

対話システムにおけるユーモア語句自動生成手法に関する検討 A Study on Generating Humor Phrases for Chat Dialogue Systems

岩倉 亮介[†] 吉川 大弘[†] ジメネス フェリックス[†] 古橋 武[†]
Ryosuke Iwakura Tomohiro Yoshikawa Felix Jimenez Takeshi Furuhashi

1. はじめに

人間とロボット/コンピュータが自然言語を用いて会話を行う、自動対話システムの研究が近年注目されている。自動対話システムとしては、Apple 社の「Siri」や NTT Docomo の「しゃべってコンシェル」などが代表例として挙げられる。これら対話システムは、タスク指向型と非タスク指向型に大別される。

タスク指向型対話システムは、対話を通じてユーザの質問や要求に対して適切な情報を提供するなど、特定のタスクの達成を目的とする。一方、非タスク指向型対話システムは、特定のタスクの達成を目的とせず、自由な対話である雑談によってユーザを楽しませることを目的とする。このような非タスク指向型対話システムにおいては、対話の継続性が重要とされるが、対話が単調だと飽きやすいという問題点がある。

これに対し宮澤らは、ユーザが継続的に対話を続けたいと感じるためには、「ユーモア表現」を含む発話の有効であることを示している[1]。このユーモア表現を含む発話の生成としては、ユーザ入力した文に含まれる名詞に着目し、Twitter から収集したデータに対して word2vec を利用して単語間類似度を考慮する手法が提案されている[2]。また、ユーザとの対話履歴から、ユーザの発話極性がポジティブ/ニュートラル/ネガティブのいずれかを判断し、ユーモアの生成に考慮する手法も提案されている[3]。

しかし、ユーザに面白さを与えるようなユーモア表現の生成は、現状まだ不十分であるといえる。その原因として、面白さの要因についての分析が十分にできていないことが挙げられる。ユーモア表現の生成には面白さを生み出す機構が必要であるが、面白さの要因が分からなければ、機構の作成も困難である。そこで本稿では、ユーモア表現における面白さの要因分析を行う。具体的には、Twitter より取得したデータを利用して生成したユーモアに対して分類を行い、要因の分析に利用する。また、シソーラスを用いて意味的な関係性を考察する。

2. 評価データ

2.1 概要

ユーモア表現としては、物語ユーモアや駄洒落など様々なあるが、今回は単純に 2 種類の名詞を助詞の「の」で接続した句を考える。これは「名詞₁+の+名詞₂」の形で表され、2 つの名詞間の関係性のみが面白さに影響することになる。

2.2 生成方法

1. Twitter より取得したツイートデータに対して Mecab を用いて形態素解析を行い、「名詞₁+の+名詞₂」の組み合わせを全て抽出する。
2. 対象とする名詞を選択し、選択した名詞が名詞₂となっている組み合わせの中から、Twitter 上での出現頻度の最も低い名詞₁を選択する。

2.3 評価方法

2.2 の方法で生成した組み合わせに対して、面白さの評価を行った。データには、Twitter から取得した約 2000 万ツイートを用了。名詞₂として「ぬいぐるみ」を選択し、生成された 900 個の組み合わせに対して、面白い/面白くないかを筆者が主観的に評価した。

3. 結果と分類

3.1 結果

表 1 に、面白いと判断された組み合わせの例を示す。評価した 900 個の内訳は、面白い:75 個、面白くない:825 個であった。

表 1:面白い/面白くない例

面白い	面白くない
ステルスモードのぬいぐるみ	お気に入りのぬいぐるみ
債務超過のぬいぐるみ	おもちゃ屋のぬいぐるみ
愛と勇気のぬいぐるみ	Aザラシのぬいぐるみ

3.2 分類

評価によって得られた面白い組み合わせについて、それぞれに対する名詞₁を分類することで、面白さの要因分析を行った。なお、分類は分析者の主観により行った。分類結果を図 1 に示す。

どの単語でも面白くなりうる	ぬいぐるみでのみ面白い	リズムとして面白い
ステルスモード ポリペプチド系 魔女になる前 メタリック調 世界最高齢38歳 前代未聞 起爆粘土性 葱たっぷり	債務超過 90歳老人 グルメツアー中 ダイエット中 ヤクザ 69代 うまれたて 基本的人権 悲劇的運命 慎重姿勢 毎日仕事 試合終了直後 アマ球界屈指 反逆 宗派別 熱い男達	愛と勇気 そんじょそこら 何かしら 命懸け 墓だらけ アイドルスター男
名詞そのものが面白い		
坊さんだらけ 黄門頼り でぶ丸出し 中身全部アヒル 人類最強兄弟 魔王降臨戦争		

図 1:面白い名詞₁の分類結果

[†]名古屋大学工学研究科 Graduate School of Engineering
Nagoya University

分類によって、面白さの要因として以下の 3 種類があることがわかった。

- (a) どんな名詞₂に対しても面白くなりうる場合
- (b) 名詞₂が「ぬいぐるみ」である時に面白くなる場合
- (c) リズムとして面白い場合

(a)の場合は、どんな名詞₂に対しても面白さにつながりやすいため、汎用性の高い名詞といえる。(b)の場合は、今回選択した「ぬいぐるみ」に対してのみ面白くなりうるため、特定の場面に特化した名詞といえる。(c)の場合は、名詞₂との関係性でなく、音韻により面白さを感じる名詞といえる。また、名詞そのものが面白い場合もあるが、分類としては(a)に含まれると考えられる。

4. シソーラスを用いた分析

3.において、名詞₁に対して分類を行うことで、ユーモアの要因分析を行った。ここでは、名詞₁がもつ特徴を検討するため、分類語彙表[4]を用いた調査を行う。また、4.2における追加データは、名詞₁を分類語彙表に含まれるものからランダムに選択することで生成した。

4.1 分類語彙表

分類語彙表は、単語を意味によって分類/整理したシソーラスであり、約 10 万語の単語に対して区分がされている。例として、分類語彙表での「ぬいぐるみ」に対する分類を表 2 と図 2 に示す。

表 2: 「ぬいぐるみ」に対する分類

類	部門	中項目	分類項目	分類番号	見出し	読み
体	生産物	道具	遊具・置物・像など	1.4570	縫いぐるみ	ぬいぐるみ

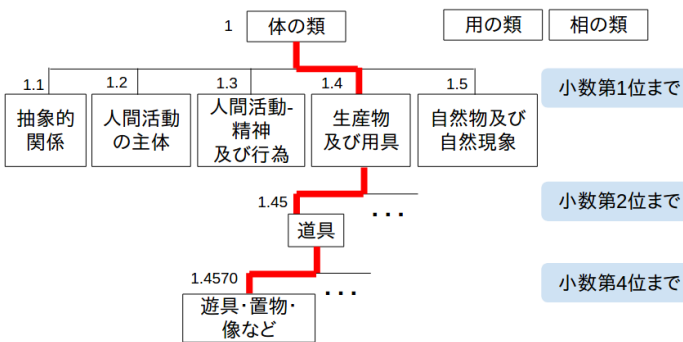


図 2: 全体における「ぬいぐるみ」に対する分類

4.2 分類語彙表を用いた分析

ユーモアの一つにダブルミーニングと呼ばれる表現を用いたものがある。また、分類語彙表には、1 つの語で 2 種類以上の意味が登録されているものがある。例えば昆布であれば、「食料」と「植物」などである。そこで、面白いとされた名詞₁に対して、分類語彙表中にある意味の種類数を分析することで、何か特徴が見られないか確認を行った。結果を図 3 に示す。グラフの縦軸は、その種類数における面白い名詞の数を、面白い/面白くない名詞の数の和で

割った割合となっている。なお、分析を行うデータを増加させるため、新たに 2681 個の組み合わせに対して評価を行い、面白い:122 個、面白くない:2559 個を追加した。

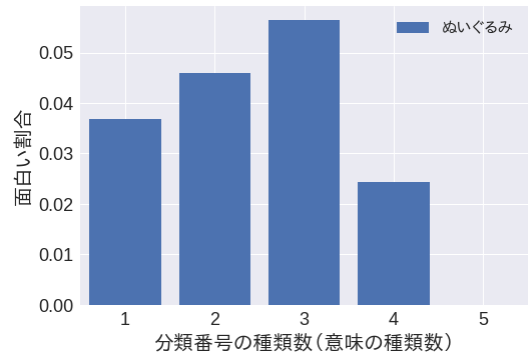


図 3: 意味の種類数と面白い割合

4.3 考察

図 3 より、意味が 1 種類の場合よりも 2 種類、3 種類の方が面白い割合が高くなっていることがわかる。そのため、意味の種類が増えると面白くなりやすい傾向がある可能性がある。意味が 4 種類の場合の割合は最も小さくなっているが、これは、面白い名詞₁の中で意味が 4 種類あるものが 1 個しかなかったため、精度としてばらつきが生じていることが原因だと考えられる。

5. まとめ

Twitter データを用いて作成したユーモアに対して評価を行い、面白いとされた名詞の分類を行った。分類により、2 つの名詞を用いる単純な組み合わせでは、面白いとされる要因には大別して 3 種類あることがわかった。また、分類語彙表を用いた分析を行った結果、意味の多い語は面白くなりやすい傾向がある可能性を示した。

今後の課題としては、「ぬいぐるみ」以外の語に対する組み合わせの評価及び分析と、分類語彙表を用いたユーモア表現の生成に対する検討が挙げられる。

謝辞

本研究は、文部科学省科学研究費(基盤研究(B), No.16H02889)の助成を受けたものです。

This work was supported by MEXT KAKENHI (Grant-in-Aid for Scientific Research (B), No.16H02889).

参考文献

- [1] 宮澤幸希, 他 "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 95.1 (2012): pp.27-36.
- [2] 藤倉将平, 小川義人, 菊池英明. "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." 人工知能学会全国大会論文集 29 (2015): pp.1-4.
- [3] 松井辰哉, 萩原将文. "発話極性を考慮したユーモアを有する非タスク指向型対話システム." 日本感性工学会論文誌 14.1 (2015): pp.9-16.
- [4] 分類語彙表. "国立国語研究所." (2004).