 "62. 'time-stamp\_VLSI\_range\_finder' → 時間スタンプVLSIレンジファインダー''\n"
- 266 "63. 'spatial\_ununiformity' → 空間的不均一性''\n"
- 267 "64. 'PSRF' → P S R F''\n"
- 268 "65. 'surface\_reflection' → 表面反射度''\n"
- 269 "66. 'modulated\_PSRF' → 変調型P S R F''\n"
- 270 "67. 'proposed\_PSRF' → 提案P S R F''\n"
- 271 "68. 'hydrogen\_bond' → 水素結合''\n"
- 272 "69. 'hydrogen\_bond\_networks' → 水素結合ネットワーク''\n"
- 273 "70. 'water\_molecules' → 水分子''\n"
- 274 "71. 'hydrogen\_bond\_move\_collectively' → 水素結合''\n"
- 275 "72. 'intermittent\_movement' → 間欠運動''\n"
- 276 "73. 'energy\_fluctuations' → エネルギー揺らぎ''\n"
- 277 "74. 'pH' → p H''\n"
- 278 "75. 'pH7' → p H 7''\n"
- 279 "76. 'insulator' → 強誘電体''\n"
- 280 "77. 'silicon\_substrate' → シリコン基板''\n"
- 281 "78. 'single\_crystal\_thin\_films' → 単結晶膜''\n"
- 282 "79. 'ferroelectrics' → 絶縁体''\n"
- 283 "80. 'silicon\_substrate' → シリコン基板''\n"
- 284 "81. 'γ'-alumina' → γ - アルミナ''\n"
- 285 "82. 'functional\_thin\_films' → 機能性薄膜''\n"
- 286 "83. 'CVD\_and\_MBE\_methods' → C V D法とM B E法''\n"
- 287 "84. 'CVD\_methods' → C V D法''\n"
- 288 "85. 'MBE\_methods' → M B E法''\n"
- 289 "86. 'alumina\_film' → アルミナ膜''\n"
- 290 "87. 'multilayer\_thin\_films' → 多層薄膜''\n"
- 291 "88. 'various\_sensors' → 各種のセンサ''\n"
- 292 "89. 'electrification\_problem' → 帯電問題''\n"
- 293 "90. 'phase\_difference\_electron\_microscope' → 位相差電子顕微鏡''\n"
- 294 "91. 'phase\_plate' → 位相板''\n"
- 295 "92. 'electrification\_problem' → 帯電問題''\n"
- 296 "93. 'fine\_processing\_technology' → 微細加工技術''\n"
- 297 "94. 'phase\_plate' → 位相板''\n"
- 298 "95. 'FIB(focus\_ion\_beam\_equipment)' → F I B (フォーカスイオンビーム装置)''\n"

- 299 "96. 'immanent\_fouling\_within\_an\_electron\_microscope' → 電顕内  
在的汚れ''\n"
- 300 "97. 'new-model\_electron\_microscope\_system' → 新型電子顕微鏡装  
置''\n"
- 301 "98. 'electrification\_problem' → 帯電問題''\n"
- 302 "99. 'organic\_substance\_sample' → 有機物試料''\n"
- 303 "100. 'sandwich\_carbon\_coating' → サンドイッチ炭素コーティング  
''\n"
- 304 "101. 'thin\_film\_processing' → 薄膜加工''\n"
- 305 "102. 'phase\_plate' → 位相板''\n"
- 306 "103. 'electrification\_fouling' → 帯電汚れ''\n"
- 307 "104. 'tip\_gate\_dielectrics' → 先端ゲート誘電体''\n"
- 308 "105. 'Hafnium-related\_thin\_film\_oxides' → ハフニウム関連薄膜酸  
化物''\n"
- 309 "106. 'field\_effect\_transistors\_(FET)' → フィールド効果トランジ  
スタ ( F E T ) ''\n"
- 310 "107. 'HfO2' → H f O 2 ''\n"
- 311 "108. 'SiNx' → S i N x ''\n"
- 312 "109. 'physical\_properties' → 物理特性''\n"
- 313 "110. 'k\_nano-laminate\_materials' → k ナノラミネート材料''\n"
- 314 "111. 'deposition\_process' → 沈着プロセス''\n"
- 315 "112. 'annealing\_process' → アニールングプロセス''\n"
- 316 "113. 'surface\_analysis\_method' → 表面分析法''\n"
- 317 "114. 'composition\_profile' → 組成プロフィール''\n"
- 318 "115. 'band\_offset' → バンドオフセット''\n"
- 319 "116. 'XPS' → X P S ''\n"
- 320 "117. 'bandgap' → バンドギャップ''\n"
- 321 "118. 'ESEELS\_method\_(elastic\_scattering\_electronic\_energy\_loss  
spectrum)' → E S E E L S 法 ( 弾性散乱電子エネルギーロス  
スペクトル ) ''\n"
- 322 "119. 'band\_offset' → バンドオフセット''\n"
- 323 "120. 'bandgap' → バンドギャップ''\n"
- 324 "121. 'Hf/(Hf+Si)\_ratio' → H f / ( H f + S i ) 比''\n"
- 325 "122. 'nano-laminate' → ナノラミネート''\n"
- 326 "123. 'nuclei' → 原子核''\n"
- 327 "124. 'pairing\_density\_distribution' → 対密度分布''\n"
- 328 "125. 'Hartree-Fock-Bogoliubov\_method' → H a r t r e e - F o c  
k - B o g o l i u b o v 法''\n"
- 329 "126. 'nuclei' → 原子核''\n"
- 330 "127. 'axial\_symmetric' → 軸対称な''\n"
- 331 "128. 'quadrupole\_deformed' → 四重極変形した''\n"
- 332 "129. 'density\_distribution' → 密度分布''\n"
- 333 "130. 'pairing\_density\_distribution' → 対密度分布''\n"
- 334 "131. 'surface\_type\_pairing\_correlation' → 表面型対相関''\n"

```

335     "132. 'Angular dependences' → '角度依存性', '\n"
336     "133. 'radial axis' → '動径軸', '\n"
337     "134. '102Zr' → '1 0 2 Z r', '\n"
338     "135. 'pairing density distribution' → '対密度分布', '\n"
339     "136. 'intermediately' → '中程度に', '\n"
340     "137. 'magnetic phase diagram' → '磁気状態図', '\n"
341     "138. 'single crystals' → '単結晶', '\n"
342     "139. 'Bi2Sr2CaCu2O8 δ + (- Bi2212)' → 'B i 2 S r 2 C a C u 2 O
343           8 + δ ( B i - 2 2 1 2 )', '\n"
344     "140. 'Josephson vortex flow resistance' → 'J o s e p h s o n 渦
345           糸フロー抵抗', '\n"
346     "141. 'overdoped' → '過剰添加', '\n"
347     "142. 'doped' → '過剰添加', '\n"
348     "143. 'upper and lower phase boundaries' → '上と下の相境界', '\n"
349     "144. 'periodic oscillation' → '周期的振動', '\n"
350     "145. 'flow resistance' → 'フロー抵抗', '\n"
351     "146. 'three-dimensional regular Josephson vortex lattice phase
352           ' → '三次元の規則的 J o s e p h s o n 渦糸格子相
353           ', '\n"
354     "147. 'lower boundary' → '下界', '\n"
355     "148. 'doping' → 'ドーピング', '\n"
356     "149. 'three-dimensional orderly vortex phase' → '三次元規則的渦
357           糸相', '\n"
358     "150. 'anisotropy' → '異方性', '\n"
359     "151. 'phase boundary' → '相境界', '\n"
360     "152. 'formation of Josephson vortex lattice' → 'J o s e p h s
361           o n 渦糸格子形成', '\n"
362     "153. 'Josephson vortices' → 'J o s e p h s o n 渦糸', '\n"
363     "154. 'two-dimensional quasi-orderly state' → '二次元の準規則的
364           状態', '\n"
365     "155. 'boundary' → '境界', '\n"
366     "156. 'phase transition' → '相転移', '\n"
367     "157. 'Josephson vortex lattice phase' → 'J o s e p h s o n 渦糸
368           格子状態', '\n\n"
369     f"翻訳対象文章:\n{translated_prompt}"
370     )
371     res = openai.chat.completions.create(
372         model='gpt-3.5-turbo',
373         messages=[{'role': 'user', 'content': prompt}],
374         temperature=0,
375         top_p=1,
376         n=1
377     )
378     return res.choices[0].message.content

```

```
371
372 #プロンプトの定義
373 prompt2 = ChatPromptTemplate.from_messages(
374     [
375         (
376             'system',
377             """与えられた文章を日本語に翻訳してください。解説や追加の情報は不要です。翻訳以外の情報を提供しないでください。
378
379
380             """,
381         ),
382         ('placeholder', '{messages}')
383     ]
384 )
385
386 #エージェントの作成
387 agent = create_react_agent(
388     model=ChatOpenAI(model='gpt-3.5-turbo'),
389     tools=[ask_gpt_with_terms],
390     state_modifier=prompt2
391 )
392
393 #でモデルとを紐付ける bind_toolsLLMtool
394 #chain = prompt2 | ChatOpenAI().bind_tools([ask_gpt])
395 #から関数の呼び出し結果を取得 LLM
396 #result = chain.invoke({'messages':['']})
397
398 # のテスト API
399 prompt1 = input('翻訳したい文章を入力してください。 \n')
400 #print(ask_gpt.invoke(prompt1 + を日本語にして'))
401
402 #エージェントの実行
403 result = agent.invoke({'messages':[prompt1]})
404 print(result['messages'][-1].content)
```

---