

# 非タスク指向対話システムにおけるユーモア応答生成手法

## Humor Utterance Generation Method for Non-Task-Oriented Dialogue System

藤倉 将平\*<sup>1</sup>    小川 義人\*<sup>1</sup>    菊池 英明\*<sup>1</sup>

Shohei Fujikura    Yoshito Ogawa    Hideaki Kikuchi

\*<sup>1</sup>早稲田大学人間科学学術院

Faculty of Human Sciences, Waseda University

In this study, we propose a humor utterance generation method which compatible with dialogue system, to increase desire of sustainability. Dialogue system retrieves Modifier-Noun pairs from Twitter as knowledge and attempts to choose the most humorous reply, which derives high incongruity explained by incongruity-resolution model, using word similarity. We look the differences among individuals by SHQ and MSHS, through evaluation, and confirmed validity of the method, high-incongruity reply is significantly effective against the low-incongruity and random reply is significantly effective against the low-incongruity in a limited subjects, SHQ greater than 8 and MSHS under 15.

### 1. まえがき

対話システムにおける要素技術の高度化に伴って日常の場面で利用される状況が整いつつある。日常で用いられる対話システムにおいては、ユーザが長期間利用し続けられる、対話継続欲求を高めるデザインが求められる。

従来の対話システム研究では、話題を増やして多様な対話シーンに対応することでインタラクションの新奇性を高める試みがなされてきた。しかし、継続性の評価は十分に行われていなかった。インタラクションの新奇性を高めるデザインは短期的にはユーザの興味を引くものの、長期的な利用においては飽きられてしまう恐れがある。加えて、話題を増やすには人手による大量のデータからの場当たりの生成によるしかなく、ユーザが対話を続けたいと感じる条件が明確になっていないため、長期的に新奇性を維持することが難しい。

そこで宮澤ら [宮澤 12] は、人同士の対話を参考に、毎回の対話において「次回も続けたい」と感じる要因を分析することで、対話継続欲求の高いシステムのデザイン手法の確立を目指してきた。第一に人同士の対話を円滑化させる「社会性の高い対話」に着目し、会話分析によって要因を明確にした。第二に相手の関心を引く「ユーモア表現を含む会話」に着目し、ユーモア発話を行う対話システムを構築して評価実験を行った。その結果、「対話において相手の発話行動を限定しないこと」、「相手の話を聞いている実感を与えること」、「人工物であることを生かした意外性の高いユーモアを使うこと」が有効であることが明確になった。

対話継続欲求を高める手法は、幅広い対話システムに適用可能であることが望ましい。対話システムには、ロボットを代表とした実体を持つ対話システムと、Apple の Siri を代表とした、実体をもたない対話システムが存在する。従って、宮澤らによる会話分析から得られた知見をそのまま幅広い対話システムに適用することは難しい。一方で、実体の有無が問題にならず、言語情報だけから生成できるユーモアも存在する。従って、言語情報により生成可能なユーモアの活用は幅広い対話システムに適用が可能である。本研究では言語情報によるユーモアを扱うことでユーザの対話継続欲求を向上させることを目指す。

以下では、2. で先行研究について概説し、3. では、ユーモア応答を生成する手法を提案し、4. でその評価実験について述べ、5. で結論を述べる。

### 2. 先行研究

#### 2.1 ユーモア

「不適合-解決モデル [Suls 72]」では、不適合のみではユーモアが生じず、不適合の解決が行われることでユーモアが生成されると説明している。刺激をユーモアとして受容者が解釈する場合、他の感情が想起されてはいけないことが知られている。「不適合-解決モデル」では「解決」としての条件を加えることで、ユーモアが生成される条件をより明確にしている。

単語間の関連度を用いる Ambiguity と Distinctiveness の 2つの観点から、駄洒落の面白さを説明できる [Kao 13] ことが明らかになっている。不適合の強さを示す Distinctiveness の計算の中で、置換される単語と置換後の単語間の関連性が小さいほどにユーモアの受容性が向上している。従って、単語間の関連度は不適合の強さを表していると想定される。一方で、人手による評定実験から単語間の関連性を求めているため、狭く限られた範囲でしかユーモアを生成することしかできない。

#### 2.2 対話システム

特定のタスクに依存しない幅広い雑談を行う雑談対話システムが提案されている。用例ベース応答手法を用いた手法では統計的応答手法 [稲葉 12] などが提案されている。文脈から素性抽出を用いた教師あり機械学習を活用することで、対話システムにおける応答文章を決定する。また、概念の連想を用いて対話を行う連想応答手法 [吉村 13] も提案されている。常識的な連想を活用し、対話を行うことで自然な対話を行うことを目指している。しかし、これらの雑談対話システムの評価は対話の自然さを対象としたものであり、面白さの要因について考慮していない点が本研究とは異なる。

Twitter のデータは広範囲にわたる話題をカバーしており、多量のデータ収集は比較的容易である。加えて、対話形式になりやすいという性質があるため、Twitter のデータを用いた研究は数多く報告されている。稲葉らは非タスク指向型対話システムの発話の自動生成を目的として、Twitter から任意の話題語に関する発話を生成する手法 [稲葉 14] を提案している。

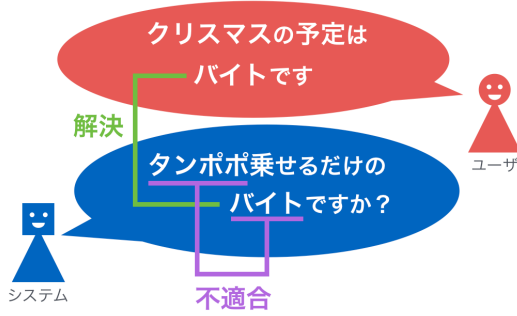


図 1: ユーモア生成の仕組み

駄洒落を含む様々なユーモアを扱う対話システム [吉田 14] では、ユーモアを扱わない対話システムと比較して、対話継続性や満足度において有効であることが明らかになっている。一方で、面白さの要因の解明には及んでいない。また、ユーザのイメージを覆すしりとりシステムとユーモア雑談システム [宮澤 12] においてもユーザの対話継続欲求が向上することが確認されている。ロボットという実態に期待の心象を抱かせ、ロボットが言わないと想定される発話を活用することで、抱かせた心象と異なる刺激を与え、ユーモアを生成している。しかし、日常の場面で利用される対話システムは、スマートフォン上で音声対話を扱ったアシストサービスのように、必ずしも実態のあるロボットのような人工物であるというわけではない。従って、人工物であることを生かした意外性によるユーモア生成手法は幅広い対話システムに応用することはできない。

### 3. 提案手法

2. を踏まえた上で、ユーモアを生成する手法を提案する。対話システムに対するイメージを覆す手法によるユーモアでは実体の有無が問題となるため、幅広い対話システムに応用できない。従って、言語情報によるユーモア生成を活用することで、幅広い対話システムに応用可能となることを目指す。

提案手法ではユーザの発話から名詞を抽出し、予め知識として用意してある先行表現と名詞を活用することでユーザに聞き返す。図 1 は、提案手法においてどのようにユーモアが生じているかを表している。提案手法は「不適合-解決モデル」によってユーモアを生じている。システム発話において、「タンポポ」と「バイト」の間で不適合が生じている。一般的な感覚から、2つの単語が関連することがないためだと推測される。システム発話のみでは、不適合のみが生じているため、奇妙な文章となる。しかし、対話において、ユーザが「バイト」について話した、という納得の再解釈としての「解決」の仮定を経ることで、対話システムによる奇妙な文章をユーモアとして受け取る。

対話システムが応答を生成するには知識が必要になる。本研究では、面白いと捉えられやすい多様な表現が多く存在し、様々なトピックに及ぶ多量の文章データを比較的容易に集めることのできる、マイクロブログの一つである Twitter を知識のリソースとして使用した。提案手法はユーザの発話に名詞が含まれている場合のみ有効である。しかし、雑談対話において広範囲に渡って対応できる範囲である。

図 2 はシステム全体のフローを表している。Twitter からツイートを集め、文章抽出、形態素解析を行い、先行表現と名詞の関係性を抽出し、発話候補として活用する。また、一般

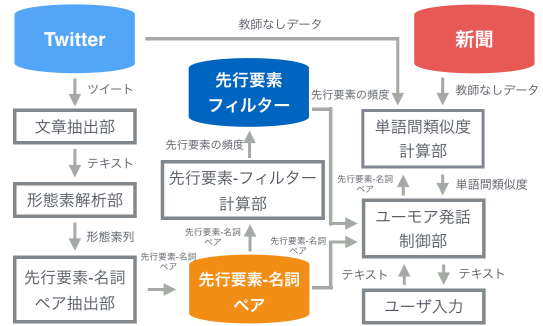


図 2: システム全体のフロー

的な表現を取り除くため、先行要素フィルターを経由する。また、ユーザ入力から抽出した名詞と発話候補における先行表現との間で単語間類似度計算を行うことで、発話候補から応答に用いる発話を選択する。

また、本研究で扱うユーモアは、テキスト刺激によって不適合が生じ、なおかつ受容者が面白いと感じるものと定義する。

#### 3.1 知識のデータ形式

Twitter から得られた文章から、どのように知識を抽出するのか記述する。

1. Twitter のツイートから抽出した文章に対して形態素解析を行う。実装は kuromoji を、辞書は IPADIC を用いた。なお、本研究では連続する名詞を一つの名詞として変換し単独の形態素として扱う。
2. 形態素列から「感動詞、フィラー」を取り除き、「接続詞、接続助詞」で分割し、それぞれの形態素列の「係助詞、格助詞」で文章を主語節と述語節に分割する。
3. 主語節と述語節のそれぞれに対し、先行要素と名詞のペアを取得し保存する。以降では先行要素-名詞ペアとする。

#### 3.2 単語間類似度計算

本研究においても、単語間類似度を用いて不適合の強さを表す。面白さの調査 [Kao 13] では単語間の関連性を人手によって評価しているため、狭く限られた範囲でしか面白さを説明できない問題点を取りあげた。従って、提案手法では多量の文章データから教師なし学習によって単語のベクトル表現を出力 [Mikolov 13] する word2vec [Řehůřek 10] を用いて、コサイン類似度として単語間類似度を算出することで、広い範囲で面白さを説明できるようにすることを目指す。学習には Twitter の 1700 万ツイート、毎日新聞データ集の 91 年から 94 年及び、96 年から 2001 年のデータを用いた。

### 4. ユーモア生成の評価実験

本節では、どのような発話がユーザにユーモアとして受容されるかを調査した。駄洒落における実験 [Kao 13] では、単語間の類似度は不適合の強さを表していると想定され、単語の関連性が低いほどにユーモアを受容性が向上していた。従って、提案手法においても、不適合を強く起こす発話を選択することで、よりユーモアとして受容されると仮定できるだろう。

#### 4.1 実験で扱うユーモア生成手法

以下の式 (1), (2) において対話システムが応答で扱う先行要素を選択する手法と、式 (1), (2) で取得できる先行要素を

図 3: 評価実験で用いたページ

除き、ランダムで取得して応答に用いる手法 (*Rand*) の、合計 3 つを実験に用いる。

$$LowSimItem = \arg \min_{m \in M} sim(noun, f(m)) \quad (1)$$

$$HighSimItem = \arg \max_{m \in M} sim(noun, f(m)) \quad (2)$$

$$f(m) = \arg \min_{w \in m} sim(noun, w)$$

$M$  は先行要素-名詞ペアにおける、ユーザの入力から得られる名詞が先行要素-名詞ペアの名詞部分に存在する全ての先行要素の集合を表している。 $m$  は  $M$  における任意の先行要素を表している。 $w$  は  $m$  に対して形態素解析を行った結果から「助詞」、「助動詞」、「動詞」を除いた全ての形態素を表す。また  $sim(noun, w)$  は 3.2 において構築した word2vec のコサイン類似度による出力を表している。*LowSimItem* を応答に用いる手法を *LowSim* とする。一方で、*HighSimItem* を応答に用いる手法を *HighSim* とする。また、文章の最後の名詞を活用することでよりユーモアとして受容されやすくなる知見 [Valitutti 13] を活用し、本研究においてもユーザ発話の最後の名詞を扱う。

本実験では、選択した先行要素と、ユーザの入力から得られた名詞で接続を行い、「ですか?」という語尾を接続し、応答文章とする。

## 4.2 実験手順

被験者によって面白さへの評価が異なることが想定されるため、被験者はユーモア測定尺度 [宇恵 08] におけるユーモア刺激の受け手として関連のある項目である、ユーモアへの気付きを表す質問項目 (SHQ)、ユーモアに対する好悪を表す質問項目 (MSHS) について答える。本研究では、それぞれの尺度における質問項目の選択肢に対して、左から順に 0 点から 3 点、あるいは 0 点から 4 点を配点し、足しあわせた得点をその被験者のその尺度における得点とした。

対話実験は図 3 で示すインターフェイスを用いて web 上で行った。被験者が文字入力によって対話システムへの発話を入力すると、それに対して *LowSim*、*HighSim*、*Rand* のそれぞれの手法による発話が画面に表示される。このやりとりを 1 対話とする。先行研究 [宮澤 12] を参考に、ユーモアを感じた程度を 5 段階で評定させた。被験者はユーモアを感じた程度の評定を行いながら、50 回対話を繰り返す。被験者は日本語を母国語とする 20 名となった。被験者には「名詞を含む発話を行うこと」、「質問を行わないこと」、特に「対話システムの能力について質問しないこと」を教示した。また、順序効果を考

表 1: *LowSim* の得点値が高い例

入力	応答	点	手法	類似度
隣の家でうさぎを飼ってた	NOV うさぎですか?	5	<i>LowSim</i>	0.00
隣の家でうさぎを飼ってた	茶色のうさぎですか?	4	<i>HighSim</i>	0.28
隣の家でうさぎを飼ってた	かわいいうさぎですか?	4	<i>Rand</i>	0.25

表 2: *HighSim* の得点値が高い例

入力	応答	点	手法	類似度
ぬいぐるみを買いました。	貴重なぬいぐるみですか?	4	<i>LowSim</i>	-0.07
ぬいぐるみを買いました。	クマのぬいぐるみですか?	5	<i>HighSim</i>	0.42
ぬいぐるみを買いました。	思い出のぬいぐるみですか?	3	<i>Rand</i>	0.12

表 3: *Rand* の得点値が高い例

入力	応答	点	手法	類似度
メガネをかけます。	中破メガネですか?	2	<i>LowSim</i>	-0.03
メガネをかけます。	メガネ探したらメガネですか?	1	<i>HighSim</i>	1.00
メガネをかけます。	黒ぶちメガネですか?	4	<i>Rand</i>	0.30

慮して、それぞれの対話システムの応答はインターフェイスの上段、中段、下段に毎回ランダムに表示した。対話システムの知識にはランダムに収集した 1700 万ツイートを用いた。先行要素-名詞ペアは約 450 万となった。先行要素フィルターの閾値  $\alpha$  は経験から 4500 と決定した。

## 4.3 結果

20 名の被験者が 3 つの手法に対して 50 回の評定を行ったため、実験から 3000 の対話を評定したデータが得られた。図 4 は実験の結果を表している。wilcoxon rank sum test の Holm 補正による多重比較により、*Rand* と *HighSim*、*Rand* と *LowSim* の間において共に有意差 ( $p < 0.01$ ) が確認された。実験の結果は、単語間類似度が低い発話候補を選択する手法と単語間類似度が高い発話候補を選択する手法はランダムに発話候補を選択する手法と比較してユーモアとして受容されにくい手法であることを示している。

## 4.4 考察

表 1, 2, 3 はそれぞれの応答手法において得点値が高く評価された例を表している。*LowSim* では、ユーザにとって突飛に感じられる応答があったためにユーモアとして受容される機会が減少している。また、*HighSim* では一般的な表現が多いことが確認できる。一方で、被験者の興味のある話題であるため、ユーモアとして受容されやすい応答がいくつか生成された。

### 4.4.1 ユーモアを感じ方の異なり

被験者によってユーモアを感じ方が異なることが予想されるため、追調査を行った。ユーモアへの気付きが十分高いと考慮される被験者に限定して調査を行った。一般的な語が扱われる範囲で調査を行うため、実験の結果から得られたデータに対して、NTT データベースシリーズ日本語の語彙特性 [成昭 99] に登録されている単語に限定して再調査を行った。評価の対象となるデータは 1250 となった。

図 5 は、SHQ が 8 以上かつ、MSHS が 16 未満を示した 4 名の被験者での調査結果を表している。*LowSim* でのユーモアの評定値平均は 3.24 となった。wilcoxon rank sum test の



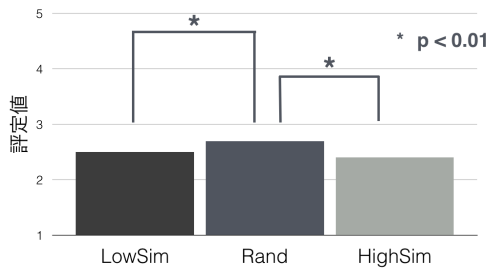


図 4: 各手法によるユーモア平均得点値の比較

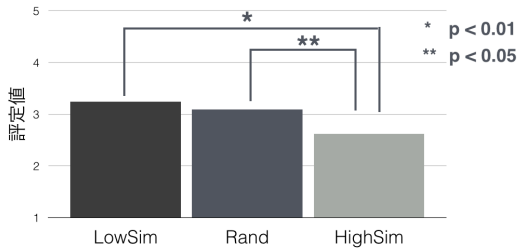


図 5: SHQ が 8 以上, MSHS が 16 未満の被験者における各手法によるユーモア平均得点値の比較

Holm 補正による多重比較により, *LowSim*, と *HighSim*, また *Rand* と *HighSim* の間に有意差が確認された. *LowSim* と *Rand* の間には有意差は確認されなかった. ユーモアへの気付きが一定上の被験者に対しては, 単語間類似度が低い発話候補を選択する手法とランダムに発話候補を選択する手法は単語間類似度が高い発話候補を選択する手法と比較してユーモアとして受容されやすい手法であることを示している.

全ての被験者における MSHS の得点は 10 以上であり, 全ての被験者がユーモアを好ましいものとみなす傾向があった. MSHS に制限をかけない, SHQ が 8 以上の 7 名の被験者においても同様な有意差が確認されたが, *LowSim* におけるユーモアの評定値平均が 2.61 と低下した. 従って, MSHS が尺度の最大値である 16 を示した被験者は, ユーモアへの評定が厳しくなる傾向があることが判明した.

## 5. まとめ

本研究では, ユーザの対話継続欲求を向上させるために対話システムによるユーモア応答生成手法を提案してきた. 提案手法では, 幅広い対話システムに適用可能にするために, 対話システムに対するイメージを扱わない, 言語情報によるユーモア生成を行った. 加えて, 多量の文章データから教師なし学習によって単語のベクトル表現を出力する word2vec を用いて, コサイン類似度として単語間類似度を算出することで, 広い範囲で面白さを説明できるようにすることを目指した.

単語間類似度を導入し, 不適合の強さを表現することで, ユーザにとってよりユーモアとして受容されやすい発話を選択できるかどうか調査を行った. 実験の結果から, 全ての被験者平均においては, 不適合の弱い発話を選択する手法と不適合の強い発話を選択する手法はランダムに発話候補を選択する手法よりもユーモアとして受容されにくい手法であることが示された. 一方で, ユーモアへの気付きが高い被験者において, 不適合の強い発話を選択する手法は不適合の弱い発話を選択する手法と

比較して, 有効であることが示された.

今後の課題として, ユーモアへの気付きが低いユーザがユーモアを受容するために, 対話システムに必要な機能を検証することが挙げられる.

## 参考文献

- [Kao 13] Kao, J., Levy, R., and Goodman, N.: The Funny Thing about Incongruity: A Computational Model of Humor in Puns, in *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, pp. 728–733 (2013)
- [Mikolov 13] Mikolov, T., Sutskever, I., Chen, K., Corrado, G., and Dean, J.: Distributed Representations of Words and Phrases and their Compositionality (2013)
- [Řehůřek 10] Řehůřek, R. and Sojka, P.: Software Framework for Topic Modelling with Large Corpora, in *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pp. 45–50, Valletta, Malta (2010), ELRA
- [Suls 72] Suls, J. M.: A two-stage model for the appreciation of jokes and cartoons: An information-processing analysis, in Goldstein, J. H. and McGhee, P. E. eds., *The psychology of humor: Theoretical perspective and empirical issues*, pp. 81–100, New York: Academic Press (1972)
- [Valitutti 13] Valitutti, A., Toivonen, H., Doucet, A., and Toivanen, J. M.: "Let Everything Turn Well in Your Wife": Generation of Adult Humor Using Lexical Constraints, in *ACL (2)*, pp. 243–248, The Association for Computer Linguistics (2013)
- [稲葉 12] 稲葉 通将, 平井 尚樹, 鳥海 不二夫, 石井 健太郎: 非タスク指向型対話エージェントのための統計的応答手法, 電子情報通信学会論文誌, Vol. J-95-D, No. 6, pp. 1390–1400 (2012)
- [稲葉 14] 稲葉 通将, 神園 彩香, 高橋 健一: Twitter を用いた非タスク指向型対話システムのための発話候補文獲得, 人工知能学会論文誌, Vol. 29, No. 1, pp. 21–31 (2014)
- [字惠 08] 字惠 弘: ユーモア測定尺度の作成, 関西福祉科学大学紀要, Vol. 11, pp. 31–40 (2008)
- [吉村 13] 吉村 枝里子, 芋野 美紗子, 土屋 誠司, 渡部 広一: 知的会話処理における連想応答手法, 人工知能学会論文誌, Vol. 28, No. 2, pp. 100–111 (2013)
- [吉田 14] 吉田 裕介, 萩原 将文: 複数の言語資源を用いたユーモアを含む対話システム, 知能と情報, Vol. 26, No. 2, pp. 627–636 (2014)
- [宮澤 12] 宮澤 幸希, 常世 徹, 榊井 祐介, 松尾 智信, 菊池 英明: 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, Vol. J-95-A, No. 1, pp. 27–36 (2012)
- [成昭 99] 成昭 天野, 公久 近藤: 日本語の語彙特性第 1 期, 三省堂 (1999)