

発話を受けた聞き手の端的な感想に基づく 例えツッコミ文の生成

関 陽介^{1,a)}

受付日 2023年4月14日, 採録日 2023年11月7日

概要: 対話システムを利用することで、ユーザは時間や場所を問わず対話的な情報収集が可能になる。ただし、素っ気ない応答や無機質なインタフェース等によりシステムへの親近感が低くなる場合、利用機会の減少や対話持続の困難さが懸念される。そこで、本研究では対話を通してユーモアを喚起してシステムへの親しみを感じさせるために、ツッコミ手法の中で生成する難易度が高い型である例えツッコミを対象として、発話を受けた聞き手の端的な感想に基づく例えツッコミ文の生成手法を提案する。具体的には、TwitterのTweet・Reply文に対して構文解析や依存構造解析、関連度の算出等により学習資源を収集して、発話文に対するツッコミ対象の推論機能を開発する。また、インターネット上で公開されている例文を基に言語モデルを用いて例えツッコミ文を拡充して、ツッコミ対象に基づく例えツッコミ文の生成機能を開発する。主に高校生を対象にした評価実験の結果、学習資源を収集するための各処理方法は適切なツッコミ対象の推論に有効であり、例えツッコミ文を交えた雑談対話はユーモアが喚起されてシステムへの親しみを感じられること等が示された。

キーワード: ユーモア生成, 例えツッコミ, 対話システム

Generation of Tatoe-tsukkomi Sentences Based on the Listener's Straightforward Impression of the Utterance

YOSUKE SEKI^{1,a)}

Received: April 14, 2023, Accepted: November 7, 2023

Abstract: This study proposes a method for generating tatoe-tsukkomi sentences based on the listener's straightforward impression of the utterances, targeting tatoe-tsukkomi, to evoke humor and give users familiarity with the systems through dialogues. Specifically, training data based on syntactic analysis and dependency structure analysis, relevance calculations for Tweet and Reply sentences on Twitter are collected, and an inference function of the tsukkomi targets against the utterances is developed. In addition, tatoe-tsukkomi sentences are expanded based on example sentences on the Internet by a language model, and a function to generate tatoe-tsukkomi sentences based on the tsukkomi-targets is developed. The results of evaluation experiments, mainly targeting high school students, showed that each processing method for collecting training data was effective in inferring appropriate tsukkomi targets, and that chat dialogues with tatoe-tsukkomi evoked humor and showed familiarity with the systems.

Keywords: humor generation, tatoe-tsukkomi, dialogue systems

1. はじめに

問合せ対応の自動化や業務負担の軽減等を目的として、

多様な分野に対話システムが導入されている。たとえば、行政機関ではゴミの分別案内サービス^{*1}や観光案内サービス^{*2}、学術機関では入学希望者や在学生の情報収集を支援

¹ 徳島大学高等教育研究センター
Research Center for Higher Education, Tokushima University, Tokushima 770-8501, Japan

^{a)} seki@tokushima-u.ac.jp

^{*1} 横浜市イーオのごみ分別案内, <https://www.city.yokohama.lg.jp>

^{*2} 徳島県阿波ナビ, <https://www.awanavi.jp/>

する対話システム^{*3,*4}がある。対話システムを利用することで、ユーザは時間や場所を問わず対話的な情報収集が可能になる。ただし、素っ気ない応答や無機質なインタフェース等によりシステムへの親近感が低くなる場合、利用機会の減少や対話持続の困難さが懸念される。

ユーモアは心理学的観点で人とシステム間の親和な関係の成立に必要とされており [1], [2], 継続的な対話にはユーモアが重要な役割を果たす [3]。また、様々なユーモア生成手法 [4], [5], [6], [7], [8], [9], [10], [11], [12] が提案されており、ユーモアによるシステムへの親しみを高める効果が示されている。寄席演芸の1つである漫才を参考にした場合、ユーモアはボケやツッコミにより生成される。前者はある話題において意図的に間違いや勘違いを言葉にして笑いを生み出す流れを作る行為であり、後者はボケ役の間違いを指摘して笑いを生み出す行為である。

ツッコミには指摘ツッコミやノリツッコミ等の10種類の型があり、ツッコミを入れる対象は子供の意見等の発言、猫が走る等の行動、寝ぐせがついた髪型等の見た目、白黒テレビ等の物、賑やかな図書館等の環境に大別される [13]。本研究では対話システムの入出力対象をテキストとするため、文字情報として表現しやすく対話の中で頻繁に扱われる発言を対象とする。ツッコミの型の中には、生成する難易度が高い型として例えツッコミがある。例えツッコミは発言内容や発言への指摘等（以下、ツッコミ対象）に対して、何かに例えてツッコミを入れる手法である。たとえば、“息子は足が速いから将来は金メダルが取れるな。”という発言に対して、ツッコミ対象を“大袈裟”とした場合、“魔法瓶ぐらい大袈裟！”という例えツッコミ文が成立する。このようなツッコミ対象は発言を受けた者の印象や考えに基づくため、本研究では聞き手の端的な感想をツッコミ対象として例えツッコミ文を生成する。多様な発話がされる雑談対話において、発話内容に応じて例えツッコミ文を交えることで、ユーモアを喚起できユーザとシステム間の親和な関係の成立が期待できる。

本研究では対話を通してユーモアを喚起してシステムへの親しみを感じさせるために、発話を受けた聞き手の端的な感想に基づく例えツッコミ文の生成手法を提案する。具体的には、Twitter^{*5}のTweet・Reply文に対して構文解析や依存構造解析、関連度の算出等により学習資源を収集して、発話文に対するツッコミ対象の推論機能を開発する。また、インターネット上で公開されている例文を基に言語モデルを用いて例えツッコミ文を拡充して、ツッコミ対象に基づく例えツッコミ文の生成機能を開発する。

本研究の貢献は以下になる。

- Twitter から学習資源を収集して、発話文に対するツッコミ対象の推論機能を開発すること。
- 公開されている例文を基に言語モデルを用いて例えツッコミ文を拡充して、ツッコミ対象に基づく例えツッコミ文の生成機能を開発すること。
- 例えツッコミ文を生成する雑談対話システムを開発すること。
- 主に高校生を対象にした評価実験により提案手法の有効性や、雑談対話に例えツッコミ文を交えた影響を評価すること。

2. 関連研究

自然言語処理や人工知能等の分野では、システム技術としての対話システムの導入、実践報告やデザイン研究が増加している [14]。また、対話を通じた多様なユーモア生成手法が報告されている。

ボケとなる文を生成する研究としては、WordNetを用いた単語間の類似度や応答文の例文検索等による、与えられた同音異義語とその関連語を含んだ駄洒落の生成手法 [4] や、音韻辞書等を用いて特定の単語を音韻やアクセントの位置が類似する別の単語に変換してユーモア文を生成する手法 [5]、発話文に表れる単語間の係り受けに着目して、曖昧な参照先単語を用いてユーモアを喚起する疑問文の生成手法 [6] がある。また、編集距離や概念距離、出現回数等を用いて、発話文の一部に対してユーモアのある聞き間違いをして聞き返すボケをする対話型エージェント [7], [8] や、Wikipedia や分類語彙表等から作成されたユーモア表現の変換候補辞書を用いた疑似的な聞き間違い文の生成手法 [9] がある。

ツッコミ文を生成する研究としては、EDR 電子化辞書や Word2vec 等を用いてユーモア発話に対する同義語・動作ツッコミ単語等で構成されたツッコミ文を生成する手法 [10] や、聞き間違いえた内容の文脈を考慮して、マスク付き言語モデルである Masked Language Model (MLM) を用いたノリツッコミ文の生成手法 [11] がある。

一方で、これらのボケやツッコミに関する手法では聞き手の感想は考慮されず、主に発話文に表れる単語に着目してユーモア文が生成されている。関連研究の中には、聞き手の感想等を対象として例えツッコミ文を生成する手法 [12] がある。この研究では、感想等の推論モデルを作成するために、Twitter から Tweet・Reply 文を取得して、発話文に聞き手の感想等をラベル付けした学習資源を収集している。ただし、感想等を表現する単語として Reply 文で最初に現れる形容詞を用いており、発話内容に関連しない単語が多く選ばれる可能性が高い。

対話エージェントが例えツッコミを行うためには、ツッコミ対象の特定とその対象に基づく例えツッコミ文の生成が求められる。深層学習を用いることでツッコミ対象

*3 徳島大学とくぼん talk, <https://taiwa.honbu.tokushima-u.ac.jp/taiwa/ui/toku>

*4 佐賀大学 AI チャットボット, <https://sdcunivchat.qabot.jp/>

*5 Twitter, <https://twitter.com/>

の推論モデル等を作成できるが、学習資源を収集する必要がある。既存の学習資源としては、文章に対して喜怒哀楽等の主観的・客観的な感情をラベル付けしたコーパス [15], [16], [17] がある。また、対話や発話ごとに感情 [18] や意図 [19], [20] をラベル付けしたコーパスがある。しかし、本研究では発話を受けた聞き手の端的な感想を対象にしている。そのため、前述した学習資源 [12] も含めて既存のコーパスを本研究では使用できないため、新たな資源を収集する必要がある。

3. 提案手法

3.1 方向性の検討

例えツッコミ文の生成には、発話文に対する適切なツッコミ対象が求められる。画像や音声等に対応したマルチモーダル対話とは異なり、本研究はテキストを対象にしているため、発話文からツッコミ対象を推論する必要がある。ソーシャルセンサに関する研究では、Twitter の分析により道路交通情報や震源地の推論方法 [21], [22] が提案されており、信頼性が高い情報を収集できることが示されている。Tweet 文に対する Reply 文には、聞き手の感想になりうる単語が含まれる場合がある。たとえば、前述した“息子は足が速いから将来は金メダルが取れるな。”という Tweet 文に対して、Reply 文が“それは大袈裟じゃない？”の場合、文中に表れる“大袈裟”をツッコミ対象として利用できる。そこで、関連研究 [12] と同様に Twitter から取得した Tweet・Reply 文を用いて学習資源を収集して、ツッコミ対象の推論モデルを作成する。本研究では、物事の状態や性質を表す形容詞・形状詞をツッコミ対象とする。また、Tweet 文を発話文として、Reply 文から取得した形容詞・形状詞をラベル付けに用いる。

生成する例えツッコミ文は、多様なツッコミ対象に対応させる必要がある。日本のお笑い芸人が発言した例えツッコミ文が、例文としてインターネット上で公開されている。たとえば、くりいむしちゅー上田普也の“しつこいよお前は、エイリアン 2 か！”やフットボールアワー後藤輝基の“奥歯のほうで猫死んでますよ！”がある。このような例えツッコミ文を本研究で利用することは可能ではあるが、公開件数には限りがある。近年、GPT4^{*6}等の言語モデルが発展しており、文書生成技術を用いることで例えツッコミ文を拡充することは可能である。そこで本研究では、取得した例文を基に言語モデルを用いて例えツッコミ文を拡充して、ツッコミ対象に基づく例えツッコミ文を生成する仕組みを検討する。

例えツッコミ文を生成する流れを図 1 に示す。(A) 発話文に対するツッコミ対象の推論機能では、発話文から聞き手の端的な感想であるツッコミ対象が推論される。たと

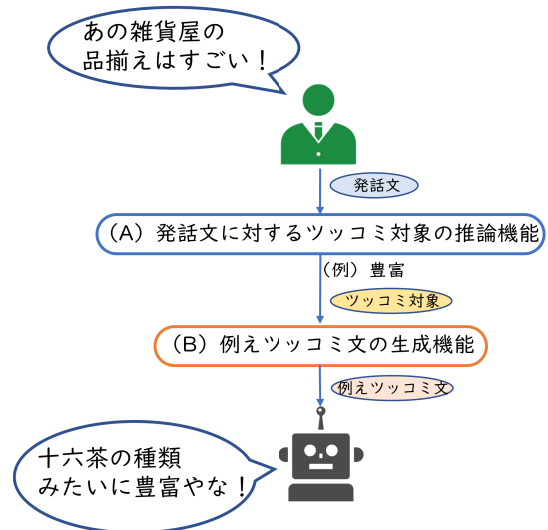


図 1 例えツッコミ文を生成する流れ
Fig. 1 The flow to generate tatoe-tsukkomi sentences.

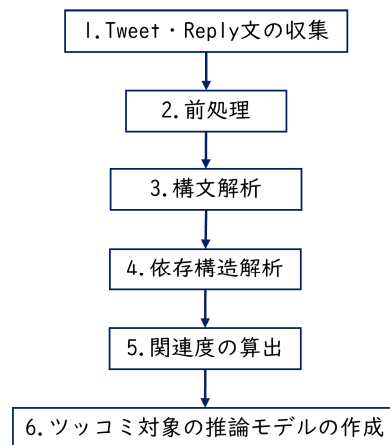


図 2 ツッコミ対象の推論モデルを作成する流れ
Fig. 2 The flow to make the inference function of tsukkomi targets.

えば、発話文が“あの雑貨屋の品揃えはすごい！”の場合、ツッコミ対象として“豊富”が推論される。(B) 例えツッコミ文の生成機能では、推論されたツッコミ対象に基づき例えツッコミ文が生成される。たとえば“豊富”を入力した場合、“十六茶の種類みたいに豊富な！”が生成される。

3.2 発話文に対するツッコミ対象の推論

3.2.1 Tweet・Reply 文の収集と前処理

ツッコミ対象の推論モデルを作成する流れを図 2 に示す。学習資源を収集するために、2022 年 3 月 22 日から 2022 年 6 月 15 日までを対象にして、Twitter API^{*7}を用いて 42,144,194 件の Tweet・Reply 文を収集した(図 2 の 1)。

収集した Tweet・Reply 文の前処理(図 2 の 2)として、英数字の半角・小文字変換や絵文字・顔文字の削除、画像・

*6 GPT4, <https://openai.com/gpt-4>

*7 Twitter API, <https://developer.twitter.com/ja/docs>

動画 URL, 疑問文, 否定文, 英文が含まれる Tweet 文の除外等を行った。また, 形容詞・形状詞が含まれない Reply 文はラベル付けに用いられないため, 対応する Tweet 文も含めて除外した。Twitter は自由発話为前提であり, 正確な文法は必ずしも求められないため, Tweet・Reply 文には文法誤りがある場合がある。たとえば, “多くの人が乗っているエレベータでオナラでた!” や “コンビニでパン買った。” の場合, 前者は “オナラ (が) でた!”, 後者は “パン (を) 買った。” が文法として正しい文になる。このような誤りは後述する構文解析 (図 2 の 3) に影響を与える可能性が高い。そこで, 本研究では助詞を対象として, MLM を用いて Tweet・Reply 文の文法誤りを訂正する。MLM は文脈を考慮して特定箇所の単語を推論できる。たとえば “多くの人が乗っているエレベータでオナラ <MASK> でた!” の場合, <MASK> の位置にふさわしい単語として助詞の “が” 等が出力される。そこで, 主語・述語間と目的語・述語間を補完対象として, これらに用いられる名詞, 動詞, 形容詞, 形状詞の単語間に助詞がない場合に <MASK> を挿入して助詞を補完した。品詞の特定には日本語の自然言語処理ライブラリである GINZA^{*8} を用いた。収集した Tweet・Reply 文のうち, 370,618 件が補完対象となった。精度を確認するために, 補完された文から無作為に抽出した 1,000 件を手で確認^{*9}した結果, 898 件は正しく処理されていた。正しく補完された例としては, “天気が悪いね。” や “毎日ご飯を作るのも嫌だよねえ” になる。一部, 正確に補完されなかったが 90% の文は正しく補完されたため, 処理されたすべての文を用いた。

3.2.2 構文解析による文意把握が困難な Tweet 文の除外

話題に関する語が省略されることで, 文意の把握が困難な Tweet 文がある。たとえば “あるといいですねえ〜。” という Tweet 文は, 主題の未提示により文意が把握できないため, 適切なツッコミ対象のラベル付けが困難になる。仮に “あるといいですねえ〜” に “素敵” がラベル付けされた学習資源を用いた場合, “不幸があるといいですねえ〜” や “時間があるといいですねえ〜” 等の発話文に対して, 推論モデルでは不適切なツッコミ対象が出力される可能性が高い。そのため, 文意の把握が困難な Tweet 文は除外することが望ましい。

主語・述語や目的語・述語が抽出できれば, 大まかではあるが Tweet 文の文意を把握できる。たとえば, “先日訪問した徳島県は寒かった。” の主語・述語は “徳島県, 寒かった” に, “昨日スーパーでサンマを買った。” の目的語・述語は “サンマ, 買った” になる。そこで, Tweet 文を構文解析して主語・述語または目的語・述語が抽出できない Tweet 文とそれに対応する Reply 文を除外する (図 2 の 3)。本

処理により, 37,361,384 件の Tweet・Reply 文を除外した。

3.2.3 依存構文解析によるツッコミ対象になりうる単語の特定

Reply 文には Tweet 文の内容に係らない形容詞・形状詞が用いられる場合がある。たとえば, Tweet 文が “あのコンビニの店員はとても親切です。” で, Reply 文が “私もそう思うけど, あそこはとても辛いカレーを売ってるね。” の場合, Reply 文の “辛い” は “カレー” に係っており, ツッコミ対象としては不適切である。そこで, ツッコミ対象になりうる単語 (以下, 候補) を依存構造解析により特定する (図 2 の 4)。本解析では GINZA を用いて Universal Dependencies (UD) の枠組み [23] で単語間の依存関係を解析する。図 3 に依存構造解析の例を示す。UD では, 品詞タグが付与された 2 語間に方向を持った依存構造ラベルが付与され, root は文全体の主辞であることを表す。図 3 では, root の形状詞である “豊富” から advcl ラベル (副詞的修飾節) で “種類” に係っている。なお, 候補として選ばれた形容詞・形状詞はすべて見出し語として扱う。

形容詞・形状詞は目的語や修飾語等, 多様な目的で用いられる。そこで, 依存構造解析から候補の特定方法を検討するために, 図 2 の 4 まで処理された 968,391 件の Reply 文を対象に, 形容詞・形状詞の用法を調査した。表 1 に調査結果^{*10}を示す。形容詞の平均利用数 (標準偏差, 以下略) が 0.88 (0.75) で, 依存構造ラベルの上位 3 件は root が 45.94%, advcl が 26.35%, acl (形容詞的修飾節) が 17.75% であった。形状詞の平均利用数が 0.55 (0.67) で, 依存構造ラベルの上位 3 件は root が 37.82%, advcl が 28.71%, acl が 18.25% であった。そこで, 解析で注目する依存構造ラベルは大部分を占める root, advcl, acl とする。また, 主辞である root は Reply 文の主要語であるため, root を基準に候補を決定する。

“声を掛けたら他人でした。” という Tweet 文を例にして, 候補を特定するための定義と解析例を図 4 に示す。青字の単語は選ばれた候補を示しており, (A)~(F) の 6 つの定義に従い解析を行う。まず, root の形容詞・形状詞を候補とする (図 4 の A)。たとえば, “それは恥ずかしい。” の場合, root の形容詞である “恥ずかしい” が候補となる。ただし, root であっても別の話題に関する感想である可能性があるため, 代名詞を除き主語が Reply 文に含まれる場合は対象外とする (図 4 の B)。

次に, root が形容詞・形状詞でない場合, root から acl または advcl で係る形容詞・形状詞を候補とする (図 4 の C)。たとえば, “恥ずかしい奴やな。” の場合, root の “奴” から acl で係る “恥ずかしい” を候補とする。“恥ずかしくなるわ” の場合, root の “なる” から advcl で係る “恥ずかしい^{*11}” を候補とする。

^{*8} GINZA, <https://megagonlabs.github.io/ginza/>

^{*9} 本研究では, 本学構成員 2 人で調査して両者の見解が異なる場合は相談して決定する作業を手による確認とする。

^{*10} 本研究では小数点以下第 3 位を四捨五入している。

^{*11} “恥ずかしく” を見出し語に変換している。

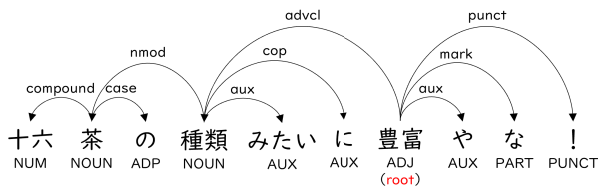


図 3 依存構造解析の例

Fig. 3 Example of dependency structure analysis.

表 1 依存構造ラベルが占める割合

Table 1 Percentage of dependency structure labels.

No.	形容詞		形状詞	
	ラベル	割合	ラベル	割合
1	root	45.94	root	37.82
2	advcl	26.35	advcl	28.71
3	acl	17.75	acl	18.25
4	dep	2.49	dep	3.32
5	compound	1.23	nmod	2.79
6	obl	1.22	obl	2.75
7	ccomp	1.10	amod	1.92
8	nmode	0.88	compound	1.38
9	nsubj	0.79	ccomp	0.92
10	advmod	0.57	advmod	0.82

Reply 文には複数の形容詞・形状詞が表れる場合があるが、(A), (B), (C) に従い候補を特定する。たとえば、“忙しいからシンプルなコメントやけど恥ずかしいわ。”の場合、“忙しい”と“恥ずかしい”が形容詞になる。このような場合は、(A) に従い root の“恥ずかしい”を候補とする(図 4 の D)。“狭いエレベーターで恥ずかしい奴やな。”のように root が形容詞・形状詞でない場合、(C) に従い root から acl または advcl で係る形容詞・形状詞である“恥ずかしい”を候補とする(図 4 の E)。“恥ずかしすぎて熱くなるわ。”のように root から acl または advcl で複数の形容詞・形状詞に係る場合、これら形容詞・形状詞である“恥ずかしい”と“熱い”を候補とする(図 4 の F)。

3.2.4 関連度によるツッコミ対象の判定

候補にはツッコミ対象として不適切な形容詞・形状詞が選ばれる可能性がある。たとえば、図 4 の上部に示した“声を掛けたら他人でした”という Tweet 文に対して Reply 文が“古い瓶やな”の場合、(C) に従い“古い”が候補になるが、Tweet 文の内容にはまったく関連しない。そのため、各候補がツッコミ対象になるかを判定する必要がある。そこで、Tweet 文の主語・述語や目的語・述語に用いられる名詞・動詞・形容詞・形状詞と、候補との関連度を算出する(図 2 の 5)。いずれか 1 つでも関連度が閾値を超えた場合、候補は Tweet 文の内容に関連がありツッコミ対象として適切であると判断する。関連度の算出には、分布仮説に基づき文脈に現れるトピックを考慮して単語間の関連度を算出する fastText^{*12}を用いる。分散表現の算出

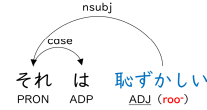
*12 fastText, <https://fasttext.cc/>

【Tweet文】

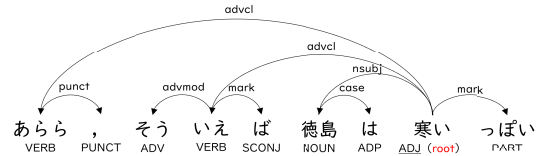
声を掛けたら他人でした

【Reply文】

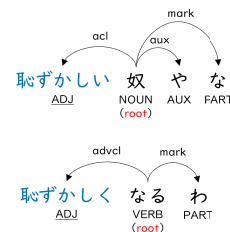
(A) rootの形容詞・形状詞



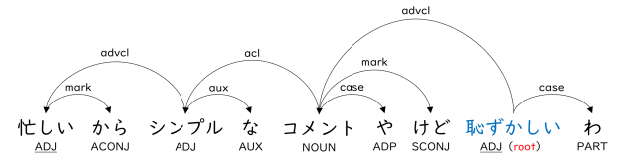
(B) rootの形容詞・形状詞で代名詞を除き主語がある場合は対象外



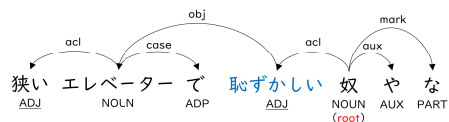
(C) rootからacl (形容詞的修飾) またはadvcl (副詞的修飾) で係る形容詞・形状詞



(D) 複数の形容詞・形状詞がある場合、rootの形容詞・形状詞



(E) 複数の形容詞・形状詞がある場合、rootからaclまたはadvclで係る形容詞・形状詞



(F) 複数の形容詞・形状詞がある場合、rootからaclまたはadvclで係る複数の形容詞・形状詞

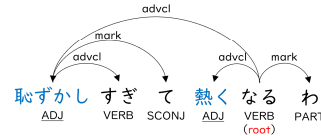


図 4 候補を特定するための定義と解析例

Fig. 4 Defines and analysis examples to determine tsukkomi target candidates.

には、Wikipedia 日本語記事全文から作成されたモデルを用いた。図 5 に関連度の算出例を示す。図 5 の上部において、Reply 文の候補である“恥ずかしい”と、Tweet 文の“声”、“掛け”、“他人”との関連度は 0.26, 0.16, 0.43 になる。

閾値を設定するために、図 2 の 4 まで処理された Tweet・Reply 文から無作為に抽出した 1,000 件を対象に、関連度を算出して閾値を調査した。閾値の値域は 0.01~0.40 で、

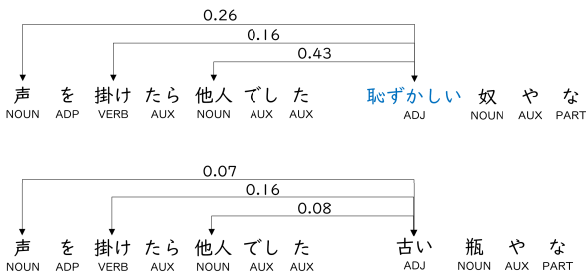


図 5 関連度の算出例

Fig. 5 Examples of relevance calculation.

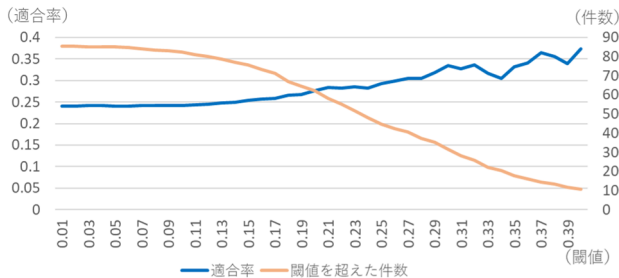


図 6 適合率と閾値

Fig. 6 Precision and threshold.

閾値を超えた単語がツッコミ対象として適切であるかを人手で確認して、適合率と閾値を超えた件数を調べた。たとえば、閾値 0.01 では真陽性が 206 件で偽陽性が 650 件になり適合率は 0.24 になる^{*13}。図 6 に調査結果を示す。閾値の上昇にともない適合率が高まる傾向にあるが、閾値を超えた件数は減少している。たとえば、適合率と閾値を超えた件数は、閾値が 0.01 の場合は 0.24 と 856、閾値が 0.4 の場合は 0.37 と 107 であった。閾値が高い場合は学習資源を十分に収集できないため、閾値を超えた件数が 300 件以上かつ適合率が最も高い 0.3 を本研究の閾値とした。図 5 の上部では、“恥ずかしい”と“他人”の関連度が閾値を超えているため、“恥ずかしい”をツッコミ対象とする。一方で、図 5 の下部では、“古い”と“声”、“掛け”、“他人”との関連度はすべて 0.3 以下であり閾値を超えないため“古い”は除外される。なお、Reply 文に複数の閾値を超えた候補がある場合は、関連度が最も高い単語をツッコミ対象として用いる。

図 2 の 5 までにおいて、Tweet・Reply 文は 333,473 件で 1,387 種類のツッコミ対象が収集された。収集されたツッコミ対象のうち、全体に占める割合が高い上位 10 件のツッコミ対象を表 2 に示す。最も多くを占めたツッコミ対象は“嬉しい”であり、次に“可愛い”や“美味しい”が多く占めていた。最後に、収集した学習資源を用いて、BERT^{*14}の事前学習モデルをファインチューニングして、ツッコミ対象の推論モデルを作成した(図 2 の 6)。

*13 適合率は真陽性を真陽性と偽陽性の和で割ることで求められる。

*14 BERT

https://huggingface.co/docs/transformers/model_doc/bert

表 2 ツッコミ対象が全体に占める割合

Table 2 Percentage of tsukkomi target to the total.

ツッコミ対象	全体に占める割合
嬉しい	14.56%
可愛い	8.59%
美味しい	4.68%
好き	3.45%
凄い	3.41%
楽しい	3.28%
面白い	2.41%
素敵	2.17%
怖い	2.02%
大事	1.89%

3.3 例えツッコミ文の生成

インターネット上で公開されている例文 305 件に対して、MLM を用いて例えツッコミ文を拡充する。具体的には、取得した例文の名詞、形容詞、形状詞の内、いずれか 1 つを別の単語に変換して新たな例えツッコミ文を生成する。たとえば、例文が“中森明菜の声くらいの細さ!”の場合、“中森明菜の声くらいの<MASK>!”から MLM により“上手さ、強さ、美しさ、高さ、綺麗さ、凄さ、滑らかさ、忠実さ、セクシーさ、柔らかさ、元気さ”等の単語を取得して、“中森明菜の声くらいの上手さ!”等の例えツッコミ文を生成する。生成した例えツッコミ文にラベル付けるツッコミ対象は、図 4 の定義に従い例えツッコミ文から特定された形容詞・形状詞とする。たとえば“弱弱しいな、トマトしか食べてないの?”では、“弱弱しい”がツッコミ対象になる。ツッコミ対象が特定できない例えツッコミ文は除外する。事前調査の結果、MLM の推論件数の増加にともない、生成元の文には適さない単語が多く出力される傾向が見られたため、MLM の推論件数は 20 件とした。以上の処理により、例えツッコミ文 4,847 件が生成されたが、最後に人手で確認^{*9}して“自動車の声くらいの細さ”のように、単語の組合せが明らかに不自然と両者が判断した 22 件の例えツッコミ文を除外した。

最終的にツッコミ対象は 1,020 種類で例えツッコミ文は 4,825 件になった。3.2 節で示した上位 10 件のツッコミ対象(表 2)を参考にして、MLM により生成された例えツッコミ文の例を表 3 に示す。変換された単語を下線で示す。たとえば、ツッコミ対象が“嬉しい”の場合は“金メダルを取るくらい嬉しい!”が、“楽しい”の場合は“コナンの毛利探偵の推理くらい楽しい!”が生成された。

1 つのツッコミ対象に対して複数の例えツッコミ文が紐づく場合がある。たとえば、“固い”というツッコミ対象には、“フランスパンぐらい固い!”や“コンクリートぐらい固い!”等が紐づいている。そこで、例えツッコミ文の生成機能としては、まず例えツッコミ文とツッコミ対象を DB に登録する。そして、推論されたツッコミ対象を検索

表 3 生成した例えツッコミ文の例

Table 3 Examples of generated tateo-tsukkomi sentences.

ツッコミ対象	例えツッコミ文
嬉しい	金メダルを取るぐらい嬉しい！
可愛い	寝起きの赤ちゃんぐらい可愛い！
美味しい	ごはんのにせられても怒らないぐらい美味しい！
好き	林家パー子の笑いのツボくらい好き！
凄い	凄い。ジャニーズでデビューするぐらいの驚き！
楽しい	コナンの毛利探偵の推理ぐらい楽しい！
面白い	杉村太蔵のコメントぐらい面白い！
素敵	ドラえもののエンディングみたいに素敵！
怖い	ミステリーの結末ぐらい怖い！
大事	寝起きのポカリくらい大事！

表 4 開発に使用した主なソフトウェア

Table 4 A list of used main software.

種類	ソフトウェア
仮想化ソフトウェア	VMware ESXi 6.7.0
OS	CentOS 7.4, Ubuntu 18
Http サーバ	Apache 2.4.6, Flask 2.0.3
データベース	MariaDB 5.5.52
プログラミング言語等	PHP 5.4.16, Python 3.6.8, jQuery 3.5.1, HTML
その他	Docker 20.10.21

条件として、無作為に選ばれた例えツッコミ文を出力する。

3.4 システム開発

提案手法に基づく機能を実装した非タスク指向型のユーモア生成対話システムを開発した。雑談対話モデルは公開されている日本語 Transformer Encoder-decoder 対話モデル^{*15}を用いた。本システムは 2 台のサーバを用いて開発した。1 台目のサーバには、Web システムとして対話画面等を VMware ESXi に構築した CentOS 上に、PHP や jQuery 等を用いて開発した。パソコンに加えてスマートフォンの利用率も高いことを想定し、レスポンスデザインにより数種類の端末向け画面を設計した。雑談対話機能とツッコミ対象の推論機能、例えツッコミ文の生成機能は、2 台目の GeForce RTX 3090 を搭載したサーバに開発した。具体的には、応答文を返す WebAPI として、Ubuntu 上の Docker 環境に python 等を用いて各機能を開発した。開発に使用した主なソフトウェアを表 4 に、開発したシステムの参考画面を図 7 に示す^{*16}。生成された例えツッコミ文は、雑談対話モデルで生成される応答文の前に表示する。対話は下から上に流れており、青色の文字列が生成された例えツッコミ文になる。例えツッコミ文は主語・述語または目的語・述語がある発話文を対象に生成する。暫定的に例えツッコミ文はユーザが 3 回発話するたびに 1 回生成

^{*15} 日本語 Transformer Encoder-decoder 対話モデル

<https://github.com/nttcs/nttcs-japanese-dialog-transformers>

^{*16} 視認性を高めるために、図の切り取りや拡大等の加工をしている。



図 7 ユーモア生成対話システムの参考画面

Fig. 7 A screenshot of the dialogue system generating humor.

するが、今後の評価を通して生成頻度を決定する。なお、応答文の吹き出しには、ユーザによる評価を目的として、Good・Bad ボタンを設置している。

4. 評価実験

提案手法の有効性を評価するために評価実験を実施した。4.1 節では、ツッコミ対象の推論機能 (3.2 節) により推論されたツッコミ対象の適切さを明らかにするために評価を行った。4.2 節では、例えツッコミ文の生成機能 (3.3 節) により生成された例えツッコミ文のユーモアの喚起具合等や、雑談対話に例えツッコミ文を交えた影響を明らかにするために評価を行った。被験者は大学職員 2 人、クラウドソーシングで募った高校生 10 人と大学生 4 人の計 16 人になる。なお、高校生と大学生 14 人との面識はないが、正確な評価をするため被験者に本実験の目的を伝えているため、ピグマリオン効果等のバイアスが発生している可能性がある。また、本研究では主に高校生を対象として評価実験を行ったため、本実験で得られた結果の解釈には注意が必要である。

4.1 ツッコミ対象の推論機能の比較評価

図 2 の 1, 2 で収集・前処理された Tweet・Reply 文を対象に、ツッコミ対象の推論モデル P1~P4 を作成して比較評価をした。各モデルの作成に用いた学習資源の詳細を表 5 に示す。P1 の発話文は、前述の収集・前処理された Tweet 文を使用している。P2 は関連研究 [12] と同じモデルであり、P4 は本研究で提案したモデル (3.2 節) になる。学習資源は図 2 の 5 まで処理された件数と同じ 333,473 件

表 5 各モデルの作成に用いた学習資源
Table 5 Training data of four models in evaluation.

モデル	発話文	ツッコミ対象
P1	Tweet 文	Reply 文の最初に表れる形容詞・形状詞
P2	構文解析 (図 2 の 3) で一部除外された Tweet 文	Reply 文の最初に表れる形容詞・形状詞
P3	構文解析 (図 2 の 3) で一部除外された Tweet 文	依存構造解析 (図 2 の 4) で選ばれた形容詞・形状詞
P4	構文解析 (図 2 の 3) で一部除外された Tweet 文	依存構造解析 (図 2 の 4) と関連度 (図 2 の 5) で選ばれた形容詞・形状詞

表 6 各モデルで推論されたツッコミ対象の例
Table 6 Examples of tsukkomi targets inferred by each model.

発話文	P1	P2	P3	P4
生まれ故郷の北海道に帰ろうかなあ.	素敵	美味しい	素敵	寂しい
私も今日は母に電話してみようかな.	大きい	悲しい	大事	大事
先日ようやく憧れのゲーム制作会社に就職が決まったんです!	好き	好き	楽しい	素敵
私は専門学校に通う女の子だよ.	大丈夫	可愛い	素敵	可愛い
私もたまには制服以外の服が着たいなあ.	多い	多い	楽しい	難しい
楽しみにしていた旅行が中止になった.	残念	残念	残念	残念
つわりで寝ていたら夫がご飯を作ってくれた.	大事	大丈夫	偉い	優しい
口コミ通りの美味しいお店だった.	凄い	素敵	美味しい	美味しい
おすすめのホラー映画を観た.	怖い	好き	怖い	怖い
夫が相談もなく勝手に冷蔵庫を買い替えてしまった.	大事	偉い	凄い	つらい

として、表 5 の内容を除きすべて同じ条件でモデルを作成した。P1・P2 間で構文解析により文意の把握が困難な Tweet 文を除外する処理方法 (図 2 の 3), P2・P3 間で依存構造解析により候補を特定する処理方法 (図 2 の 4), P3・P4 間で関連度によりツッコミ対象を決定する処理方法 (図 2 の 5) を評価した。インターネット上で公開されている対話データ*17から、無作為に抽出した発話文を用いて各モデルでツッコミ対象を推論した。被験者別に異なる 10 件の発話文と推論されたツッコミ対象を提示して、各対象が適切であるかの評価を被験者に依頼した。評価尺度は 5 段階 (1: 適切でない~5: とても適切である) とした。各被験者の評価件数は 1 モデルにつき 10 件で、各モデルは 160 件の評価がされた。検定はウィルコクソンの符号付き順位和検定による多重比較法 (基準値 0.05 とした BH 法 [24]) を用いた。

各モデルで推論されたツッコミ対象の例を表 6 に示す。たとえば、発話文が“生まれ故郷の北海道に帰ろうかなあ”の場合、P1 では“素敵”, P2 では“美味しい”, P3 では“素敵”, P4 では“寂しい”が推論された。評価実験の結果を図 8 に示す。4 以上の評価の割合は、P1 が 35%, P2 が 35%, P3 が 43.13%*10, P4 が 53.75%であり、P4 が最も高くなった。検定の結果、有意水準は P1・P2 間が 0.11, P1・P3 間が 0.02×10^{-1} , P1・P4 間が 0.01×10^{-1} , P2・P3 間が 0.41×10^{-1} , P2・P4 間が 0.02×10^{-1} , P3・P4 間が 0.24×10^{-1} となり、P1・P2 間以外で有意差が認めら

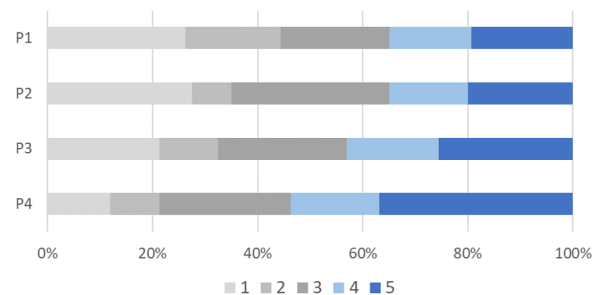


図 8 推論モデルの評価結果
Fig. 8 Evaluation results of inference models.

れた。

P1・P2 間は有意差が認められなかったが、2 以下の評価の割合は P1 が 44.38%で P2 は 35%であった。構文解析により文意の把握が困難な Tweet 文を除外することで、P2 の推論精度が P1 より若干ではあるが高まった可能性がある。P2・P3 間では、依存構造解析により Tweet 文の内容には係らない形容詞・形状詞を除外してツッコミ対象になりうる単語を特定したことで、P3 ではより適切な推論がされたと判断できる。P3・P4 間では、候補の中から Tweet 文と関連性が高い単語がツッコミ対象として選ばれたことで、P4 ではより適切な推論がされたと判断できる。以上より、構文解析においては有意差が認められなかったが、学習資源を収集するための各処理方法は、より適切なツッコミ対象の推論に有効であったと考えられる。

P4 で評価値が高かった発話文/ツッコミ対象としては、たとえば“今日も早起きして一仕事してきましたよ!/偉い”や“家族を養ってくれた父を心から尊敬してるんだ。/素晴らしい”, “先日ようやく憧れのゲーム制作会社に就職が決

*17 評価実験で用いた対話データ
<https://github.com/nttclab/japanese-dialog-transformers/blob/main/README-jp.md>

まったんです！/素敵”がある。これらのツッコミ対象は発話内容に対して聞き手の端的な感想を正しく表現できており、適切な例えツッコミ文の生成につなげることができる。

P4で評価値が低かった例としては、“給料のほとんどを貯金にまわしています。/安い”がある。“安い”がラベル付けされた Tweet 文には“税金”や“賃金”等のお金に関する単語が多く含まれていた。“貯金にまわしています。”からは“偉い”が推論されたため、発話文の“貯金”が不適切な推論の要因になった可能性が高い。他には、“東京に住んだら出会いのチャンスが増えそうな気がする！/好き”がある。“好き”がラベル付けされた Tweet 文の内容は多様であったが、“東京”や“出会い”等の単語も多く含まれていたため、P4では発話文に関連がないツッコミ対象が推論された可能性がある。

このような不適切な推論の改善方法としては、たとえばシステムの利用によるユーザ評価がある。試行段階で本システムの評価を依頼したり、実運用を通してフィードバックを集めたりすることで、推論精度の改善につなげられる。前述した給料に関する発話文は、P2とP3では“素晴らしい”が推論されておりP4より評価値が高くなった。他にも発話文によってはP4より他モデルが適切に推論されていた。P4のモデルを単独で使わず、複数のモデルを組み合わせた推論手法を検討することは可能である。たとえば、関連度の閾値(3.2.4項)を高めたモデルや評価実験で用いたモデル等を用いてアンサンブル学習をすることで、推論精度を高められる可能性がある。

同じ発話文とツッコミ対象であっても、被験者により評価値が異なる場合があった。たとえば、“そして変わらず女性が苦手だよ。/同じ”において、評価値が1または5に分かれていた。これは、被験者自身の女性への苦手意識の有無が大きく影響した可能性がある。実際、評価値1は女性、評価値5は男性が選択していた。受け手によりツッコミ対象の適切さが異なるため、たとえばユーザの属性情報に基づく学習資源の収集やツッコミ対象の推論、評価が分かれるツッコミ対象の除外等、多様なユーザを考慮した設計が必要と考えられる。

4.2 例えツッコミ文の生成機能と雑談対話に例えツッコミ文を交えた影響の評価

例えツッコミ文の生成機能(3.3節)を評価するために、公開されている対話データから無作為に発話文を抽出した。そして、被験者別に異なる発話文10件と、P4(4.1節)で推論されたツッコミ対象に基づき生成された例えツッコミ文10件を、被験者に提示して評価を依頼した。評価項目は以下Q1とQ2として、評価尺度は5段階(1:感じなかった~5:とても感じた)とした。各被験者の評価件数は1項目につき10件で、項目別に例えツッコミ文は160件の評価がされた。次に、雑談対話に例えツッコミ文を交

えた影響を評価するために、被験者に開発した対話システム(3.4節)を用いて作成した対話動画の視聴と、例えツッコミ文がない場合との比較を前提とした以下Q3~Q5の評価を依頼した。評価尺度は前述した5段階を用いた。各被験者の評価件数は1項目につき1件で、項目別に対話動画は16件の評価がされた。対話動画では、公開されている対話データを参考に入力された発話文6件に対して、応答文6件と例えツッコミ文2件が返される。Uをユーザ、Sをシステム、生成された例えツッコミ文を下線で表した対話例を表7に示す。複数の対話パターンの動画を作成して、被験者は無作為に選ばれた1つの動画を視聴する。

Q1 ユーモアを感じたか？

Q2 発話文に対する例えツッコミ文として成立しているか？

Q3 より対話システムに対して親しみを感じたか？

Q4 より対話にユーモアを感じたか？

Q5 より対話を継続したいと感じたか？

生成された例えツッコミ文の例を表8に示す。たとえば、発話文が“まあそのおかげで趣味を満喫できてますけどね”の場合、“えっ！、化学調味料がいっさい入っていないくらい素晴らしい！”が生成された。評価結果を図9に示す。Q1とQ2における4以上の評価の割合は、Q1が56.25%でQ2が45%であり、ややユーモアが感じられてはいるが例えツッコミの成立・不成立の判断は困難となる結果であった。評価値が高い発話文/例えツッコミ文の例としては、“私も今日は母に電話してみようかな/えっ！、メモぐらい大事！”がある。ツッコミ対象は“大事”であり、“母に電話”と“メモ”の大事さは内容や関連性が理解しやすく、例えツッコミが成立して評価値が高くなったと考えられる。

評価値が低い例としては、“私はサーフィンが得意なのでよくサーフィンをします。/えっ！、泥棒猫が仲間にいるぐらい楽しい！”がある。ツッコミ対象として“楽しい”が推論されたが、“サーフィン”と“泥棒猫が仲間”の楽しさの関連性が理解しにくく、例えツッコミが成立しなかったと考えられる。提案手法ではツッコミ対象に紐づく例えツッコミ文を無作為に選んでいる。“楽しい”に紐づく他の例えツッコミ文として、“ブラジルのサンバぐらい楽しいよね！”がある。この例えツッコミ文が選ばれた場合、“サーフィン”と“サンバ”の楽しさは内容や関連性が理解しやすく、高評価が得られた可能性がある。このように、適切なツッコミ対象が推論されても発話文に対して例えツッコミ文が成立しない場合がある。そのため、たとえば発話文と例えツッコミ文間の意味的なつながりを分析する等、発話内容を考慮した適切な例えツッコミ文の選択方法を検討する必要がある。

Q3~Q5における4以上の評価の割合は、Q3が93.75%、Q4が68.75%、Q5が62.5%であった。対話を通してユーモアが喚起されたことで、システムへの親しみが感じられ

表 8 生成された例えツッコミ文の例

Table 8 Examples of tatoe-tsukkomi sentences generated by the proposed method.

発話文	例えツッコミ文
まあそのおかげで趣味を満喫できてますけどね.	えっ！, 化学調味料が一切入っていないくらい素晴らしい！
私も今日は母に電話してみようかな.	えっ！, メモぐらい大事！
先日ようやく憧れのゲーム制作会社に就職が決まったんです！	えっ！, 映画のエンディングみたいに素敵！
私は専門学校に通う女の子だよ.	えっ！, 天使ぐらい可愛い！
私はパートナーと二人暮らしです.	えっ！, 人生のエンディングみたいに素敵！
私は外国語を趣味でやっています.	えっ！, 焼肉ぐらい楽しい！
危険運転をする車に遭遇した.	えっ！, エイリアン 2 ぐらい怖い！
予防接種をしなかったがインフルエンザが猛烈に流行ってきた.	えっ！, 箱根駅伝でいったら権太坂ぐらい嫌い！
元気な頃の祖父の動画を見た.	えっ！, チャーシュー麺にチャーシュー追加するくらい面白い！
赤ちゃんの泣き声が大きくて, 元気に育ちそう.	えっ！, 犬のような可愛さ！

表 7 評価で用いた対話例

Table 7 Dialogue examples used in the evaluation.

発話・応答内容 (簡略化)
U: こんにちは.
S: こんにちは！
U: 将来の夢はある？
S: うん, まだ定まっていなくて. あなたは？
U: 私はいつか庭付きの一戸建てに住むのが夢です.
S: ドラえもんエンディングみたいに素敵！夢が叶うといいね！
U: ありがとう. でも高いからお金を貯めないといけないね.
S: そうですね. わたしもいつか庭付きの一戸建てに住みたいなあ.
U: あと, 猫を 10 匹飼いたい.
S: 10 匹か, 賑やかになりそうだね.
U: キャットフード代が高くなりそうだけど.
S: ブランドの子供服ぐらい高いよね！キャットフードは高級品だね.

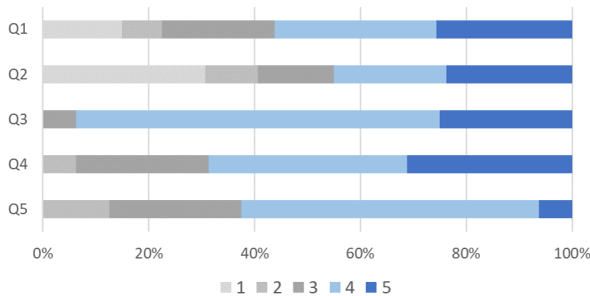


図 9 例えツッコミ文の評価結果

Fig. 9 Evaluation results of tatoe-tsukkomi sentences.

たと判断できる. 一方で, Q3 と Q4 の評価値が高くて Q5 に 1 や 2 をつけた評価者が数名いた. これはユーモアが感じられた一方で, 例えツッコミ文の生成頻度の高さが対話を妨げた可能性がある. 開発した対話システムでは, 暫定的に 3 回の発話ごとに例えツッコミ文が 1 回生成されるが, 生成回数を減らしたり対話の流れに沿って生成したりする等, 対話を妨げない仕組みが求められる.

4.3 今後の課題

聞き手によりツッコミ対象に感じる適切さは異なるが, 例えツッコミ文も同様にユーザを感じるユーモアの程度は

異なる. 同一条件ではなく個々の嗜好や属性を考慮することで, よりユーモアが喚起される例えツッコミ文を生成できる可能性がある. 嗜好等を把握する方法としては, たとえばシステム利用前のアンケートがある. 芸能やスポーツ等の好きなジャンル, 年齢層, 性別等を確認して, これらの結果に基づき例えツッコミ文を生成することで, ユーモア性を高められる可能性がある.

本研究では直前の発話文からツッコミ対象を推論しており, 過去の発話内容は考慮していない. 発話履歴を用いると, たとえば“昨日, 魚を買った”, “めっちゃ高かった”と発話が続いた場合, 高い魚を買ったことに対してツッコミ対象を推論できる. このように文脈を考慮してユーモア表現を生成することで, より自然な対話を実現できる. また, 開発した試作システムの入出力はテキストが対象であるが, 映像や音声等に対応させることも可能である. マルチモーダル対話により, たとえば画像認識によりユーザがユーモアを感じているかの視覚的把握や, ユーザの音声が高いつまみ場合は“サビを歌ってるときぐらい声大きい!”等の聴覚的要素に対するツッコミが可能になる.

本研究ではツッコミ対象を形容詞・形状詞に限定したが, 名詞等の他の品詞も対象にすることで, より多様な例えツッコミ文が生成できる. たとえば, 発話文が“息子はスポーツや勉強ができて性格も非常によい.”でツッコミ対象が名詞の“万能”になる場合, “アロエと並んだな!”という例えツッコミ文を生成できる. また, ツッコミ対象が動詞の“できる”の場合, “大谷翔平の様に何でもできるな!”が生成できる. ツッコミ対象の決定方法を検討する必要があるが, 名詞や動詞等も対象に加えることで, 多様な発話文への対応が可能になる. また, 本研究では Twitter から約 3 カ月分の Tweet・Reply 文を収集したが, 学習資源の件数を増やすことでツッコミ対象の推論精度を高められる. そのため, より適切な例えツッコミ文を生成するためにも, 質的・量的観点で継続的に学習資源を収集したい.

ChatGPT^{*18}等の大規模言語モデルを用いた対話システ

*18 ChatGPT, <https://openai.com/blog/chatgpt>

ムが普及しており、このようなサービスを用いても例えツッコミ文を生成することは可能である。ChatGPT 等と比較すると本研究は例えツッコミ文の生成に特化しており、自然言語処理ライブラリや言語モデル等の活用、さらには独自に実施した調査結果に基づき、ツッコミ対象の推論機能 (3.2 節) と例えツッコミ文の生成機能 (3.3 節) を設計した。そのため、生成された例えツッコミ文の精度改善において、開発したシステム (3.4 節) は不適切な出力結果の原因究明や機能単位での改修が行いやすい。また、本研究では例えツッコミ文を公開されている例文に MLM を用いて独自に作成しているため、少ない資源から多様な表現を出力することが可能である。精度比較やさらなる改善点の把握に向けて、ChatGPT 等により生成された例えツッコミ文、さらには技能を持った者が作成した文も対象として、提案手法との比較評価を行いたい。

5. まとめ

本研究では、対話を通してユーモアを喚起してシステムへの親しみを感じさせるために、発話を受けた聞き手の端的な感想に基づく例えツッコミ文の生成手法を提案した。具体的には、Twitter の Tweet・Reply 文に対して構文解析や依存構造解析、関連度の算出等により学習資源を収集して、発話文に対するツッコミ対象の推論機能を開発した。また、インターネット上で公開されている例文を基に言語モデルを用いて例えツッコミ文を拡充して、ツッコミ対象に基づく例えツッコミ文の生成機能を開発した。

主に高校生を対象にした評価実験の結果、学習資源を収集するための各処理方法は適切なツッコミ対象の推論に有効であり、例えツッコミ文を交えた雑談対話はユーモアが喚起されてシステムへの親しみが感じられること等が示された。一方で、発話文に対して例えツッコミ文をより成立させるためには、発話内容を考慮した適切な例えツッコミ文の選択方法を検討する必要がある。また、ユーモア性を高めるためには、ユーザ個々の嗜好や属性を考慮したユーモア生成や、音声や映像等に対応させたマルチモーダル対話、対象品詞の拡張等が今後の課題としてあげられる。

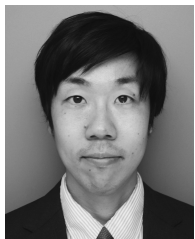
例えツッコミ文の生成頻度等の検討が必要ではあるが、本研究で開発したユーモア生成対話システムは高齢者や入院患者等、会話希望者の話し相手を親和的に務められる。例えツッコミに限定するとユーザは飽きを感じやすくなるが、漫才のように聞き間違いやノリツッコミ等の多様なユーモア表現を用いることで、よりユーザを楽しませることが可能になる。そのため、ボケとツッコミに関する他のユーモア生成機能も本システムに実装して、研究成果の実用化を目指したい。

謝辞 本研究は JSPS 科研費 19K14317 の助成を受けたものです。

参考文献

- [1] Khooshabeh, P., McCall, C., Gandhe, S., Gratch, J. and Blascovich, J.: Does It Matter If a Computer Jokes, *Proc. CHI'11*, pp.77–86 (2011).
- [2] Luger, E. and Sellen, A.: Like Having a Really Bad PA: The Gulf Between User Expectation and Experience of Conversational Agents, *Proc. CHI'16*, pp.5286–5297 (2016).
- [3] 宮澤幸希, 常世 徹, 榊井祐介: 音声対話システムにおける継続欲求の高いインタラクションの要因, 電子情報通信学会論文誌 A, Vol.95, No.1, pp.27–36 (2012).
- [4] He, H., Peng, N. and Liang, P.: Pun Generation with Surprise, *Proc. NAACL'19*, pp.1734–1744 (2019).
- [5] Valitutti, A., Toivonen, H., Doucet, A. and Toivanen, J.M.: Let Everything Turn Well in Your Wife: Generation of Adult Humor Using Lexical Constraints, *Proc. ACL'13*, pp.1734–1744 (2013).
- [6] Tinholt, H.W. and Nijholt, A.: Computational Humour: Utilizing Cross-Reference Ambiguity for Conversational Jokes, *Proc. WILF'07*, pp.477–483 (2007).
- [7] 呉 健朗, 中原涼太, 長岡大二, 中辻 真, 宮田章裕: ボケて返す対話型エージェント, 日本バーチャルリアリティ学会論文誌, Vol.23, No.4, pp.231–238 (2018).
- [8] 呉 健朗, 長岡大二, 中原涼太, 宮田章裕: 文のトピックを考慮した単語置換によるユーモア発話を行う対話型エージェント, 情報処理学会論文誌, Vol.61, No.1, pp.113–122 (2020).
- [9] 関 陽介: 用例に基づく疑似的な聞き間違いによりユーモア表現を生成する対話システム, 人工知能学会論文誌, Vol.36, No.2, pp.C-KA3.1–10 (2021).
- [10] 呉 健朗, 富永詩音, 武藤佑太, 宮田章裕: 複数対話型エージェントの役割分担によるユーモア生成システム, 情報処理学会論文誌, Vol.61, No.8, pp.1353–1362 (2020).
- [11] Seki, Y.: A Dialogue System Making Humor by Nori-Tsukkomi, *Proc. 2022 IEEE 11th Global Conference on Consumer Electronics*, pp.893–894 (2022).
- [12] Seki, Y.: Dialogue System to Make Humor by Tatoes-Tsukkomi, *Proc. 2022 Tenth International Symposium on Computing and Networking Workshops*, pp.147–150 (2022).
- [13] 村瀬 健: 最強のコミュニケーション ツッコミ術, 祥伝社新書 (2015).
- [14] Chen, H., Liu, X., Yin, D. and Tang, J.: A Survey on Dialogue Systems: Recent Advances and New Frontiers, *ACM SIGKDD Explorations Newsletter*, Vol.19, No.2, pp.25–35 (2017).
- [15] Buechel, S. and Hahn, U.: EMOBANK: Studying the Impact of Annotation Perspective and Representation Format on Dimensional Emotion Analysis, *Proc. 15th Conference of the European Chapter of the Association for Computational Linguistics*, Vol.2, pp.578–585 (2017).
- [16] Kajiwara, T., Chu, C., Takemura, N., Nakashima, Y. and Nagahara, H.: WRIME: A New Dataset for Emotional Intensity Estimation with Subjective and Objective Annotations, *Proc. 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp.2095–2104 (2021).
- [17] Mohammad, S. and Bravo-Marquez, F.: Emotion Intensities in Tweets, *Proc. 6th Joint Conference on Lexical and Computational Semantics (*SEM 2017)*, pp.65–77 (2017).
- [18] Rashkin, H., Smith, E.M., Li, M. and Boureau, Y.-L.: Towards Empathetic Open-domain Conversation Mod-

- els: A New Benchmark and Dataset, *Proc. 57th Annual Meeting of the Association for Computational Linguistics*, pp.5370–5381 (2019).
- [19] Li, Y., Su, H., Shen, X., Li, W., Cao, Z. and Niu, S.: DailyDialog: A Manually Labelled Multi-turn Dialogue Dataset, *Proc. 18th International Joint Conference on Natural Language Processing*, Vol.1: Long Papers, pp.986–995, Asian Federation of Natural Language Processing (2017).
- [20] Hsu, C.-C., Chen, S.-Y., Kuo, C.-C., Huang, T.-H. and Ku, L.-W.: EmotionLines: An Emotion Corpus of Multi-Party Conversations, *Proc. 11th International Conference on Language Resources and Evaluation* (2018).
- [21] Sakaki, T., Okazaki, M. and Matsuo, Y.: Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors, *Proc. 19th International Conference on World Wide Web, WWW '10*, pp.851–860 (2010).
- [22] 榎 剛史, 柳原 正, 那和一成, 松尾 豊: Twitter を用いた道路交通情報の抽出, 電子情報通信学会論文誌 D, Vol.J98-D, No.6, pp.1019–1032 (2015).
- [23] 浅原正幸, 金山 博, 宮尾祐介, 田中貴秋, 大村 舞, 村脇有吾, 松本裕治: Universal Dependencies 日本語コーパス, 自然言語処理, Vol.26, No.1, pp.3–36 (2019).
- [24] 松田真一: FDR の概説とそれを制御する多重検定法の比較, 計量生物学, Vol.29, No.2, pp.125–139 (2008).



関 陽介 (正会員)

2008 年甲南大学工学部情報システム工学科卒業。2014 年徳島大学大学院博士前期課程修了。2017 年同大学大学院博士後期課程単位修得退学。博士 (工学)。2008 年 4 月から IT 系企業での勤務を経て、2011 年 3 月から徳島大学情報センター特任助教。2017 年 5 月より同大学総合教育センター (現, 高等教育研究センター) 特任研究員。特任講師を経て 2022 年 4 月より徳島大学高等教育研究センター准教授。対話システム, 大学入学者選抜等の研究に従事。電子情報通信学会, 人工知能学会, 日本教育工学会各会員。