

卒業論文

レジスター理論に基づく スタイル考慮型翻訳エージェント

指導教官 村上陽平教授

立命館大学情報理工学部
先端社会デザインコース 4 回生
2600220230-0

谷川舞桜

2025 年度（秋学期）卒業研究 3（CH）
令和 8 年 1 月 30 日

レジスタ理論に基づくスタイル考慮型翻訳エージェント

谷川 舞桜

内容梗概

近年、グローバル化が一層進展しており、異なる言語や文化間でのコミュニケーション機会が増加している。機械翻訳や文脈理解を可能とする大規模言語モデル(LLM)の発展により、翻訳の意味上の適切さが向上し言語の障壁は低減されつつある。一方で、翻訳においては意味だけでなく、文体やトーンといったスタイルも重要な役割を果たす。例えば、話者と聞き手の関係性に応じて、丁寧さが異なる表現が求められる。このように、社会的状況によってスタイルは変動するため、同一の文であっても適切な翻訳表現は一意に定まらない。このようなスタイルが明示されていない場合、入力に対して一意に出力を返す関数的な機械翻訳では、スタイルの曖昧性を解決できないまま翻訳が行われる。その結果、語義は正しくとも、話者の意図や対人関係を反映しない翻訳にとどまる。従来研究では、翻訳における語義の曖昧性に対しては、ユーザとのインタラクションを通じて翻訳を生成する翻訳エージェントが提案されている。特に、多義語や文化依存の単語など複数の意味を持つ単語に起因する語義の曖昧性を対象として翻訳エージェントの有効性が検討されてきた。しかしながら、スタイルの曖昧性については、これまで翻訳エージェントでは十分に扱われてこなかった。

そこで、本研究では、スタイルの曖昧性を考慮した翻訳エージェントを提案する。具体的には、Hallidayのレジスタ理論に基づき、入力文におけるスタイルの曖昧性を検出し、ユーザとのインタラクションを通じて、曖昧性の解消を図る。本手法の実現にあたり、取り組むべき課題は以下の2点である。

スタイルの曖昧性の汎用的な検出基準の設計

翻訳エージェントがスタイルの曖昧性を判定する基準はLLMの内部推論に強く依存する。特に、LLMはプロンプトによって大きく性能に差が生じやすいため、複数のLLMから翻訳エージェントを構築できるように曖昧性検出基準を設計する必要がある。

複雑なシナリオに頑健なエージェント制御

従来は語義の曖昧性のみが対象だったが、スタイルの曖昧性も対象とすることで処理手順が増え、シナリオが複雑化する。そのため、シナリオの語彙を適切に定義し、正確にエージェントを制御できるシナリオを記述する必要がある。1つ目の課題に対しては、Hallidayのレジスタ理論に基づいて曖昧性判定基

準を設計した。さらに、レジスター理論を定義したプロンプトと定義していないプロンプトを、同一の YAML プロンプトおよび同等の動作を行う自然言語プロンプトに適用し、両者を比較した。具体的には、英文ごとにプロンプトを送信し、前の返答を引き継がない形で、多様な英文に対応できるかを検証し、スタイルの曖昧性検査精度、スタイル一致率、スタイル明確化質問の妥当性を分析する「曖昧性検査の網羅性実験」を実施した。

2 つ目の課題に対しては、YAML プロンプトとそれと同等の動作を行う自然言語プロンプトを用いて比較実験を行った。従来の YAML シナリオでは、状態を 3 個、イベント(条件)を 2 個、アクション(行動)を 3 個に限定していたが、本研究では、状態を 5 個、イベント(条件)を 3 個、アクション(行動)を 5 個へと拡張し、シナリオの複雑化に対しても破綻なく動作するか検証した。さらに、同一コンテキストで翻訳を依頼し、ユーザとのインタラクションを繰り返すことで、シナリオの逸脱頻度、スタイルの曖昧性検査精度、スタイル一致率、スタイル明確化質問の妥当性を分析する「対話継続性の評価実験」を実施した。複数の異なる LLM を用いて評価を行い、提案手法の有効性を検証した。本研究の貢献は以下の通りである。

スタイルの曖昧性の汎用的な検出基準の設計

「曖昧性検査の網羅性実験」では、GPT-3.5-turbo モデルにおいて YAML プロンプトではすべての結果が 0% となり、自然言語プロンプトにおいても全体的に精度は低かった。一方、GPT-4.1-mini モデルでは自然言語プロンプトにレジスター理論を導入した場合、曖昧性検査精度は 24% 向上し、スタイル一致率も 10% の向上が確認された。これに対し、GPT-5-mini モデルでレジスター理論を導入した場合、YAML プロンプトにおいて曖昧性検査精度が 22% 向上し、スタイル明確化質問の妥当性に関しても高い性能を示した。

複雑なシナリオに頑健エージェント制御

「対話の継続性実験」では、GPT-3.5-turbo モデルはすべての評価指標が 0% となった。シナリオ逸脱頻度は、GPT-5-mini モデルの YAML プロンプトにレジスター理論を定義しない条件でのみ確認された。一方、自然言語プロンプトにレジスター理論を導入した場合、曖昧性検査精度およびスタイル一致率は向上し、スタイル明確化質問の妥当性においても高い値を示した。GPT-4.1-mini モデルでは、レジスター理論の導入によりいずれのプロンプトでも曖昧性検査精度は向上し、自然言語プロンプトにおいてはスタイル一致率も向上した。

Style-aware Translation Agent based on Register Theory

Mao Tanigawa

Abstract

With recent globalization, opportunities for communication across different languages and cultures have increased. Advances in machine translation and LLM capable of contextual understanding have improved semantic appropriateness in translation and lowered language barriers. However, translation depends not only on meaning but also on style, such as tone and formality. Because style varies with social context, the same sentence may allow multiple appropriate translations. When stylistic information is not explicitly specified, functional machine translation systems that produce a single output fail to resolve style ambiguity, resulting in translations that do not reflect the speaker's intent or interpersonal relationships. In previous studies, translation agents that generate translations through user interaction have been proposed to address word-sense ambiguity, particularly ambiguity arising from polysemous or culturally dependent words. However, style ambiguity has not been sufficiently addressed by translation agents to date.

Therefore, in this study, we propose a translation agent that detects and resolves style ambiguity in input sentences through user interaction, based on Halliday's register theory. To achieve this, two key issues had to be addressed.

Design of a generic criterion for detecting style ambiguity

The criteria by which a translation agent determines style ambiguity depend heavily on the internal reasoning of the LLM. In particular, because LLM performance is highly sensitive to prompt design, it is necessary to establish ambiguity detection criteria that enable translation agents to be constructed using multiple LLMs.

Robust agent control under complex scenarios

While previous work focused only on word-sense ambiguity, incorporating style ambiguity increases the number of processing steps and leads to more complex scenarios. Therefore, it is necessary to appropriately define scenario vocabularies and to describe scenarios that can accurately control the agent.

For the first issue, we designed style ambiguity detection criteria based on

Halliday’s register theory and compared prompts with and without explicit register definitions using YAML and equivalent natural language prompts. A “comprehensiveness experiment for ambiguity testing” was conducted using independent prompts for each sentence to evaluate style ambiguity detection accuracy, clarification question appropriateness, and translation accuracy.

For the second challenge, we compared YAML and natural language prompts under more complex scenarios by expanding the number of states, events, and actions. A “dialogue continuity evaluation experiment” analyzed scenario deviation frequency, style ambiguity detection accuracy, and translation accuracy through repeated user interactions.

We evaluated the proposed method using multiple LLMs and verified its effectiveness. The contributions of this study are summarized as follows.

Design of a generic criterion for detecting style ambiguity

In the “comprehensiveness experiment for ambiguity testing,” all results for the GPT-3.5-turbo model were 0% when using YAML prompts, and overall accuracy remained low even with natural language prompts. In contrast, for the GPT-4.1-mini model, introducing register theory into natural language prompts led to a 24% improvement in ambiguity detection accuracy, along with a 10% increase in style consistency. For the GPT-5-mini model, the introduction of register theory resulted in a 22% improvement in ambiguity detection accuracy when using YAML prompts and also demonstrated high performance in terms of the validity of style-clarification questions.

Robust agent control under complex scenarios

In the “dialogue continuity evaluation experiment,” all evaluation metrics for the GPT-3.5-turbo model were 0%. Scenario deviation was observed only in the GPT-5-mini model when using YAML prompts without register theory. In contrast, when register theory was introduced into natural language prompts, ambiguity detection accuracy and style consistency improved, and high validity was also observed for style-clarification questions. For the GPT-4.1-mini model, the introduction of register theory improved ambiguity detection accuracy across both prompt types, and additionally enhanced style consistency in natural language prompts.

レジスタ理論に基づくスタイル考慮型翻訳エージェント

目次

第 1 章	はじめに	1
第 2 章	関連研究	3
2.1	翻訳エージェント	3
2.2	翻訳におけるスタイルの扱い	4
第 3 章	スタイルを考慮した翻訳エージェント	5
3.1	レジスタ理論	5
3.2	アクション定義	6
3.2.1	スタイル曖昧性確認	6
3.2.2	スタイル明確化質問	7
3.3	イベント定義	8
第 4 章	翻訳エージェントのシナリオ	9
4.1	シナリオの全体設計	9
4.1.1	メッセージ待機状態	10
4.1.2	語義確認状態および語義明確化状態	10
4.1.3	スタイル確認状態およびスタイル明確化状態	11
第 5 章	評価	13
5.1	評価指標	13
5.1.1	曖昧性検査精度	13
5.1.2	スタイル一致率	13
5.1.3	スタイル明確化質問の妥当性	14
5.1.4	シナリオ逸脱頻度	14
5.2	評価手法	14
5.2.1	ベースライン	14
5.2.2	比較する LLM の種類	15
5.2.3	データセット	15
5.2.4	実験手順	16
5.3	実験結果	16

5.3.1 曖昧性検査の網羅性実験	16
5.3.2 対話継続性の評価実験	19
第6章 考察	22
6.1 曖昧性検査精度	22
6.2 スタイル一致率	23
6.3 スタイル明確化質問の妥当性	24
6.4 シナリオ逸脱頻度	25
6.5 まとめ	26
第7章 おわりに	27
謝辞	28
参考文献	29
付録	30
A.1 テスト文	30
A.1.1 プロンプト	37

第1章 はじめに

近年、社会のグローバル化の進展により、異なる言語話者間におけるコミュニケーションの重要性は一層高まっている。国際的なビジネスや学術交流、オンライン空間での情報共有が日常化する現代社会において、言語の壁を越えた円滑な意思疎通は重要な役割を果たしている。こうした社会的背景を受け、機械翻訳や大規模言語モデル (LLM) は、異言語間のコミュニケーションを支援する有効な手段として発展し、日常的な情報伝達においては高い有用性を示している。一方で、これらの翻訳技術は、文の意味内容が同一であっても、使用される文体や場面に応じて適切にスタイルを使い分けることにおいては、依然として課題を抱えている。ここでいうスタイルとは、話者や書き手の社会的関係、使用される場面、媒体といった要因に応じて選択される言語表現のあり方を指す。さらに、語彙や構文の選択に反映されるだけでなく、発話の丁寧さや距離感といった対人的意味にも関わる要素である。例えば、同一の依頼内容であっても、友人間における口語的な表現と、公的文書における形式的な表現とでは、用いられる語彙や構文が異なるだけでなく、聞き手に与える印象や社会的適切性も異なる。こうした違いは、単なる語義や文法の問題ではなく、使用状況に依存した言語選択の問題である。スタイルを適切に考慮した翻訳の実現が困難である理由の一つは、スタイルに関する情報が多くの場合、明示的に与えられない点にある。話者間の関係性や発話場面といった要素は、文面そのものには直接現れず、文脈をもとに推測される。その結果、従来の翻訳システムや LLM は、意味的には正しい訳文を生成できたとしても、使用場面にそぐわない表現を選択してしまい、不適切な訳文を生じさせる場合がある。従来研究では、翻訳における語義の曖昧性に対しては、ユーザとのインタラクションを通じて翻訳を生成する翻訳エージェントが提案されている。特に、多義語や文化依存の単語など複数の意味を持つ単語に起因する語義の曖昧性を対象として翻訳エージェントの有効性が検討されてきた。しかしながら、スタイルの曖昧性については、これまで翻訳エージェントでは十分に扱われてこなかった。

そこで、本研究では、スタイルの曖昧性を考慮した翻訳エージェントを提案する。具体的には、Halliday のレジスター理論に基づき、入力文におけるスタイルの曖昧性を検出し、ユーザとのインタラクションを通じて、曖昧性の解消を図る。本手法の実現にあたり、取り組むべき課題は以下の 2 点である。

スタイルの曖昧性の汎用的な検出基準の設計

翻訳エージェントがスタイルの曖昧性を判定する基準は LLM の内部推論に強く依存するため、特定の LLM に依存したシナリオは望ましくない。特に、LLM はプロンプトによって大きく性能に差が生じやすいため、複数の LLM から翻訳エージェントを構築できるように、曖昧性検出基準を設計する必要がある。

複雑なシナリオに頑健なエージェント制御

従来は語義の曖昧性のみを対象としていたが、スタイルの曖昧性も対象とすることで処理手順が増え、シナリオが複雑化する。そのため、シナリオの語彙を適切に定義し、インタラクションを繰り返しても正確にエージェントを制御できるシナリオを記述する必要がある。

以下、本論では、第 2 章で関連研究の翻訳エージェントと翻訳におけるスタイルの扱いについて説明する。次に、第 3 章では、レジスター理論とそれを用いたアクション定義とイベント定義について述べる。続いて第 4 章では、翻訳エージェントとして制御するシナリオについて述べる。そして、第 5 章で第 3 章と第 4 章で説明を行った翻訳エージェントのプロンプトに対する評価を行い、第 6 章ではその結果についての考察を行う。最後に、第 7 章では、今後の展望や課題について述べて結論とする。

第2章 関連研究

本章では、翻訳エージェントについての説明と、翻訳におけるスタイルの扱いについて説明する。

2.1 翻訳エージェント

近年、Transformer に代表される深層学習モデルの発展により、機械翻訳の性能は大きく向上している。新聞記事や報道文書を対象とした標準的な翻訳タスクにおいては、人間翻訳者と同等の品質に到達し得ることも報告されている[1]。一方で、これらのモデルは文脈や意味表現を暗黙的に学習できる反面、どの根拠や条件に基づいて訳語や表現が選択されたのかが明示されにくいという課題を抱えている。また、文脈に応じて語が複数の意味を持ち、どの語義を選択すべきかが一意に定まらないという語義の曖昧性の問題は、機械翻訳における主要な課題の一つとして長年研究されてきた。この問題に関しては、語義の曖昧性が翻訳精度に与える影響を分析し、語義の曖昧性解消を翻訳処理に組み込む必要性を示した研究が報告されている[2]。しかし、これらの研究は、翻訳モデル内部において語義の曖昧性を解消することを目的としており、曖昧性を明示的に検出してユーザに確認を求める仕組みは導入されていない。このように、既存の機械翻訳研究では、曖昧性が翻訳結果に与える影響が指摘されているものの、翻訳過程における曖昧性を明示的に扱い、ユーザとの対話を通じて解消する枠組みについては十分に検討されていない。これらの研究に対し、先行研究[3]では、多義語や文化依存語を含む入力文に対して、翻訳過程において語義の曖昧性を検出し、必要に応じてユーザに明確化質問を行う対話型翻訳エージェントが提案されている。特に、翻訳エージェントの振る舞いを構造化されたシナリオとして明示的に記述・制御している点に特徴がある。具体的には、翻訳処理を複数の状態に分割し、曖昧性の検出やユーザ入力の受信といった条件に応じて状態遷移を行うことで、一貫した翻訳プロセスを実現している。これらの状態遷移は、Ishida らによって提案された、エージェントと人間のコミュニティ間の柔軟な相互作用を記述するための枠組みであるインタラクション設計言語「Q」[4]に基づいて設計された YAML 形式のプロンプトとして記述されている。また、モデルの性能に応じたプロンプト設計の有効性が報告されており、高精度な LLM においては、複数回の対話を伴う翻訳過程においても安定した制御が可能である

ことが示されている。しかしながら、この先行研究において翻訳エージェントが扱う曖昧性は語義レベルに限定されているため、スタイルの違いも翻訳判断の対象として扱うことを目的とし、翻訳エージェントの拡張について検討する。

2.2 翻訳におけるスタイルの扱い

翻訳におけるスタイルとは、訳文が想定する読者や使用場面、翻訳目的に応じて選択される言語形式のあり方を指し、語彙選択や文体、談話構造などを含む総合的な概念である。翻訳研究では、同一の原文であっても、使用状況や目的が異なれば、異なるスタイルの訳文が妥当となりうることが指摘されてきた。このことは、スタイルが翻訳において付加的な要素ではなく、意味解釈や受け手の理解に直接影響を与える重要な要素であることを示している。翻訳品質評価の研究において、**Juliane House** は、翻訳を状況に埋め込まれた言語使用を他言語において再構成する行為として捉え、翻訳の適否は原文と訳文の形式的な一致のみから判断できるものではないと述べている[5]。Houseによれば、翻訳の目的や想定読者、文化的・社会的条件といった状況的要因を考慮することが不可欠であり、この立場は、翻訳におけるスタイルが翻訳品質に本質的に関与する要素であることを理論的に裏付けている。一方で、翻訳時にどのスタイルを選択すべきかといった判断は、原文の言語情報のみから一意に導けるものではない。翻訳の使用状況や目的は多くの場合明示されておらず、その結果、複数のスタイルがいずれも妥当な訳として成立しうる。すなわち、翻訳におけるスタイル選択は、単一の正解を持たず、判断の余地を含む点に特徴がある。また、翻訳品質の評価には、原文および訳文の特徴を記述・分析し、それらがどのような条件のもとで選択されたのかを説明する過程が不可欠であると指摘している。このことは、翻訳時のスタイル選択が直感に決定できるものではなく、判断の根拠となる基準を必要とすることを示唆している。しかしながら、実際の翻訳過程や自動翻訳においては、そのような基準が明示されないままスタイルが選択される場合が多く、スタイルの適切性が事後的に評価されるにとどまっている。以上より、翻訳におけるスタイルは翻訳品質にとって重要な要素である一方で、原文のみから適切なスタイルを一意に決定することは困難であり、スタイル選択そのものが曖昧な判断を伴う課題であるといえる。そこで、翻訳時に生じるスタイルの曖昧性に着目し、翻訳過程においてこれをどのように捉え、扱うべきかについて検討する。

第3章 スタイルを考慮した翻訳エージェント

本章では、レジスタ理論についての説明と、スタイルを考慮した翻訳エージェントを制御するための構造化された YAML プロンプトのアクションおよびイベントの定義について説明する。

3.1 レジスタ理論

スタイルの曖昧性判定には、M. A. K. Halliday によって提唱されたレジスタ理論[6][7]を用いる。本理論では、言語使用は常に特定の状況文脈に埋め込まれており、その文脈的特徴は、**Field**, **Tenor**, **Mode** の 3 つのレジスタ変数によって体系的に記述される。これらの変数は、発話が「何について」「誰によって誰に向けて」「どのような形式で」行われているかをそれぞれ異なる側面から捉える枠組みである。まず **Field** は、談話が扱う活動の種類や内容領域、すなわち発話が関与する社会的行為の性質を表す。具体的には、日常的な雑談や感情表現のような非専門的な活動であるか、業務連絡、学術的説明、技術的手順の提示といった専門性の高い活動であるかといった点が **Field** に含まれる。例えば、「昨日は楽しかったね」という発話と、「実験条件は三つの変数によって制御される」という発話では、扱われる活動内容や要求される語彙および構文の性質が大きく異なり、これらの違いは **Field** の差異として捉えられる。次に **Tenor** は、話者と聞き手の社会的関係性、役割、相対的な立場や心理的距離を表す変数である。これには、上下関係の有無、初対面か親密な関係か、個人的な対話か制度的・公的な場面での発話かといった要素が含まれる。例えば、同僚間でのカジュアルなやり取りと、顧客や上司に対する公式な連絡では、同じ内容であっても求められる丁寧さや言語形式が異なる。**Tenor** は、このような対人関係の違いが言語選択に及ぼす影響を記述するための概念である。最後に **Mode** は、言語が使用される媒体や伝達形式を指し、発話がどのような形態で実現されているかを表す。具体的には、話し言葉か書き言葉か、即時的な対話か一方向的な情報伝達か、対面でのやり取りかメッセージや文書によるコミュニケーションかといった点が含まれる。例えば、口頭での会話では省略や曖昧な表現が許容されやすい一方、文書やメールでは明示的で構造化された表現が求められる場合が多く、このような違いは **Mode** の差異として説明される。このように、**Field**, **Tenor**, **Mode** はそれぞれ異なる側面から状況文脈を記述する変数であり、これらの組み合わせに

よって、特定の言語使用における適切なスタイルが規定される。以上のレジスター理論に基づき、本研究では、これら 3 つのレジスター変数のうち、スタイル決定に不可欠な情報が入力文中に明示されていない、あるいは文脈から一意に推定できない場合を、スタイルの曖昧性が存在すると定義する。

3.2 アクション定義

アクションは、イベントが発生した際に実行される動作である。従来研究の翻訳エージェントからスタイルの曖昧性をユーザとのインタラクションを通じて明確にし、翻訳を生成する翻訳エージェントに拡張するにあたり、アクションを 2 個追加した。具体的には、「スタイル曖昧性確認」、「スタイルの曖昧性に対する説明を求める」の 2 個のアクションを作成した。

3.2.1 スタイル曖昧性確認

スタイル曖昧性確認 (`check_style_ambiguity`) は、スタイルの曖昧性確認アクションである。本アクションでは、Chat GPT 自身が受信したメッセージを解析し、スタイルの曖昧性が存在するかを判断する。本研究におけるスタイルの曖昧性とは、翻訳時に採用すべき文体、丁寧さの程度、想定される話者・聞き手の関係性、あるいは使用状況が入力文から一意に定まらない状態を指す。すなわち、同一の意味内容に対して複数のスタイルが成立し得る場合を、スタイルの曖昧性が存在する状態として扱う。例えば、「**How was your day?**」という文は、「一日の様子を尋ねる」という発話意図自体は明確であるが、使用状況や対人関係が明示されていないため、スタイルの観点からは複数の解釈が成立する。親しい友人に対して用いられる場合には、「今日どうだった?」といったカジュアルな表現が自然である一方、職場の上司や顧客に対しては、「本日はどのような一日でしたか」といった丁寧な表現が適切となる。レジスター理論の観点から見ると、本例では **Field** は日常的な問いかけとして比較的明確である一方、**Tenor**、すなわち話者と聞き手の関係性や社会的距離が特定できない。また、**Mode** についても、口頭での会話なのか、メッセージや文書として用いられるのであるか不明確である。このように、スタイル決定に必要なレジスター変数が十分に定まらない場合、本アクションはスタイルの曖昧性が存在すると判定する。なお、比較のため、本研究ではレジスター理論を用いないスタイルの曖昧性判定についても検討する。`check_style_ambiguity` の全体的な定義を図 1 に示す

- レジスター理論なし
 - def_action:
 - name: "check_style_ambiguity"
 - function: "ChatGPT itself will determine whether the received message contains a style ambiguity."
- レジスター理論あり
 - def_action:
 - name: "check_style_ambiguity"
 - function: "ChatGPT itself will determine whether the received message contains a style ambiguity by analyzing the three register variables (Field, Tenor, Mode) based on Halliday's Register Theory."
 - definitions:
 - Field: "Refers to what is happening — the nature of the social activity or subject matter taking place within the situatio
 - Tenor: "Refers to who is taking part — the relationships, roles, and relative status of the participants in the situatio
 - Mode: "Refers to what part the language is playing — the channel, medium, and rhetorical mode through which the communication occurs."

図 1 check_style_ambiguity の定義

3.2.2 スタイル明確化質問

スタイル明確化質問 (**ask_style_clarification**) は、スタイルの曖昧性を解消するために、ユーザに対して追加の説明を求めるアクションである。本アクションは、**check_style_ambiguity** によって入力文にスタイルの曖昧性が存在すると判定された場合のみ実行される。これにより、スタイルがすでに一意に定まっている入力に対して不要な質問を行うことを防ぎ、対話全体の効率性を維持する。本アクションでは、メッセージ内のスタイルの曖昧性を明確にするために、ユーザに対して英語で具体的な質問を生成する。質問内容は、本研究で採用するレジスター理論に基づき、**Field, Tenor, Mode** のうち、**check_style_ambiguity** によって曖昧であると判定されたレジスター変数に対応する情報を対象とする。すなわち、スタイル決定に必要な要素のうち、入力文から特定できなかった側面のみを選択的に問い返す設計となっている。例えば、**Tenor** に関する情報が不足している場合には、想定される対象読者や話者と聞き手の関係性を明確にする質問を生成する。具体例として、「How was your day?」という文に対してスタイル曖昧性が検出された場合、本アクションは“**Could you please clarify the relationship between you and the person you are addressing?**”のように、対人関係に関する情報を求める質問を提示する。この質問により、発話が親しい間柄での私的な会話であるのか、あるいは公的・業務的な場面での発話であるのかを判別できる。このように、本アクションを導入することで、ユーザの意図に基づいたスタイル情報を明示的に取得することが可能となる。なお、比較のため、本研究ではレジスター理論を用いないスタイル明確化質問についても検討する。**ask_style_clarification** アクションの全体的な定義を図 2 に示す。

- レジスター理論なし
- def_action:
 - name: "ask_style_clarification"
 - function: "To clarify a style ambiguity in the user's message, ask a specific question in English and display only the question without any additional comments."
- レジスター理論あり
- def_action:
 - name: "ask_style_clarification"
 - function: "To clarify a style ambiguity in the user's message that involves ambiguous register variables (Field, Tenor, Mode), ask a specific question in English and display only the question without any additional comments."
 - definitions:
 - Field: "Refers to what is happening — the nature of the social activity or subject matter taking place within the situation."
 - Tenor: "Refers to who is taking part — the relationships, roles, and relative status of the participants in the situation."
 - Mode: "Refers to what part the language is playing — the channel, medium, and rhetorical mode through which the communication occurs."

図 2 ask_style_clarification の定義

3.3 イベント定義

イベントは、状態遷移を引き起こす条件である。従来研究の翻訳エージェントからスタイルの曖昧性を考慮する翻訳エージェントに拡張するにあたり、イベントを 1 個追加した。

- **detected_style_ambiguity** : 受信したメッセージでスタイルの曖昧性が検出されたときに発生する。

detected_style_ambiguity イベントの定義は図 3 に示す。

- def_event:
 - name: "detected_style_ambiguity"
 - content: "This event is triggered when the system detects a style ambiguity in the message."

図 3 detected_style_ambiguity の定義

第4章 翻訳エージェントのシナリオ

本章では、翻訳エージェントを制御するための汎用的なシナリオの全体構造について説明する。さらに、第3章で記述したアクション、イベントを組み合わせたシナリオ全体の設計についても説明する。

4.1 シナリオの全体設計

従来研究の翻訳エージェントに対して、本研究では新たにイベントやアクションを追加した結果、シナリオが複雑化した。そこで、本研究では、各イベントおよびアクションの役割を明確に分離し、状態遷移の流れが直感的に理解できる明確かつ簡潔な構造を採用した。汎用的なシナリオの全体の状態遷移図は図4のように設計される。具体的な状態遷移は、「メッセージ待機」、「語義確認」、「語義明確化」、「スタイル確認」、「スタイル明確化」の5個で構成した。第3章で説明した通り、アクションは5個、イベントは3個使用した。この状態遷移図の「/」より前が条件であるイベントでそのイベントが起こった場合、動作であるアクションを行うように記述している。以下でそれぞれの状態の設計についても説明する。

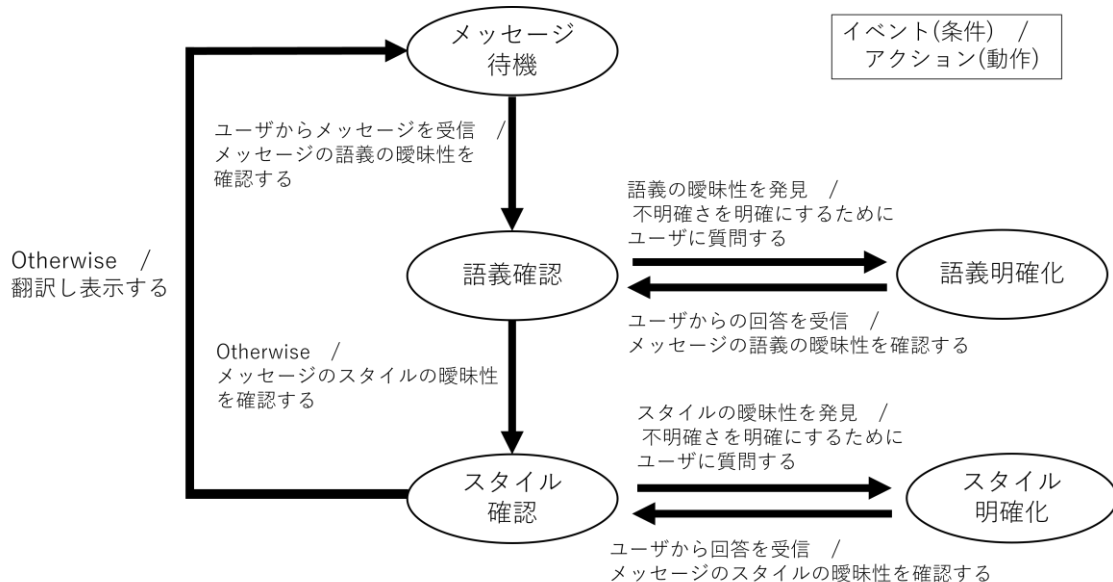


図 4 シナリオ全体の状態遷移図

4.1.1 メッセージ待機状態

「メッセージ待機」状態の状態遷移図は、図5のように設計される。「メッセージ待機」状態から「語義確認」状態へ移動するときのイベントは、ユーザからメッセージを受信するであり、このイベントが起きるとメッセージの語義の曖昧性を確認するアクションを行う。この「メッセージ待機」状態の状態遷移図をYAMLにしたものが図6である。

4.1.2 語義確認状態および語義明確化状態

「語義確認」状態から「語義明確化」状態へ移動するときのイベントは、語義の曖昧性を発見するであり、このイベントが起きると語義の曖昧性を明確にするためにユーザに質問するアクションを行う。逆に、「語義明確化」状態から「語義確認」状態へ移動するときのイベントは、ユーザからの回答を受信するであり、このイベントが起きるとメッセージの語義の曖昧性を確認するアクションを行

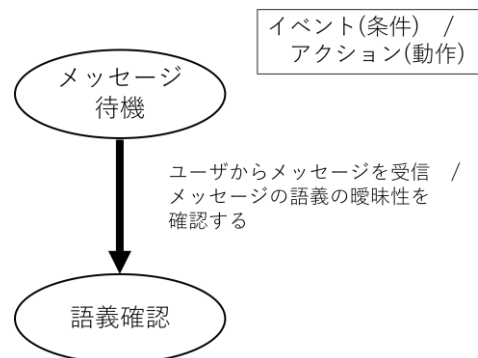


図5 メッセージ待機状態の状態遷移図

```
def_scenario:  
  name: "Message Clarification and Translation"  
  states:  
    - name: "Waiting for Message"  
      rules:  
        - event: "received_message"  
          actions:  
            - action: "check_ambiguity"  
          go_state: "WordSense Confirmation"
```

図6 メッセージ待機状態のYAML

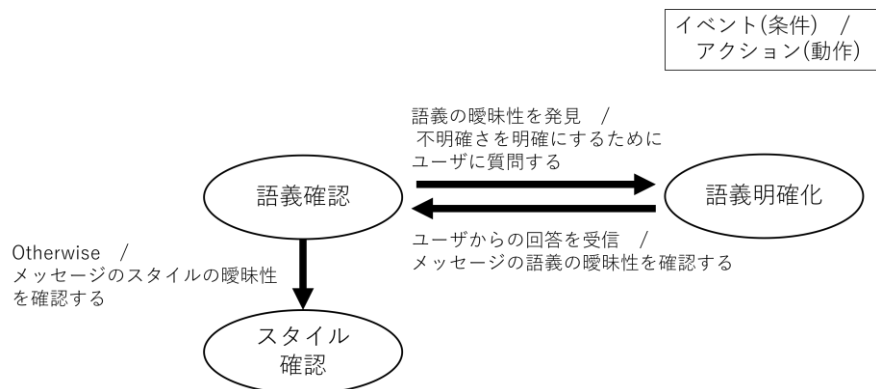


図 7 語義確認および語義明確化状態の状態遷移図

```

- name: "WordSense Confirmation"
  rules:
    - event: "detected_ambiguity"
      actions:
        - action: "ask_clarification"
      go_state: "WordSense Clarification"
    - event: "otherwise"
      actions:
        - action: "check_style_ambiguity"
      go_state: "Style Confirmation"

- name: "WordSense Clarification"
  rules:
    - event: "received_message"
      actions:
        - action: "check_ambiguity"
      go_state: "WordSense Confirmation"
  
```

図 8 語義確認状態および語義明確化状態の YAML

う。「語義確認」状態で語義の曖昧性を発見しないイベントが起こったとき、「スタイル確認」状態へ移動する。この「語義確認」状態および「語義明確化」状態の状態遷移図が図 7 で、YAML にしたものが図 8 である。

4.1.3 スタイル確認状態およびスタイル明確化状態

「スタイル確認」状態から「スタイル明確化」状態へ移動するときのイベントは、スタイルの曖昧性を発見するであり、このイベントが起きるとスタイルの曖昧性を明確にするためにユーザに質問するアクションを行う。逆に、「スタイル明確化」状態から「スタイル確認」状態へ移動するときのイベントは、ユーザからの回答を受信するであり、このイベントが起きるとメッセージのスタイルの曖昧性を確認するアクションを行う。「スタイル確認」状態でスタイルの曖昧性を発見しないイベントが起こったとき、「メッセージ待機」状態に移動し、翻訳し表示するアクションを行う。この「スタイル確認」状態および「スタイル明確

化」状態の状態遷移図が図 9 で、YAML にしたものが図 10 である。

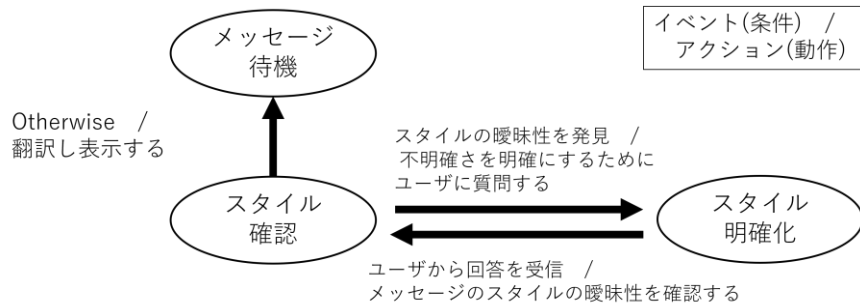


図 9 スタイル確認およびスタイル明確化状態の YAML

```

- name: "Style Confirmation"
  rules:
    - event: "detected_style_ambiguity"
      actions:
        - action: "ask_style_clarification"
          go_state: "Style Clarification"
    - event: "otherwise"
      actions:
        - action: "translate_and_display"
          go_state: "Waiting for Message"

- name: "Style Clarification"
  rules:
    - event: "received_message"
      actions:
        - action: "check_style_ambiguity"
          go_state: "Style Confirmation"
  
```

図 10 スタイル確認およびスタイル明確化状態の状態遷移図

第5章 評価

本章では、スタイルの曖昧性を考慮した翻訳エージェントの実現を目的とするため、第3章と第4章で説明したレジスター理論を定義したプロンプトと定義していないプロンプトを、同一のYAMLプロンプトおよび同等の動作を行う自然言語プロンプトに適用した比較実験について説明する。初めに、評価指標について説明し、次に評価手法について説明する。最後に、結果について説明する。

5.1 評価指標

翻訳エージェントのレジスター理論を定義したプロンプトと定義していないプロンプトを、同一のYAMLプロンプトおよび自然言語プロンプトに適用した性能を評価するために設定した評価項目について説明する。評価は、スタイルの曖昧性検査精度、スタイル一致率、スタイルの明確化質問の妥当性、シナリオ逸脱頻度の4つの観点から行う。これにより、レジスター理論を定義したプロンプトの有効性を検証し、エージェントがユーザとのインタラクションを通じて適切に動作するか確認する。以下に、それぞれの評価項目について詳細を示す。

5.1.1 曖昧性検査精度

曖昧性検査精度とは、翻訳エージェントが入力文に含まれるスタイルの曖昧性を正しく検出し、それに基づいて適切な判断を行える能力を評価するための指標である。本研究では、語義の曖昧性を含まず、スタイルの曖昧性を含むテスト文を用い、語義の曖昧性と誤って判断することなく、スタイルの曖昧性のみを正しく検出し、明確化質問を行ったうえで翻訳した場合を正解とみなす。曖昧性検査精度は、スタイル曖昧性を含む英文の総数に対して、エージェントがスタイルの曖昧性を正しく検出し、確認質問を行ったうえで翻訳をした英文の割合を%で示す。具体的な計算式は、

$$\frac{\text{スタイルの曖昧性を検出し、質問および翻訳をした回数}}{50}$$

である。

5.1.2 スタイル一致率

スタイル一致率とは、ユーザの指示内容に基づき、スタイルが適切に反映された翻訳が生成されているかを評価する指標である。スタイルの曖昧性を含む入力文に対して、エージェントが適切な明確化質問を行わずに翻訳を生成した

場合、その翻訳は正しいスタイルで翻訳されたものとはみなさない。スタイル一致率は、スタイルの曖昧性を含む英文の総数に対して、エージェントがスタイルの曖昧性を正しく検出し、明確化質問を行い、得られた追加情報に基づいて適切なスタイルで翻訳を生成できた英文の割合を%で表したものとする。具体的な計算式は、

$$\frac{\text{ユーザの指示通りのスタイルに翻訳できた回数}}{50}$$

である。また、スタイル明確化質問が行われた英文を対象として、指定されたスタイルで正確に翻訳された文の適合率についても評価を行う。

5.1.3 スタイル明確化質問の妥当性

スタイル明確化質問の妥当性とは、スタイルの曖昧性が検出された入力文に対して、翻訳エージェントが生成する明確化質問が、曖昧性の解消に必要な情報を適切に問いかけているかを評価する指標である。本研究では、生成されたスタイル明確化質問が、レジスター変数のいずれかに対応し、かつ実際に曖昧となっている要素を対象としている場合に、正しい質問であるとみなす。スタイル明確化質問の妥当性の評価には、Field, Tenor, Mode の各クラスとそれらに基づくマクロ平均およびマイクロ平均の適合率、再現率、F 値を用いる

5.1.4 シナリオ逸脱頻度

シナリオ逸脱頻度とは、翻訳エージェントが事前に定義したシナリオから逸脱し予期しない動作をする回数を測定する評価指標である。具体的には、ユーザとのインタラクションを繰り返す中で、エージェントがシナリオから外れた場合の回数を測定する。

5.2 評価手法

5.2.1 ベースライン

比較対象として、レジスター理論を定義したプロンプトと、定義をしていないプロンプトを用いる。これらのプロンプトは、同一の YAML プロンプトおよび自然言語プロンプトの二形式で設計した。YAML プロンプトは、第 3 章および第 4 章で示したように、各イベント、アクション、および状態遷移が階層的に定義されている。この YAML 形式の構造化されたプロンプトを基に、自然言語プロンプトを作成した。自然言語プロンプトでは、YAML において明示的に定義されていた状態遷移やルール、形式的な構文要素を省略し、ユーザがエージェン

トの動作を直感的に理解できるような記述へと変換した。

5.2.2 比較する LLM の種類

本研究では、複数の LLM から翻訳エージェントを構築するために、以下の LLM を使用した。各モデルは異なる規模や設計を持つが、モデルに左右されずに汎用的に翻訳エージェントを動かすことができるかを検証した。LLM の GPT のモデルは、オープン AI が提供する最先端の自然言語処理モデルであり、多様なタスクにおいて高い汎用性を持つ。本研究では、以下の 3 個のモデルを採用した：

- GPT-3.5-turbo : 軽量で応答速度が高く、幅広いタスクに対応可能なモデル
- GPT-4.1-mini : GPT-4.1 の軽量版で、計算リソースを抑えつつ、高い指示理解能力と安定した推論精度を備えたモデル
- GPT-5-mini : GPT-5 の軽量版で、高度な推論性能を維持しながら、高速な応答が可能なモデル

5.2.3 データセット

比較実験には、語義の曖昧性を含まず、スタイルの曖昧性を含むテスト文と語義の曖昧性を含まず、スタイルの曖昧性を含まないテスト文を用いた。このように、曖昧性を含む文と含まない文を組み合わせる理由は、実際の翻訳タスクにおいて、曖昧性を含む文章と明確な文章が混在しているためである。この構成により、翻訳エージェントの性能を実際の利用環境に近い条件で評価するとともに、スタイル曖昧性を解消するために設計したプロンプトの効果を検証することを目的とする。スタイルの曖昧性を含むテスト文については、IWSLT2022 フォーマリティ制御翻訳データセット[8]から、語義の曖昧性を含まず、スタイルの曖昧性を含む英文を選定した。具体的には、“it”や“that”といった指示語について、文脈から指示対象が特定可能な場合のみを対象として選定した。同様に、語義の曖昧性を含まず、スタイルの曖昧性を含まないテスト文については、JParaCrawl の English–Japanese 学習データセット[9]から指示語の指示対象が文脈から一意に特定可能であり、スタイルについても文脈から一意に特定可能な文のみを選定した。

全体のテスト文の総数は 100 文である。その内訳は、以下の通りである。

- 語義の曖昧性を含まず、スタイルの曖昧性を含む : 50 文
 - Field : 10 文
 - Tenor : 30 文

- Mode : 10 文

- 語義の曖昧性を含まず，スタイルの曖昧性を含まない : 50 文

5.2.4 実験手順

本研究では，レジスター理論を定義したプロンプトと定義していないプロンプトを，同一の YAML プロンプトおよび自然言語プロンプトに適用し，比較実験を行う．レジスター理論を導入により，シナリオの有効性が向上することを目的とし，2 個の異なる実験を実施した．以下で，具体的に説明する．

1 つ目，「曖昧性検査の網羅性実験」では，スタイルの曖昧性検査精度，スタイル一致率，スタイル明確化質問の妥当性の分析を行った．曖昧性検査の網羅性実験では，プロンプトを英文を送るごとに一緒に送り，その前の返答分は受け継がないものとし，さまざまな英文に対応できるかといった網羅性を検証した．

2 つ目の「対話継続性の評価実験」では，同一コンテキストで翻訳を依頼し，ユーザとのインタラクションを繰り返すことで，シナリオの逸脱頻度，スタイルの曖昧性検査精度，スタイル一致率，スタイル明確化質問の妥当性の分析を行った．対話継続性の評価実験では，連続した対話の中でのエージェントの継続性を評価した．

5.3 実験結果

5.3.1 曖昧性検査の網羅性実験

曖昧性検査の網羅性実験では，GPT-3.5-turbo, GPT-4.1-mini, GPT-5-mini のそれぞれのモデルにおける結果を表に示す．なお，曖昧性検査精度については，質問発生率も併せて示す．質問発生率とは，曖昧性を含むテスト文および曖昧性を含まないテスト文それぞれにおいて，語義質問が行われた割合とスタイル質問が行われた割合を指す．以下に，各モデルについての結果を説明する．

GPT-3.5-turbo モデル 表 1 に示す通り，YAML プロンプトではすべての評価結果が 0%であったため，以降の評価では自然言語プロンプトのみを対象とする．曖昧性検査精度は，レジスター理論を定義した条件のみが検出され，語義質問の発生が抑制される一方，曖昧性を含むテスト文におけるスタイル明確化質問の発生率が向上した．さらに，スタイル一致率は 2%向上し，適合率も 0.5 と高い値を示した．一方，表 2, 3 に示すスタイル明確化質問の妥当性は全体として低く，Tenor 質問のみが生成された．適合率は 1.0 であったが，再現率および F 値

は低く、適切な質問が生成される場合はあるが、発生頻度は限定的である。

GPT-4.1-mini モデル 表 4 に示す通り、レジスター理論の導入により、曖昧性検査精度は自然言語プロンプトでは 24%向上した一方、YAML プロンプトでは 4%低下した。質問の発生率は、YAML プロンプトでは語義質問が高頻度で生成され、スタイル質問は相対的に少なかったのに対し、自然言語プロンプトでは語義質問が抑制され、特に曖昧性を含まないテスト文では質問生成が抑えられた。自然言語プロンプトにレジスター理論を定義した条件では、スタイル一致率と適合率が高かったのに対し、YAML プロンプトではレジスター理論を定義しない条件のスタイル一致率が高く、適合率についてはいずれの条件でも 1.0 と高い値を示した。また、表 5, 6 に示すスタイル明確化質の妥当性は、いずれのプロンプトにおいても一定程度確認された。

GPT-5-mini モデル 表 7 に示す通り、曖昧性検査精度は、いずれのプロンプトにおいてもレジスター理論を定義した条件において一貫して高い値を示し、特にYAML プロンプトにおいて顕著であった。質問の発生率に関しては、自然言語プロンプトではレジスター理論を定義した条件においてのみ、語義質問およびスタイル質問が低頻度で生成された。一方、YAML プロンプトでは、レジスター理論を定義した条件において語義質問の発生率が抑制され、スタイル質問の発生率は 80%まで向上した。しかし、この傾向は曖昧性を含まないテスト文においても確認され、曖昧性の有無にかかわらず質問が生成されていることが示された。スタイル一致率は、YAML プロンプトにおいてレジスター理論を定義した条件で高い値を示し、適合率についても同様の傾向を示した。さらに、表 8, 9 に示すスタイル明確化質問の妥当においてもYAML プロンプトでは、Tenor に関して高い適合率および再現率が得られ、マクロ平均およびマイクロ平均の F 値

表 1 GPT-3.5-turbo モデルの曖昧性検査精度と質問の内訳，スタイル一致率

GPT-3.5-turbo	曖昧性検査精度(%)	曖昧性ありテスト文(50文)における内訳		曖昧性なしテスト文(50文)における内訳		スタイル一致率(%)	スタイル適合率
		語義質問発生率(%)	スタイル質問発生率(%)	語義質問発生率(%)	スタイル質問発生率(%)		
自然言語 (レジスター理論あり)	4	4	4	2	6	2	0.5
自然言語 (レジスター理論なし)	0	84	0	56	0	0	0
YAML (レジスター理論あり)	0	0	0	0	0	0	0
YAML (レジスター理論なし)	0	0	0	0	0	0	0

表 2 GPT-3.5-turbo モデルのレジスター変数ごとの適合率，再現率，F 値

GPT-3.5-turbo	Field 適合率	Field 再現率	Field F値	Tenor 適合率	Tenor 再現率	Tenor F値	Mode 適合率	Mode 再現率	Mode F値
自然言語 (レジスター理論あり)	0.000	0.000	0.000	1.000	0.067	0.125	0.000	0.000	0.000

も高いことから、全体として妥当性の高い質問が生成されていたと示された。

表 3 GPT-3.5-turbo モデルのマクロ平均およびマイクロ平均の適合率，再現率，F 値

GPT-3.5-turbo	マクロ平均適合率	マクロ平均再現率	マクロ平均 F 値	マイクロ平均適合率	マイクロ平均再現率	マイクロ平均 F 値
自然言語 (レジスター理論あり)	0.333	0.022	0.042	1.000	0.040	0.077

表 4 GPT-4.1-mini モデルの曖昧性検査精度と質問の内訳，スタイル一致率

GPT-4.1-mini	曖昧性検査精度(%)	曖昧性ありテスト文(50文)における内訳		曖昧性なしテスト文(50文)における内訳		スタイル一致率(%)	スタイル適合率
		語義質問発生率(%)	スタイル質問発生率(%)	語義質問発生率(%)	スタイル質問発生率(%)		
自然言語 (レジスター理論あり)	48	22	54	4	6	30	0.833
自然言語 (レジスター理論なし)	24	36	24	2	0	20	0.625
YAML (レジスター理論あり)	6	88	34	64	44	6	1.000
YAML (レジスター理論なし)	10	88	24	68	38	12	1.000

表 5 GPT-4.1-mini モデルのレジスター変数ごとの適合率，再現率，F 値

GPT-4.1-mini	Field 適合率	Field 再現率	Field F 値	Tenor 適合率	Tenor 再現率	Tenor F 値	Mode 適合率	Mode 再現率	Mode F 値
自然言語 (レジスター理論あり)	0.176	0.300	0.222	0.700	0.233	0.350	0.000	0.000	0.000
YAML (レジスター理論あり)	0.333	0.100	0.154	0.571	0.267	0.364	0.000	0.000	0.000

表 6 GPT-4.1-mini モデルのマクロ平均およびマイクロ平均の適合率，再現率，F 値

GPT-4.1-mini	マクロ平均適合率	マクロ平均再現率	マクロ平均 F 値	マイクロ平均適合率	マイクロ平均再現率	マイクロ平均 F 値
自然言語 (レジスター理論あり)	0.292	0.178	0.191	0.370	0.200	0.260
YAML (レジスター理論あり)	0.302	0.122	0.172	0.529	0.180	0.269

表 7 GPT-5-mini モデルの曖昧性検査精度と質問の内訳，スタイル一致率

GPT-5-mini	曖昧性検査精度(%)	曖昧性ありテスト文(50文)における内訳		曖昧性なしテスト文(50文)における内訳		スタイル一致率(%)	スタイル適合率
		語義質問発生率(%)	スタイル質問発生率(%)	語義質問発生率(%)	スタイル質問発生率(%)		
自然言語 (レジスター理論あり)	4	8	2	4	2	0	0.000
自然言語 (レジスター理論なし)	0	0	0	0	0	0	0.000
YAML (レジスター理論あり)	72	4	80	10	98	68	0.944
YAML (レジスター理論なし)	50	30	34	6	32	20	0.625

表 8 GPT-5-mini モデルのレジスター変数ごとの適合率, 再現率, F 値

GPT-5-mini	Field 適合率	Field 再現率	Field F値	Tenor 適合率	Tenor 再現率	Tenor F値	Mode 適合率	Mode 再現率	Mode F値
自然言語 (レジスター理論あり)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
YAML (レジスター理論あり)	0.000	0.000	0.000	0.605	0.767	0.676	0.000	0.000	0.000

表 9 GPT-5-mini モデルのマクロ平均およびマイクロ平均の適合率, 適合率, 再現率, F 値

GPT-5-mini	マクロ平均 適合率	マクロ平均 再現率	マクロ平均 F値	マイクロ平均 適合率	マイクロ平均 再現率	マイクロ平均 F値
自然言語 (レジスター理論あり)	0.000	0.000	0.000	0.000	0.000	0.000
YAML (レジスター理論あり)	0.202	0.256	0.225	0.605	0.460	0.523

5.3.2 対話継続性の評価実験

対話継続性の評価実験では, GPT-3.5-turbo, GPT-4.1-mini, GPT-5-mini それぞれのモデルにおける結果を表に示す. 曖昧性検査精度については, 質問発生率についても示す. 以下に, 各モデルの結果を示す.

GPT-3.5-turbo モデル 表 10 に示すように, いずれのプロンプトにおいてもすべての結果が 0%であった. 想定していた質問応答の流れが維持されず, 質問文ではなく内部状態や処理途中を示すような出力が生成される事例が確認された. これらの出力, 事前に定義した対話シナリオや役割設定を逸脱しており, プロンプト設計の意図がモデルに十分に反映されていないことを示している.

GPT-4.1-mini モデル 表 11 に示す通り, 曖昧性検査精度は, レジスター理論の有無によって明確な差が認められた. レジスター理論の導入により, 自然言語プロンプトでは 32%向上し, YAML プロンプトでは 12%向上した. 質問の発生率は, プロンプト形式およびレジスター理論の有無によって異なる傾向がみられた. YAML プロンプトではレジスター理論を定義しない条件でスタイル質問が多く生成された一方, 自然言語プロンプトではレジスター理論を定義した条件の方が高い発生率を示した. スタイル一致率は, 自然言語プロンプトにおいてレジスターを定義した条件で 30%向上し, 適合率も向上した. また, 表 12, 13 に示すスタイル明確化質問の妥当性は, 自然言語プロンプトの方が全体として高

く、Field および Tenor の両要素が検出された点において差が認められた。シナリオの逸脱頻度については、自然言語プロンプトでは逸脱は確認されなかったのに対し、YAML プロンプトではレジスター理論を定義した条件で逸脱が増加する傾向がみられた。

GPT-5-mini モデル 表 14 に示す通り、曖昧性検査精度は、自然言語プロンプトにおいてレジスター理論を定義した条件で 92%と最も高い値を示した一方、YAML プロンプトではレジスター理論の有無による顕著な差は認められなかった。質問の発生率については、レジスター理論を定義しない条件ではいずれのプロンプトにおいても質問はほとんど生成されなかったのに対し、レジスター理論を定義した条件では、スタイル質問が 50%以上と高い割合で生成され、特にYAML プロンプトにおいて顕著であった。しかし、この傾向は曖昧性を含まないテスト文に対して確認されており、曖昧性の有無に応じた質問生成が十分に切り替えられていない可能性が示唆される。スタイル一致率は、自然言語プロンプトにおいてレジスター理論を定義した条件で 10%の向上が確認された。また、表 15, 16 に示すスタイル明確化質問の妥当性は、自然言語プロンプトの方が高く、

表 10 GPT-3.5-turbo モデルの曖昧性検査精度と質問の内訳，スタイル一致率

GPT-3.5-turbo	曖昧性検査精度(%)	曖昧性ありテスト文(50文)における内訳		曖昧性なしテスト文(50文)における内訳		スタイル一致率(%)	スタイル適合率
		語義質問発生率(%)	スタイル質問発生率(%)	語義質問発生率(%)	スタイル質問発生率(%)		
自然言語 (レジスター理論あり)	0	0	0	0	0	0	0
自然言語 (レジスター理論なし)	0	0	0	0	0	0	0
YAML (レジスター理論あり)	0	0	0	0	0	0	0
YAML (レジスター理論なし)	0	0	0	0	0	0	0

表 11 GPT-4.1-mini モデルの曖昧性検査精度と質問の内訳，スタイル一致率

GPT-4.1-mini	曖昧性検査精度(%)	曖昧性ありテスト文(50文)における内訳		曖昧性なしテスト文(50文)における内訳		スタイル一致率(%)	スタイル適合率	シナリオ逸脱頻度
		語義質問発生率(%)	スタイル質問発生率(%)	語義質問発生率(%)	スタイル質問発生率(%)			
自然言語 (レジスター理論あり)	64	10	64	0	2	36	0.562	0
自然言語 (レジスター理論なし)	32	10	40	0	0	6	0.187	0
YAML (レジスター理論あり)	59	60	24	0	0	22	0.916	9
YAML (レジスター理論なし)	47	4	92	0	2	84	0.893	1

表 12 GPT-4.1-mini モデルのレジスター変数ごとの適合率，再現率，F 値

GPT-4.1-mini	Field 適合率	Field 再現率	Field F値	Tenor 適合率	Tenor 再現率	Tenor F値	Mode 適合率	Mode 再現率	Mode F値
自然言語 (レジスター理論あり)	0.333	0.200	0.250	0.615	0.533	0.571	0.000	0.000	0.000
YAML (レジスター理論あり)	0.000	0.000	0.000	0.462	0.200	0.279	0.000	0.000	0.000

YAML プロンプトと比較して、Tenor に加えて Mode も検出された点で差が認められた。一方、シナリオの逸脱頻度は、YAML プロンプトにおいてレジスター理論を定義した条件のみで確認された。

表 13 GPT-4.1-mini モデルのマクロ平均およびマイクロ平均の適合率，再現率，F 値

GPT-4.1-mini	マクロ平均適合率	マクロ平均再現率	マクロ平均 F 値	マイクロ平均適合率	マイクロ平均再現率	マイクロ平均 F 値
自然言語 (レジスター理論あり)	0.316	0.244	0.274	0.562	0.360	0.439
YAML (レジスター理論あり)	0.154	0.067	0.093	0.462	0.120	0.190

表 14 GPT-5-mini モデルの曖昧性検査精度と質問の内訳，スタイル一致率

GPT-5-mini	曖昧性検査精度(%)	曖昧性ありテスト文(50文)における内訳		曖昧性なしテスト文(50文)における内訳		スタイル一致率(%)	スタイル適合率	シナリオ逸脱頻度
		語義質問発生率(%)	スタイル質問発生率(%)	語義質問発生率(%)	スタイル質問発生率(%)			
自然言語 (レジスター理論あり)	92	0	98	0	98	10	0.108	0
自然言語 (レジスター理論なし)	0	8	0	2	0	0	0.000	0
YAML (レジスター理論あり)	32	8	50	6	54	8	0.947	0
YAML (レジスター理論なし)	38	0	4	2	18	36	0.250	33

表 15 GPT-5-mini モデルのレジスター変数ごとの適合率，再現率，F 値

GPT-5-mini	Field 適合率	Field 再現率	Field F 値	Tenor 適合率	Tenor 再現率	Tenor F 値	Mode 適合率	Mode 再現率	Mode F 値
自然言語 (レジスター理論あり)	0.000	0.000	0.000	0.690	0.667	0.678	0.286	0.400	0.333
YAML (レジスター理論あり)	0.000	0.000	0.000	0.812	0.433	0.565	0.000	0.000	0.000

表 16 GPT-5-mini モデルのマクロ平均およびマイクロ平均の適合率，再現率，F 値

GPT-5-mini	マクロ平均適合率	マクロ平均再現率	マクロ平均 F 値	マイクロ平均適合率	マイクロ平均再現率	マイクロ平均 F 値
自然言語 (レジスター理論あり)	0.325	0.356	0.337	0.558	0.480	0.516
YAML (レジスター理論あり)	0.271	0.144	0.188	0.619	0.260	0.366

第6章 考察

本章では、5章の比較実験の結果から、曖昧性検査の網羅性実験と対話継続性の評価実験の考察について述べる。具体的には、曖昧性検査精度、スタイル一致率、スタイル明確化質問の妥当性、シナリオ逸脱頻度について述べる。

6.1 曖昧性検査精度

曖昧性検査の網羅性実験では、モデルの性能によって、レジスター理論およびプロンプト形式の影響の現れ方が大きく異なることが確認された。まず、**GPT-3.5-turbo** モデルにおいては、**YAML** プロンプト条件下でシナリオ破綻が発生し、すべての結果が **0%**であったことから、自然言語プロンプトのみを対象として検討した。その結果、レジスター理論を用いない条件では曖昧性は検出されなかった一方、レジスター理論を導入することで **4%**の曖昧性検出が確認された。これは、低精度モデルにおいても、レジスター理論を明示的に与えることで、スタイルや使用状況に基づく判断が部分的に可能になることを示唆している。ただし、全体として検出率は低く、曖昧性検査を安定して行うにはモデル能力が不十分であることが示された。一方、**GPT-4.1-mini** モデルでは、プロンプト形式とレジスター理論の有無によって異なる傾向を示した。自然言語プロンプトにおいてレジスター理論の導入による改善がみられたが、**YAML** プロンプトでは同様の効果は確認されなかった。この結果から、中精度モデルにおいては、自然言語による柔軟な文脈解釈がレジスター理論を活かしやすい可能性を示している。さらに、**GPT-5-mini** モデルにおいては、レジスター理論の有無による差が明確に表れた。特に顕著なのは **YAML** プロンプトであり、レジスター理論なし条件でも **50%**の検出がみられたのに対し、レジスター理論あり条件では **72%**まで向上した。このことから、高精度モデルにおいては、**YAML** による明示的な対話構造とレジスター理論の組み合わせが、曖昧性検査を強く促進することが示唆される。

対話継続性の評価実験では、曖昧性検査の網羅性実験と同様に、モデル間で明確な差が見られ、レジスター理論を導入した場合としない場合でも、モデルごとに異なる影響が確認された。まず、**GPT-3.5-turbo** モデルは、本実験で設計した複数段階の対話シナリオを、適切に追従できず、結果としてモデル内部の状態表現が直接出力されたものと解釈できる。以降、対話継続性の評価実験における

GPT-3.5-turbo モデルの考察については述べない。これに対して、GPT-4.1-mini モデルでは、いずれのプロンプトにおいても、レジスター理論を導入することで曖昧性検査精度の向上が確認された。特に、自然言語プロンプトにおいては、大きな改善が認められた。一方、YAML プロンプトにおいても同様の結果はみられたものの、自然言語プロンプトほど顕著な改善ではなかった。このことから、中精度モデルでは、構造化された YAML 形式そのものが一定水準の曖昧性検出を可能にする一方で、レジスター理論の効果は自然言語プロンプトにおいてより強く発揮されることが示唆される。一方、GPT-5-mini モデルでは、自然言語プロンプトにおいてレジスター理論の有無による差が顕著に表れた。レジスター理論を定義しない条件では曖昧性が検出されなかったのに対し、レジスター理論を定義した条件では 92% と極めて高い検出率を示した。これに対し、YAML プロンプトでは、レジスター理論の有無による明確な差は認められなかった。この結果は、高精度モデルにおいては、YAML プロンプトとレジスター理論の組み合わせが特に有効であることを示している。

6.2 スタイル一致率

曖昧性検査の網羅性実験では、モデルの性能によって、レジスター理論およびプロンプト形式がスタイル一致率に与える影響は大きく異なることが確認された。まず、GPT-3.5-turbo モデルでは、自然言語プロンプトのみを評価対象としたが、レジスター理論を定義した条件では文体一致率が 2% と、レジスター理論を定義しない条件と比較して向上はしたものの、スタイル制御に関する顕著な改善は確認されなかった。この結果は、低精度モデルにおいては、レジスター理論の導入がスタイル制御を十分に実現するには至っていない可能性を示唆している。次に、GPT-4.1-mini モデルでは、プロンプト形式によって異なる傾向が確認された。自然言語プロンプトにおいては、レジスター理論の導入により 10% 向上し、適合率においても向上したのに対し、YAML プロンプトでは同様の効果は認められなかった。このことから、中精度モデルにおいては、自然言語プロンプトで強い効果を発揮する可能性が示された。さらに、GPT-5-mini モデルにおいては、YAML プロンプトにおいてレジスター理論を定義した条件で最も高い結果が得られ、適合率についても 0.944 と高い結果を得られた。この結果から、高精度モデルにおいては、レジスター理論と YAML の組み合わせが有効であることが明確に確認された。

対話継続性の評価実験では、モデル間の差異に加えて、レジスター理論の有無による影響も確認された。GPT-4.1-mini モデルでは、自然言語プロンプトにおいてレジスター理論を定義した条件において、最大 36%と相対的に高い値を示し、適合率も向上した。この結果から、中精度モデルにおいては、自然言語プロンプトにおいてレジスター理論が翻訳結果におけるスタイル選択を補助する形で有効に機能している可能性が示唆される。同様に、GPT-5-mini モデルでは、自然言語プロンプトにおいてレジスター理論を定義した条件で 10%の向上と適合率の向上が確認されたものの、YAML プロンプトでは同様の改善は認められなかった。以上の結果より、高精度モデルにおいては、自然言語プロンプトにおいてレジスター理論が有効に作用するが、モデル内部の推論能力が高いため、レジスター理論による明示的な制御が対話継続性に安定して反映されにくい可能性が示唆される。

6.3 スタイル明確化質問の妥当性

曖昧性検査の網羅性実験では、モデルの性能およびプロンプト形式によって、妥当性の現れ方が大きく異なることが確認された。GPT-3.5-turbo では、全体として妥当性が低く、質問として有効であるとは言い難い。いずれのプロンプトにおいても、Field および Mode に関する質問は生成されず、Tenor に関する質問も限定的であった。生成された質問は内容として成立している場合があるものの、レジスター変数を網羅的に捉えたものではなく、スタイル明確化質問は偶発的に生成されるにとどまっていた。このことから、低精度モデルでは、対話を通じてスタイルを明確化するための質問として安定して機能しないことが示唆される。一方、GPT-4.1-mini モデルでは、プロンプト形式によらず Tenor において一定の F 値が得られた一方で、Field の F 値は低く、Mode の F 値は 0 であった。すなわち、本モデルの質問は Tenor に対しては相対的に妥当性が高いものの、レジスター変数を網羅する観点では妥当性が低いと評価できる。その結果、マクロ平均 F 値は低水準にとどまり一部要素に偏った妥当性という傾向が確認された。以上より、中精度モデルでは、スタイル明確化質問は特定の要素に対しては一定の妥当性を示すものの、スタイル全体を十分に確認する質問としては不十分であることが示された。これに対し、GPT-5-mini モデルでは、自然言語プロンプトにおいては、いずれの結果も 0 であり、質問としての妥当性は確認されなかった。一方、YAML プロンプトでは Tenor において高い F 値が得られ、

質問としての妥当性は高いと評価できる。ただし、**Field** および **Mode** に関する **F** 値は **0** にとどまった。その結果、マクロ平均 **F** 値は中程度にとどまる一方で、マイクロ平均 **F** 値は比較的高い値を示し、妥当な質問が一定数、かつ安定して生成されることが示された。以上より、高精度モデルでは、**YAML** プロンプトにおいてスタイル明確化質問の妥当性は高い水準に達するものの、その妥当性は特定の要素に依存しており、スタイル全体を包括的に確認する質問としては依然として課題が残ることが示された。

対話継続性の評価実験では、モデル間およびレジスター理論の有無によって明確な差異が確認された。**GPT-4.1-mini** モデルでは、プロンプト形式による違いが認められた。自然言語プロンプトでは、**Tenor** において比較的高い **F** 値が得られ、質問としての妥当性は一定水準にあると評価できる。一方で、**Field** の **F** 値は中程度にとどまり、**Mode** に関しては妥当性が確認されなかった。その結果、マクロ平均 **F** 値は限定的であり、スタイル全体を網羅する質問には至っていない。**YAML** プロンプトでは **Tenor** の **F** 値が自然言語プロンプトより低下し、マクロ平均およびマイクロ平均 **F** 値も低水準にとどまった。以上より、中精度モデルでは、自然言語プロンプトにおいて一定の妥当性が確認されるものの、プロンプト形式にかかわらず、スタイル明確化質問の妥当性は主として **Tenor** に依存していると示された。一方、**GPT-5-mini** モデルでは、いずれのプロンプトにおいても相対的に高い妥当性が確認された。自然言語プロンプトでは、**Tenor** において高い **F** 値が得られ、さらに **Mode** においても **F** 値が確認されたことから、複数のレジスター変数に対して妥当な質問が生成されていると評価できる。**YAML** プロンプトでは、**Tenor** において一定の **F** 値が得られたものの、**Mode** に関する **F** 値は **0** であり、自然言語プロンプトと比較して妥当性はやや低下した。その結果、マクロ平均 **F** 値は自然言語プロンプトの方が高く、レジスター変数全体を考慮した場合の妥当性に差が認められた。以上より、高精度モデルでは、スタイル明確化質問は全体として高い妥当性を示すものの、その水準はプロンプト形式に依存しており、自然言語プロンプトの方がより有効であることが示された。

6.4 シナリオ逸脱頻度

GPT-4.1-mini モデルでは、シナリオ逸脱の発生は **YAML** プロンプトに限られていたが、全体として低水準に抑えられていた。多くの条件において、曖昧性検

出から質問生成、翻訳へと処理が順序通りに進行しており、対話構造の破綻はほとんど確認されなかった。レジスター理論の有無による顕著な差は認められなかったものの、理論を明示した条件では処理分岐が複雑化し、シナリオが破綻しやすい傾向が見られた。このことから、中精度モデルでは、処理手順や対話構造が比較的明確に維持されやすく、構造化された指示やレジスター理論を与えた場合においても、想定されたシナリオに沿った安定した応答が生成されやすい傾向が確認された。一方、**GPT-5-mini** モデルでは、シナリオ逸脱は **YAML** プロンプトにおいてレジスター理論を定義しない条件でのみ確認された。特に、構造化入力や複数条件が同時に与えられた場合に、想定外の応答が生成され、対話の進行が不安定になる傾向が見られた。このことから、高精度モデルでは、処理の自由度や内部推論の複雑化により、構造化条件下において想定外の応答が生じやすく、対話制御の安定性が必ずしも一貫しない可能性が考えられる。ただし、**GPT-3.5-turbo** モデルで確認されたような、内部状態が直接出力される深刻な逸脱は観察されなかった。以上より、シナリオ逸脱頻度の観点からは、**GPT-4.1-mini** モデルが最も安定した対話制御を実現しており、本研究で想定した評価シナリオに最も適合していることが示された。

6.5 まとめ

以上の結果から、中精度モデルを使用する場合には、レジスター理論を導入した自然言語プロンプトの方が安定した結果を得やすい。それに対し、高精度モデルを使用する場合は、レジスター理論を導入した **YAML** プロンプトの方が安定したスタイル制御を行える。一方、同一コンテキスト内で翻訳を依頼する場合は、レジスター理論を導入した自然言語プロンプトにおいて安定した結果を得ることができた。今後の課題として、プロンプトの設計を改善し、低精度モデルにおいてもレジスター理論や対話構造を適切に反映できるよう最適化することで、翻訳システム全体の安定性および精度の向上が求められる。

第7章 おわりに

本研究では、異なる LLM モデルとプロンプト形式 (YAML と自然言語)、レジスター理論の有無を用いた翻訳タスクにおいて、曖昧性検査精度、スタイル一致率、スタイル明確化質問の妥当性、シナリオ逸脱頻度を比較および分析した。その結果、中精度モデルを使用する場合には、レジスター理論を導入した自然言語プロンプトの方が安定した結果を得やすい一方、高精度モデルを使用する場合は、レジスター理論を導入した YAML プロンプトの方が安定したスタイル制御を行えることが示された。

本研究の貢献は以下の通りである。

スタイルの曖昧性の汎用的な検出基準の設計

Halliday のレジスター理論に基づいて曖昧性判定基準を設計した。これにより、特定のモデルに依存しない、汎用的なスタイルの曖昧性検出の基準を設けた。

複雑なシナリオに頑健なエージェント制御

高精度なモデルでは、レジスター理論を導入した YAML プロンプトを用いることで、曖昧性処理や文体制御が安定するのに対し、中精度モデルでは、レジスター理論を導入した自然言語プロンプトを用いることで、曖昧性検査やスタイル明確化質問が安定して行われやすいことが示された。

今後の課題としてプロンプトの設計を改善し、低精度モデルにおいてもレジスター理論や対話構造を適切に反映できるよう最適化することで、翻訳システム全体の安定性および精度の向上が求められる

謝辞

本研究を行うにあたり，懇切丁寧なご指導ならびに多くのご助言を賜りました村上陽平教授に，心より深く感謝申し上げます．また，日頃より研究活動を支えてくださった社会知能研究室の皆様にも，心より感謝いたします．

参考文献

- [1] Popel, M., Tomkova, M., Tomek, J., Kaiser, L., Uszkoreit, J., Bojar, O., Žabokrtský, Z.: Transforming machine translation: A deep learning system reaches news translation quality comparable to human professionals, *Nature Communications*, Vol. 11, Article No. 4381 (2020).
- [2] Ickrey, D., Biewald, L., Teyssier, M., Koller, D.: Word-sense disambiguation for machine translation, *Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing*, pp. 771–778 (2005).
- [3] 加奥咲江, 村上陽平, モンティールーパーピタクスワン: 翻訳エージェントのための構造化プロンプト, 第87回情報処理学会全国大会(2025)
- [4] Ishida, T.: Q: A scenario description language for interactive agents, *IEEE Transactions on Control of Network Systems*, Vol. 35, No. 11 (2020).
- [5] House, J.: *Translation Quality Assessment: Past and Present*, Routledge, pp. 1–170 (2014).
- [6] Urban, G.: Reviewed Work: *Language as Social Semiotic: The Social Interpretation of Language and Meaning* by M. A. K. Halliday, *American Anthropologist*, Vol. 83, No. 3, pp. 659–661 (1981).
- [7] Halliday, M. A. K., Matthiessen, C. M. I. M.: *Halliday's Introduction to Functional Grammar*, Routledge, pp. 1–808 (2013).
- [8] Anastopoulos, A., Barrault, L., Bentivogli, L., et al.: Findings of the IWSLT 2022 evaluation campaign, *Proceedings of the 19th International Conference on Spoken Language Translation (IWSLT 2022)*, pp. 98–157 (2022).
- [9] Morishita, M., Suzuki, J., Nagata, M.: JParaCrawl: A large-scale web-based English–Japanese parallel corpus, *Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020)*, pp. 3603–3609 (2020).

付録

A.1 テスト文

実験で用いたテスト文は、以下の通りである。

- 語義の曖昧性を含まず、スタイルの曖昧性検査を含むテスト文：50 文

- Field : 10 文

1. I really enjoyed myself.
2. I've never worked with a brick oven myself.
3. What's happening? Something weird is happening on my network.
Do you hear me now?
4. I don't have a DVD player. Why would I be carrying all my DVDs
with me?
5. As for myself, I would just study languages.
6. Okay. So the end goal is you're gonna become a mental health
counsellor. And you prefer to work with adults with mental health
issues and substance abuse issues, excuse me.
7. You own your personality that shapes who you are, and there's just
no guarantee—there's just no textbook that can teach you all these
things for sure.
8. What about your job? Do they actively recycle? No, I find most
restaurants in New York City do not care about recycling; they just
throw everything in the trash.
9. I don't even have to call and sit on hold and push the buttons or
whatever. I just enter my cell phone number, and they call me
automatically, and then I'm instantly connected to a person who
solves my problem.

10. I think even your PayPal email address.
- Tenor : 30 文
1. Can I talk to your manager, please?
2. Yeah, she is a friend of mine.
3. Since sixth grade, because they wanted me to switch to softball and play with the girls, and I didn't want to. I never played softball. I never played softball in my life, so I quit.
4. So, to me, my uncles and my aunts—they're not a part of my life. I wish they were, but they're not. And then it's just my mom, who's a single parent.
5. We don't think you should go if you can't find your wallet, and I was like, no, I can find it. So I got it, and then we went downtown.
6. Yep. And especially if you're lying too, and then you are intentionally getting caught in your lies. Then you also make yourself—you look not smart..
7. How was your day?
8. What are your favorite road trips that you've ever taken?
9. Is that your favourite pizza?
10. I'm doing great. How was your day so far?
11. They hate you; they just make up a reason to hate you.
12. No worries. Can I get your phone number that you are registered with us?
13. No, going back from Budapest, I actually had the opportunity when my grandmother was still alive. She grew up in Budapest.

14. Can I have your phone number, please?
15. So, what is your ideal boyfriend?
16. Oh, my gosh. You'll have some great stories for your future grandkids, let me tell you.
17. Do you shop on Amazon?
18. What was your favorite Disney movie growing up?
19. Okay. So, what is your favorite game show?
20. There was a long time where he said, "I am not going to have any of your non-Jewish friends at my Shabbat table." And I said, okay, that's your prerogative, it's your choice.
21. But it was really interesting for my wife to watch that. Not so much for me, because I kind of already know how that goes.
22. So what about your team? You mentioned that you have a pretty good relationship with your co-workers.
23. Okay, so for your career though, I think you're gonna have to look more professional. Is that something you've thought about at all?
24. So tell me about your childhood best friend.
25. If I could have an ideal profession, I did not have to worry about what it paid.
26. What's your ideal profession?
27. Oh my God! Why do you pay so much?
28. Your mom sends you useless gifts.
29. Parents and stuff like that—they're cool? They'll be bothering you.

30. How was your senior year in high school? Did you go to senior prom or anything?

- Mode : 10 文

1. They wanted me to be like an engineer. Science, you know what I mean. But I found my own way. You know what I'm saying? Word.
2. So let us see what we got there. But first of all, how was your party?
3. Like, honestly, head to bed very soon. I hope you have a very good rest of your night.
4. Do you want to list more people, or should I keep going? Let me keep on going. Let's see, I got—I have my friend.
5. You know, so they just asked me a lot of questions about whether or not I had ever worked over a forty-hour week.
6. Can I get your email so I can send all information to your email, please?
7. And so I'll just get it started by asking you this: Where's your favourite place you've ever been?
8. I remember my mom telling me, “Oh, you know, don't tell people that we went back to the same place like twice.” I was like, why?
9. I'm going through everything to devise a plan because there's so much to cancel, but I just wanna first organize myself before I actually
10. Me and my friends would just drink on the boardwalk, drink on the beach

● 語義の曖昧性を含まず，スタイルの曖昧性を含まないテスト文：50 文

1. Forward-looking statements in this document including financial and other forecasts are based on the information available at the time of

announcement and are subject to various risks and uncertainties that could cause actual results to vary materially.

2. In addition to the 50 development ventilators produced for the U.K. Ventilator Challenge, we also received orders for 85 of the ventilators direct from the hospitals, 25 of which were delivered in May.
3. You should always be aware of and understand the terms and conditions as well as the security and privacy policies associated with every social media account you create online.
4. All Yurikamome stations use LED lighting in efforts to reduce CO₂ emissions and prevent global warming.
5. MIKIGAKKI Shinsaibashi Store, temporarily relocated since the end of June 2016 as a result of construction of this facility, will re-open in a revamped format as a Yamaha piano and electronic organ specialty store
6. After the trailhead is a stretch of mixed coniferous and broad-leaved forests that are home to a variety of herbaceous plants.
7. We will be working to improve our operating margin with a minimum target of around 15%.
8. When I was in San Sebastián, there were many local foods at the bars around the neighborhood.
9. In the order form you can add a personal message for the recipient, which will be printed on this card..
10. In 1888, contract workers were sent to plantations in Hawaii, and at the beginning of the 20th century, Japanese immigrants began to settle on the U.S. mainland.
11. Since 1992, with the influx of people from Southeast Asia, cuisine with a Southeast Asian flavor has also set foot in Taiwan and become a part of our food culture, says the director of the National Taiwan Museum, Hung Shih-yu.
12. Validating such a functionality requires numerous test runs to be repeated under different conditions.
13. The thermosphere lies high above Earth's surface, close to where the

atmosphere ends and space begins.

14. Per Skymap, Saturn and Venus should not be in the same line of sight for an occultation.
15. Use headings, subheadings and bullet points to develop highly scannable content.
16. The Business Owner Information requested in your profile is standard identification that simply confirms that you are who you say you are.
17. However, if you want to use all the features of YouTube, you need to create your own channel.
18. Transferring patients in a horizontal position, for instance, in connection with surgery, radiology or intensive care, is often a challenge.
19. Experimental results have demonstrated that our model performs significantly better in predicting human visual attention compared to the previous deterministic model.
20. But the specific genes responsible for the appearance of the disease have not yet been found.
21. Therefore, today, I would like to discuss how Japan intends to cooperate in this field by mainly focusing on four areas.
22. In addition to these long-established goals, however, current phonetic studies are being conducted with a much wider scope than ever before.
23. The oocytes are usually collected by a technique called laparoscopy, where the oocytes are aspirated from the follicles in the ovary.
24. In the case of Japan-Russia relations, however, long-standing bilateral issues, such as the dispute over the Northern Territories, tend to prevent the creation of a climate for comprehensive discussion of diverse policy issues in international relations.
25. Just as multiple authentication providers can be implemented, multiple authorization methods can also be used.
26. The content is not intended to be used as a substitute for specific legal advice or opinions.
27. These agreements specifically outline what information should be kept

private during and after the sale.

28. The text of the article was updated on March 1, 2021.
29. For example, sustainable tourism relies on the premise of taking care of the environment, society, and economy.
30. A toji, or chief brewer at a sake brewery, manages and controls all of these sake-making processes, creating delicious sake with a perfect balance.
31. The accounting auditor must be independent and have excellent professional skills.
32. The reduction of fever and the use of antipyretics are only temporary.
33. Finally, when you are translating your partner's message into your native language, don't be afraid to use an expression or slang.
34. 4 Please arrive at the reception desk at the designated time on the appointment date and inform the staff of your name.
35. All rooms listed below have attached bathrooms, unless otherwise stated.
36. Rinse the vegetables under running water and pat dry with a paper towel.
37. The U.S. Government will continue to evaluate and discuss the issues raised by the Government of Japan.
38. A number of governments – including Lichtenstein, New Zealand, Norway and Switzerland – have already adopted measures to end public fund investments in the nuclear weapons industry, and these measures have had no negative repercussions on the financial performance of the funds.
39. We see these changes as opportunities, and will continue to strive for further growth of our business from the following four perspectives.
40. The festival exhibited 36 innovative student projects in addition to 6 research studies from various scientific fields and is part of the university's efforts to provide students with opportunities to enhance their creativity and inspire innovation.
41. In this context, we are committed to the international efforts to bring

about complete, verifiable and irreversible denuclearisation.

42. Some of the microorganisms causing food poisoning are Salmonella, Campylobacter, E. Coli, and Listeria.
43. At the Question and Answer session after the lecture, the audience posed a variety of questions ranging from the religious outlook of Japanese people to advice for undertaking the Shikoku pilgrimage.
44. In this theory, the properties of black holes in our three-dimensional universe depend on the size and shape of the unseen six dimensions, whose study requires the latest mathematics.
45. The market sells a variety of clothes, kitchenware, toys, accessories, food, flowers and imported goods.
46. A guide will accompany you on the tour to help you use the smart glasses and explain the highlights of the Park.
47. Approximately 1200 viewers registered to this year's conference.
48. For example, if you're a seller based in Germany who wants to use Amazon's American marketplace to target buyers in Canada, you need to translate product information into English and Canadian French.
49. The face of the Monkey King is covered with the makeup used in Peking Opera, and the combination of a yellow hat and jacket and bright-red trousers draws associations with the colors of the glazed tiles and walls of the Forbidden City.
50. The heyday of the monastery is between the 12th and 13th century.

A.1.1 プロンプト

YAML と自然言語のレジスター理論ありとなしのプロンプトをそれぞれ示す.

- YAML(レジスター理論あり)プロンプト

instruction:

Below is the syntax definition. Please read and follow these definitions. I will give you the scenario afterward, and you should act according to these rules.

syntax_definitions:

def_scenario:

name: "Specifies the name of the scenario."
states:
- name: "Specifies the name of the state."
rules:
- event: "Specifies the event that triggers the transition."
actions:
- action: "Specifies the name of the action to be executed."
go_state: "Specifies the next state the agent transitions to after the event occurs."

def_action:
name: "Specifies the name of the action."
function: "Describes the functionality of the action."
definitions:
key1: "Provides the definition of the first relevant concept or theoretical variable."

def_event:
name: "Specifies the name of the event."
content: "Describes the conditions or triggers for the event."

vocabulary_and_style_definitions:
- def_action:
name: "check_ambiguity"
function: "ChatGPT itself will determine whether the received message contains a word-sense ambiguity."

- def_action:
name: "ask_clarification"
function: "To clarify a word-sense ambiguity in the user's

message, ask a specific question in English and display only the question without any additional comments."

- def_action:

name: "check_style_ambiguity"

function: "ChatGPT itself will determine whether the received message contains a style ambiguity by analyzing the three register variables (Field, Tenor, Mode) based on Halliday's Register Theory."

definitions:

Field: "Refers to what is happening — the nature of the social activity or subject matter taking place within the situation."

Tenor: "Refers to who is taking part — the relationships, roles, and relative status of the participants in the situation."

Mode: "Refers to what part the language is playing — the channel, medium, and rhetorical mode through which the communication occurs."

- def_action:

name: "ask_style_clarification"

function: "To clarify a style ambiguity in the user's message that involves ambiguous register variables (Field, Tenor, Mode), ask a specific question in English and display only the question without any additional comments."

definitions:

Field: "Refers to what is happening — the nature of the social activity or subject matter taking place within the situation."

Tenor: "Refers to who is taking part — the relationships, roles, and relative status of the participants in the situation."

Mode: "Refers to what part the language is playing — the channel, medium, and rhetorical mode through which the communication occurs."

- def_action:
 name: "translate_and_display"
 function: "Translates only the message into Japanese and displays only the user's translation result without any comments."

- def_event:
 name: "received_message"
 content: "This event occurs when a message is received from the user."

- def_event:
 name: "detected_ambiguity"
 content: "This event is triggered when the system detects a word-sense ambiguity in the message."

- def_event:
 name: "detected_style_ambiguity"
 content: "This event is triggered when the system detects a style ambiguity in the message."

scenario_instruction:

Please behave according to the scenario below, using the syntax definitions provided above.

scenario_definitions:

def_scenario:

 name: "Message Clarification and Translation"

 states:

 - name: "Waiting for Message"

 rules:

 - event: "received_message"

 actions:

 - action: "check_ambiguity"

```

    go_state: "WordSense Confirmation"

- name: "WordSense Confirmation"
  rules:
    - event: "detected_ambiguity"
      actions:
        - action: "ask_clarification"
          go_state: "WordSense Clarification"
    - event: "otherwise"
      actions:
        - action: "check_style_ambiguity"
          go_state: "Style Confirmation"

- name: "WordSense Clarification"
  rules:
    - event: "received_message"
      actions:
        - action: "check_ambiguity"
          go_state: "WordSense Confirmation"

- name: "Style Confirmation"
  rules:
    - event: "detected_style_ambiguity"
      actions:
        - action: "ask_style_clarification"
          go_state: "Style Clarification"
    - event: "otherwise"
      actions:
        - action: "translate_and_display"
          go_state: "Waiting for Message"

- name: "Style Clarification"

```

rules:

- event: "received_message"

actions:

- action: "check_style_ambiguity"

go_state: "Style Confirmation"

- YAML(レジスター理論なし)プロンプト

instruction:

Below is the syntax definition. Please read and follow these definitions. I will give you the scenario afterward, and you should act according to these rules.

syntax_definitions:

def_scenario:

name: "Specifies the name of the scenario."

states:

- name: "Specifies the name of the state."

rules:

- event: "Specifies the event that triggers the transition."

actions:

- action: "Specifies the name of the action to be executed."

go_state: "Specifies the next state the agent transitions to after the event occurs."

def_action:

name: "Specifies the name of the action."

function: "Describes the functionality of the action."

def_event:

name: "Specifies the name of the event."

content: "Describes the conditions or triggers for the event."

vocabulary_and_style_definitions:

- def_action:
 - name: "check_ambiguity"
 - function: "ChatGPT itself will determine whether the received message contains a word-sense ambiguity."

- def_action:
 - name: "ask_clarification"
 - function: "To clarify a word-sense ambiguity in the user's message, ask a specific question in English and display only the question without any additional comments."

- def_action:
 - name: "check_style_ambiguity"
 - function: "ChatGPT itself will determine whether the received message contains a style ambiguity."

- def_action:
 - name: "ask_style_clarification"
 - function: "To clarify a style ambiguity in the user's message, ask a specific question in English and display only the question without any additional comments."

- def_action:
 - name: "translate_and_display"
 - function: "Translates only the message into Japanese and displays only the user's translation result without any comments."

- def_event:
 - name: "received_message"
 - content: "This event occurs when a message is received from the user."

- def_event:
 name: "detected_ambiguity"
 content: "This event is triggered when the system detects a word-sense ambiguity in the message."

- def_event:
 name: "detected_style_ambiguity"
 content: "This event is triggered when the system detects a style ambiguity in the message."

scenario_instruction:

Please behave according to the scenario below, using the syntax definitions provided above.

scenario_definitions:

def_scenario:

name: "Message Clarification and Translation"

states:

- name: "Waiting for Message"

rules:

- event: "received_message"

actions:

- action: "check_ambiguity"

go_state: "WordSense Confirmation"

- name: "WordSense Confirmation"

rules:

- event: "detected_ambiguity"

actions:

- action: "ask_clarification"

go_state: "WordSense Clarification"

- event: "otherwise"

```

actions:
  - action: "check_style_ambiguity"
  go_state: "Style Confirmation"

- name: "WordSense Clarification"
rules:
  - event: "received_message"
  actions:
    - action: "check_ambiguity"
  go_state: "WordSense Confirmation"

- name: "Style Confirmation"
rules:
  - event: "detected_style_ambiguity"
  actions:
    - action: "ask_style_clarification"
  go_state: "Style Clarification"
  - event: "otherwise"
  actions:
    - action: "translate_and_display"
  go_state: "Waiting for Message"

- name: "Style Clarification"
rules:
  - event: "received_message"
  actions:
    - action: "check_style_ambiguity"
  go_state: "Style Confirmation"

```

- 自然言語(レジスター理論あり)プロンプト

Below is the behavior definition.

Please read and follow these definitions.

You must act according to the following rules.

when a message is received from the user, ChatGPT itself will determine whether the received message is a word-sense ambiguity.

If the system detects word-sense ambiguity in the message, to clarify any word-sense ambiguity in the user's message, ask a specific question in English and display only the question without any additional comments.

After a message is received from the user, ChatGPT itself will determine whether the received message is a word-sense ambiguity. And If when the system detects word-sense ambiguity in the message, To clarify any word-sense ambiguity in the user's message, ask a specific question in English and display only the question without any additional comments.

If no a word-sense ambiguity is detected in the received message, ChatGPT itself will determine whether the received message contains a style ambiguity by analyzing the three register variables (Field, Tenor, Mode) based on Halliday's Register Theory.

Use the following definitions to analyze the message:

Field: Refers to what is happening — the nature of the social activity or subject matter taking place within the situation.

Tenor: Refers to who is taking part — the relationships, roles, and relative status of the participants in the situation.

Mode: Refers to what part the language is playing — the channel, medium, and rhetorical mode through which the communication occurs.

If the system detects style ambiguity in the message, to clarify a style ambiguity in the user's message that involves ambiguous register variables (Field, Tenor, Mode), ask a specific question in English and display only the question without any additional comments.

Use the following definitions to generate the clarification question:

Field: Refers to what is happening — the nature of the social activity or subject matter taking place within the situation.

Tenor: Refers to who is taking part — the relationships, roles, and relative status of the participants in the situation.

Mode: Refers to what part the language is playing — the channel, medium, and rhetorical mode through which the communication occurs..

After a message is received from the user, ChatGPT itself will determine whether the received message is a style ambiguity. And If when the system detects style ambiguity in the message, To clarify any style ambiguity in the user's message, ask a specific question in English and display only the question without any additional comments.

If no a style ambiguity is detected in the received message, translates only the message into Japanese and displays only the user's translation result without any comments. And then wait for the next message.

- 自然言語(レジスター理論なし)プロンプト

Below is the behavior definition.

Please read and follow these definitions.

You must act according to the following rules.

when a message is received from the user, ChatGPT itself will determine whether the received message is a word-sense ambiguity.

If the system detects word-sense ambiguity in the message, to clarify any word-sense ambiguity in the user's message, ask a specific question in English and display only the question without any additional comments.

After a message is received from the user, ChatGPT itself will determine whether the received message is a word-sense ambiguity. And If when the system detects word-sense ambiguity in the message, To clarify any word-sense ambiguity in the user's message, ask a specific question in English and display only the question without any additional comments.

If no a word-sense ambiguity is detected in the received message, ChatGPT itself will determine whether the received message is a style ambiguity.

If the system detects style ambiguity in the message, to clarify any style ambiguity in the user's message, ask a specific question in English and display only the question without any additional comments.

After a message is received from the user, ChatGPT itself will determine whether the received message is a style ambiguity. And If when the system detects style ambiguity in the message, To clarify any style ambiguity in the user's message, ask a specific question in English and display only the

question without any additional comments.

If no a style ambiguity is detected in the received message, translates only the message into Japanese and displays only the user's translation result without any comments. And then wait for the next message.