

対話システムにおけるユーモア語句の分析に関する一検討

岩倉 亮介¹ 吉川 大弘¹ ジメネス フェリックス¹ 古橋 武¹

概要：近年，自動対話システムの研究が注目を集めている．非タスク指向型対話システムでは，タスクを特定せず雑談を行うため，長期間に渡って継続的に対話を続けたいとユーザが感じるとシステムが必要とされる．一方，対話継続性の向上には，ユーモア表現を含む発話が有効であることが，従来研究で示されている．しかし，従来のユーモア生成手法では，生成されたユーモアに対して，実際にユーザが面白いと感じる割合が低いという課題がある．そこで本稿では，非タスク指向型対話システムにおける発話生成を目的として，ユーモアが面白いと判断される要因の分析とユーモア語句生成手法に対する検討を行う．

1. はじめに

人間とコンピュータが自然言語を用いて対話を行う，自動対話システムの研究が注目を集めている．自動対話システムとしては，Apple 社の「Siri」^{*1} や NTT Docomo の「しゃべってコンシェル」^{*2} 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する情報を提供するなど，特定のタスク達成を目的とする．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」を含む発話が有効であると示している [1]．また，ユーモア表現を含む発話の生成としては，ユーザが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]．さらに，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，ユーザが面白いと感じ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十分な分析が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工夫が必要であるが，面白さの要因が分からなければ，工夫の考案も困難となる．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，シソーラスを用いて作成した単語の組み合わせが面白いかな否かを評価し，ユーモアの要因分析に利用する．

2. シソーラスを用いた分析

2.1 分析対象

対話で用いるユーモアの形式としては，長文による物語性のあるものと，短文による単語間の関係性を利用したものが考えられる．本稿では，より基礎的な検討を行うため，単語間の関係性を利用したユーモアを対象とする．中でも，単語間の関係性を最も単純化した形として，2 種類の名詞を助詞の「の」で繋いだ，「名詞₁ + の + 名詞₂」という形式のユーモアを扱う．

2.2 分類語彙表

「名詞₁ + の + 名詞₂」の形式において，2 種類の名詞間での特徴を検討するため，分類語彙表 [4] を用いる．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 1 と図 1 に示す．

2.3 ユーモアの生成と評価

評価対象とするユーモアの生成を行う．分類語彙表において名詞が属する，小数第 1 位までの分類番号 1.1 ~ 1.5 の 5 区分から，1 区分につき 3 種類ずつ名詞を選択し，計 15 種類を名詞₂ とする．さらに，各名詞₂ に対して，小数第

¹ 名古屋大学

Nagoya University

^{*1} <http://www.apple.com/jp/ios/siri/>

^{*2} https://www.nttdocomo.co.jp/service/shabette_concier/

表 1 「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

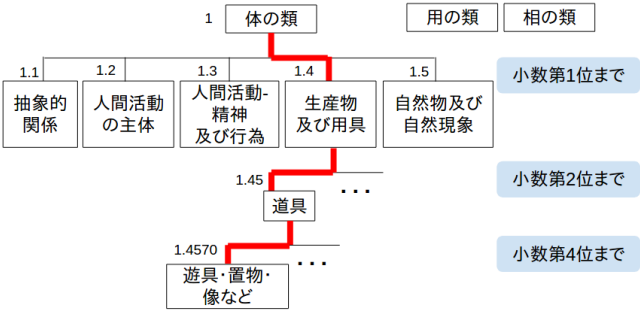


図 1 分類語彙表全体で見た「ぬいぐるみ」に対する分類

表 2 使用した名詞₂と分類番号

レベル	1.1101
登場	1.1210
誕生日	1.1633
自分	1.2020
世界	1.2600
学校	1.2629
情報	1.3123
動画	1.3240
イベント	1.3510
チケット	1.4040
おでん	1.4310
電車	1.4650
台風	1.5150
猫	1.5501
身体	1.5600

2 位までの分類番号 1.10 ~ 1.57 の 43 区分から、1 区分につき 20 種類ずつランダムに名詞₁ を選択し、860 個の組み合わせを作成する。つまり、評価対象は $860 \times 15 = 12,900$ 個となる。なお名詞₂ には、Twitter データにおいて、1.1 ~ 1.5 のそれぞれで出現回数の多い名詞上位 50 からランダムに選択した。使用した名詞₂ とその分類番号を表 2 に示す。これら 12,900 個の組み合わせに対して、面白いが面白いかないかの評価を筆者一名が行った。

2.4 結果

ユーモアの評価結果を図 2 に示す。全体として、名詞₁ が分類番号 1.1, 1.3 に属する場合に面白い割合が高くなっていることがわかる。そのため、若干ではあるが、これらの名詞が使用された際には面白くなりやすい傾向があると考えられる。また分類番号について、名詞₂ が 1.3 で名詞₁ が 1.5 の場合、名詞₂ が 1.4 で名詞₁ が 1.5 の場合、名詞₂ が 1.5 で名詞₁ が 1.3/1.4 となる場合に、それぞれ面白い割合が高くなっていることがわかる。つまり、1.3 と 1.5、1.4 と 1.5、となる名詞の組み合わせの場合に、面白くなり

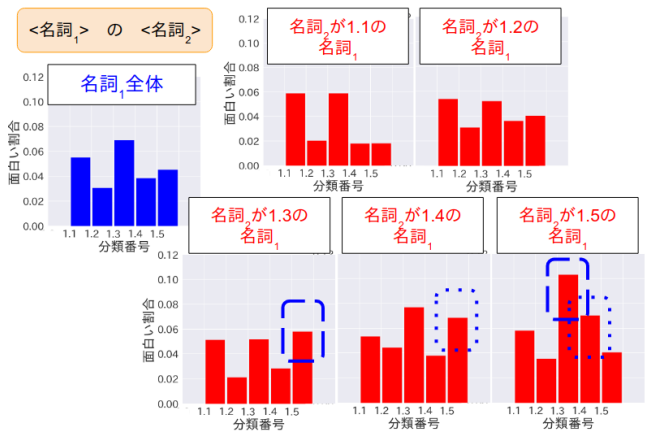


図 2 分類語彙表を用いた評価の結果

表 3 面白いとされた例

名詞 ₁ の区分	名詞 ₂ の区分	面白いとされた例
1.3	1.5	病欠の台風
1.5	1.3	先祖返りのイベント
1.4	1.5	透かし織りの身体
1.5	1.4	強肩のおでん

やすい傾向があると考えられる。なお、1.3 は笑顔や同情などの人間活動/精神/行為、1.4 は模型や手紙などの生産物/用具、1.5 は電波や水蒸気などの自然物/自然現象となる名詞が属している。また、面白いとされた例を表 3 に示す。

3. おわりに

ユーモアの要因分析のため、分類語彙表から作成した様々な名詞の組み合わせに対して、面白さの評価を行った。評価の結果、名詞単体やその組み合わせによって、面白くなりやすい傾向をもつ分類があることを確認した。

今後の課題としては、評価人数の増加、及び各評価者の共通性と独自性に着目した分析、さらにユーモア表現の生成に対する検討が挙げられる。

謝辞 本研究の一部は、科学研究費補助金（基盤研究（B）、No.16H02889）の助成を受けたものです。

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] 国立国語研究所: 分類語彙表 (2004).