信頼できる言語理解 AI 実現のためのフレーム知識の創出と活用

慶應義塾大学　理工学部
小原　京子

１．はじめに

自然言語を扱う生成AI（言語理解AI）には「信頼」の実現が必須だが、ChatGPTを代表とする大規模言語モデル（LLM）に基づく現在の言語理解AIは「信頼」を実現しているとは言い難い。しばしばハルシネーション（不正確な情報の生成）を起こすからである。人々が実社会で言語理解AIを安心して利用するためには、その信頼性の抜本的な向上が不可欠である。

本研究では信頼できる言語理解 AI の実現には、１）人間の持つ経験に基づく暗黙知を資源化する、２）構造的な形式知を資源化する、３）LLM と人間を協働させ相補的な関係を築く、４）LLM を高度化する、の４つの技術が必要と考え、主に１）の人間の持つ暗黙知の資源化を行い、さらに３）の LLM と人間との協働の可能性を検討した。

２．研究の背景

信頼できる言語理解AIとは、ハルシネーションを起こさず、出力の根拠を示すことができる生成AIを指す。本研究では、言語理解AIに必要は「信頼」の要素とは、1) Relevance: ある問題を解けるなら、その問題や解の要素となるモノ・コトを知っている；2) Composability: ある複雑な問題を解けるなら、その要素問題も解ける；3) Consistency: ある問題を解いたら、常に同様の方法で同じ解答を返す；4) Coherence: ある問題を解いたら、類似の問題を類似の方法で解ける；5) Awareness: 答えを知らないことを知っている、という仮説を立てた。

人間は言葉を聞くと、その言葉にまつわる自分の多くの経験を一般化した場面（フレーム）を想起する。人間は様々なフレームに関する知識（フレーム知識）を使って言葉の意味を理解する。たとえば、「地点Aにいる人がある物を地点Bに送った」という文を聞くと、日本語話者は「その人」は地点Aにいると理解する。これは「モノを送る」ことに関するフレーム知識を日本語話者が持っており、「モノを送る」という事象においてはその人はモノとともに移動しないということを知っているからである。また、「地点Aにいる人がある人を地点Bに送った」という文を聞くと、日本語話者は「その人」は地点Bにいると理解する。これは、「人を送る」ことに関するフレーム知識を日本語話者が持っており、「人を送る」という事象においてはその人は「送る対象である人」と共に移動するということを知っているからであ

る。

人間の持つこのようなフレーム知識は暗黙知であり、人間は日常生活において自らがもつこのような知識を意識することはない。これに対して、現在の言語理解AI はある単語が与えられた際に次に来る単語の確率をもとに文章を生成している。つまり、人間の持つフレーム知識が言語理解AIに組み込まれていないのである。言語理解AIがハルシネーションをしばしば起こすのは、人間の持つフレーム知識を言語理解AIが持っていないことに起因すると言える。言語理解AIが、ハルシネーションを起こさず、そのアウトプットの質を人間が常に「信頼」できるレベルまで引き上げるためには、AIに人間の持つ言葉の意味に関するフレーム知識を取り込む必要がある。

3. 研究の目的

本研究では、ハルシネーションを起こさず、出力の根拠を示すことができる「信頼できる言語理解AI」の実現を目指した。

4. 研究の内容

本研究では信頼できる言語理解 AI の実現には、１）人間の持つ経験に基づく暗黙知（フレーム知識）を資源化する、２）構造的な形式知を資源化する、３）LLM と人間を協働させ相補的な関係を築く、４）LLM を高度化する、の４つの技術が必要と考えた（図１参照）。本研究では、主に１）の人間の持つ暗黙知の資源化を行った。また、３）の LLM と人間との協働についても検討した。

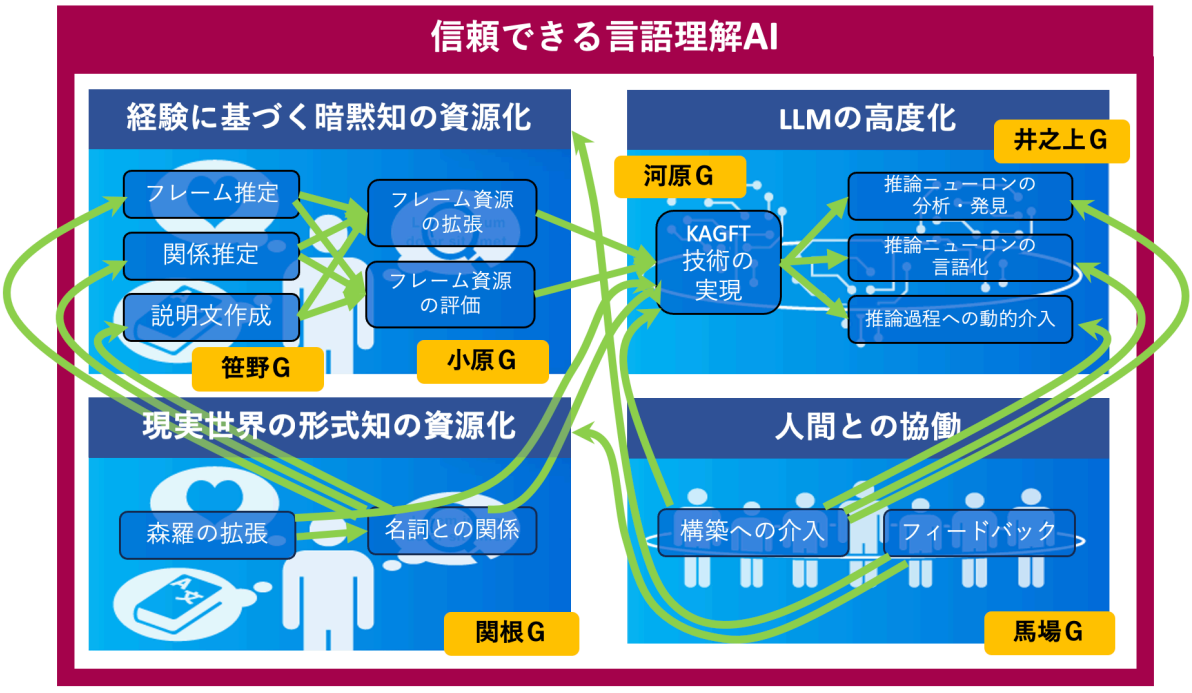

図１．信頼できる言語理解 AI のための４つの技術

５．研究の方法

●

研究代表者は日本語話者の持つ語の意味に関する知識「フレーム知識」（人間の経験に根ざした、言葉の意味に関する背景知識）を、認知言語学の枠組みに則って網羅した言語資源「日本語フレームネット」を構築してきた[1]。上記１）のフレーム知識の資源化として、具体的には日本語フレームネットの拡張を行った。日本語フレームネット上の意味フレーム知識の規模を英語フレームネット[2]のそれと同規模に拡張するため、200意味フレームに日本語単語を登録し用例を蓄積した。

上記３）のLLMと人間との協働の検討については、共同研究者笹野氏のグループが行った、LLMを活用したフレーム推定の結果を、話者の持つフレーム知識に相応するかどうかという観点から評価した。

６．成果

上記「２．研究の背景」で述べたとおり、本研究では、言語理解AIに必要は「信頼」の要素とは、1）Relevance：ある問題を解けるなら、その問題や解の要素となるモノ・コトを知っている；2）Composability：ある複雑な問題を解けるなら、その要素問題も解ける；3）Consistency：ある問題を解いたら、常に同様の方法で同じ解答を返す；4）Coherence：ある問題を解いたら、類似の問題を類似の方法で解ける；5）Awareness：答えを知らないことを知っている、であるとあらかじめ仮説を立てた。本研究を遂行した結果、「フレーム知識」とそれに基づく論理・推論機構はこれらの「信頼」の要素を実現できる、ということが明らかとなった。

研究代表者の具体的な成果は以下のとおりである。

- Ohara (2024) Ohara, Kyoko. What are interactional frames doing in Constructions? . Talk given at the 13th International Conference on Construction Grammar. University of Gothenburg, Sweden. August 26, 2024.
- Ran Iwamoto and Kyoko Ohara. (To Appear). "How to Improve Generative AI with Frame Name Annotation in Prompts." IFNW2025.
- 国際フレームネットワークショップ 2025(International FrameNet Workshop 2025 (IFNW2025))2025 年 3 月 7、8 日．慶應義塾大学日吉キャンパスにて、国立国語研究所との共同開催予定．(https://www.globalframenet.org/ifnw-2025)
- 慶應義塾大学 GIC センター・理工学部総合教育セミナー “Human Language Understanding and Generative AI:An Exploration of the Uses of AI in University Classes” 開講．(https://www.gic.keio.ac.jp/courses/interview_03.html)

７． まとめ

本研究では、信頼できる言語理解 AI の実現を目的として、人間の持つ言葉の意味に関する知識「フレーム知識」の創出を行い、それを LLM と協働させる可能性を検討した。言語理解

AI の信頼性の向上に向けての最大の課題であるハルシネーションを解消させるために、フレーム知識を LLM に組み込むことが有効であることが明らかとなった。

謝辞

本研究を遂行するにあたり、公益財団法人　天野工業技術研究所から多大なご支援を頂きました。ここに記して謝意を示します。また、共同研究者の関根聡氏(理化学研究所革新知能統合研究センターチームリーダー・国立情報学研究所特任教授)、河原大輔氏（早稲田大学・国立情報学研究所特任教授)、笹野遼平氏（名古屋大学)、馬場雪乃氏（東京大学)、井之上直也氏（北陸先端科学技術大学院大学)、鈴木久美氏（国立情報学研究所特任教授）にも感謝の意を表します。

参考文献

1)　日本語フレームネット. https://jfn.st.hc.keio.ac.jp/ja/home-ja/

2)　FrameNet. https://framenet.icsi.berkeley.edu