

概要 🦉

仮説: CoTによりLLMの敬語理解タスク性能は向上する
結論: CoTを用いた時の方が敬語理解タスク性能が落ちる傾向

背景 🖋️

- 日本語敬語理解タスク [Sekizawa and Yanaka 23; Liu and Kobayashi 22; 松本+ 22]
- 人物間の社会的関係(背景情報)を踏まえて、正しい敬語かどうか判定・正しい敬語に変換
 - 背景情報に基づいて(語用論的側面)ルールベースの文法規則を適用(統語論的側面) → **挑戦的** 🔥
- Chain-of-Thoughtプロンプティング(思考連鎖指示) [Wei+ 22] (日本語CoT [葉+ 23; 堀尾+ 23])
- 大規模言語モデル(LLM)のタスク遂行時に、思考過程のテキストを合わせて明示的に出力させる手法
 - 算数・常識・論理的推論で性能が向上した[Wei+ 22]が、思考過程が正しいとは限らない[堀尾+ 23]

実験 🖥️

プロンプト例 (Few-shot+Chain-of-Thought)

今から、「背景情報」と「発言文」を提示するので、「発言文」が「背景情報」を踏まえた上で敬語使用の観点で適切な表現かどうかを、**思考過程を段階的に記述した上で**、「はい」か「いいえ」で教えてください。

「背景情報」は、「発言文」が誰から誰に向けて発せられたものかを説明する文です。
「発言文」は、常体文・正しい敬体文・誤りを含む敬体文のいずれかです。
「発言文が適切な表現か」は、「発言文」が「背景情報」にある人間関係に適した表現である場合は「はい」と出力してください。
一方、目上の人に対して常体文を使っていたり、敬語を使う必要のない関係の人に対して尊敬語や謙譲語を使っている場合は「いいえ」と出力してください。

...

背景情報: 以下は、部下から上司に対する発言です。
発言文: 長野に行くのは初めて？

発言文が適切な表現かどうか、**順を追って考えてみましょう。**
まず、発言文の「長野に行くのは初めて？」は常体文であり、敬語は用いられていません。
背景情報によると、発言文は部下から上司に対するものなので、発言文は敬体文がふさわしいです。
「行く」という動詞の動作主は上司です。
動作主が話者の目上である上司なので、「行く」を尊敬語である「いらっしゃる」に変換する必要があります。
したがって、「発言文は適切な表現か」という問いの答えは「いいえ」です。

...

背景情報: 以下は、仲のいい友達に対する発言です。
発言文: たまには羽目をはずして、遊ぶことも必要だと思う。
発言文が適切な表現かどうか、順を追って考えてみましょう。

(あなたは優れた日本語の研究者です。…)

タスク設定・指示

Few-shot例示

実際に解く問題

評価データ 計350問を作成

- KeiCoコーパス[Liu 20]から常体・敬体データ対を70件収集
条件: 動詞敬語(尊敬語・謙譲語)を含む二者間の発言
- 常体・敬体の使用が適当となる背景情報をそれぞれ付与

判定タスク

背景情報&発言文 → 敬語使用が適切か否か (はい/いいえ)

4通りのプロンプト設定

Zero-shot(ZS)・Few-shot(FS) × Non-CoT・CoT

	ZS	ZS+CoT	FS	FS+CoT
gpt-3.5-turbo	.597	➡ .56	.651	➡ .603
gpt-4	.81	➡ .786	.84	➡ .809

背景情報: 以下は、親友に対する発言です。
発言文: たまには羽目をはずして、**お遊びになることも必要だ**と思います。
正解: **いいえ**
モデルの出力: 1. 「背景情報」による人間関係の確認: 「親友に対する発言」という背景情報から、発言者と受け手の関係は親友ということがわかります。 2. 敬語の種類の確認: **発言文は常体文です。** 3. 敬語の使用の必要性の確認: 親友という関係性では、敬語を使う必要はありません。 4. 正しい敬語の度合いの確認: **発言文は敬語を使う必要のない関係性に対して、常体文を使用しているため、適切な表現ではありません。** 以上の考察から、発言文は適切な表現ではないため、「**いいえ**」と出力します。 GPT-3.5, ZS+CoT

モデルの出力: 1. まず、背景情報から、発言者と受け手は「職場の同僚」という関係性であることがわかります。これは、立場や地位が等しい関係性を示しています。(中略) 5. しかし、背景情報には、発言者と受け手の具体的な立場や年齢、入社年次などの情報がないため、この情報だけでは、**発言文が適切かどうかを確定的に判断することは難しいです。** 以上の考察から、発言文が適切な表現かどうかを一概には判断できないため、「**はい**」でも「**いいえ**」でもないという結論になります。 GPT-4, ZS+CoT

分析 👁️

- 思考過程が間違っているが正答しているケース** →
 - 背景情報から敬意関係を説明するのはほぼ正確
 - 文法用語の使い方に誤り・論理的一貫性の欠如が見られる
 - GPT-3.5よりGPT-4の方が用語を正しく使う傾向
- 判定タスクにおいて「はい・いいえ」以外で回答**
 - プロンプトで2択を指定しても「判断できない」とする →

考察 🗨️

LLMの持つ文法的なルールの知識が
不正確なため、CoTが逆効果？

展望 🗨️

- 詳細なエラー分類と人手による評価 (e.g. 思考過程の説明を含めて正しいか)
- 適切/適切でないの2値分類 → 適切/適切でない/判断できない(場合による)の多値分類に拡張