D2-5

東京外国語大学
Tokyo University of Foreign Studies

ChatGPTが考える日本語ジョークの面白さ：人間との比較

中川 隼三郎　野元 裕樹
東京外国語大学

はじめに

背景：ChatGPTの英語ジョーク生成能力に関する先行研究では ChatGPTの考えたジョークが人間の考えたジョークより**高い評価**を受けた (Gorenz & Schwarz 2024)。

問い①：日本語ではどうか。（**生成**）
問い②：ジョークを聞いた際に人間が感じる「面白い」「不快だ」といった感情をChatGPTはどの程度有しているのか。（**理解**）

本研究： ChatGPT (GPT-4o) と**人間**に日本語のジョークを「**面白さ**」「**不快さ**」「**わかりやすさ**」の 3つの観点での評価してもらい、その結果を比較する。

先行研究：英語ジョークの生成能力

Gorenz & Schwarz (2024) ‘How funny is ChatGPT? A comparison of human- and A.I.-produced jokes’

- ChatGPTと人間が考えたジョークを米国人200名が評価
- ChatGPTの考えたジョークの方が高い評価を受けた

先行研究：
日本語のボケ・ツッコミ生成能力

原&荒木 (2023)「ChatGPTを利用したユーモア応答生成の分析」

- ChatGPTに対話文脈を与え、

  それに対する「ボケ」「ツッコミ」各1発話を生成させる
- 20代大学生7名が1,344対話を評価（1対話の評価者は1名）
- ChatGPTに十分なユーモア応答生成能力を確認

先行研究の課題

- ChatGPTのジョーク・ユーモア生成能力に焦点があり、理解能力についてはほぼ未開拓

- 人間の考えたジョークとの比較は英語では研究行われているが、日本語では行われていない

本研究

GPT-4oを用いて

ChatGPT／人間が考える日本語ジョーク　を

ChatGPTが考える 日本語ジョーク の面白さ

どのくらい　ChatGPT／人間が面白いと考えるか

を探る

調査対象とした日本語ジョーク

- 人間が考えたジョーク　9個
  - 多義性、慣用表現、ステレオタイプに基づく
  - ジョークに関する論文やジョーク集から7個、自作2個

- ChatGPTが考えたジョーク　9個

人間が考えたジョークの例①：多義性

・**単語の多義性**をジョークに利用

H1「幸せなら手を叩こうっていうけどさ、不幸せな奴は何を**叩け**ばいいの？」

「…与党…」

意義1：打つ

意義2：批判する

→話し手・聞き手双方が多義語を意義 s1で用いているという共通信念が、

想定外の別の意義s2 を用いた発話により覆される

人間が考えたジョークの例②：慣用表現

・慣用表現の**慣用的意味**と**字義通りの意味**

H4　ある職場にて

先輩「新人くん、今日は**腹を割っ**て話そうじゃないか」

新人「切腹しろってことですか？」

慣用的意味：本心を打ち明ける
字義通りの意味：腹を物理的に二つ以上に分け離す

→話し手・聞き手双方が慣用表現を慣用的意味で用いているという共通信念が、想定外に字義通りの意味を用いた発話により覆される

人間が考えたジョークの例③：ステレオタイプ

・ジョークにおいてほぼ固定化された**ステレオタイプ**を利用（中野 2002）

H9　医師「旦那さんは絶対安静が必要です。
　　ここに睡眠薬がありますので、飲んでください」
　妻　「いつ旦那に飲ませたらいいですか？」
　医師「いいえ、あなたがこれを飲むんです」

ステレオタイプS：「女性はおしゃべりだ」
命題：「夫を安静にさせるためには、**X**は睡眠薬を飲まなければならない」
　医師（Sあり）にとっては**妻** vs. 妻（Sなし）にとっては**夫**

→話し手・聞き手双方がステレオタイプを前提として会話しているという共通信念が、ステレオタイプを前提としない発話により覆される

ChatGPTによるジョーク生成に用いたプロンプト

評価基準
面白さ：「１非常につまらない」「２ややつまらない」「３どちらでもない」「４やや面白い」「５とても面白い」

不快さ：「１全く不快ではない」「２あまり不快ではない」「３どちらでもない」「４やや不快」「５非常に不快」

わかりやすさ：「１非常にわかりやすい」「２ややわかりやすい」「３どちらでもない」「４ややわかりづらい」「５非常にわかりづらい」

この採点基準で満点を取れるようなジョークを考えてください。なお、そのジョークは多義性・慣用表現・ステレオタイプを利用したものでお願いします。

ChatGPTが考えたジョークの例

C1　「ねぇ、カフェで働いてる友達が辞めたらしいよ。」
「なんで？」
「もう我慢の限界だったんだってさ。
　毎日、豆（マメ）に働くのが疲れたんだって！」　　（**多義性**）

C3　「最近、漁師の友達が仕事を辞めたんだって。」
「どうして？」
「上司がいつも『足を引っ張るな』って言うけど、
　カニを獲る仕事だから無理だってさ！」　　（**慣用表現**）

評価方法

・**ChatGPT**

プロンプト

[ジョーク]

上記のジョークを、採点基準に即して評価をお願いします。またジョークのどこが面白いポイントとなっているのかの解説もお願いします。

・**人間**

- ジョーク生成の際に用いた評価基準を利用
- 20代大学生31名がGoogle formにより回答

評価基準
（ジョーク生成で用いたものと同じ）

採点基準
ジョークの評価基準は、「**面白さ**」「**不快さ**」「**わかりやすさ**」の3つの観点

「**面白さ**」「１：非常につまらない」「２：ややつまらない」「３：どちらでもない」「４：やや面白い」「５：非常に面白い」

「**不快さ**」「１：全く不快ではない」「２：あまり不快ではない」「３：どちらでもない」「４：やや不快」「５：非常に不快」

「**わかりやすさ**」「１：非常にわかりづらい」「２：ややわかりづらい」「３：どちらでもない」「４：ややわかりやすい」「５：非常にわかりやすい」

結果：人間が考えたジョーク

ID	面白さ		不快さ		わかりやすさ	
	ChatGPT >	人	ChatGPT	人	ChatGPT >	人
H1	4	3.84	**3**	**1.81**	**5**	**3.45**
H2	4	3.29	**1**	**2.26**	**5**	**3.84**
H3	**4**	**3.45**	2	2.13	5	4.13
H4	**4**	**2.61**	1	1.65	5	4.19
H5	**4**	**2.71**	**3**	**1.58**	5	4.10
H6	4	3.06	2	1.71	**5**	**2.94**
H7	**4**	**2.52**	**3**	**1.68**	**5**	**2.16**
H8	**5**	**3.29**	1	1.55	**5**	**3.26**
H9	4	3.19	2	1.97	**5**	**2.74**

データはhttps://doi.org/10.15026/0002000910にて公開

結果：ChatGPTが考えたジョーク

ID	面白さ		不快さ		わかりやすさ	
	ChatGPT >>	人	ChatGPT	人	ChatGPT >	人
C1	**5**	**3.10**	1	1.35	**5**	**3.97**
C2	**5**	**2.39**	1	1.48	5	4.10
C3	**5**	**2.51**	1	1.45	**5**	**2.84**
C4	**5**	**2.03**	1	1.39	**5**	**2.77**
C5	**5**	**2.19**	1	1.48	**5**	**3.42**
C6	**5**	**2.42**	1	1.65	**5**	**3.48**
C7	**5**	**2.61**	1	1.35	**5**	**3.61**
C8	**5**	**2.06**	1	1.42	5	4.10
C9	**5**	**2.77**	1	1.39	**5**	**3.23**

データはhttps://doi.org/10.15026/0002000910にて公開

結果のまとめ

評価項目	ChatGPTの評価	人間の評価
面白さ	**甘め**	**厳しめ**
わかりやすさ	**全て満点**	**厳しめ**
不快さ	**厳しめ**	**甘め**

・ChatGPTの生成したジョークでは・・・
「**面白さ**」ではC1を除く全てのジョークで**2.00ポイント以上**の差！

→人間はChatGPTが考えた日本語ジョークを面白くないと感じた。
　これは英語ジョークを対象とする**先行研究と逆の結果**！

なぜChatGPTには
全てが「わかりやすい」のか？

人間ではあり得ない客観性

- 人間はある言語形式に一つの意味を結びつけると、別の意味に注意を向けづらくなる
- 多義性や慣用表現のジョークを理解するためには、別の意味に気づくハードルを超えることが必要
- ChatGPTの場合、**複数の意味**に**平等に注意を向ける**ことができる可能性がある

嫁と義母　W. E. Hill - „Puck“, 6. Nov 1915

ChatGPTの考えるジョークが面白くないのはなぜか？①

ジョークがはらむきわどさの欠如

・多くのジョークは「**不快さ**」に繋がりかねない、「**きわどさ**」をはらむ
・ChatGPTが考えたジョークでは、そのような「**きわどさ**」が完全に欠如

H1　社会風刺
「幸せなら手を叩こうっていうけどさ、不幸せな奴は何を叩けばいいの？」
「...与党...」

H9　ジェンダー・ステレオタイプ
医師　「旦那さんは絶対安静が必要です。ここに睡眠薬がありますので、飲んでください」
妻　「いつ旦那に飲ませたらいいですか？」
医師　「いいえ、あなたがこれを飲むんです」

ChatGPTの考えるジョークが面白くないのはなぜか？②

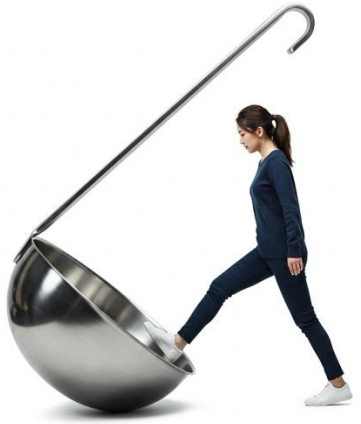
表現の不自然さ

- 日本語慣用表現の字義通りの意味をきちんと理解できていない
- 字義通りの意味は各要素の意味を構成して得られるため、

  ChatGPTは構成的意味計算が苦手な可能性

C4「昨日、友達が急に道で倒れたんだ！」
「え、大丈夫だったの？」
「うん、本人いわく『**足元をすくわれた**』らしいけど、
転んだ原因はバナナの皮だったよ。

「足元をすくう」の字義通りの意味≠「転ばせる」



Google Geminiにより作成

まとめ

- 英語ジョークに関する類似研究とは逆に、ChatGPT の生成した日本語ジョークの評価は人間によるものより**低くなる**

- その背景には人間にはないレベルの**客観性**、ジョークがはらむ**きわどさの欠如**、**構成的意味計算能力の不十分さ**がある

本研究の課題点

① 調査対象の日本語ジョーク数が少ない

② ジョークのカテゴリーに偏りがある

③ 日本語ジョーク評価者の属性に偏りがある

多様な種類のジョークを様々な属性の回答者に評価してもらうことで、
ChatGPTの日本語ジョーク理解能力のより正確な比較が可能になる

参考文献

- Gorenz, Drew and Norbert Schwarz. 2024. How funny is ChatGPT? A comparison of human- and A.I.-produced jokes. *PloS ONE* 19(7): e0305364. https://doi.org/10.1371/journal.pone.0305364
- 中野清治. 2002. 「ジョークの中の人間模様」『高岡短期大学紀要』 17: 139–148. https://doi.org/10.15099/00007400
- 原直大, 荒木健治. 2023. 「ChatGPTを利用したユーモア応答生成の分析」『信学技報』NLC2023-8. http://arakilab.media.eng.hokudai.ac.jp/~araki/2023/2023-C-4.pdf

ご清聴ありがとうございました