対話システムにおける画像からのユーモア発話の自動生成とそれによる対話継続欲求の向上

二又 航介　　藤倉 将平　　菊池 英明
早稲田大学 人間科学部
{f-e@toki., 0spiral1@asagi., kikuchi@}waseda.jp

1 はじめに

近年 Apple の Siri や Softbank の Pepper など，ヒトと対話的なインタラクションを行う対話システムが数多く提案されている．日常生活で用いられる対話システムやエンターテイメント性が要求される対話システムではユーザーを飽きさせない，対話継続欲求の高いデザインを採用する必要がある．宮澤らは人同時のインタラクションを参考に，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対話システムのデザイン方法の確立を目指してきた [7]．分析の結果「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活かした意外性の高いユーモアを利用すること」が有効であると提言した．

以上の知見を参考に，藤倉らはユーモア発話を用いることでユーザーの対話継続欲求の向上を目指した [1]．しかし，ユーザーの言語情報に基づきユーモア発話を生成する試みは，ユーザーが発話をやめた時点で閉ざされてしまう．ユーモア発話は言語情報のみならず，視覚情報などからも生成可能である．視覚情報を用いることで，ユーザーの言語情報に依らないユーモア発話が生成可能であると想定される．さらに視覚情報を用いて対話システムがユーモアを自ずから生成することができれば，ユーザーが発話を行う必要がないため，ユーザーが対話システムの利用をやめてしまってもユーモア生成の可能性が閉ざされない．したがって対話システムによって生成されたユーモア発話に対して，ユーザーが面白さを感じることができれば，対話システムが再び利用される可能性が残されるだろう．

そこで本研究では，画像からユーモアを想起させる発話文を生成することで，ユーザーの対話継続欲求を向上させることを目指す．図 1 に，本研究において提案するシステムがユーモア発話を生成する流れを示す．初めに，システムがユーザーなどの写真を撮影する．次にシステムが撮影した画像から「ジャージを着たサボテン男が立っていますね！」などのユーモア発話を生成し，ユーザーに提示する．システムによって生成されたユーモア発話に対して面白さを感じることができれば，ユーザーの対話継続欲求が向上すると想定される．

2 先行研究

ユーモアに関する認知的メカニズムを説明する理論として「不適合理論 [4]」がある.「不適合理論」には，ユーモアの生成因を不適合そのものに求める「不適合モデル [4]」や，その解決に求める「不適合-解決モデル [5]」がある.「不適合モデル」は，期待の心象と実際の刺激が異なることにより発生する不適合そのものによってユーモアが引き起こされると考えるモデルである．例えば，一発芸やナンセンスなジョークなどが「不適合モデル」に対して当てはまりが良いとされる．一方「不適合-解決モデル」は，不適合に対して論理的な脈絡を発見することにより解決の過程が導かれ，その結果ユーモアが引き起こされると考えるモデルである．例えば，漫才の「ボケ」に対する「ツッコミ」はボケが生じさせた不適合を解決する手がかりを与える

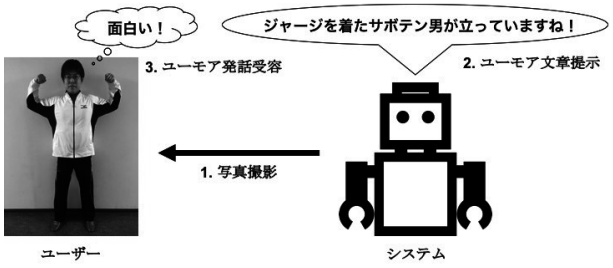

図 1: システムによるユーモア発話生成の流れ

役割を持っていると解釈される.「不適合-解決モデル」は解決という概念の導入により「不適合モデル」に比べ，より多くのユーモア現象を説明できる [8]．また，曖昧さ耐性の低い個人ほど「不適合モデル」より「不適合-解決モデル」を好むという結果が明らかになっている [3].

ユーモア現象は，単語間類似度の観点からも説明できる．Kao らは，ambiguity と distinctiveness の観点から駄洒落の面白さを説明しており，実験の結果から置換される対象の単語と置換後の単語が同じ文脈で発生しやすく，かつ置換される単語と置換後の単語間の類似度が小さいほどユーモアの受容性が向上すると示した [2]．したがって，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る．

3　提案手法

前章の先行研究で述べたように「不適合-解決モデル」は「不適合モデル」と比較すると，より多くのユーモア現象を説明することができ，かつ曖昧さ耐性の低い個人に好まれるため，より多くの人にユーモアとして受容されると想定される．そこで，本研究では「不適合-解決モデル」に基づき，画像からユーモアを想起させる発話文を生成する．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るため [2]，生成されるユーモア文章における単語と画像に写る対象を表す単語の類似度が小さいほどユーモアの受容性が向上すると想定される．しかし，単語間の類似度を考慮するだけでは，解決の過程が導かれないため「不適合-解決モデル」を当てはめたユーモアとして受容されないだろう．そこで，画像間類似度を計算することで解決の過程を導く．画像に写る対象と生成される文章における単語を表す画像の画像間類似度が高ければ，両者が類似しているという観点から解決の過程が導かれると想定される．つまり，生成される単語と画像に写る物体の単語間類似度が低いことにより不適合が生じ，それらの画像間類似度が高いことにより解決の過程が導かれる．

本研究において提案する類似度計算の過程を図 2 に示す．初めに，画像を入力し，システムによって予め用意されたオブジェクトデータと画像間類似度を比較し，入力画像に写る物体と類似する物体を選択する．オブジェクトデータは「イヌ」,「カンガルー」,「カカシ」,「サボテン」など様々な物体の画像と，その画像に対応する名詞のペアから構成される.「不適合-解決モデル」の観点から，画像間類似度が高いほど，解決

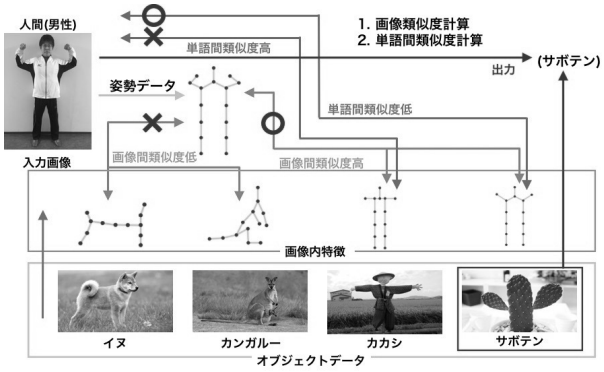

図 2: 類似度計算の過程

の過程が容易になると想定されるため，入力画像とオブジェクトデータを比較した際，画像間類似度が高いと想定される「カカシ」,「サボテン」を選択する．次に，単語間類似度を計算する．単語間類似度が低いほど，不適合が強まると想定されるため，画像間類似度の計算過程で選択された「カカシ」,「サボテン」のうち，入力画像を表す名詞である「人間」との単語間類似度が低いと想定される「サボテン」を最終的に選択する.

画像間類似度の計算は，大規模クラスの画像分類用に学習された Convolutional Neural Network を用いることにより実現される．一方単語間類似度の計算は，word2vec により単語のベクトル表現を取得し，コサイン類似度を用いる.

類似度計算の結果を用いて文章内の単語を他の単語と置換することで，文章がユーモアとして受容されると想定される．提案手法では，Neural Image Caption[6] を用いて画像に写る状況を説明するキャプションを生成する．Neural Image Caption により生成されたキャプション内の主語を，類似度計算の過程から導出した単語と置換したものをユーモア発話文として扱う．Neural Image Caption により生成された文章に対して単語の置換を行うことで，画像と文章の間に不適合が生じると想定される.

4　ユーモア発話評価実験

本研究において提案した手法によりユーモアの受容性が向上するかどうか検証するため，ユーモア発話の評価実験を行った．以下簡略化のため，Neural Image Caption による発話文，低画像間類似度かつ低単語間類似度の発話文，低画像間類似度かつ高単語間類似度の発話文，高画像間類似度かつ低単語間類似度の発話文，高

表 1: 各発話文におけるユーモア平均得点

Original	*LL caption*	*LH caption*	*HL caption*	*HH caption*
1.5	1.7	1.7	1.8	1.7

画像間類似度かつ高単語間類似度の発話文をそれぞれ *Original*，*LL Caption*，*LH Caption*，*HL Caption*，*HH Caption* と記述する．実験は以下の手順で進めた．初めに，ユーモアを含まない *Original* をベースラインとして，*LL Caption*，*LH Caption*，*HL Caption* 及び *HH Caption* との間にユーモアの受容性に変化が見られるかどうか検証した．次に，*LL Caption*，*LH Caption*，*HL Caption* 及び *HH Caption* の間にユーモアの受容性に変化が見られるかどうか検証した．

実験では被験者 20 名に *Original*，*LL Caption*，*LH Caption*，*HL Caption*，*HH Caption* の各発話文を 50 文，計 250 文の面白さの程度を 5 段階で評定させた．それぞれの発話文におけるユーモアの平均得点及びユーモアの平均得点分布をそれぞれ表 1，図 3 に示す．表 1 及び図 3 から *LL Caption*，*LH Caption*，*HL Caption*，*HH Caption* は *Original* よりユーモアの平均得点が高いことが読み取れる．同様に，*HL Caption* はその他発話文よりユーモアの平均得点が高いことが読み取れる．

初めに，ユーモアを含まない *Original* をベースラインとして，*LL Caption*，*LH Caption*，*HL Caption*，*HH Caption* との間にユーモアの受容性に差が見られるかどうか検証した．各被験者の各発話文に対するユーモア得点の頻度分布は正規分布に従っていなかったため，ノンパラメトリック検定を行った．Wilcoxon signed rank test を行った結果，*Original* とその他全ての発話文との間で有意な差が見られた ($p < .001$)．したがって，提案手法含む画像間類似度及び単語間間類似度を調整した発話文ではユーモアの受容性が向上することが明らかになった．

次に，*LL Caption*，*LH Caption*，*HL Caption* 及び *HH Caption* の間にユーモアの受容性に変化が見られるかどうか検証した．分析には各被験者の各発話文に対する平均値を用いた．対応ありの二要因分散分析を行った結果，画像間類似度において，ユーモアの受容性に有意な差が見られた ($p < 0.05$)．単語間類似度及び交互作用効果には有意な差が見られなかった．

以上の結果から，画像間類似度の高いユーモア発話文を選択する手法はユーモアの受容性を向上させることが明らかになった．

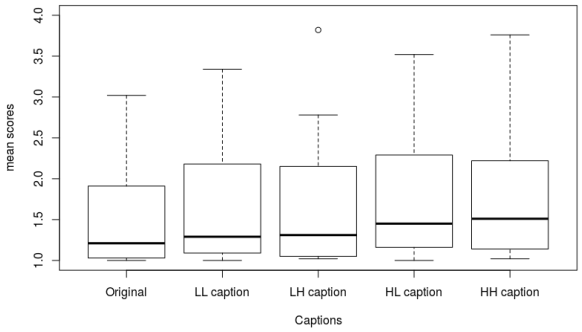

図 3: 各発話文におけるユーモア平均得点分布

表 2: 各発話文における対話継続欲求の度合いの平均得点

Original	*HL Caption*	*HH Caption*
2.3	2.4	2.4

5 システム評価実験

本研究において提案した手法によりユーザーの対話継続欲求が向上するかどうか検証するため，システムの評価実験を行った．前章のユーモア発話評価実験において *HL Caption* 及び *HH Caption* によりユーモアの受容性が向上することが明らかになったため，これら 2 つの発話文及びベースラインとなる *Original* を用いて対話継続欲求の度合いについて検証を行った．実験では，被験者 20 名に *Original*，*HL Caption*，*HH Caption* の各発話文を 50 文，計 150 文を提示することにより対話継続欲求の度合いを調査した．各発話文の評定では，該当する発話文を生成するシステムを継続して利用してみたいかどうか 5 段階で評価させた．それぞれの発話文における対話継続欲求の度合いの平均得点及び対話継続欲求の度合いの平均得点分布をそれぞれ表 2，図 4 に示す．表 2 及び図 4 から *HL Caption* 及び *HH Caption* は *Original* より対話継続欲求の度合いの平均得点が高いことが読み取れる．各被験者の各発話文に対する対話継続欲求の度合いの頻度分布は正規分布に従っていなかったため，ノンパラメトリック検定を行った．Wilcoxon signed rank test を行った結果，*Original* と *HL Caption*，*Original* と *HH Caption* の間に対話継続欲求の度合いについて有意な差が見られた ($p < .001$)．したがって，画像間類似度の高いユーモア発話文を選択する手法は対話継続欲求を向上させることが明らかになった．

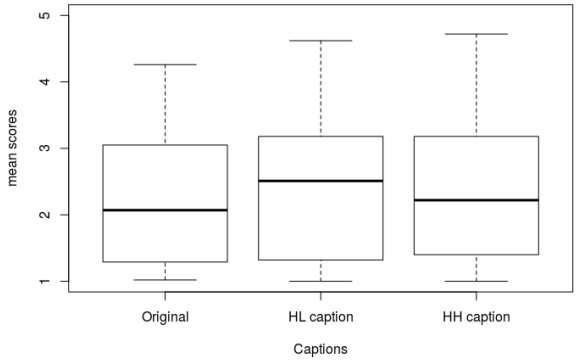

図 4: 各発話文における対話継続欲求の度合いの平均得点分布

6 まとめ

本研究では，ユーザーの対話継続欲求を向上させるために，対話システムにおいて画像からユーモアを想起させる発話文を自動生成する手法を提案してきた．提案手法では，画像間類似度及び単語間類似度を導入することにより，ユーザーの発話に依らない「不適合-解決モデル」に基づいたユーモア発話文を自動生成した．

画像間類似度及び単語間類似度を導入することにより，ユーザー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，画像間類似度が高いユーモア発話文を選択する手法は，画像間類似度が低いユーモア発話文を選択する手法よりもユーモアとして受容されやすい手法であることが示された．さらに，画像間類似度が高い発話文を選択する手法により，対話継続欲求が向上するかどうか調査を行った．実験の結果から，画像間類似度が高いユーモア発話文を選択する手法は，ユーモアを含まない発話文を選択する手法よりもユーザーの対話継続欲求を向上させる手法であることが示された．したがって，画像間類似度が高いユーモア発話文を生成することで，ユーザーの対話継続欲求が向上することが明らかになった．

提案手法含む画像間類似度の高い発話文をユーザーに提示することにより，ユーモアの受容性及び対話継続欲求が向上することが明らかになった．しかし，単語間類似度の低い発話文を用いることにより，ユーザーのユーモア受容性及び対話継続欲求を高めることはできなかった．今後の課題として，単語間類似度の高低差によりユーモアの受容性及び対話継続欲求の向上性を説明できる手法を模索する．単語間類似度の高低差によりユーモアの受容性及び対話継続欲求の向上性を説明することができれば，よりユーザーの対話継続欲求を向上させうる発話候補を選択できることが期待される．

参考文献

[1] Shohei Fujikura, Yoshito Ogawa, and Hideaki Kikuchi. Humor utterance generation for non-task-oriented dialogue systems. In *Proceedings of the 3rd International Conference on Human-Agent Interaction*, HAI '15, pp. 171–173. ACM, 2015.

[2] Justine T. Kao, Roger Levy, and Noah D. Goodman. A computational model of linguistic humor in puns. *Cognitive Science*, 2015.

[3] W. Ruch and F Hehl. Intolerance of ambiguity as a factor in the apprecia- tion of humour. *Personality and Individual Differences*, Vol. 4, No. 5, pp. 443–449, 1983.

[4] Jerry Suls. Cognitive processes in humor appreciation. In Jeffrey H. Goldtein and Paul E. Mcghee, editors, *Handbook of Humor Research*, pp. 39–57. Springer-Verlag, 1983.

[5] Jerry M. Suls. A two-stage model for the appreciation of jokes and cartoons: An information-processing analysis. In *The psychology of humor*, pp. 81–100. Academic Press Inc, 1972.

[6] Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. Show and tell: A neural image caption generator. *CoRR*, Vol. abs/1411.4555, , 2014.

[7] 宮澤幸希, 常世徹, 桝井祐介, 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.

[8] 佐藤大幸. ユーモア経験に至る認知的・情報的仮定に関する検討:不適合理論における 2 つのモデルの統合に向けて. *Cognitive Studies*, Vol. 14, No. 1, pp. 118–132, 2007.