

業界用語と駄洒落を使用したユーモア応答文の生成*

香川 茂樹, 唐澤 博†

山梨大学 工学部‡

E-mail:{kagawa, karasawa}@jewel.yamanashi.ac.jp

1 はじめに

近年, エンタテインメントロボットの開発が盛んに行われているが, ユーモアの計算機処理は人工知能分野における基礎研究としてだけでなく, ヒューマンフレンドリーなインタフェースを開発する応用研究としても重要である [1].

人工知能の基礎研究において, 駄洒落やことば遊びの計算機処理を指向する研究がなされている [1][2] が, それらによって生成されるユーモアは表現の制約が多く, 対話システムへの応用には至っていない. また対話システムとして笑いやユーモアに注目した研究 [3], 製品も存在するが, それらはそれぞれ独立したシステムになっており, 他システムへの応用は困難だと考えられる. これらの理由から筆者らは, ユーモアを生成する機能を独立して開発し, それを既存の自然言語対話理解システムと組み合わせることでユーモアを含んだ応答文を生成する研究を進め, システムが業界用語を使用する方法 [4] を提案してきた. 本稿ではその方法に基づいて開発したシステムと新たに設計している駄洒落を含んだ応答文の生成方法について述べる.

本稿は, 2章で笑いの理論についてふれ, 3章で現在開発しているユーモア生成システムについて説明する. 4章でシステムの評価を行い, 5章でまとめを述べる.

2 笑いの理論

現在までに提唱された笑いの心的メカニズムには, 優越の理論, ズレの理論, 放出の理論の三つの有力な理論があるが, これらはそれぞれ相違しており, それらを統一する原理は提唱されていない. そこで本研究では理論にはこだわらず種々雑多な笑いを構造別に分類し, パターン化したもの [5] の中から, 比較的既存の自然言語対話理解システムに適用しやすいと思われる業界用語と駄洒落 [4] の2つのユーモアを生成対象とする.

3 ユーモア生成システム

ユーモア生成システムは自然言語対話理解システムとは切り離してサーバ化し, ユーモアの計算機処理部分を独立に開発しながら, 初期の段階から自然言語対話理解システムと組み合わせることで, 基礎研究を実用システムに応用していく際の隔たりを小さくし, 同時に自然言語対話における効率的なユーモアの生成法を模索していく. また, サーバとして常駐させておくことで要求があった場合にのみユーモアを生成することができ, フォーマルとインフォーマルな応答を使い分けることが可能になる.

*Generation of humorous response containing pun and special word by natural language dialogue system

†Shigeki Kagawa, Hiroshi Karasawa

‡Yamanashi University, 4-3-11 Takeda, Kofu, Yamanashi 400-8511, Japan

3.1 業界用語の使用

自然言語対話理解システムが業界用語を使用できるようユーモア生成システムを開発した. システムの構成を図1に示す.

ユーモア生成システムへの入力自然言語対話理解システムへのユーザーの入力文, システムの応答文の2つで, これらは話題同定部の入力ファイルに記録され, セッション中の過去の文脈から現在の話題を計算できる状態を常に保つ.

話題同定は自然言語対話理解システム側の応答文が入力されると行われる. 話題同定には, 角田らが場面によって変わる多義語の曖昧性解消を目的として提案している連想システム WAVE (Winner Associative Voting Engine) [6] を場面同定を行うように一部改良, 実装し開発用にツール化した CWAVE [7] を用いた. CWAVE は文セットの入力に対し, 一文ごとの話題の推移と話題の強さを示す活性値を出力する.

話題を同定した後は業界用語データベース (若者用語: 140語, 警察用語: 136語, 音楽用語: 56語) の中から同定した話題の業界用語を参照し, 自然言語対話理解システムの応答文中に業界用語と置換できる一般語が含まれていれば, それらを置換して自然言語対話理解システムに出力する. 最後に, 業界用語に置換された応答文は指定された自然言語対話理解システムのチャンネルを通して出力される.

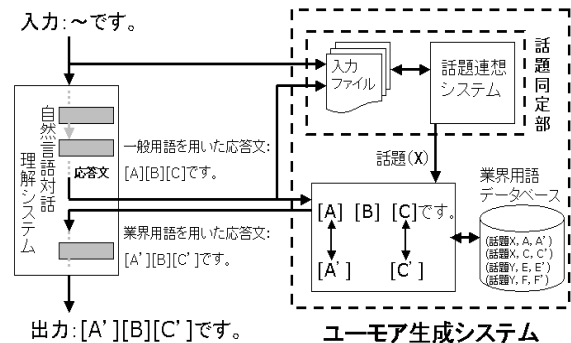


図1: ユーモア生成システム (業界用語)

3.2 実行結果

ユーモア生成システムを動作させて自然言語対話理解システムと対話を行ったときの実行結果を表1に示す.

システムは4文目と8文目で「真犯人」を「本ボシ」(警察用語), 「就職活動」を「就活」(若者用語)と, 同定された話題の業界用語を使用して応答している.

下線は話題を活性させる単語を示しているが, 4文目と8文目の「本ボシ」と「就活」については元々の「真犯人」と「就職活動」という単語が話題を活性させている.

表 1: 業界用語を含むシステムの発話

	対話文	同定された話題	活性値
Def	S>やあ、こんにちは、私はガイアです。	なし	0
1	Y>唐澤は真犯人です。	警察	0.714
2	S>分かりました。	警察	0.357
3	Y>唐澤は何ですか？	警察	0.179
4	S>本ボシです。	警察	0.804
5	Y>私は就職活動をしています。	若者	0.676
6	S>分かりました。	若者	0.338
7	Y>私は何をしていますか？	若者	0.169
8	S>就活です。	若者	0.760

(S:システム発話, Y:ユーザー発話)

3.3 駄洒落の生成

自然言語対話理解システムが駄洒落を含んだ応答文を出力する場合のユーモア生成システムの処理を図 2 に示す。

ユーモア生成システムへ自然言語対話理解システムの応答文が入力されると、応答文は形態素解析され、形態素列の中から駄洒落に変換する形態素（駄洒落キー）が決定される。本システムでは形態素列中の一般名詞を駄洒落キーとしている。次に筆者らの研究室で運用されている類音語辞書から駄洒落キーと音素的に似ている語を 100 個取り出し、駄洒落として出力する候補を得る。類音語辞書は音素類似度 [1] をコストとし、EDR 日本語単語辞書 [8] に記載されている語すべてを DP マッチングした上位 100 語が記述されている。駄洒落候補は駄洒落キーと共起スコアを計算し、候補の中で駄洒落キーと最も結束生の強いものが選定される。選択された駄洒落候補は応答文中の駄洒落キーと置換され、自然言語対話理解システムに出力される。

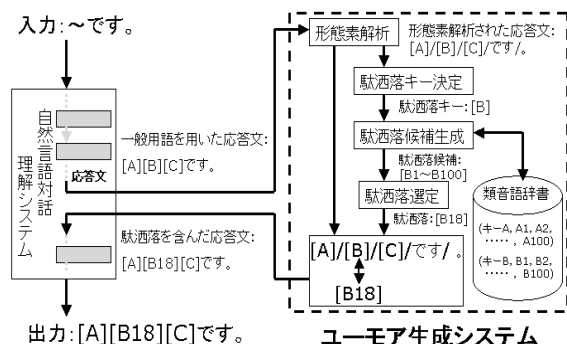


図 2: ユーモア生成システム（駄洒落）

4 評価

4.1 実験方法

本システムによって生成された業界用語を使用した文章にユーモアを感じるか、アンケートによる主観評価実験を行った。本システムは自然言語対話理解システムと組み合わせて使用するものであるが、筆者らの研究室で運用している自然言語対話理解システムは一般ユーザーと自由な対話ができる程度の性能を有していないこと、ユーモア生成システム本来の性能を測る必要があることから、HP 上からテキストデータを集集し、それを直接ユーモア生成システムに入力した結果と入力

前の元データの 2 つをランダムに被験者に提示した。テキスト 4 種類について実験を行い、2 つのおもしろさの違いを大学生 45 人に評価させた。

4.2 実験結果

主観評価実験の有効回答に対する集計結果を表 2 に示す。

表 2: 主観評価実験の集計結果

回答	テキスト			
	1	2	3	4
出力文の方がおもしろい	26 (57.8%)	24 (54.5%)	6 (13.6%)	27 (62.8%)
入力文の方がおもしろい	3 (6.7%)	8 (18.2%)	8 (18.2%)	6 (14.0%)
おもしろさは変わらない	16 (35.6%)	12 (27.3%)	30 (68.2%)	10 (23.3%)
合計	45	44	44	43

テキスト 1,2,4 に対する回答の分布は似ており、5~6 割の被験者が元の文章よりユーモア生成システムを通して一般語を業界用語に置換した文章の方がおもしろいと回答した。テキスト 3 に関しては約 7 割の被験者がユーモア生成システムを通して文章のおもしろさは変わらないと回答し、その他の回答数はほぼ同程度であった。

4.3 考察

実験結果より、以下のことが言える。

- ユーモア生成システムにより文中の一般語を業界用語に置換すると文章のおもしろみは大きくなるか変化しない
- また、おもしろみの変化は入力される文章の内容によって左右されることが示唆される

以上のことから、文章によっては 6 割以上の方がユーモア生成システム（業界用語）を使用することでおもしろみを感じるようになっており、この結果は人工知能の基礎的分野を対話システムに応用していく萌芽的研究としては有意義なものであると考える。

5 おわりに

本稿では、ユーモアの計算機処理として既存の自然言語対話理解システムの応答文に業界用語・駄洒落を含ませる方法と実際に開発したシステムの評価を報告した。本研究は未だシステムの開発途中であり、駄洒落の生成法の妥当性や実験 / 評価方法の信頼性の確立などが今後の課題である。

参考文献

- [1] キム, 滝沢: 日本語ダジャレなぞなぞ生成システム”BOKE”, 人工知能学会誌, Vol.13, No.6, pp.920-927(1998).
- [2] 松澤, 堀, 金杉, 阿部: こたば工学入門, 人工知能学会誌, Vol.15, NO.3, pp.446-455(2000).
- [3] 石井鈴木, 岡田, Campbell: 漫才の生成とその創発的計算, 情報研報, Vol.97, No.66, pp.59-66(1997).
- [4] 香川, 唐澤: 自然言語対話におけるユーモアを含んだ応答文の生成, 情報処理学会第 65 回全国大会講演論文集 (2), 6M-3, pp.173-174(2003).
- [5] 安理 岳村: 笑いの構造 - バルクソンから吉本新喜劇まで, 近代文藝社 (1995).
- [6] 角田, 田中: PDAI&CD に基づく意味の学習および文脈依存の多義性解消, 電子情報通信学会技術研究報告, Vol.DE93-1, pp.1-8(1993).
- [7] 藤田, 唐澤: 場面同定システムの開発と談話理解, 情報処理学会第 65 回全国大会講演論文集 (2), 4M-1, pp.159-160(2003).
- [8] EDR. 電子化辞書 1.5 版仕様説明書, 日本電子化辞書研究所.