

話題遷移に適応した駄洒落ユーモア統合型対話システムの性能評価

Evaluation of a Topic-Sensitive Dialog System Integrated with Pun Humor

谷津 元樹¹, 荒木 健治¹

Motoki Yatsu¹, Kenji Araki¹

¹北海道大学大学院 情報科学研究科

¹Graduate School of Information Science and Technology, Hokkaido University

1 はじめに

聴覚と言語能力及び発声器官を有するヒトにとって、音声言語を用いた情報交換は最も容易な意思疎通の手段の一つである。このため、機器の操作に音声言語を取り入れようとする音声対話技術の研究開発が長く続けられている。現状でも、スマートフォンやPC等の様々な情報機器への応用が実現し、その代表的な例としてはSiri®, Cortana®, シャベってコンシェル ④などがある。

ユーモアへの応答を示す能力は、対話システムに実装することにより、応答の多様性を生むばかりでなく、話し手（ユーザ）のユーモア表出によるストレスの緩和などの効果が期待できる。実際に、上述のシステムは駄洒落などのユーモアに対する応答を行えることも知られている。例えば、「布団が吹っ飛んだ」を入力発話とした場合のSiriの応答は、図1のようになる。

ユーモアが笑いを引き起こす原因の研究については、不適合あるいは不一致 (incongruity) の存在が聴者に面白さを感じさせるという理論が古くから提唱され [1]、現在でも根強い支持を集めている [2][3]。

しかし、対話の関連性を重視する立場からは、話題の変遷に応じて駄洒落の認識・生成を行う頻度を調節することにより、自然かつ状況に対して適切な対話を行うことができるという仮説が考えられる。反対に、駄洒落の認識・生成は、文脈とは関係なく無秩序に行うほうが印象に良い影響を与えるということが、上述の不適合理論からは想定される。

これらの仮説を検証するため我々は、この駄洒落ユー

モアの認識・生成と発話動作に、システムの話題への適応がどのような影響を与えるのかについての調査を行っている。本稿では現段階で得られている結果及び考察を述べる。

初めに2章で関連する研究を挙げる。3章では、実験のために実装した質問応答が可能な雑談対話システム、及び雑談対話システムに付与した駄洒落ユーモアの認識・生成能力及び話題への適応能力について述べる。4章では提案手法に対して行った実験について、5章では実験の結果に対する考察、最後にまとめと今後の課題について述べる。

2 関連研究

駄洒落ユーモアについては、不適合理論に基づいた対話システムに関する藤倉らによる研究が存在する [4]。ユーモアを生成する対話システムは上述のものほかに、吉田ら [5]、ディバワラ [6] により実装されその有効性が実証されている。

しかし、ユーモアの認識手法を対話システムに実装した例は我々の知るところでは確認されていない。ユーモアの生成及び認識を同一のシステム上で実現することによって、人工知能としてのユーモア感覚を実現し、ユーザによる印象に良い影響を与えることができると考えられる。

また、ユーザに向けてユーモアを表出ないしユーザの発したユーモアを認識するタイミングについて明確化するデータを与えている研究も少ない。ディバワラの研究では感情語からユーザの感情を推測し、ユーモア生成に適切な感情の状態となった際に駄洒落を生成するアプローチがとられているが、有効性を実証したデータは未だ提供されていない。感情の分析に用いることの可能な感情語が、常にユーザ発話より得られるとは限らないのがその原因と考えられる。

一方、我々はすでにユーザの発話から特定のトピックへの関連性を計測する手法を用いて対話に内部情報を与える手法を提案している [7]。この手法は、ユーザによる



図 1: Siri との対話例。

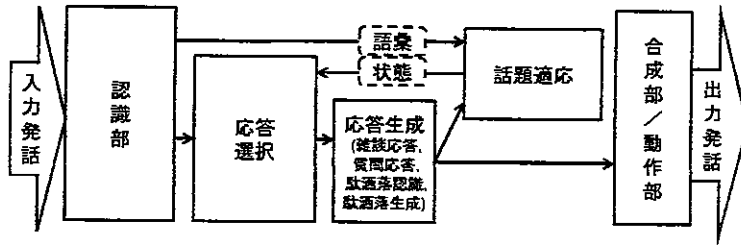


図 2: システムの構成.

表 1: システムの状態及び 2 種の駄洒落動作の発生確率.

状態	駄洒落認識	駄洒落生成
s_1 : 「食事」への関連度高	0.05	0.2
s_2 : 「食事」への関連度低	0.2	0.05

入力発話における任意の内容語(名詞, 動詞, 形容詞)を特定トピックとの関連性計測に用いることができるという利点がある。そのため, 本稿における提案手法においては, この手法を用いて対話の 2 つの状態を設定し駄洒落認識及び生成の頻度を調整し, 実験を行う。

3 提案手法及び実験システム

1 章において述べた目的のため, 雑談対話システムに話題適応及び駄洒落の認識・生成機能を統合し, 小型ロボット RAPIRO¹への実装を行った。音声認識には Google Speech API²を, 音声合成には Open JTalk³を用いた。システム全体の構成図を図 2 に示す。

3.1 対話システムの機能

基本システムとして, 株式会社 NTT ドコモの提供する雑談対話 API[8]を用いて雑談対話システムを構成した。ここで, 疑問詞(誰, どこ, どう, いつ, なぜ等)を含む発話は, 質問応答発話として認識され, 質問応答システム [9]による応答が行われる。また, 3.2 に述べる手法で算出された話題への関連性の状態に基づいて, 3.3 に述べる手法により駄洒落の認識及び生成を行う。

3.2 話題適応

本実験システムは, 以下に述べる方法で入力発話と「食事」トピックとの関連性を計測する。まず, 入力発話を形態素解析し内容語(名詞, 動詞及び形容詞)を抽出する。抽出した内容語とトピックを表す単語「食事」との, 日本語

WordNet[10]における意味類似度を Leaco 法 [11]により計測する。このようにして累積値を超えているか否かを, 発話の入力毎に累積値を超えている状態においては, 式 (1) で表

$$f(t) = e^{-at} \cos bt$$

式 (1) における t は, 累積値が閾値を超えたターン数であり, $t \geq 0$ である。パラメタ予備実験を行い最適な値を実験的に求める。洒落の認識及び生成において, 累積値が閾値を超える状態とそうでない状態とで認識及び生成を変化させる。この変化の手法については, 3.3 に述べる。関数を用いてターンの経過毎に累積させる。

入力発話との関連性の計測対象を「食事」定めた理由は以下に述べる通りである。まず駄洒落ユーモアの生成手法が食べ物の語彙に関するものであることが主な理由である。そして「好き」というキャラクター付けは対話親近感を持ち接するために妥当であると考えられるトピックを用いる場合と比較して 4.1 に示す雑談対話への導入も容易になると考えられる。

3.3 駄洒落の認識及び生成

駄洒落ユーモアの認識は, 発話の入力が駄洒落検出モジュールを用い駄洒落を検出することによって行われる。駄洒落の認識モジュールは, 子音間の音韻類似度 of-words を素性とする教師あり機械学習に SVM を用いる。駄洒落の検出が行われた場合, 「それ」という命題文にランダムにモダリティ表現を付与し答発話を入力する。

駄洒落ユーモアの生成は, (1) 入力文の改変(同義語置換), (2) なぞなぞ, (3) 駄洒落データベースからのタイプ別の駄洒落を生成するシステムを用いる。それぞれのタイプの生成手法は下記の通り。

¹<http://www.rapiro.com/ja/>

²<https://cloud.google.com/speech/>

³<http://open-jtalk.sourceforge.net/>

入力文の改変 入力発話中の内容語について、音韻の類似した食べ物・飲み物の名称をシソーラスより抽出し、入力発話中の内容語を食べ物に関連した語に置換していく。複数の内容語に対して置換に成功した場合、テンプレートを用いて出力する。例は以下の通りである。

例

入力：人間の言葉で話ができるシステムの研究

「最近、私は人間の言葉で話ができるシステムの研究っていう感じでしたよ。」

「俺の場合、天井の言葉でバナナができるシステムの研究って感じたけど。」

「何言ってるんだよ！」

なぞなぞ 入力発話中の内容語(語1)について、音韻の類似した食べ物・飲み物・菓子の名称(語2)をシソーラスを用いて抽出する。抽出に成功した場合、「(語1)は(語2)でも、<修飾子>(語1)はなーんだ?」「何?」「(語2)」「<修飾子>には『食べれる』『ゴクゴク飲める』及び『甘〜い』のいずれかが挿入される」というテンプレートを用いて出力する。例を以下に示す。

例

入力：職人の技

「職人は職人だけど、食べれる職人はなーんだ?」

「何?」「冷凍職人」

駄洒落データベース 入力発話中の内容語と一致する語を持つ文を駄洒落データベースより検索し、ヒットした文を応答発話とする。例を以下に示す。

例

入力：印象の評価

印象と言えば、印象がいいんしょ。って言いたくなるよね。

生成モジュールが駄洒落の出力に失敗した場合は、「聞こえない、聞こえない…」という発話を応答とする。

駄洒落の認識及び生成は、一定の確率により確率的に行われる。その確率は、3.2で述べた話題適応動作により、「食事」トピックへの対話の関連性の高さに応じたシステムの内部状態 $\{s_1, s_2\}$ により変化する。表1に状態に応じた各動作の発生確率を示す。

4 印象評価実験

本章では、1章において述べた仮説を検証するために行った印象評価実験の実施の方法及び結果について述べる。

表 2: 実験条件の一覧。

条件	話題適応	駄洒落認識	駄洒落生成
c ₀	-	-	-
c ₁	-	-	○
c ₂	-	○	-
c ₃	-	○	○
c ₄	○	-	-
c ₅	○	-	○
c ₆	○	○	-
c ₇	○	○	○

4.1 実験の方法

実験は、3名の20歳代の理系男子大学院生を被験者として行った。

被験者は、音声によってロボットと1条件につき10分間程度の対話を行った。表2に示す条件c₀~c₇のうち、条件c₄は、話題適応が通常の雑談発話に対し影響を与えることがなく、条件c₀と同一の出力となるため、c₄を除く7項目について実験を行った。対話の開始時には、ロボットが自己紹介を行い、「食べるのが大好きである」という内容の発話を出力した。これは、被験者とロボットの対話の話題が食事に関するものに近づくことを期待するため、対話の導入をロボットが行うものとしたことによる。

各条件での対話の終了後、被験者は以下に示す設問にアンケートフォームを用いて回答した。回答手法には、1点を「全く同意しない」、7点を「強く同意する」とする7段階のリッカート尺度[14]を用いた。

- Q1. ロボットからの応答は、自然なものだった。(応答の自然性)
- Q2. ロボットからの応答は、日本語として通じるものだった。(応答の文法性)
- Q3. ロボットは、色々な知識を持っているように感じた。(応答の知識量)
- Q4. ロボットが「それは駄洒落」などと言った時、自然な応答だと感じた。(駄洒落認識の適切さ)
- Q5. ロボットが駄洒落を言った時、自然な応答だと感じた。(駄洒落生成の適切さ)

ただし、Q4.Q5は、駄洒落認識または駄洒落生成応答がシステムによりなされなかった場合には回答不要とした。

表 3: Welch の t 検定の結果.

設問	話題適応	駄洒落認識	駄洒落生成
Q1(自然性)	B	B	C
Q2(文法性)	C	C	C
Q3(知識量)	C	C	C
Q4(駄洒落認識)	C	-	C
Q5(駄洒落生成)	C	C	-

各列の条件の 2 グループの平均得点を \bar{x}_{ON} (機能あり), \bar{x}_{OFF} (機能なし) としたとき, 表中の記号は次を表す: A. 有意差 ($\bar{x}_{ON} > \bar{x}_{OFF}$), B. 有意差 ($\bar{x}_{ON} < \bar{x}_{OFF}$), C. 有意差なし, - 比較不可. なお, 有意水準は $p < 0.10$ とする.

4.2 実験の条件及び結果

実験を行った 7 つの条件は, 話題適応, 駄洒落認識, 駄洒落生成のそれぞれの有無により 2 グループに分けられる. 2 グループの実験結果について, 各設問の値の平均値を求め, Welch の t 検定を行った. この結果を表 3 に示す.

また, 設問 Q1, Q2, Q3 の得点平均値の大きい順に並び替えた条件を, 各設問の平均得点とともに表 4 に示す.

4.3 対話例

本実験において行われた対話の一部を表 5 及び表 6 に示す. 図内にはターン数及び実験システム内で動作したモジュール名を記載した.

5 考察

4 章の実験で得られた結果の分析と, 今後の改善に必要と考えられる事項について述べる.

表 3 より, 話題適応及び駄洒落認識を用いた場合には, 応答の自然性は向上せず, むしろ低下することが示唆される. また, 他の全ての項目(文法性・知識量・駄洒落認識の適切さ・駄洒落生成の適切さ)に関しては, 想定した各機能の動作が評価値に有意な影響を与えていることは統計的には確認されなかった.

このことから, 応答の自然性に着目すると, システムが話題適応に応じた形で駄洒落認識・生成を行わないほうがユーザにより良い印象を与えることが示唆され, かつ, 表 3 の「駄洒落認識」に対し Q1 が B を示していることから, 駄洒落認識発話そのものにも応答の自然性の面で課題があることが示された. 1 章において述べた仮説について考えると, 不適合理論から推測される, 駄洒落の表出においてはランダム性がより良い印象を与えるという仮説を支持する結果が示されたと考えられる.

表 4 において, 話題適応及び駄洒落の両機能を一切含まない条件 c_0 が設問 Q1~Q3 の平均値で最上位となつて

表 4: 実験で用いた 7 条件 (設問 Q1, Q2, Q3 の得点平均値順) 及び各設問の平均得点.

条件	Q1(自然性)	Q2(文法性)	Q3(知識量)	平均値
c_0	3.33	4.33	4.00	3.89
c_7	3.00	4.00	4.00	3.67
c_1	3.33	3.33	3.67	3.44
c_6	2.33	4.33	3.33	3.33
c_2	2.33	3.67	3.33	3.11
c_3	2.67	3.00	3.00	2.89
c_5	2.00	3.33	2.67	2.67

いる. 仮に各駄洒落モジュールの機能に潜在的有効性があったとしても, システムとしての統合性から付加の一切無い条件が最も高い印象を与えており, ベースとなった雑談システムと各モジュールの連携に課題が残っていたことが推測される.

表 6 における駄洒落認識 (下線部) は, 音声認識の誤りに対し作動しているため, 不自然なものとなっている.

6 まとめ

本稿では, 駄洒落への反応及び生成を行う頻度を対話の話題の変遷に従って自動的に調節する, 質問応答の可能な非タスク指向型対話システムを提案し, その有効性の確認を行った.

実験の結果より, 話題適応による駄洒落の認識及び生成の頻度の調節が必ずしもユーザへの印象の向上には繋がらないことが判明した. このことを踏まえると, ランダムネスに基づいた駄洒落の認識及び生成が他の実行頻度の調節手法に対しても優位であるか否かを検証する必要があると考えられる.

さらに, 実験システムの構築においてベース雑談システム及び駄洒落認識・生成モジュールの統合をもう一度見直す必要があると考えられる. 実験における印象の低下の一要因と考えられるのは, 複数の情報源を利用した駄洒落生成応答に要する時間の長さである. 同一入力発話に対する応答所要時間は, 駄洒落生成は 5 割平均で 12.92 秒を要し, 雑談応答 (2.86 秒) の 4.51 倍であった. キャッシュの利用などにより, 極力通常の雑談対話と同程度のレスポンスの早さを実現する必要がある.

以上のような課題は残るが, 我々は今後も駄洒落の認識を含め, ユーモアを理解する対話能力の研究を続けていく予定である.

参考文献

- [1] Lisa Glebatis Perks. The ancient roots of humor theory. *Humor*, Vol. 25, No. 2, pp. 119-132, 2012.

表 5: システムとの対話例 (駄洒落生成のある例)。

T	話者 (モジュール名)	発話
1	システム (対話導入) ユーザ	やあこんにちは！ぼくはラビロ。食べることが大好きなロボットです。一緒にお話ししましょう！ あなたは人間ですか
2	システム (雑談応答) ユーザ	秘密です。 親は誰ですか
3	システム (質問応答) ユーザ	やはり答えは「智か可愛いく」でございます。 元気がないですね
4	システム (雑談応答) ユーザ	午後から天気が崩れるのです。 雨は好きですか
5	システム (雑談応答) ユーザ	少し離れていますね。 雨は好きですか
6	システム (駄洒落生成)	雨やあめと言え、この雨、あめえ〜って言いたくなるよね。

表 6: システムとの対話例 (駄洒落認識のある例)。

T	話者 (モジュール名)	発話
1	ユーザ	研究は楽しいですか
2	システム (雑談応答) ユーザ	はい、飛行機に乗った事ありますよ。 天気は辛いですか†
3	システム (雑談応答) ユーザ	はい、飛行機に乗った事ありますよ。 黒鉄扉†
4	システム (雑談応答) ユーザ	飛んでるのがいいですね。 電車は好きですか
5	システム (雑談応答) ユーザ	たまに電車にのりますよ。 ゴースト電車はどちらがいいですか†
6	システム (駄洒落認識) ユーザ	あら、今は駄洒落。 飛行機と電車どっちが好き
7	システム (雑談応答) ユーザ	飛行機は危ないですね。 飛行機に乗りますか
	システム (雑談応答)	はい、飛行機に乗った事ありますよ。

* T はターン数を表す。

† † のあるユーザ発話は音声認識誤りである。

- 2] Kevin N Francisco, Johann-Paulo L Ruste, Ria A Sagum, Yun Seop Song, and Darl Sander G Sy. Incongruity Theory Applied in Dynamic Adaptive Game Artificial Intelligence. *International Journal of Future Computer and Communication*, Vol. 2, No. 5, pp. 499–504, 2013.
- 3] Arthur Asa Berger. Why We Laugh and What Makes Us Laugh: The Enigma of Humor. *Europe's Journal of Psychology*, Vol. 9, No. 2, pp. 210–213, 2013.
- 4] 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, Vol. 29, pp. 1–4, 2015.
- 5] 吉田裕介, 萩原特文. 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.
- 6] Pawel Dybala, Michal Ptaszynski, Rafal Rzepka, and Kenji Araki. Humorized Computational Intelligence towards User-Adapted Systems with a Sense of Humor. In *Workshops on Applications of Evolutionary Computation*, pp. 452–461. Springer, 2009.
- 7] Motoki Yatsu, Rafal Rzepka, and Kenji Araki. A Domain Analytic Method in Modular-Designed Reflexive Agent. In *Linguistic and cognitive approaches to dialogue agents, AISB/IACAP*, pp. 25–30, 2012.
- [8] Takeshi Yoshimura. Casual Conversation Technology Achieving Natural Dialog with Computers. *NTT DOCOMO Technical Journal*, pp. 16–21, 2014.
- [9] 龜山憲祐, 荒木健治, 木村泰知. 質問応答・システムにおける知識選択規則の自動探知の有効性について. 情報処理学会研究報告自然言語処理 (NL), Vol. 2007, No. 35, pp. 85–90, 2007.
- [10] Francis Bond, Timothy Baldwin, Richard Fothergill, and Kiyotaka Uchimoto. Japanese SemCor: A sense-tagged corpus of Japanese. In *Proceedings of the 6th Global WordNet Conference (GWC 2012)*, pp. 56–63, 2012.
- [11] Claudia Leacock and Martin Chodorow. Combining Local Context and WordNet Similarity for Word Sense Identification. *An Electronic Lexical Database*, pp. 265–283, 1998.
- [12] 谷津元樹, 荒木健治. 子音の音韻類似性及び SVM を用いた駄洒落検出手法. 知能と情報: 日本知能情報 ファジィ学会誌, Vol. 28, No. 5, (in print), 2016.
- [13] Jonas Sjobergh and Kenji Araki. Evaluation of a Humor Generation System by Real World Application with ¥500,000 to Win. In *Proceedings of the Linguistic And Cognitive Approaches To Dialog Agents Symposium*, 2010.
- [14] Rensis Likert. A technique for the measurement of attitudes. *Archives of psychology*, 1932.