HOKKAIDO UNIVERSITY

Title	統合型対話システムにおける話題適応及び駄洒落ユーモア処理に関する研究
Author(s)	谷津, 元樹
Degree Grantor	北海道大学
Degree Name	博士(情報科学)
Dissertation Number	甲第12645号
Issue Date	2017-03-23
DOI	https://doi.org/10.14943/doctoral.k12645
Doc URL	https://hdl.handle.net/2115/65779
Type	doctoral thesis
File Information	Motoki_Yatsu.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

博 士 論 文

統合型対話システムにおける話題適応及び駄洒落ユーモア処理に関する研究

谷津 元樹
北海道大学 大学院情報科学研究科

目次

第 1 章　序論 1
1.1　研究の背景 2
1.2　研究の目的 3
1.3　研究の独創性 4
1.4　本論文の構成 4
参考文献 6

第 2 章　汎用的な対話システムにおける話題変遷の捕捉 7
2.1　はじめに 7
2.1.1　タスク指向型対話 8
2.1.2　非タスク指向型対話 8
2.1.3　汎用的な対話システムにおける話題遷移検出の必要性 . 10
2.1.4　対話の 2 状態モデル 10
2.1.5　本章の構成 11
2.2　汎用的対話生成手法 12
2.2.1　はじめに 12
2.2.2　発話集合からの応答選択ルールによる発話生成 12
2.2.3　入力発話の特徴に基づく応答発話生成手法の選択 13
2.2.4　汎用的対話システムの構成 13
2.2.5　応答発話選択における外部モジュール 13

2.2.6 フィルタモジュール 13
2.2.7 応答選択モジュール 13
2.2.8 発話生成モジュール 14
質問応答発話生成 (タグ QA) 14
非タスク指向型応答発話生成 (タグ NT) 14
ドメイン選択応答発話生成 (タグ RU) 15
タスク指向型応答発話生成 (タグ Dn) 15
2.2.9 外部モジュールの呼出とタグ 16
2.2.10 入出力 16
2.3 ドメイン選択発話生成手法 16
2.3.1 対象ドメイン 17
2.3.2 キーワード集合 17
2.3.3 目的ベクトル $\boldsymbol{a}$ 18
2.3.4 WordNet における意味類似度 19
2.3.5 使用者の目的の明確度および選択発話生成 19
2.4 定性的対話評価実験 20
2.4.1 はじめに 20
2.4.2 対話印象評価実験 20
2.4.3 実験方法 21
2.4.4 印象評価項目 21
2.4.5 形容詞対項目 21
2.4.6 発話選択精度評価実験 22
2.4.7 実験方法と結果 22
2.5 考察 26
2.5.1 概要 26
2.5.2 印象評価及び発話選択精度について 26
2.5.3 SD 法評価値について 26
2.5.4 ドメイン選択発話生成の適切性評価項目について 29

2.6 おわりに 29
2.6.1 今後の課題 29
参考文献 30

第 3 章 駄洒落ユーモアの検出 34
3.1 はじめに 34
3.2 対象とする駄洒落の類型 36
3.2.1 駄洒落の主な類型 37
3.2.2 併置型駄洒落における種表現・変形表現 37
3.2.3 各類型の構成比 39
3.2.4 併置型駄洒落を検出の対象とする理由 39
3.3 関連研究 41
3.4 分析対象のデータ 42
3.5 提案手法 43
3.5.1 分類器の構築，素性の選択 44
3.5.2 音韻類似度の計算 46
3.5.3 学習データおよびテストデータの規模 48
3.6 検出性能評価実験 49
3.6.1 カーネル関数間の比較 50
3.6.2 正例比率の変化に対する各素性セットの頑健性の評価 . 51
3.6.3 語彙素性の有効性に対する分析 51
3.7 考察 51
3.7.1 提案手法の有効性 52
3.7.2 ルールベース素性 (b),(c) 及び (d) の追加による効果 . . 54
3.7.3 語彙素性 (a) 追加による効果 55
3.7.4 重畳型駄洒落の検出性能 56
3.7.5 駄洒落を含まない文の音韻的性質 56
3.7.6 まとめ 57

3.8 おわりに 59
参考文献 60

第 4 章 対話における駄洒落ユーモアの認識/生成と話題適応 63
4.1 はじめに 63
4.2 関連研究 64
4.3 提案手法及び実験システム 66
4.3.1 対話システムの機能 66
4.3.2 話題適応 66
4.3.3 駄洒落の検出及び生成 67
4.4 印象評価実験 69
4.4.1 実験の方法 69
4.4.2 実験の条件及び結果 70
4.4.3 対話例 71
4.5 考察 71
4.6 まとめ 72
参考文献 74

第 5 章 結論 77

謝辞 80

研究業績 81

表目次

2.1 ドメイン選択対話の例 11
2.2 本研究で用いたモジュールおよびタグの一覧 17
2.3 印象評価項目 22
2.4 印象評価実験の結果 (5 段階評価) 22
2.5 形容詞対項目 (右辺は高得点項目) 23
2.6 選択精度評価実験の 3 段階評価値と尺度 24
2.7 応答精度評価値 (3 段階評価) の分布 24
2.8 統合型対話システムの対話例 25
2.9 統合型対話システムの発話タグ毎の発話精度評価平均値 28
2.10 SD 法印象評価値データの因子分析結果 28

3.1 駄洒落の類型とその構成比 39
3.2 駄洒落を含む文の取得元の Web サイト 40
3.3 実験対象のデータセット 42
3.4 性能評価実験において用いた素性セット 50
3.5 素性セット F_{15} に対する各カーネルの検出性能 50
3.6 ルールベース素性単体による検出性能 52
3.7 重畳型駄洒落の検出性能 56
3.8 Mann-Whitney の U 検定による音韻類似度スコアの比較 . . . 57

4.1 システムの状態及び 2 種の駄洒落動作の発生確率 65
4.2 実験条件の一覧 69
4.3 実験条件間の評価値の比較結果 71
4.4 実験で用いた 7 つの条件及び各設問の平均得点 72
4.5 ユーモア機能付きシステムとの対話例（駄洒落生成のある例） . 73
4.6 ユーモア機能付きシステムとの対話例（駄洒落検出あり） . . . 74

図目次

2.1 統合型対話システムの概要 15
2.2 目的ベクトルの差分 $\boldsymbol{\Delta a}$ の算出アルゴリズム 18
2.3 SD 法印象評価の結果データのクラスター分析結果の樹形図 . . 27

3.1 併置型駄洒落における種表現と変形表現の例 38
3.2 検出システムの処理の流れ 42
3.3 検出性能評価に用いる尺度 47
3.4 正例数/負例数比に対する線形カーネル SVM の F 値のプロット 53

4.1 Siri との対話例 63
4.2 ユーモア機能付き統合型対話システムの構成 65

第 1 章

序論

コミュニケーションの不足に悩む人の数は近年増加傾向にある．その原因のうち最大と考えられるものは，会話のできる機会の減少である．

第一に，一人暮らし (独居世帯) の高齢者が増加し続け，またそのような高齢者が会話をする機会が少ないことが挙げられる．内閣府による高齢社会白書 [1] は，65 歳以上の高齢者の世帯数が近年増加傾向をたどっていることを指摘している．また同白書によれば，夫婦のみの世帯を合わせると，高齢者のみでの生活を送る世帯の高齢者のいる総世帯数に対する割合は 50% を超える．同白書の 2013 年の統計 [2] では，そのような高齢者において，二世代以上が同居する家庭に比べ，会話の頻度が「2 日〜3 日に 1 回」以下である人の割合が顕著に高いという調査結果が示されている．

また，若年世代の間でも会話をする機会を忌避する動きが観察されている．1990 年代より 2000 年代には携帯電話のメール [3]，2010 年代には Facebook, LINE などのソーシャルメディアの普及が爆発的に進み [3]，ユーザの間では「いつでも誰かとつながれる」状況が定着した．その反動として，若年層を中心に「つながれる」状態の解消への不安が恐怖感へと転化し，そのために逆に現実世界における「つながり」を回避する人々が増えているという指摘 [4] がある．つまり，コミュニケーションを図る手段が余剰になったことが，最も基本的な会

話などのコミュニケーションを行う機会を奪っている状況を生み出しているといえる．

このため，会話による意思の疎通の可能な相手として機能し，そして正の感情を喚起させてくれる存在に対する需要が高まっていると考えられる．

1.1　研究の背景

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®，しゃべってコンシェル®などがある．

一方，ユーモア感覚の向上が心理的ストレスへの対処能力を増進し，それによって，健康に対して良い効果が得られることがいくつかの研究において示唆されている [5]．ユーモアへの感受性が高くユーモアを日常的に用いることへの抵抗感が少ない人々は，そのような傾向をもたないグループと比較して，日常的に感じている心理的ストレスの強度の調査で得られる平均値が有意に低かったという調査結果がある [6]．これらの研究においては，ユーモアはストレスに対処する有効な手段の一つと見なされている [7]．

ユーモアの使用が，心理的ストレスが原因となる病気の症状を和らげていることを示した研究もある．ユーモアの感受性に関する評価指標が高いグループと低いグループとの間で，呼吸器系の器官を感染症から守っている抗体の唾液中の含有量の時間変化を計測すると，ユーモアの感受性の高いグループの方が有意に減少量が小さいという実験結果 [8] がある．これは，ユーモアに日常的に親しむことにより心理的ストレスが緩和され，免疫機能の活性度がより高くなっているためであると考えられている．また，最近の研究では，消防士を被験者とした実験により，強い心理的ストレスに曝されることが多い職業の人々では，ユーモアによりストレスに対処する能力の高い人は低い人と比較して，心的

外傷後ストレス障害 (PTSD)・燃えつき症候群から被る影響が有意に小さいことを示した研究 [9] がある．加えて, ユーモアの引き起こす笑いには, 社会的なつながりを維持しコミュニティの連係を深める作用も確認されている [10].

対話システムがユーモア機能をもつことにより，人間の話し相手がいない状況でも，ユーザがユーモア感覚向上をはかることが可能となる．対話システムとのユーモアのある対話を通し，ユーザの健康維持に寄与できる可能性がある．

このような背景から, 対話システムへのユーモアセンスの実装が試みられている．ユーモアセンスにおいては, ユーモアへの感受性とユーモア発話を適切に行う能力が必然的に大きな重みを占める．ユーモアを表出することが適切な状況において，最適な内容のユーモアを表出するためには, 状況及び内容の適切性を判断する能力が必須と考えられる．しかし, 現状で提案されているユーモア機能付き対話システムには, ユーザから入力されたユーモアの理解の結果を考慮するものは存在しない．したがって, 現状の対話システムの行う対話において, 完全なユーモアセンスが実現されているとは言い難い.

また, ユーモア理解の手法が確立できたとしても, 対話, 特に雑談という行為においては話題や感情などの話者間及び話者内の心的状態の変遷が伴う．ユーモアに対して常に同様な反応を表出したり, 常に一定のタイミングでユーモア発話を行うことは明らかに適切ではない．このため, 心的状態の変遷に応じてユーモア認識・発話の手段を変化させることが求められる.

1.2　研究の目的

本学位論文において述べる研究（以降，本研究と記す）の目的は, 統合型対話システムにおいて, 話題の変遷への適応に基づく対話戦略とユーモア理解・生成能力との間の適切な相互作用を実現することである．すなわち, 雑談を主とした対話を行う能力をもつ対話システムが, 現在進行中の対話が特定の話題にどれだけ関連したものになっているかに応じて, ユーモアの理解・生成の手段を適切に変化させることが可能なシステムを実現することである.

1.3　研究の独創性

本研究の独創性は，以下に示す通りである．

- 従来のユーモア生成機能付き非タスク指向型対話システムにおいて，話題の変遷への対応及びユーモアを発するタイミングを対話履歴より推定する能力を持ち合わせたものは存在しない．本研究においては，適切なユーモア生成を行う頻度を，対話履歴の名詞表現の概念の意味類似性に基づく話題適応（詳細は第 **2** 章および第 **4** 章において述べる）により算出している．これにより，音声認識誤り等により生じるノイズに対する，テキストレベルでの頑健性が原理的に保証される．
- ユーモア機能付き対話システムに関する従来研究は，ユーモアの生成を行うものが主である．駄洒落のような短文ユーモアの認識及び生成を両立して行うことのできる手法は，これまでに発表されていない．本学位論文において提案するユーモア機能付き統合型対話システムは，駄洒落ユーモアの生成及び認識の双方を適切な頻度において行う能力を持つものである．同様に初の試みである，同対話システムにおいて話題適応に基づいてユーモア機能の適切な頻度を決定する手法も，ユーモアに対する複数の処理を同一のシステム内で行うことがその前提となっている．

1.4　本論文の構成

本学位論文は 5 つの章からなる．構成を以下に示す．

第 **1** 章では, 研究の背景となる問題及び関連研究について, 文献を示しながら考察を行った．

第 **2** 章では, 話題適応発話の生成機能をもつ統合型対話システムについて述べる．対話ドメインの遷移を検出する手法について述べ, 対話システムの構成を説

明し, 応答精度評価実験及び印象評価実験の結果について考察する. 統合型対話システムとは, 複数手法で応答を生成し入力発話に対し最適な応答を選択して出力する非タスク指向型対話システムを指す. 実験の結果, 提案手法が既存の非タスク指向型対話システムの対話の自然性を低下させずに, 話題適応発話の生成が可能であることが示された. 実験において扱った対話例の分析を通して今後の課題を検討する.

第 **3** 章では, 駄洒落ユーモアの検出手法について述べる. 検出の対象としたのは, 構成比の調査において最も多数存在することが確認された「併置型」の駄洒落を含む文である. 併置型駄洒落は, 元となる種表現と, 種表現が音韻的な変化を受けた変形表現の 2 つの部分を含む. 初めに, 併置型駄洒落のルールベースでの検出を行った. 検出ルールは, 駄洒落文内の種表現・変形表現の間の子音音韻類似度の閾値に基づくものである. 次に, 文に含まれる語彙情報に加え, 上記の子音音韻類似度に基づく検出ルールの結果を素性とし機械学習の一つである Support Vector Machine (SVM) を用いた併置型駄洒落の検出手法を提案し, 提案手法と前述のルールベースの検出手法との比較およびパラメータの変化に対する性能の特性の調査を行った.

第 **4** 章では, 本研究の目的である, 話題遷移に対応したユーモア機能付き対話システムの性能評価について述べる. ユーモア認識・生成の発生頻度を現状の対話と話題キーワードの関連度に基づいて変動させることにより, 対話の印象を向上させることが可能と考えられる. そこで, 第 **2** 章および第 **3** 章で述べた統合的対話システムに第 **3** 章の駄洒落ユーモア認識手法及び駄洒落生成手法を加えた統合型対話システムの変動 (話題適応) に対する印象評価値の変化の調査を行った.

最後に第 **5** 章において, 本研究を総括し、残された課題及び今後の研究の対象について述べる.

参考文献

[1] 日本国政府内閣府. 平成 28 年版 高齢社会白書. 政府刊行物.

[2] 日本国政府内閣府. 平成 26 年版 高齢社会白書. 政府刊行物.

[3] 日本国政府総務省. 平成 27 年版 情報通信白書. 政府刊行物.

[4] 今川民雄, 山口司, 渡辺舞. 人とのつながりとこころ—人と社会を見通す心理学. ナカニシヤ出版, 2015.

[5] 椙本知子, 山崎勝之. 対人ストレスユーモアコーピングが敵意, 意識的防衛性と抑うつに及ぼす影響. 心理学研究, Vol. 82, No. 1, pp. 9–15, 2011.

[6] Millicent H Abel. Humor, stress, and coping strategies. *Humor–International Journal of Humor Research*, Vol. 15, No. 4, pp. 365–381, 2002.

[7] Rod A Martin. *The psychology of humor: An integrative approach*. Academic press, 2007.

[8] Rod A Martin and James P Dobbin. Sense of humor, hassles, and immunoglobulin A: evidence for a stress-moderating effect of humor. *The International Journal of Psychiatry in Medicine*, Vol. 18, No. 2, pp. 93–105, 1989.

[9] Michael Sliter, Aron Kale, and Zhenyu Yuan. Is humor the best medicine? The buffering effect of coping humor on traumatic stressors in firefighters. *Journal of Organizational Behavior*, Vol. 35, No. 2, pp. 257–272, 2014.

[10] Barbara Fraley and Arthur Aron. The effect of a shared humorous experience on closeness in initial encounters. *Personal Relationships*, Vol. 11, No. 1, pp. 61–78, 2004.

第 2 章

汎用的な対話システムにおける話題変遷の捕捉

2.1 はじめに

最近，音声対話システムの研究開発全般がますます進展している．一般の消費者向け分野の携帯電子機器などへの応用についても同様のことがいえる．筆頭に挙げることのできる例として，Apple 社の Siri，Microsoft 社の Cortana などが挙げられる．

この傾向は，情報機器のもつユーザビリティ・煩雑さといった指標がその煩雑化に伴って重要視されていることによるものと考えられる．機械があるタスクの達成においてヒトと同等の言語能力を備えているとき，そのドメインにおいて他者のモデルを機械に当てはめることが可能となる [1]．この場合，その機械と人間との間に自然な HRI (*Human-Robot Interaction*) が成立する [2]．

その実現手段として，コンピュータ上に対話システムを構築することが考えられる．対話システムは，使用者の発した発話文の入力に対して何らかの処理を加え，音声又は画面表示などの出力を行うシステムである [3]．対話システムとの対話は，ある役割 (タスク) の達成を目的とした対話と，そうでない対話とに

大別できる．前者をタスク指向型対話，後者を非タスク指向型対話と呼称する．

2.1.1　タスク指向型対話

タスク指向型対話の一例として，近年は音声対話処理技術を利用した観光案内が実現されている [4]．タスク指向型対話システムは，入力発話に対して当該システムのもつタスクが属するドメイン (対話のトピックの範囲) にのみ属する応答を出力する能力をもつ．

また，タスク指向型対話機能の代表的な例として，質問応答システムを対話に組み込む研究が続けられている．質問応答システムとは，自由な質問文を入力すると適切な応答を試みるシステムであり，人名・地名・組織名及び年号などの名詞概念を問うファクトイド型と，理由・方法及び手続きなどを答えるノンファクトイド型に分類される．

タスク指向型対話システムの行動，すなわちどのような発話行為を行うかの選択を行う対話管理機能は，経験的に得られたパラメータから確率的に行う手法が採用されることが多い．対話管理に用いる対話のモデルとして，初めにマルコフ決定過程 (MDP) が仮定された [5]．MDP を用いる場合には，直前の入力発話 (行動選択に必要な情報の入力源) により成立した状態 (得られている情報の組合せ) に対して，システムが選択する行動により得られる報酬 (対話の成立により正の値となる) が最大化するように，強化学習を用いて各々の行動の選択される確率を調整する．これを発展した対話モデルとして，音声認識及び言語理解の不確実性を反映する部分観測マルコフ決定過程 (POMDP) が後に採用され [6]，音声認識誤りや言語理解誤りの影響下での行動選択の精度が向上した．

2.1.2　非タスク指向型対話

一方，非タスク指向型対話はタスクの達成を目的としない対話である．非タスク指向型対話においては，発話が属するドメイン (話題) が変遷するため，一般にオープンドメイン対話，すなわち対象とする分野を指定しない対話に包

含される．非タスク指向型対話を行うシステムである非タスク指向型対話システムは最近その研究の裾野を広げている．非タスク指向型対話システムの一例は Weizenbaum による ELIZA [7] である．ELIZA は，機械による応答が応答ルールによって人間による発話を表現することを可能にした最初の例である．ELIZA は人間の発するあらゆる種類の発話に対し返答するように，そのルールセットが設計されている．

ELIZA によるルールベース手法を拡張し，Artificial Intelligence Markup Language (AIML) によるルールセットの記述を可能にした手法 (A.L.I.C.E.; Artificial Linguistic Internet Computer Entity) が Wallace により提案され [8]，ルールベースの手法が確立された．審査員及び一般参加者による評価の最も高い非タスク指向型対話システムに贈られる Loebner 賞 [9] を 2016 年に受賞したシステム *Mitsuku*[*1] は AIML を用いて人手で構成されたものであり，抜本的には ELIZA と同様な手法を用いている．河原 [10] も指摘しているように，印象評価の面では，現状の非タスク指向型対話システムにおいては一問一答のルールベース手法に代わる対話管理手法は確立されておらず，飛躍的な進歩が未だ起こっていない状況にあるといえる．

前節で述べた POMDP に基づく行動選択手法は，非タスク指向型対話にも適用可能である．しかしながら，オープンドメインを対象とする場合には，ユーザとシステムの属する環境及びそれに対し選択し得る行動の組合せ数は膨大なものとなり，行動選択の優先順位を決定する関数の学習には多大な量の実対話データを要する．限定されたドメインにおけるカジュアルな対話を目的とした対話管理機能は提案されているものの [11]，汎用的といえる広範なドメインに対応可能な手法は，実対話データの入手可能性の制約から未だ実現されていない．ただし，ユーザとの対話例から自律的にルールを学習してゆく手法 [12] [13] [14] も近年提案されており，今後の研究が注目されている領域である．

また，POMDP を用いた行動決定はマルコフ性に基づくものである．すなわ

[*1] `http://www.mitsuku.com`

ち，行動決定の方法はあくまでも直前のターン (話者交替) におけるユーザの状態及びそれに対するシステムの応答に依存する．従って，長期的な文脈における累積的な話題の変遷を捉えることは困難である．本章において提案する手法では，**2.1.4** にて述べる対話の二つの状態のモデルを設定し，**2.3** にて述べるドメイン選択発話生成手法を用いて累積的な話題の変遷に対応する．

2.1.3　汎用的な対話システムにおける話題遷移検出の必要性

2.1.1 に述べたように，観光案内システム等のタスク指向型対話システムの利用は，タスクが明示的であることが前提となる．タスクの達成によりシステムとの対話が終了するため，当該システムを利用する前後の状況との接続性や連続性は考慮されない．従って，対話ロボットや仮想エージェントに実装される単一の対話システムが非タスク指向型対話及びタスク指向型対話を行う場合，話題 (対話のドメイン) の遷移とともに非タスク指向型対話及び各々のタスク指向型対話の間の切り換えを行う必要があることが予想される．

2.1.4　対話の 2 状態モデル

上記の問題を緩和するため，人間の対話との類似に基づき対話モデルを設定した [15]．対話モデルにおいては，汎用的なシステム内で，使用者の目的の明確化・曖昧化に伴いオープンドメインと特定ドメインとの間の状態遷移が行われる．人間による対話の例を表 2.1 に挙げる．この対話では，話者 A,B が話題の自由な対話を行っている．表 2.1 中の対話では，B が A の行いたいことを対話履歴から推定し，対話の末尾において提案している．

このような対話を自然な対話の例としてあげることができる．特に，6 発話目のような応答発話はシステムの内部の知識によりドメインの判定・選択を行うシステム主導型のものである．2 状態の間の遷移には，このようなオープンドメイン対話においてドメインの選択を促進する発話が必要であるといえる．これまでの対話エージェントの研究において，このようなシステム主導型の発話を

表 2.1　ドメイン選択対話の例

A	1	寒い
B	2	寒いな
A	3	午後には晴れて暖かくなるそうだけど
B	4	そうなんだ
A	5	とにかくなにか暖まるものが欲しいね
B	6	コーヒーでも飲む？
A	7	ああ，それいいね
B	8	じゃあ，喫茶店に行くか

行う手法の発表を著者は確認していない．

2.1.5　本章の構成

本章では，表 2.1 における B が 6 発話目に行ったものと同種の発話をドメイン選択発話と呼称する．本研究の目的は，オープンドメイン対話から特定ドメインへの移行を，対話の連続性を保持しつつシステム主導により行うことである．本章では，次の 2 つの手法について提案する．

汎用的対話生成手法

対話システムをモジュール化し，オープンドメイン対話を生成することの可能な対話システムを構築した．モジュール対話システムには，非タスク指向型対話システムおよび質問応答システムが含まれる．詳細を **2.2** で述べる．**2.4** では，汎用的対話生成手法の定性的評価のために行った実験 (発話分類精度評価，印象評価実験) について述べる．

ドメイン選択発話生成手法

提案した汎用的対話機能においてドメインの選択を行うためのフィルタ，および選択発話生成モジュールを実装した．これについて **2.3** で述べる．ドメイン選択精度の評価を **2.4** での実験と同時に行い，実装された仮想的なタスク指向型対話システムに関連付けられたドメインを選択する発話の生成精度について検討した．その結果と考察について **2.5** で述べる．

2.2　汎用的対話生成手法

2.2.1　はじめに

2.1.3 に述べた考え方を具現化し，テキストベースの汎用的対話システムとした．本節では，実装した汎用的対話システム（以下，本システム）の概要を説明する．本システムは 1 節でも述べたとおり，オープンドメイン対話の中でドメイン選択発話を生成することを目的としている．このためには，次にあげるような方法が考えられた．

2.2.2　発話集合からの応答選択ルールによる発話生成

この方法は，ルールの数が応答候補数となるため，**2.1.1** で述べた ELIZA のように，有限個のルールセットに基づく応答生成手法のもつ制約を受ける．具体的には，発話生成部分 (ルールの右辺に当たる部分) に自由度が小さいといったことがあげられる．

その反面，単一セットのルール辞書の検索に基づくことから高速であるため，非タスク指向型システムへの応用としては文脈を考慮した統計的発話選択手法の研究 [16] がある．

2.2.3　入力発話の特徴に基づく応答発話生成手法の選択

発話生成手法をモジュール化し，それらの手法から入力発話に対し適切なモジュールを選択するためのルールないし選択条件を準備する．この手法では，ドメインに特化した発話生成手法を用いることができる．本研究における汎用的対話生成にこの手法を用いる．

2.2.4　汎用的対話システムの構成

提案手法のシステムの構成について述べる．システムはメインプログラムと外部モジュールで構成されている．

2.2.5　応答発話選択における外部モジュール

メインプログラムは，外部モジュールとして，発話，フィルタ，スコア付与のための各モジュールを使用する．各モジュールにはキーワードタグが人手で付与されている．以下に各モジュールの役割を示す．

2.2.6　フィルタモジュール

タグに対しスコアを付与する．具体的には，入力発話に対し最適なタグと事前に人手で決定された一定のスコアを出力する．発話の特徴に対するスコア付与に用いる．スコアは，異なるフィルタ間の優先順位の決定に用いることができる．タグとスコアの関係については **2.2.9** を参照されたい．

2.2.7　応答選択モジュール

同一タグの付与された発話モジュールによる応答発話にスコアを付与する．**4** の評価実験で用いた手法は，今井ら [17] の用いた人対人チャットログとの類似

度による評価手法，および入力文と各候補中の単語の共起に基づく関連性評価である．

2.2.8　発話生成モジュール

モジュール化された対話システムであり，1 入力発話について 1 応答発話を出力する．他の入力先を伴うモジュールもある．対話の種類 (タスク指向型／非タスク指向型，システム主導型／使用者主導型) に関わらず，システムによる応答発話を出力する．

発話モジュールは 1 つ以上のタグ（文字列）を持つ．タグについては **2.2.9** で述べる．

本研究で用いたモジュールの元となったシステム及びタグは次の通りである．

質問応答発話生成 (タグ QA)

Web 検索を知識源とする質問応答システム [18] (以下 QAC4) を用いた．このシステムは非タスク指向型の応答に対応できるという意味では非タスク指向型システムに分類できるが，汎用的対話機能におけるフィルタモジュールによるタグ付けを用いた発話生成手法の選択の試行のためにこの手法を取り入れた．

非タスク指向型応答発話生成 (タグ NT)

非タスク指向型応答の生成時には，単一の発話選択モジュールのみを用いることにより応答が単調な印象を与えることを避けるため，複数の発話生成手法の出力した発話を応答選択モジュール (**2.2.7**) により評価し，合計のスコアの最も高い発話を出力応答として選択する．

選択対象となる 3 つの発話生成手法は，高橋ら [19] による Web 検索とマルコフモデルを用いた雑談システム (以下 Maru)，樋口ら [20] の Web による単語共起頻度によりモダリティ表現を付加する雑談システム (以下 Moda)，および木村ら [12] の研究において用いられた ELIZA 型システム (以下 Eliza) である．

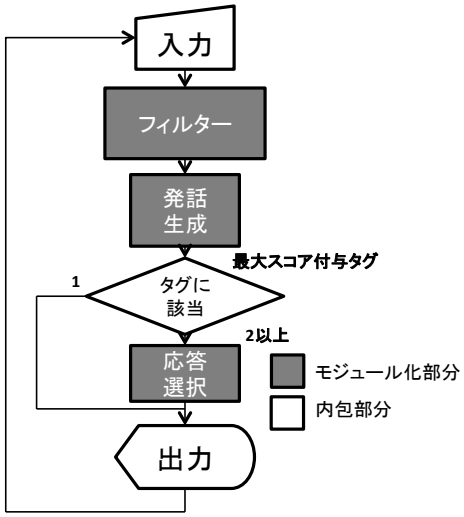

図 2.1　統合型対話システムの概要

ドメイン選択応答発話生成 (タグ RU)

ユーザ発話に含まれる自立語表現から適切なドメインを選択し，そのドメインキーワードを提示する応答発話を生成するモジュールである．詳細は **2.3** にて述べる．

タスク指向型応答発話生成 (タグ Dn)

2.3.2 で述べるキーワード集合 a_n に対応するタスク指向型応答生成手法である．**2.4** の評価実験では，モデル実装として 1 ドメインに対して 1 タイプの応答発話生成が可能な発話生成モジュールを用いた．

本研究では，発話選択モジュールの機能を試験するため，複数の発話生成手法を用いた．

2.2.9　外部モジュールの呼出とタグ

システムは入力発話に対する応答を次のような手法で出力する．メインプログラムは，図 2.1 に示すアルゴリズムに従って，使用者の入力発話から最終的な出力発話を生成する．初めに，入力発話をフィルタモジュールが評価する．戻り値のタグとスコアは，連想配列において加算される．最大スコアの付与されたタグが決定する．

システムは，最大スコアの付与されたタグに属している全ての発話モジュールからの応答を取得する．ここで取得する応答発話が複数となる場合，全ての応答を一つまたは複数の応答選択モジュールが評価する．スコアの線形和を比較し，最大スコアを持つ応答発話を出力する．応答発話が単一である場合は，その応答を採用する．

評価実験で用いたモジュールとタグの一覧を表 2.2 に示す．

2.2.10　入出力

本研究で構築したシステムは音声対話への応用を意図している．そのため，クラスオブジェクトでの実装形式をとる．ただし，単体でコンソール環境で実行が可能である．システムへの入力発話，システムの応答発話，および応答の所要時間はログファイルに記録される．

2.3　ドメイン選択発話生成手法

本節では **2.1.4** の対話の 2 状態モデルを基にドメイン選択発話生成手法を提案する．使用者が対話システムとの対話を開始するとき，しばしばその使用目的が十分に明確でないことがある．現実世界ではさらに，ある使用者の持つ目的の属するドメインはタスク指向型対話および非タスク指向型対話の両方に含まれている．提案手法はこの観察に基づき，オープンドメイン対話におけるユー

表 2.2　本研究で用いたモジュールおよびタグの一覧

モジュール種別	名称	タグ	機能
発話生成	Maru	NT	非タスク指向雑談応答
発話生成	Moda	NT	非タスク指向雑談応答
発話生成	Eliza	NT	非タスク指向雑談応答
発話生成	QAC4	QA	非タスク指向質問応答
発話生成	Recom	RU	ドメイン選択応答
発話生成	TaskTest	Dn	キーワード a_n 関連タスク指向応答
フィルタ	Nontask	NT	雑談入力発話 (スコア 1.5*)
フィルタ	QA	QA	質問入力発話検出 (スコア 2.0*)
フィルタ	WdnetS	RU	質問入力発話検出 (スコア 2.0*)
応答選択	IRCChatLog	-	チャットログとの類似度による評価
応答選択	IRCChatLogAbst	-	チャットログとの類似度による評価
応答選択	NGramHitnum	-	単語の共起頻度に基づく関連性評価

* スコアは，評価実験 (**2.4** 節) にて使用したものを記載している．

ザ発話に含まれる自立語表現とキーワード列との語彙概念辞書における意味類似度をベクトル表現化し，最大値条件から適切なドメインを選択してユーザに提示する応答発話を生成する．

2.3.1　対象ドメイン

提案手法の選択対象ドメインは，**2.2** で提案した汎用的対話機能において実装されたタスク指向型発話生成手法に関連付けられた所与のドメインである．

2.3.2　キーワード集合

キーワードは，本システムに組み込まれたタスク指向型発話モジュールと設定ファイルにおける設定により関連付けられたドメインに属する日本語の一般

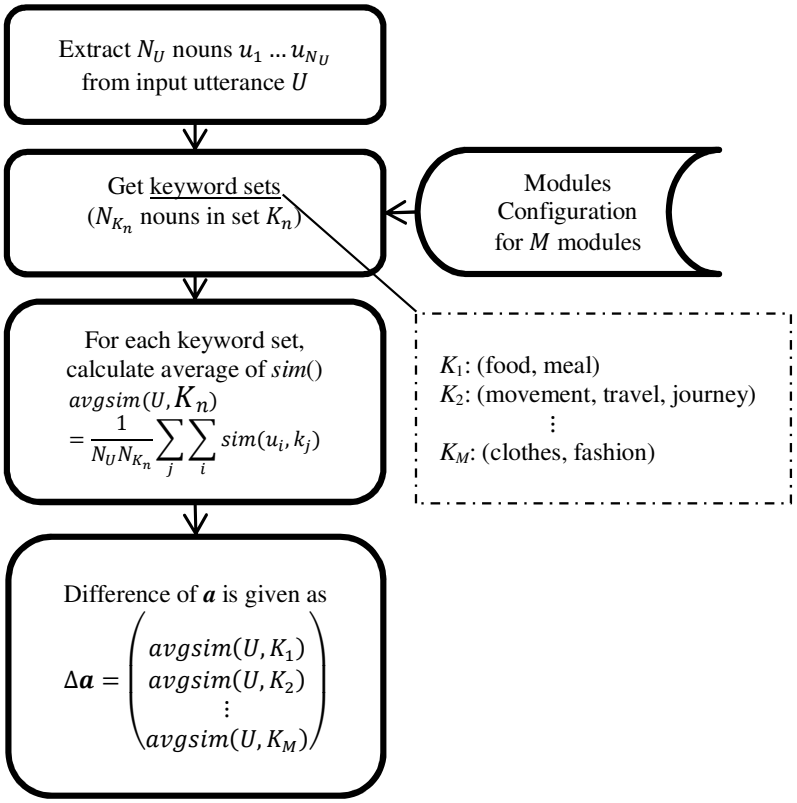

図 2.2　目的ベクトルの差分 $\boldsymbol{\Delta a}$ の算出アルゴリズム

名詞である.

2.3.3　目的ベクトル $\boldsymbol{a}$

目的ベクトル $\boldsymbol{a} = (a_1, a_2.a_3, ...)$ は使用者の現在の目的とその方向性 (aim) を表す．ここで，$a_1, a_2, ...$ はキーワード集合の各自立語表現 (但し，語彙概念辞書 日本語 WordNet [21] に概念として含まれているもの) に対応する.

2.3.4　WordNet における意味類似度

$\boldsymbol{a}$ は 意味類似度ベクトル $\boldsymbol{\Delta a}$ の累加によって算出される．$\boldsymbol{\Delta a}$ の要素は Leacock-Chodorow 法 [22] による意味類似性の尺度によって求められ，次式で表される．

$$\Delta a_i = \frac{1}{N_U N_{K_i}} \sum_{u \in U} \sum_{k \in K_i} sim(u, k) \tag{2.1}$$

$$sim(c_1, c_2) = max\left(-\log \frac{N_p}{2D}\right) \tag{2.2}$$

ここで，N_p は 2 単語 c_1, c_2 間の語彙概念辞書における最短のグラフ距離を，D は語彙概念辞書における両者の分類の深さを表す．意味類似度の算出には，Python [*2] 上で動作する Natural Language Toolkit [*3] を用い，語彙概念辞書には日本語 WordNet [21] を用いた．一連の流れを図 2.2 に示す．

2.3.5　使用者の目的の明確度および選択発話生成

使用者の目的の明確度を，目的ベクトル $\boldsymbol{a}$ のユークリッドノルム $\|\boldsymbol{a}\| = \sqrt{\sum_k a_i^2}$ を求めることにより得る．ノルム $\|\boldsymbol{a}\|$ が閾値 T を上回るターンにおいては，キーワード集合内のあるタスク指向型発話生成手法に関連付けられたドメインへの関連度が十分に高いという条件が成立する．**2.4** の評価実験においては，$T = 3.0$ とした．

このとき，フィルターモジュールはタグ 'RU' と正のスコアを出力し，選択される発話生成手法の候補がただ 1 つとなる．発話モジュールは，$\boldsymbol{a}$ における最大の値を持つ要素に関連付けられたキーワード k_i を用い，使用者がタスクを選択するための発話を出力する．

[*2] http://www.python.org

[*3] http://www.nltk.org

2.4　定性的対話評価実験

2.4.1　はじめに

本節では，汎用的対話手法の定性的評価のために行った 2 つの実験について述べる．

2.4.2　対話印象評価実験

2 における汎用的対話生成手法が被験者に与える印象および **3** でのドメイン選択発話生成手法の被験者に対する適性を調査するために行った Web を用いた実験と，アンケート調査による評価について述べる．本研究のベースラインとして，質問応答システム [18] 単体，ELIZA 型システム [12] 単体との比較を併せて行った．

ELIZA 型システム及び質問応答システム単体をベースラインとして用いた理由を述べる．**2.1.2** において述べたように，ELIZA に代表されるルールベース手法の登場後も種々の対話管理手法が検討されてきたが，現状において最も一般的かつ性能面でも優位にあると考えられる非タスク指向型対話システムはルールベース手法である．このため，非タスク指向型システムを基本システムとする提案手法の評価には，ルールベース手法に基づくシステムをベースラインに用いることが妥当と考えられる．また，前述の理由より，大規模な実対話データを学習データとする統計的な対話管理手法が確立されていないことなどから，本提案手法のシステムには直接の比較対象が存在しない．従って，提案手法におけるドメイン選択発話の有効性の評価方法としては，一般的な非タスク指向型対話システム及び単一のタスク指向型機能を提供するシステムとの比較が適当であると考えられる．

2.4.3　実験方法

被験者は，20 ターン以上の自由な対話を行った．被験者とシステムとの対話は，CGI スクリプトを用いて Web ブラウザを経由して行った．20 ターン経過後に提示されるアンケート画面にて被験者は下記の調査項目について回答した．記録した被験者の属性と分布については，次の通りである：性別 (男性 9，女性 1)，職業（研究者・学生以外 4，人工知能・自然言語処理に関連する分野の研究者または学生 3，人工知能・自然言語処理に関連しない分野の研究者または学生 2，小・中・高生 1)，年齢層（19 歳以下 1 名，20-29 歳 8 名，60 歳以上 1)，理・文の別 (人工知能・自然言語処理に関連しない分野の研究者または学生について，うち 2 名の内数は文系 1，理系 1)，対話ログの提供の有無．

2.4.4　印象評価項目

被験者は表 2.3 に示した項目を用いて印象評価を行った．その結果の評価平均値を，表 2.4 に示す．

2.4.5　形容詞対項目

被験者は，Osgood らによる意味差判別法 [23] での評価に頻繁に用いられる形容詞対 [24] を用いシステムを想起して 5 段階の評定を行った．使用した 33 組の形容詞対を表 2.5 に示す．

表 2.3　印象評価項目

項目	説明
(A)	対話を続けたいかどうか
(B)	対話が文法的に自然であるか
(C)	対話が意味的に自然であるか
(D)	システムの語彙が豊富かどうか
(E)	システムが人間らしいか
(F)	システムによる推薦が，被験者の考えていたことや悩みに対して関連が強かったか

表 2.4　印象評価実験の結果 (5 段階評価)

項目	評価平均値
A	1.75
B	1.75
C	1.63
D	2.00
E	2.13
F	1.75

2.4.6　発話選択精度評価実験

提案手法における汎用的対話能力の有効性を計量的に示すため，発話生成手法の選択精度の評価を行った．被験者はシステムと使用者の対話について，入力発話も対する応答発話の意味の妥当性を評価する．具体的には，応答発話内容の主観による内容理解可能性，文脈的容認度を 0〜3 の評価値として採用する．

2.4.7　実験方法と結果

被験者は，4 名の 24,25 才の男性理系大学院生である．評価対象とする対話は，総合定性評価実験にて収集された 150 ターンの対話である．評価は Web を

表 2.5　形容詞対項目 (右辺は高得点項目)

1	暗い	明るい
2	冷たい	暖かい
3	弱い	強い
4	陰気な	陽気な
5	軽い	重い
6	嫌いな	好きな
7	かたい	やわらかい
8	消極的な	積極的な
9	うるさい	静かな
10	不活発な	活発な
11	悪い	良い
12	不親切な	親切な
13	はげしい	おだやかな
14	苦しい	楽しい
15	地味な	派手な
16	つまらない	面白い
17	鈍い	鋭い
18	気持ちのわるい	気持ちのよい
19	頼りない	頼もしい
20	弱々しい	たくましい
21	小さい	大きい
22	ふまじめな	まじめな
23	遅い	速い
24	不愉快な	愉快な
25	不安定な	安定した
26	無口な	おしゃべりな
27	きたない	きれいな
28	だらしのない	きちんとした
29	単純な	複雑な
30	静的な	動的な
31	強情な	率直な
32	無責任な	責任感のある
33	悲しい	嬉しい

表 2.6　選択精度評価実験の 3 段階評価値と尺度

ラベル	値	意味の把握	文脈的な容認度
表現が自然 (正応答)	3	可	高
表現が不自然 (準応答)	2	可	低
意味的に誤り (誤応答)	1	不可	—
発話モジュールでのエラー	0	エラー応答	

表 2.7　応答精度評価値 (3 段階評価) の分布

正応答	準応答	誤応答	エラー
302	176	104	17
50.4%	29.4%	17.4%	2.84%

利用して行い，評価対象対話と評価フォームを表示し集計する CGI プログラムを用いて評価データを収集した．評価値及び尺度は，表 2.6 の通りである．

選択精度評価の結果を表 2.7 に示す．全評価の平均値は 2.27，標準偏差は 0.85 であった．印象評価実験で得られた対話の例を，表 2.8 に示す．

表 2.8　対話例 (User は被験者，QAC4,Recom,Maru,Eliza は表 2.2 の示す発話生成モジュール名)

話者	ターン数	発話内容
User	1	日本の首相は誰ですか
QAC4	1	野田佳彦だった
User	2	アメリカの大統領は誰ですか
QAC4	2	オバマらしいですからねぇ
User	3	今日は食堂で夕食を食べる予定です。
Recom	3	もしかして、健康, 病気に興味ありますか？
User	1	こんにちは
Recom	1	もしかして、健康, 病気に興味ありますか？
User	2	いいえ
Maru	2	ええと、ごめんなさい。よく分からないです。
User	3	健康なんか興味ありません
(エラー)	3	ごめんなさい、ちょっとエラーです。
User	4	もしかしてそっちの方が病気なんじゃないの？
QAC4	4	今日じゃないよ
User	5	母国語はなんですか？
QAC4	5	日本ですかね
User	6	日本で生まれたってこと？
Eliza	6	日本が... ってことなのかな？
User	7	マレーシア行ったことある？
Eliza	7	確かに... よね
User	8	札幌で一番美味しいラーメン屋さんはどこですか？
QAC4	8	北海道ですがね
User	9	性別はなんですか？
QAC4	9	..... 東京だなと
User	10	俺と結婚してください
Eliza	10	... ようになれるかな？
User	11	今いくつですか？
QAC4	11	..... 松田聖子だった

2.5　考察

2.5.1　概要

2.4 にて行った実験に対する考察を述べる．実験で得られた平均値は，1〜10 名の被験者の提示した値に対するものであり，これらの実験データは統計的に明確な結果が出るとされる標本数に達していない．しかし，対話システムの傾向のある側面を定性的に示すために有効な実測値であるとして分析を行った．

2.5.2　印象評価及び発話選択精度について

汎用的対話機能のパラメータの一部を固定した定性的実験は，発話そのものに注目し行う．入力がフィルタ機構による分類が正しく行われたものに正応答評価がなされる傾向があったことが，表 2.9 に示した発話タグ毎の発話精度評価平均値から読み取れる．

質問応答発話の精度においては標準的なものが得られ，また非タスク指向型対話システムの中で標準的と捉えられている ELIZA 型対話システムとほぼ同精度で質問応答機能を対話システムに付加することが可能である点において，本提案手法における汎用的対話機能が有効であると考えられる．

2.5.3　SD 法評価値について

この項目から得られた評価値データの因子分析を R[*4]を用いて行った．因子数は 4，累積寄与率は 0.70 である．結果を表 2.10 に示す．また，結果の 33 次元のデータを転置ベクトルと見なし，最遠隣法によるクラスター分析を行いデンドログラムの描画を行ったものを図 2.3 に示す．(評価値) = 6.0 で切断した場合，4 つのクラスターに分割可能であることがわかる．

[*4] http://www.r-project.org/

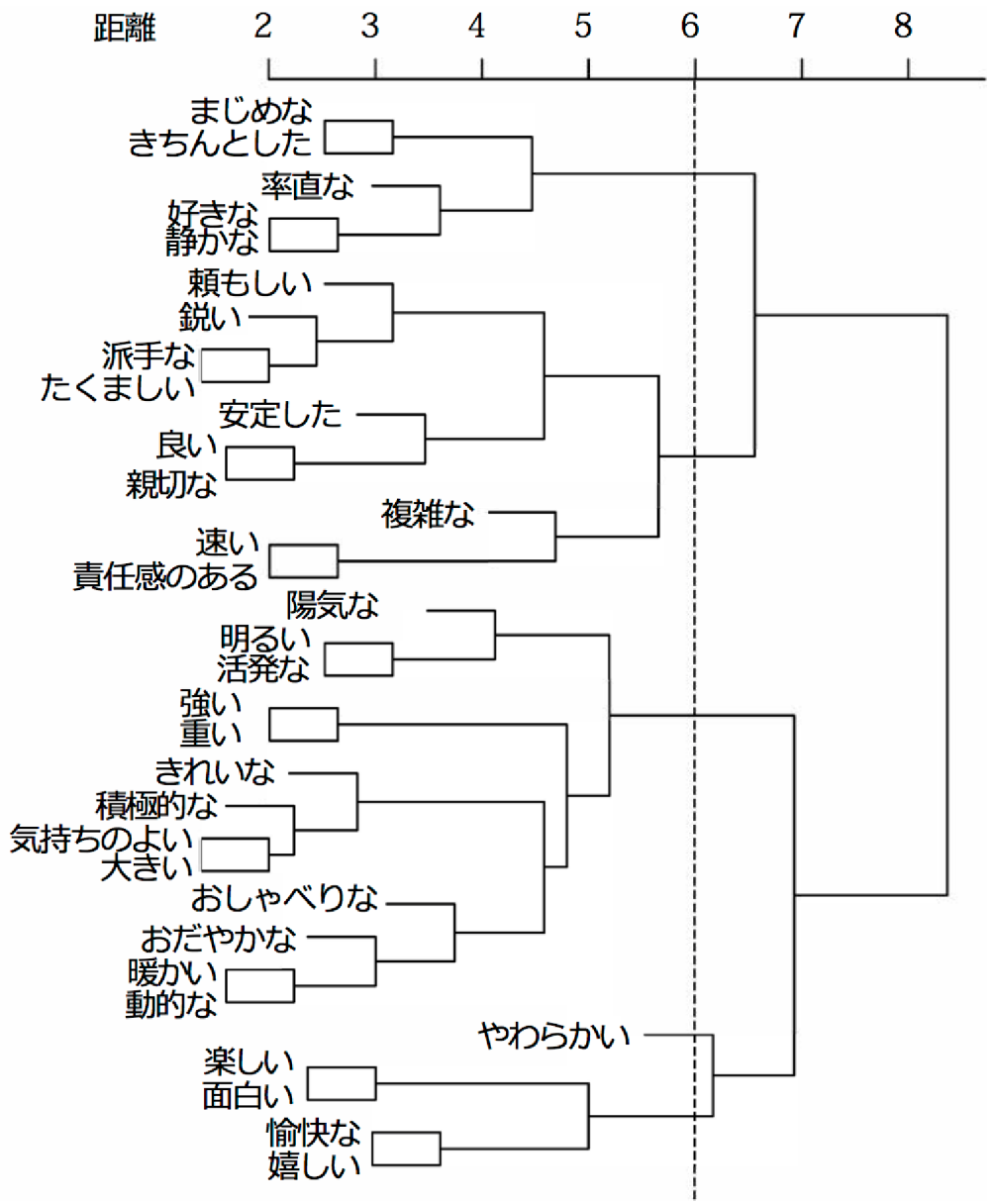

図 2.3　SD 法印象評価の結果データのクラスター分析結果の樹形図

表 2.9　発話タグ毎の発話精度評価平均値 (3 段階評価，被験者数は＊が 1 名，それ以外が 4 名)

タグ	評価平均値
NT	2.315
QA	2.270
提案手法による全発話	2.275
質問応答システム単体*	1.853
ELIZA 型システム単体*	2.390

表 2.10　SD 法印象評価値データの因子分析結果

累積寄与率	形容詞対番号	正負
0.231	6,9,17,19,22,24,28,31	＋
0.444	2,10,13,20,26,30	-
0.580	11,12,23,32	-
0.700	7,33	＋

因子分析を行った結果，実験に使用した形容詞対の群からは 4 つの因子を抽出した．これらの因子の中で、「責任感のある-無責任な」「速い-遅い」を含む一群は，クラスター分析で得られた樹形図にも含まれており，これらは機能的因子を表しているといえる．従ってこれらの因子は，印象評価実験においてベースラインシステムと比較して評価が低くなったことの原因としてシステムの処理全体に時間を要していることを示している．記録された応答生成の平均所要時間は，ベースラインの 0.163 秒に対し，提案手法システムが 33.10 秒である．

2.5.4　ドメイン選択発話生成の適切性評価項目について

印象評価における評価項目 (F) はドメイン選択発話生成の適切性を問う項目である．(F) では，ドメイン選択発話がシステムにより行われなかった被験者による評価値が低くなる傾向がみられた．

2.6　おわりに

本章では，機械と人間との間の自然な HRI の実現に向け，汎用的対話機能においてドメイン選択発話の生成を行う手法を開発した．対話システム全体が被験者に与える印象評価実験では，システムの応答速度が今後の主要な改善点となることが示された．発話選択精度評価実験においては，質問応答システム単体との比較を行った場合の，本システムにおける汎用的な対話機能による発話選択の有効性が示された．

2.6.1　今後の課題

本章での提案手法の評価実験においては，ベースラインとの部分的な優位性が示された．提案手法を改善し，人間の言語能力により近づけるためには，各種パラメータの最適化のため，さらなる評価が必要である．また，評価の客観性を保持するためには，提案手法の汎用性のもつ自由度を有効化できる手法を用いた客観的な自動評価手法が必要であると考えられる．

一方で，人間の言語行動に関しては他分野においても研究が続けられている．エージェントのとる行動が信念 (Belief)，願望 (Desire)，そして意図 (Intention) によって裏付けられているものであることを示す研究 [25] が BDI モデルとしてエージェントの設計に応用される例があり [26]，これを採り入れた使用者モデルおよびシステム主導発話生成手法の再検討が必要と考えられる．

2.3.2 でも述べたように，提案手法の対応可能なドメインを規定しているキー

ワードは現時点では実装されたタスク指向型対話システムに関連付けられた所与のものである．ユーザの公開プロフィール情報やライフログ・対話履歴からの自動獲得を行うことで，選択可能なドメインの領域が飛躍的に広がると考えられる．

また，本章ではドメイン選択を行う際の条件遷移に必要な閾値となる WordNet 意味類似度ベクトルのノルム $T = \|\boldsymbol{a}\|$ を固定して実験を行っている．対話の文脈に応じて，閾値を動的に変化させる手法などの検討が必要であると考えられる．

参考文献

[1] Thomas B. Sheridan. Eight Ultimate Challenges of Human-robot Communication. In *Proceedings of 6th IEEE International Workshop on Robot and Human Communication (RO-MAN'97)*, pp. 9–14, 1997.

[2] Terrence Fong, Charles Thorpe, and Charles Baur. Collaboration, Dialogue, Human-Robot Interaction. In *Robotics Research*, pp. 255–266. Springer, 2003.

[3] 荒木健治. 自然言語処理ことはじめ—言葉を覚え会話のできるコンピュータ. 森北出版, June 2004.

[4] Teruhisa Misu, Komei Sugiura, Tatsuya Kawahara, Kiyonori Ohtake, Chiori Hori, Hideki Kashioka, Hisashi Kawai, and Satoshi Nakamura. Modeling Spoken Decision Making Dialogue and Optimization of its Dialogue Strategy. In *Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pp. 221–224, 2010.

[5] Esther Levin, Roberto Pieraccini, and Wieland Eckert. A stochastic model of human-machine interaction for learning dialog strategies. *IEEE Transactions on speech and audio processing*, Vol. 8, No. 1, pp. 11–23, 2000.

[6] Jason D Williams and Steve Young. Partially Observable Markov Decision Processes for Spoken Dialog Systems. *Computer Speech and Language*, Vol. 21, No. 2, pp. 393–422, 2007.

[7] Joseph Weizenbaum. ELIZA—A Computer Program for the Study of Natural Language Communication Between Man and Machine. *Commun. ACM*, Vol. 9, pp. 36–45, January 1966.

[8] Richard Wallace. The elements of AIML style. *Alice AI Foundation*, 2003.

[9] Hugh Loebner. Loebner Prize. *described at: http://www. loebner. net/Prizef/loebner-prize. html. Breazeal, C*, 2001.

[10] 河原達也. 音声対話システムの進化と淘汰 (第 3 回対話システムシンポジウム). 言語・音声理解と対話処理研究会, Vol. 67, pp. 7–12, 2013.

[11] 南泰浩, 東中竜一郎, 堂坂浩二, 目黒豊美, 森啓, 前田英作. 対話行為タイプ列 Trigram による行動予測確率に基づく POMDP 対話制御. 電子情報通信学会論文誌 A, Vol. 95, No. 1, pp. 2–15, 2012.

[12] 木村泰知, 荒木健治, 桃内佳雄, 栃内香次. 遺伝的アルゴリズムを用いた帰納的学習による音声対話処理手法. 電子情報通信学会論文誌 D, Vol. 84, pp. 2079–2091, 2001.

[13] Arnaud Jordan and Kenji Araki. Evaluation of a Standalone Language-independent Dialogue Framework. In *Proceedings on the International Conference on Artificial Intelligence (ICAI)*, p. 316. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2015.

[14] Arnaud Jordan and Kenji Araki. Real-time Language-Independent Algorithm for Dialogue Agents. 知能と情報（日本知能情報ファジィ学会誌）, Vol. 28, No. 1, pp. 535–555, 2016.

[15] Motoki Yatsu, Rafal Rzepka, and Kenji Araki. A Modules-Based, Task-Navigational Dialogue System. In *Proceedings of PACLING 2011*, 2011.

[16] 平井尚樹, 稲葉通将, 鳥海不二夫, 石井健一郎. 対話エージェントのための文脈を考慮した統計的発話選択. 合同エージェントワークショップ&シンポジウム (JAWS2010) 論文集, 2010.

[17] 今井健太, ジェプカラファウ, 荒木健治. 複数の対話システムからの応答候補文を用いた最適応答文選択手法の性能評価. 情報処理学会研究報告 自然言語処理 (NL) 2010-NL-195(10), pp. 1–7, October 2010.

[18] 亀山恵祐, 荒木健治, 木村泰知. 質問応答システムにおける知識源選択規則の自動獲得の有効性について. 情報処理学会研究報告 自然言語処理 2007-NL-178(35), pp. 85–90, March 2007.

[19] 高橋瑞希, ジェプカラファウ, 荒木健治. Web 検索と単語 n-gram モデルを用いた文生成手法の性能評価. 言語処理学会第 16 回年次大会発表論文集, pp. 391–394, March 2010.

[20] Shinsuke Higuchi, Rafal Rzepka, and Kenji Araki. A Casual Conversation System Using Modality and Word Associations Retrieved from the Web. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pp. 382–390, October 2008.

[21] Francis Bond, Hitoshi Isahara, Sanae Fujita, Kiyotaka Uchimoto, Takayuki Kuribayashi, and Kyoko Kanzaki. Enhancing the Japanese WordNet. In *Proceedings of the 7th Workshop on Asian Language Resources, ACL-IJCNLP*, pp. 1–8, August 2009.

[22] Claudia Leacock, George A Miller, and Martin Chodorow. Using corpus statistics and WordNet relations for sense identification. *Computational Linguistics*, Vol. 24, No. 1, pp. 147–165, 1998.

[23] Charles E. Osgood, George Suci, and Percy Tannenbaum. *The Measurement of Meaning*. University of Illinois Press, 1957.

[24] 井上正明, 小林利宣. 日本における SD 法による研究分野とその形容詞対尺度構成の概観. 教育心理学研究, Vol. 33, No. 3, pp. 253–260, 1985.

[25] Michael E. Bratman. *Intention, Plans, and Practical Eeason*. Cam-

bridge University Press, 1999.

[26] Tibor Bosse, Zulfiqar A. Memon, and Jan Treur. A Two-Level BDI-Agent Model for Theory of Mind and its Use in Social Manipulation. In *Proc. of the AISB 2007 Workshop on Mindful Environments*, 2007.

第 3 章

駄洒落ユーモアの検出

3.1 はじめに

近年，機械がユーモアを表出・理解することに対する関心が高まっている．ユーモアへの関心は，機械によるユーモアがそれに接するユーザの生活の質 (QOL) を向上させることに対する関心，そして人工知能技術全体の発展への貢献に対するものに大きく分けられる．

ユーモアを使用する人の QOL 向上を実証する例として，社会心理学において，ユーモアを交えることによるアサーティブネス (自己表現・意見表明) の向上が報告されている [1]．また，日常的にユーモアを使用する機会の増加による健康増進の効果も期待できる．臨床心理学においては，日常生活におけるユーモアの表出によって，表出者自身の心理的ストレスが軽減する [2]，抑うつ症状が改善する [3] といった報告がある．

このような効果を，人工的なエージェント及びロボットに対するコミュニケーションにおいて得るための研究が進められている．例えば Dybala ら [4] は対話エージェントがユーモアを発する条件下では，ユーザがシステムに対し意外性を感じ，システムとの対話を継続する欲求が高められることを示した．宮澤ら [5] は，対話システムとユーザの間の対話の継続性を高める要因の調査を行っ

た．その結果，ユーモア性のある所作が継続性を高めることが示された．

これまでに，人工知能システムがユーモアを発する状況下でのユーモア生成の有効性に対しては多くの研究成果がある [6–8]．しかし，システムがユーモアを識別する手法，その意味の曖昧性を解消する手法，識別・理解の結果を利用した応答を行う手法およびそれらの及ぼす効果についての知見は未だ十分に得られているとはいえない．

ユーザはシステムに対し，入力した発話の意味が文字通りに理解され望ましい応答が得られないという不満をもつことがある．システムがユーモアを理解して適切な応答をすることにより，その不満を軽減できると考えられる．システムがユーモアを理解した (と，ユーザが考える) 上での動作を行うまでの過程は，次の 3 つの段階に分解できる．

A. ユーザによる発話をユーモアとして識別する．
B. ユーモアとして識別された場合，内容に対してスコア付けを行う (面白さ等の評価)．
C. 識別したユーモアに対し，面白さ等の評価結果に基づいた適切な応答または反応動作を行う．

ユーモアの理解を確実なものとするためには，上記の過程における **A.** の段階を高い精度で実現しなければならない．

そこで，本章の研究では韻文ユーモアである駄洒落の検出手法に着目する．駄洒落においては，**3.2.1** で述べるように一部の音素が異なるが音韻的に類似性の高い 2 区間を持つものが多くを占めていることが，**3.2.3** で述べる調査によって判明している．そこで，音韻の類似に着目した 3 つの駄洒落検出ルールを実装し (詳細は **3.5** で述べる)，これらに学習データ文の語彙素性を加えた統計的機械学習手法による分類器を用いて検出を行う．本章の研究においては，**3.5.2** で述べる Support Vector Machine (SVM) を用いる．SVM は，高い分類性能から文書分類において広く用いられる教師有り機械学習手法であり，線形分類に加え同条件下でのカーネル法を用いた非線形分類を行うことが可能である．

本章の研究の目的は，駄洒落を含む文を一般的なテキストより SVM を用いて検出する際に有効となる素性及び高精度な検出を妨げる基礎的な問題点を明らかにすることである．素性に関しては，素性単独の有効性に加えその組合せの有効性を評価する必要がある．従って，本章の研究において検証する仮説を以下に示す．

(1) 駄洒落の検出に対し，分類器に与える素性として語彙素性の適性がある．
(2) 読み仮名の完全一致，音韻類似度による類似部分検出または促音・撥音・長音の付加/脱落を考慮した音韻的類似部分検出に基づくルールベース素性の付加により，検出性能が向上する．
(3) 仮説 **(1)**, **(2)** で述べた全ての素性を用いた場合，検出性能が最大化する．

本章の構成を述べる．初めに **3.2** で検出対象となる駄洒落の類型及び基本構造について述べる．続いて **3.3** で，既存の言語的ユーモアの検出・理解のための手法の本章の研究との差異を述べる． **3.4** では本章の研究の分析に用いたデータの取得方法について述べる．**3.5** では提案する駄洒落検出手法を，音韻類似度の計算手法とともに概説する．**3.6** では提案手法の検出性能の評価実験について述べる．**3.7** において，得られた結果に関して考察し，前述の仮説 **(1)** **(3)** の検証を行う．

なお，「駄洒落」という表現の名称が指す表現の範囲には，文献による異なりがある．よって，従来の研究を把握するにあたり，駄洒落という表現の定義を確認する必要がある．そのため，先に **3.2** において対象とする駄洒落の類型を概観し，続いて **3.3** において従来の研究を検討する．

3.2　対象とする駄洒落の類型

本節では，駄洒落の類型を概観し，中でも，併置型駄洒落を主な検出の対象とする理由について述べる．

駄洒落は，言葉遊びにも分類される言語表現であり，日本語の短文ユーモアの

代表例である．駄洒落にはいくつかの類型があるが，最も基本的とされるものは，音韻的に類似するものの意味の乖離が認められる 2 つの音素列の現れた表現 (併置型駄洒落) である．**3.2.3** での調査が示すように，併置型駄洒落は駄洒落全体の数に占める割合が大きい．

併置型駄洒落は文内に 2 つの音韻的に類似した区間 (種表現及び変形表現) をもつ．併置型駄洒落において，種表現は文内の独立形態素 (自立語となる名詞・動詞・形容詞・感動詞) あるいはその句 (名詞句・動詞句・形容詞句) を，変形表現は種表現との音韻的類似性を持つ文内の任意の区間の音素列を指す．

3.2.1　駄洒落の主な類型

駄洒落は，対となる 2 音素列の音韻的な類似に基づき生成される比喩表現と捉えられる．類似する対となる音素列が，対象文内に共起する場合と，しない場合とがあり，滝澤 [9] は前者を併置型駄洒落，後者（例: 図 3.1 における (vi)）を重畳型駄洒落と定義している．併置型駄洒落における種表現及び変形表現の現れ方については **3.2.2** で述べる．理想的な例においては，音韻的類似性によるルールのみを用いて可能となる．

Kawahara [10] はこの併置型駄洒落において，種表現と変形表現の音韻が完全に一致するものを perfect puns，語音の一部の変化または脱落により一致しないものを imperfect puns と呼称している．本章の研究においては，語音の変化に柔軟な分類器を音韻類似度に基づいて構築するため，perfect/imperfect 双方に属する駄洒落を本章の研究における提案手法の検出の対象とする．これに加え，音韻類似性によるルールにより検出することが困難な重畳型駄洒落を検出の対象に含める．

3.2.2　併置型駄洒落における種表現・変形表現

本項では，併置型駄洒落のもつ構造を説明し，これを検出の対象とする理由を述べる．

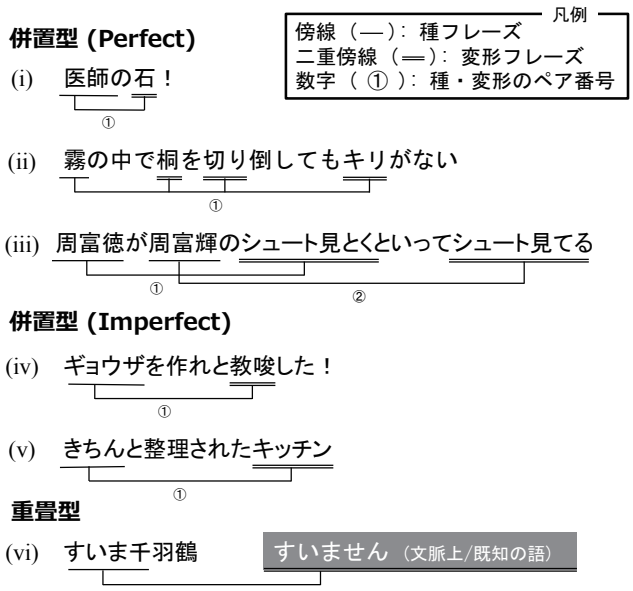

図 3.1　併置型駄洒落における種表現と変形表現の例

1 つの併置型駄洒落において，種表現および変形表現の対が少なくとも 1 対出現する．図 3.1 の文 (i) および (ii) はその例である．ただし，文 (ii) に表れるように，1 つの種表現に対し複数の変形表現が対応する場合がある．種・変形表現がいずれも交換可能な場合，便宜上先行する表現を種表現とし，その他をこれに対応する変形表現とする．

また，一文において複数の種・変形表現の対が観察される場合もある．この例を図 3.1 の文 (iii) に示す．文 (iii) は，種表現に当たるフレーズとして「周富徳」および「周富輝」という人名に対し，それぞれに対応する変形表現は「シュート見とく」および「シュート見てる」である．

3.2.3　各類型の構成比

3.2.1 で述べた 2 種の類型のそれぞれが，駄洒落の標本中に占める割合を調査した．調査者は 1 名 (著者) である．標本は，**3.4** にて述べる分析対象として取得した全駄洒落文 45,970 文より，600 文を無作為抽出した．ここで得られた駄洒落の類型別の構成を表 3.1 に示す．この結果より，併置型駄洒落が全体において高い割合を占めること，特に音韻的に類似した 2 つの区間を持つ imperfect な駄洒落の比率が大きいことがわかる．なお，抽出した駄洒落文数は次のように決定した．信頼区間 95%($\lambda = 1.96$) において，母比率 $P = 0.5$ に対する標本内誤差 d の上限を 5% とすると，標本の大きさ n の下限は式 (3.1) より求められる [11].

$$n \geq \frac{N}{(\frac{d}{\lambda}^2 \frac{N-1}{P(1-P)}) + 1} = 380.98.. \tag{3.1}$$

ここでは，式 (3.1) において求められた標本の大きさの下限及び調査の所要時間の制限を考慮して標本の大きさを $n = 600$ とした．

3.2.4　併置型駄洒落を検出の対象とする理由

併置型駄洒落を検出の対象とする理由は，第一に併置型駄洒落の全駄洒落に占める割合が大きいためである．このことは， **3.2.3** で述べた類型の構成比の調査において，併置型駄洒落の割合が 93.8% であることから確認できる．

表 3.1　駄洒落の類型とその構成比

併置型	563 (全体 93.8%)
その他の類型	33 (全体 5.50%)
取得失敗	4 (全体 0.667%)
imperfect な併置型駄洒落	521 (全体 86.8%, 併置型 92.5%)

表 3.2　駄洒落を含む文の取得元の Web サイト

サイト名 (URL)	取得件数
ダジャレナビ (http://www.dajarenavi.net/pc/i_today_index.htm	39,120
Dajare Station (http://dajare.jp	8,795
ダジャレネット (http://www.dajare.net	1,621
ひとくちダジャレ大集合 (http://www.biwa.ne.jp/~aki-ina/gyagu.html	1,067
ダジャレ集 ダジャレ事典 (http://dajareshuu.web.fc2.com	982
ダジャレの缶詰 (http://www.geocities.jp/pikumin_hiroba/dajare.html	572
駄洒落倶楽部 (http://with2.net/dajakura	428
ダジャレ広場 (http://www1.ocn.ne.jp/~origo/dazyare	303
駄洒落を言ったのは誰じゃ? (http://wtpage.info/dajare	107

第二に，併置型駄洒落でない駄洒落，すなわち重畳型駄洒落の検出は次の理由で一般に困難である．まず，重畳型駄洒落は，種表現が文内または文外の暗黙の文脈中にあると見なしうる駄洒落であるため，種表現が何であるかを変形表現及び文脈より推定する必要がある．

(vii) ブドウを食べて驚いた.「きょほぉぉぉ〜〜〜!」

例えば上記の文 (vii) では下線部「きょほぉぉぉ　!」が変形表現に当たるが，その種表現となると考えられる形態素の「巨峰」は文内には出現しない.「巨峰」との音韻的類似に基づく一致を求めるには，処理の前に，先頭の形態素「ブドウ」の関連語として同形態素を推定する必要がある．このような文脈情報の付与の必要から，重畳型駄洒落の検出のコストは他の類型に比べて大きい.

一方，併置型駄洒落は，種表現及び変形表現の両方が文内に明示されているため，推定の必要がなく検出に要するコストが小さい.

以上の考察から，本章の研究における提案手法は，併置型駄洒落の検出を行うことにより駄洒落を検出するものとする．ただし，同手法の検出対象は併置型駄洒落のみに限らず，重畳型駄洒落を含む他類型の駄洒落をも検出対象とする．また，併置型の下位類型の perfect/imperfect は，双方を検出の対象とする.

3.3　関連研究

ユーモア・笑いというテーマ全体に対しては，人工知能以外の分野でも研究が活発化してきている [12]．笑いの検出・測定という観点では，ビデオカメラによる入力映像からの表情及び音声入力から笑いの検出手法が，米国ではすでに特許出願されている [13].

駄洒落文の意味理解を目的とするシステムは，これまでに Yokogawa [14] により提案されている．しかし，実験において有効性評価に用いられた駄洒落は 100 項目に限られ，より多数のデータに対する性能が得られていない．本章の研究においては，これまでに作成されたものより規模の大きい口語文データより駄洒落を検出する手法を提案する．また，Yokogawa の手法が対象とするのは，**3.2.1** で述べた類型のうち imperfect な駄洒落のみであり，同音異義語を利用した perfect な駄洒落の検出は不可能であるのに対し，本章の研究の提案手法は perfect/imperfect に関わらず併置型駄洒落全体を検出の対象にしている．

天谷ら [15] は，単語 bag-of-words を基本とする表層情報を手がかりとした散文ユーモアの SVM 及び Naive-bayes 法を複合的に用いた識別器を構築し，Twitter [*1] 等より取得したユーモア/ノンユーモア文集合の分類実験において有効性を示した．本章の研究の提案手法は，語彙素性に加え，韻文のもつ音韻的特徴を考慮した音韻類似度に基づく素性を使用する．

また，システムがユーモアを理解するために用いる基礎的なモデルとして，ファジィ理論に基づくおかしみの誘発モデル [16] が考案されている．現時点において同モデルは，山根らによる駄洒落に類した短文ユーモア (ことわざすかし) の生成手法 [17] におけるランキング関数として使用されている．しかし，同手法においては対象文が駄洒落であることを前提に面白さの評価を行うため，評価関数を駄洒落の認識に用いることを意図していない．

[*1] `https://twitter.com`

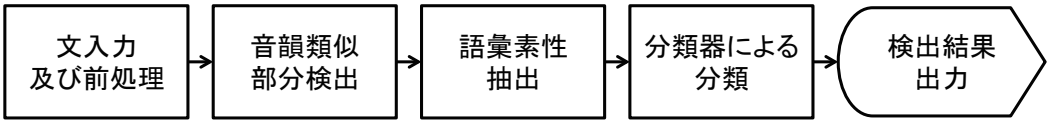

図 3.2　検出システムの処理の流れ

表 3.3　実験対象のデータセット

名称	正例文数	負例文数
開発データ	6,000	6,000
評価データ	23,954	23,954

上述のように，言語的なユーモアのうち，散文ユーモアの検出に関する研究はすでに行われているが，韻文すなわち言葉遊びとなる駄洒落などのユーモアに対する，オープンデータを対象とした，口語文等からの検出を行う手法に関する研究は，著者の知るところでは確認されていない．

3.4　分析対象のデータ

本章の研究においては，Web 資源に存在する駄洒落の集積された Web サイトより駄洒落を収集し，分類問題の対象の正例データとした．本来，駄洒落を分析するにあたり，他の言語現象と同様に，できるだけ現実世界における生起頻度や生起のための条件 (駄洒落を含む会話が発生する可能性の高いトピックなど) を反映した分析を行う必要がある．しかし，日常会話における駄洒落の使用に関しては，実コーパスやその統計的分析結果がこれまでに得られておらず，具体的な実用例においての検出を行うことが困難である．

従って，正例データの取得をスパイダーを用いて Web より行った．取得の際，重複文を削除した．データの取得元の Web サイト，取得した正例文数の合計及び重複除去後の正例文数を表 3.2 に示す．

駄洒落を含む文の検出を行うため，駄洒落を含まない文からなる負例データを準備した．負例データは 2011 年より 2012 年の間に Ameba[*2]において書かれたブログ記事を収集した YACIS コーパス [18] より以下で述べる手法を用いて抽出した． すなわち，コーパスに文として収録されている項目をプログラムを用いて無作為抽出を行った．

重複文を除いた全正例文より求めた文長の分布が正規分布に従うかについて，Shapiro–Wilk 検定 [19] を行ったところ，正規性が確認されなかった．従って，文長に操作を加えることにより，その分布に偏りの生じる可能性が高いため，取得する負例データに対する文長の制限は設けなかった．

なお，同コーパスの収録文より無作為に抽出した 2,000 文中，駄洒落を含む文が 1 文 (0.05%) 存在した．このことから，混入する正例の数は非常に少ないため，上記の手法を用いることでブログコーパスより駄洒落を含まない文集合の収集が可能である．

取得した駄洒落ならびにブログ内短文から，開発及び評価用のデータセットを構築した．各データの規模は表 3.3 に示す通りである．開発データは，**3.6** に述べる検出性能評価実験において，**3.5.1** 及び **3.5.2** で述べる SVM を用いた分類器における各カーネル関数内のパラメタ，素性値の計算時に用いられる閾値の決定，及び子音音韻類似度の算出式 (4),(5) を用いた子音ペアデータからの新類似度の事前の算出に使用した．評価データは，**3.6** の検出性能評価実験において分類器に与える学習データ及びテストデータとして用いた．

3.5 提案手法

提案手法のシステムの概要を述べる．システムは入力したテキスト (以下，入力文) を，駄洒落を含む文または含まない文の 2 クラスに分類する．システムの処理の流れを図 3.2 に示す． 前処理においては， **3.5.2** で述べる音韻類似度の

[*2] `http://ameblo.jp`

計算に適するように入力文を整形するため，入力文より長音符 (「ー」) 以外の記号を除去する．

3.5.1　分類器の構築，素性の選択

提案手法における分類器に教師有り機械学習手法である Support Vector Machines (SVM) [20] を用いる． SVM は，入力をベクトルとする 2 クラスの教師有り機械学習によるモデルを利用した分類手法である．SVM の実装は scikit-learn [21] を用いた．

入力ベクトルを $\boldsymbol{x}$, 重みベクトルを $\boldsymbol{w}$, 切片を b とすると，分離超平面を表す関数は

$$f(\boldsymbol{x}) = \boldsymbol{w} \cdot \boldsymbol{x} - b \tag{3.2}$$

式 (3.2) で表される分離超平面を，学習データの存在する空間において最適な (マージンが最大化された) 条件で $\boldsymbol{w}$ および b を決定することにより，同空間に入力される評価データを識別する．SVM では，式 (2) の $\boldsymbol{w} \cdot \boldsymbol{x}$ を置換するカーネル関数の定義によって，線形分離が困難なデータに対しても，より高次元の空間への写像による非線形の分類が可能になる．分離超平面はパラメタ α_i を伴って式 (3.3) の形になる．

$$f(\boldsymbol{x}) = \sum_{i=1}^{n} \alpha_i k(\boldsymbol{x}, \boldsymbol{x}_i) - b \tag{3.3}$$

分類器に対し学習及び分類を行う際，入力ベクトル には各素性による素性空間に対応する値を代入する．提案手法において用いることの可能な素性は，以下に揚げる語彙素性 (bag-of-words) 及び 3 つのルールベース素性である．以下の素性を組み合わせた分類器を構築することにより，**3.1** で述べた仮説 **(1)**,**(2)** を検証することができる．仮説 **(3)** に関しては下記の全ての素性を用いることにより検証が可能である．

(a): **語彙素性 (→ 仮説 (1) に対応)**

内容語 (名詞，固有名詞，一般動詞，形容詞，副詞, 感動詞) の語幹を形態素解析器 MeCab[*3]を用いて抽出することにより作成される bag-of-words 素性である．形態素辞書には IPAdic を用いる．

Bag-of-words 素性の値の決定は次のようにして行う．事前に学習データの全ての入力文より取得した bag-of-words と素性番号を対応付ける．学習時及び検出時に，システムへの入力文から同様に bag-of-words を取得し．事前に指定された素性番号に対応する内容語が得られた素性の値を 1 とする．

(b): **読み仮名列完全一致素性 (→ 仮説 (2) に対応)**

MeCab ならびに補助形態素解析器として，形態素解析器 JUMAN[*4]を用いて読み仮名列を抽出し，種表現 (独立形態素) と読み仮名が完全に一致する区間を変形表現として探索する．ただし 1 文字の読み仮名に対する一致区間が多数となることによる誤検出を防ぐため，対象の読み仮名は 2 文字以上とする．検出に成功した場合，素性値を 1 とする．

なお，他のルールベース素性 (c),(d) と異なり，素性 (b) は閾値等のパラメタを持たない.

(c): **子音の音韻類似度に基づく一致素性 (同上)**

種・変形表現の語音の一部が異なる imperfect な併置型駄洒落が対象駄洒落の 85 パーセント超を占めることから (**3.2.3**)，音韻の類似部分の検出においては語音の異なりの吸収が必要と考えられる．従って，**3.5.2** で述べる子音の音韻類似度を用いた素性を構築する．

具体的には，素性 (b) において抽出した読み仮名列に次の処理を加える．最初に拗音 (ゃ，ゅ，ょ，ゎ) が単独で一致することを避けるために，読み仮名をモーラ単位に分割し配列化する．このとき長音を二重母音に変換し，促音及び撥音のモーラに独自の記号を与える．モーラ内の子音音素と母音

[*3] http://taku910.github.io/mecab/

[*4] http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN

音素を分離して配列内に格納する．例えば「インターショップ」は {/*,i/, /*,N/,/t,a/,/*,a/,/sh,o/,/*,Q/,/p,u/} という配列に変換される．ここで，N は撥音，Q は促音を指す．その他の子音部の音素の表記はヘボン式ローマ字による音声表記と同一である．

各モーラを子音部と母音部に分離し，各形態素に対する任意の区間を探索する際に，子音部の類似度を **3.5.2** に掲げる手法で算出する．この類似度の合計を求め，閾値以上であれば素性値を正の値とする．

(d): **促音・撥音の加減を認める類似度一致素性 (同上)**

種表現内部へ促音・撥音の挿入及び脱落を許容する条件で，子音の音韻類似度に基づき素性値を決定する．例えば，駄洒落「ふとんがふっとんだ!」における種表現「ふとん」に対し，変形表現「ふっとん」において促音「っ」が挿入されている．本素性の値を決定するには，種表現あるいは変形表現における促音・撥音の追加を検出し，追加された促音・撥音を削減した状態で素性 (c) と同一の手法を用いて子音類似度スコアの算出を行う．

このようにして，これらの音素の種/変形表現間の加減が許容される．素性 (c) と同様に，類似度の合計から素性値を決定する．

3.5.2　音韻類似度の計算

システムは解析された各形態素の音韻モーラ列に対して，全体のモーラ列との音韻類似度を算出する．このとき，同じ子音については算出の対象としない．異なる子音のペアにおいて音韻類似度の高いものは駄洒落である可能性がより高く，類似度が低く母音のみが一致する子音ペアは，駄洒落ではない表現に偶然表れる可能性が高いと考えられるためである．

音韻類似度の算出手法として，駄洒落コーパスより算出した子音間の音韻類似性の指標による対数 O/E 比の概念を用いる．対数 O/E 比は，Kawahara らによる子音の心理音響的類似性の指摘 [22] の他，広く音韻論の分野で 2 音素の

$$\text{精度} = \frac{\text{検出成功した駄洒落を含む文の数}}{\text{駄洒落を含む文とした判定の総数}} \tag{3.4}$$

$$\text{再現率} = \frac{\text{検出成功した駄洒落を含む文の数}}{\text{駄洒落を含む文の総数}} \tag{3.5}$$

$$F\text{ 値} = \frac{2 \times \text{精度} \times \text{再現率}}{\text{精度} + \text{再現率}} \tag{3.6}$$

$$\text{正解率} = \frac{\text{正しく判定された文の総数}}{\text{対象データの文の総数}} \tag{3.7}$$

図 3.3　検出性能評価に用いる尺度

共起の度合の指標として用いられる．

異なる子音 p, q に対する対数 O/E 比 $OER(p,q)$ 及びそれを用いた種表現 t を含む文 S における音韻類似度 $score$ は，ある対象コーパスについて，それぞれ式 (3.8), (3.9) により定義される．

$$OER(p_1,p_2) = \log \frac{N_{p_1,p_2} + N_{p_2,p_1}}{N_{\text{pair}} \cdot P(p_1) \cdot P(p_2)} + c \tag{3.8}$$

$$score = \sum_{i=1}^{|t|} OER(p_i,q_i) \tag{3.9}$$

ただし， p, q はそれぞれ種表現，変形表現に表れる子音を，$N_{p,q}$ は p, q の順に表れる子音の対 (p,q) の対象コーパスにおける生起頻度を指す．$OER(\cdot)$ の計算において p,q は交換可能である．式 (3.9) の $|t|$ は種表現形態素 t の持つ音素数であり，変形表現候補内の音素数に等しい．N_{pair} は，対象コーパスに存在する子音のペアの総数を表す．$P(p)$ は N_{pair} 個の子音ペアにおける子音 p が含まれるものの割合である．対数 O/E 比は，N_{pair} 個の全子音ペアについて、式 (3.8) を用いて事前に算出したものを用いる．式 (3.8) における c は定数であり，同じく事前の実験により定める．

p, q が同一の子音である場合は，類似度計算の対象としない．その目的

は，$p = q$ である場合に式 (3.9) の $score$ が不当に大きくなることを回避し，imperfect な併置型駄洒落の検出率を相対的に高めることである．

式 (3.9) を用いて種表現の形態素と変形表現候補となる全区間の音韻類似度を計算し，$score$ が閾値を上回る場合，素性 (b) ないし (c) は素性値 1 をとる．

例として，次の文 (vii) における種表現となる形態素「**人魚**」と変形表現となる区間「**人情**」の間の音韻類似度を計算する．

(viii) 人魚の人情劇！
{/n,i/, /*,N/, /gy,o/, /n,o/, /n,i/, /*,N/, /j,o/, /*,u/, /g,e/, /k,i/}

初めに，子音間類似度を異なる子音に対して計算する．ここでは子音のペア [gy-j] の対数 O/E 比 $OER(/\mathrm{gy}/, /\mathrm{j}/)$ を利用する．**3.4** で構築した開発データを用いた場合，$OER(/\mathrm{gy}/, /\mathrm{j}/) = 0.1$ となり，$score = 0.1$ となる．

1 つの種表現候補に対し複数の変形表現候補が存在する場合は，最も $score$ の高いものではなく，全ての変形表現候補を等価なものとして取り扱う．また，複数の変形表現候補が最も高くなる場合も，それに従って全ての変形表現候補を変形表現として検出する．

種・変形表現検出の条件においては，種・変形表現の 1 文内の検出数や，閾値を超えた $score$ の大きさは考慮に入れないものとする．ただし，これらの指標は駄洒落の意味理解や面白さの自動評価において不可欠な表層的情報を与えると考えられる．

3.5.3　学習データおよびテストデータの規模

3.6 の実験において評価データ (全 47,908 項目) を用いて 20 分割交差検定を行った場合，教師データが 45,505 文，テストデータ) が 2,395 文となる．SVM の分類器に入力するデータ量の規模については，以下に挙げる他研究の例と比較して 2 オーダー以上の差は見られないため，適切なものと考えられる．

文献 [23] は SVM を用いた分類問題の研究の例である．この研究は，59 種類

の日本語機能表現と内容表現の分類を目的としたものであり，訓練データ 3,083 項目を用いてテストデータ 7,703 項目を分類している．また，文献 [24] の地方議会会議録からの要望表現の抽出タスクにおいては，オープンデータ 5,388 項目に対して 10 分割交差検定を実施し，提案手法の有効性を確認している．

ただし，これらの研究は分類対象が異なる点に留意する必要がある．最適なデータ規模の追加調査が必要であるが，本章の研究においては今後の課題とする．

3.6　検出性能評価実験

提案手法を実装したシステムのオープンなデータに対する検出性能の評価を行った．具体的には，まず同一の素性の組合せ (素性セット) に対し 20 分割交差検定による各カーネル関数の検出性能の比較を行った (**3.6.1**)．次に，負例データに対する正例データの比率を変化させ，正例数の減少に対する各素性セットの示す頑健性の調査を行った (**3.6.2**)．頑健性とは，負例データに対する正例データ量の比率が低くなる条件において検出性能の低下を抑えられる度合いを指す．最後に，語彙素性の有効性を分析するため，素性 (a) と素性 (b)(c)(d) の間の比較を行った (**3.6.3**)．

本節において行った実験における性能を表す尺度には図 3.3 における式 (3.4)〜(3.7) に示すものを用い，性能評価の結果値は交差検定の全結果の平均値とした．

本実験で使用した素性セットの一覧を表 3.4 に示す．

3.5.1 において述べた各カーネル関数におけるパラメタ及び素性 (c), (d) の値に影響する閾値は，**3.4** で述べた開発データ計 12,000 文に対し予備検出実験を行い，最も高い検出性能を示したものに決定した．用いた閾値を表 3.6 において示す．

また，**3.5.2** にて提案した子音の類似度を示す対数 O/E 比は，同開発データにおける正例文 6,000 文に著者による種・変形表現のアノテーションを行った

表 3.4　性能評価実験において用いた素性セット (FS)

FS	**(a)**	**(b)**	**(c)**	**(d)**	**FS**	**(a)**	**(b)**	**(c)**	**(d)**
F_1	–	–	–	○	F_9	○	–	–	○
F_2	–	–	○	–	F_{10}	○	–	○	–
F_3	–	–	○	○	F_{11}	○	–	○	○
F_4	–	○	–	–	F_{12}	○	○	–	–
F_5	–	○	–	○	F_{13}	○	○	–	○
F_6	–	○	○	–	F_{14}	○	○	○	–
F_7	–	○	○	○	F_{15}	○	○	○	○
F_8	○	–	–	–					

表 3.5　素性セット F_{15} に対する各カーネルの検出性能

素性セット	**精度**	**再現率**	**F 値**	**正解率**
線形カーネル				
F_{15}	0.851	0.865	0.858	0.857
2 次多項式カーネル				
F_{15}	0.903	0.894	0.898	0.901
RBF カーネル				
F_{15}	0.907	0.909	0.908	0.909

ものを用いて事前に算出した.

3.6.1　カーネル関数間の比較

線形カーネルを含むカーネル関数の検出性能を比較するため，素性セット F_{15} に対してカーネル毎の検出性能を求めた．学習データ及びテストデータにおける正例数と負例数の比は 1 : 1 である．この検出性能の比較結果を表 3.5 に示す.

3.6.2　正例比率の変化に対する各素性セットの頑健性の評価

雑談対話コーパスなどの実データにおいては，正例数は負例数に比べごく少量になると考えられる．そのような場合，正例数の減少に対して検出性能の低下を抑えられること (頑健性) が求められる．

各素性セットの頑健性を比較するため，予備実験において特に検出性能の高かった素性セット F_8,F_9,F_{10},F_{11},F_{12},F_{13},F_{14} を用い，評価データに対する 20 分割交差検定を，正例 : 負例 $= 0.01 : 1.0$ より $1.0 : 1.0$ となるまで正例文の構成比率を 0.01 ずつ増加させながら繰り返し行った．この結果における正例負例比率を横軸に，検出性能を示す F 値を縦軸にプロットしたものを図 3.4 に示す．計算資源の制約から，カーネル関数には学習が高速な線形カーネルを用いた．

3.6.3　語彙素性の有効性に対する分析

(a) 語彙素性は，ルールベース素性 (b),(c) 及び (d) に付加されることによって，文法・音韻上の検出ルールのみでは捉えられない語彙上の傾向を検出に反映することを可能にすると考えられる．これを実証するため，素性 (b)(c)(d) において用いられているルールの単体での検出能力と素性セット F_{15} ともに SVM を用いた分類器との検出性能の比較を行った．結果を表 3.6 に示す．なお，表 3.6 において閾値が「-」で示される場合，当該素性が閾値のパラメタを持たないことにより表記を省略している．

3.7　考察

本節では，性能評価実験の結果より考察される点について述べ，改善方法を検討する．なお，本節で平均性能値の有意差について述べている場合は，指定のない限り，20 分割交差検定で得られた平均値について，Student の t 分布を用いた片側検定を行ったものとする．有意水準は指定のない限り 1% ($p < 0.01$) で

表 3.6　各ルールベース素性単体による検出性能．*1 は表上段の各ルール素性適用時の閾値を組み合わせて用いたことを示す．下線は表上下段それぞれにおける最高値，**太字**は全体での最高値である

ルール (単体)/手法	閾値	精度	再現率	F 値	正解率
(b) 完全一致	-	0.516	0.696	0.593	0.521
(c) 子音類似度	0.28	0.589	0.612	0.600	0.592
(d) 子音類似度 + 促・撥音	0.99	0.610	0.674	0.640	0.621
線形カーネル SVM (F_7)	*1	0.645	0.785	0.708	0.677
線形カーネル SVM (F_8)	-	0.788	0.860	0.821	0.814
線形カーネル SVM (F_{15})	*1	**0.851**	**0.865**	**0.858**	**0.857**

ある．また，素性セット F_n の検出性能が F_m を上回る形での両者の有意差が存在しないことを帰無仮説とした検定を行う場合に，$\mathsf{F}_m \rightarrow \mathsf{F}_n$ と表記する．

3.7.1　提案手法の有効性

提案手法の有効性について述べる．提案手法は，ルールベースによる分類結果を素性として用い，語彙素性ともに SVM による学習に用いることにより，検出性能の向上を図るものである．有効性の評価に関して，下記の I 及び II の知見が得られた．

I. RBF カーネル使用時の検出性能の最大化

各カーネル間の性能を比較した **3.6.1** の実験では，RBF カーネルが最も高い性能を示し，次いで 2 次多項式カーネル・線形カーネルの順となった．ただし，RBF カーネル及び 2 次多項式カーネルは学習に多大な時間を要する（**3.6.1** において行った実験の場合，交差検定における学習時間はそれぞれ 3,750 秒/分割及び 2,420 秒/分割．*cf.* 線形カーネルの場合: 6.54 秒/分割）ことから，計算資源の単位所要量に対しての性能は線形カーネルの場合が最大となるといえる．

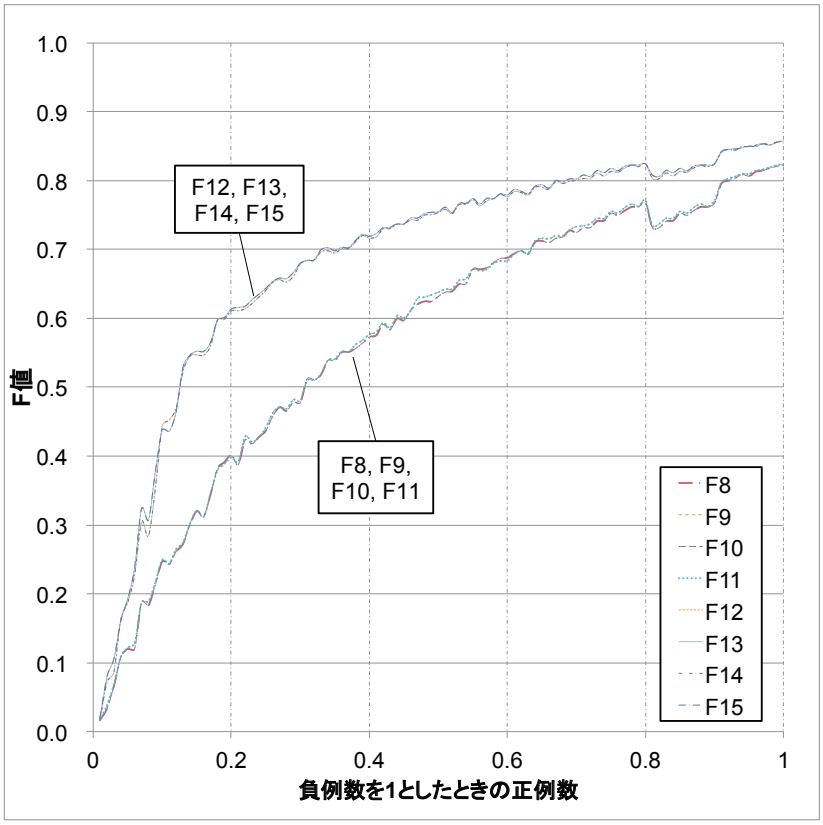

図 3.4　正例数/負例数比に対する線形カーネル SVM の F 値のプロット．(対象: 素性セット F_8〜F_{15}) F 値の差の小さい素性セットのグループをボックスで示した

II. ルールベース素性 (b) による性能向上

図 3.4 が示す通り，線形カーネルを使用した実験では，素性セット F_{12},F_{13},F_{14},F_{15} 間，及び F_8,F_9,F_{10},F_{11} 間に有意な差が認められなかった．F_8-F_{10}, F_9-F_{11}, F_{12}-F_{14}, F_{13}-F_{15} 間は (c) 促撥長音挿入を認めない子音類似度素性の有無による差が有意差の要因となるが，有意差は検出されなかった．(d) 促撥長音増減を認める音韻類似度素性 ($F_8 \rightarrow F_9$,$F_{10} \rightarrow F_{11}$,$F_{12} \rightarrow F_{13}$,$F_{14} \rightarrow F_{15}$

間) の場合も同様である．素性 (d) においては，$F_4 \rightarrow F_5$, $F_6 \rightarrow F_7$ の 2 つの組において有意差が確認されたが,．本実験においては素性 (c) 及び (d) の効果は顕著ではなかったことがわかる．

反対に素性の付加により有意な差が表れている素性は，(b) 完全一致素性 ($F_8 \rightarrow F_{12}$,$F_9 \rightarrow F_{13}$,$F_{10} \rightarrow F_{14}$,$F_{11} \rightarrow F_{15}$) である．このことから，素性 (c),(d) には改善の余地が残るものの，ルールベース素性 (b) の付加により，語彙素性 (a) のみの場合 (素性セット F_8) と比較し検出性能及び正例データ数の縮小に対する頑健性が向上することを確認できた．

3.7.2　ルールベース素性 (b),(c) 及び (d) の追加による効果

以下では，素性 (b), (c), (d) のそれぞれの追加により新たに検出が可能となった正例を挙げる．また，素性の追加により誤って分類された正例および負例についても考察する．分析対象は RBF カーネルを使用した場合の分類結果である．初めに，語彙素性 (a) に類似度素性 (b) を付加することによって取得が可能となった正例を挙げる．これらの例では，種表現の形態素における子音の音素と変形表現となった区間内の子音との異なりが見受けられる．下記の例においては，こうした音素の差異を吸収した検出に成功したと考えられる．なお，異なる子音のペアを例文右横に示す．

(ix) ユウコという子！ `[(y, *)]`
(x) ヨーグルト食べてから潜ると！ `[(y, m)]`
(xi) ハンペンを食べる時は、シャンペン！ `[(h, sh)]`

一方，語彙素性 (a) に完全一致素性 (c) を付加することによって取得に成功した正例を以下に挙げる．素性 (c) は，音素表現に分解する前での読み仮名表記に完全一致が見られる場合に正の値をとる．したがって素性 (c) では，表記は同一だが，音韻的には一致のない種表現と変形表現のペアも検出対象となる．

(xii) チーターはどうちーたーん？
(xiii) 相談してもいいそうだん！

また，類似度を用いる条件下で，促撥長音付加の場合でのみ取得に成功した例を挙げる．これらの例においては，種表現となる形態素に促音および撥音が付加されたものが変形表現の候補として検出されたと考えられる．

(xiv) ナタデココおいしくなったで、ここ！
(xv) 体重が前月比マイナスだよ。まぁ、いいなーっす！

3.6.3 の結果より，3 種のルールを素性とする F_7 を用いた SVM による検出性能は，各ルール単体による性能に対し有意差が確認されたことから，ルールベース素性を入力とした場合の機械学習を使用することの有効性を確認できる．

3.7.3　語彙素性 (a) 追加による効果

ルールベース素性のみのセット F_7 による検出に失敗した駄洒落文においては，明らかに (a) 語彙素性の作用により検出されたものも存在する．

(xvi) 一見さん、男割り！
(xvii) 君は男なんだから、もっと、お酢飲みなさい！

上記がその例である．(xvi) は **3.2.1** で掲げた重畳型駄洒落に分類され，(xvii) は併置型にも重畳型にも分類できない駄洒落文の例である．これらの検出はルールベースを用いた提案手法のシステムの想定を超えたものである．語彙素性の効力は，**3.6.3** に述べたルール単体及び SVM との比較において，SVM にルールベース素性のみを与える条件 (素性セット F_7) に比べ素性セット F_8 の条件での F 値が 0.113 ポイント高いことから明らかである．ただし，一般的に語彙素性の効力は学習データに依存する傾向があることに留意する必要がある．

表 3.7　重畳型駄洒落の検出性能

精度	再現率	F 値
1.00 (67/67)	0.475 (67/141)	0.644

3.7.4　重畳型駄洒落の検出性能

3.7.3 の例文 (xvi),(xvii) で挙げたように，併置型駄洒落に属さない駄洒落の検出は，主に (a) 語彙素性によって検出が可能となっていると考えられる．このような重畳型駄洒落の検出能力を確認するために，2.3 の構成調査において抽出した重畳型駄洒落 141 文のみのデータを準備し，同データに対する提案手法の検出性能の追加評価を行った．学習データとして，評価データより当該重畳型駄洒落と重複する文を除去した正負例各 39,885 文を用いた．また，分類器には線形カーネル及び素性セット F_{15} を使用した．本実験で得られた検出性能を表 3.7 に示す．

結果より，提案手法は重畳型駄洒落に対しても高い精度での検出を実現していることが確認できる．

3.7.5　駄洒落を含まない文の音韻的性質

駄洒落を検出するために正例・負例の分類性能を確認する実験を行うことにより，同時に駄洒落を含まない文 (負例) の音韻的な特性を考察することが可能である．その特性は，駄洒落を含む文における種・変形表現内の子音の性質と反対のものであると考えられる．すなわち，入力文から種・変形表現候補を検出した場合，子音ペアの音韻類似度の合計値が閾値に満たない文が，駄洒落を含まない文に多いと推測される．そこで，駄洒落を含まない文において検出された種・変形ペア候補内の子音ペアと，駄洒落を含む文における子音ペアの音韻類似度

表 3.8　Mann-Whitney の U 検定による音韻類似度スコアの比較

子音類似度の平均値		p 値
正例	負例	
3.07	2.31	< 0.01

の平均値に有意差があるということを仮説として，その検証を行った．

具体的には，**3.6.1** の実験において非駄洒落と正しく判定された文のうち，種・変形表現候補が検出されたものについて，子音ペアの平均類似度を求め，同じく検出に成功した駄洒落文の平均値との間に Mann-Whitney の片側 U 検定を適用し，2 つの平均値の間に 1% 水準または 5% 水準での有意差の存在の有無を確認した．この追加実験の結果を表 3.8 に示す．

結果から，前述の仮説通り，種・変形表現候補内の子音ペアの平均音韻類似度は，駄洒落を含む文と比較して含まない文のほうが有意に低いものであることが確認された．

3.7.6　まとめ

3.7.1 の l で確認した，線形カーネルが非線形分類の可能な手法に次ぐ検出性能を示したことの一つの要因として，語彙素性が冗長となっていることが考えられる．駄洒落 Web サイトに収録される駄洒落を含む文がユーザの自由投稿であり，内容語の重複のある文が多数確認されることから，学習データにおける語彙の重複がこの現象の一つの要因と考えられる．語彙の重複については，学習データにおいて語彙・概念の重複を認める限度についての検討が未だなされていないため，また本章の研究は一般的なテキストへの検出手法の適用を試みることを目的としているため，今後の課題とする．

さらに，ブログ記事以外の評価データを含む多様なデータセットの構築による，多角的な検出性能評価も必要である．ただし，新しいデータセットにどの程

度の語彙のばらつきが見られるかの推計の必要があるが，語彙の重複に対する検討と同様の理由より今後の課題となる．

加えて，ルールベース素性にも改善の余地がある．素性 (c) 及び (d) は子音間のデータに基づく共起頻度の比に基づくが，母音を用いた指標も提唱されている [25]．子音及び母音の音韻類似度を同時に評価する場合，両者の重みの調整手法，及び種・変形表現のアライメント手法を構築する必要がある．手法としては，子音類似度では母音共通部分を，母音類似度では子音共通部分を基点として種・変形表現の検出を行うことが挙げられる．この場合は共通性の確定しない音素列の存在も考慮しなければならず，多量の語彙からの推定が必要となる可能性がある．

以降では，**3.1** で述べた仮説の妥当性についての結論を検討する．

(1) 駄洒落の検出に対し，分類器に与える素性として語彙素性の適性がある．

この仮説については，**3.6.3** における実験結果より語彙素性の高い有効性が確認された．次に，

(2) 読み仮名の完全一致，音韻類似度による類似部分検出又は促音・撥音・長音の付加/脱落を考慮した音韻的類似部分検出に基づくルールベース素性の付加により，検出性能が向上する．

(3) 仮説 **(1)**, **(2)** で述べた全ての素性を用いた場合，検出性能が最大化する．

仮説 **(3)** については， 6.3 の結果において，素性セット $\mathsf{F}_8 \rightarrow \mathsf{F}_{15}$ の組に 0.037 ポイントの F 値の有意な差 ($p < 0.01$) が存在するため，支持される．

一方，**(2)** については，**3.6.2** の結果から，3 つのルールベース素性のうち，(b) の完全一致素性が相対的に強力である結果が示された．また，(c) ないし (d) の音韻類似度に基づく素性を付加した場合に性能が向上しなかった条件が存在する．例えば，$\mathsf{F}_8 \rightarrow \mathsf{F}_{10}$, $\mathsf{F}_9 \rightarrow \mathsf{F}_{11}$, $\mathsf{F}_{12} \rightarrow \mathsf{F}_{14}$ のそれぞれの素性セットの組では， F 値の有意差がみられなかった．一部の結果はこのように，(b),(c) 及び

(d) 各ルールに改善の余地のあることを示し，従って仮説 **(2)** は全ての条件では成立しないことが判明した．

3.8 おわりに

本章では，SVM による教師あり学習に基づいた駄洒落を含む文の検出において，語彙素性に加え，音韻類似度を用いた類似部分の検出及び促撥長音の挿入・脱落を考慮する手法を提案した．検出性能評価実験において，語彙素性の有効性，ルールベース素性の有効性及びそれぞれ検出成功事例を考察において確認し，駄洒落を含まない文の音韻的な性質に並んで提案手法の総合的な有効性が確認された．また，語彙素性の利用により，提案手法が併置型駄洒落に加え重畳型駄洒落に対する一定の検出性能を得ることが示された．ただし，結果の考察においてルールベース素性における改善の余地が指摘され，また語彙素性の学習データへの依存性も **3.7.3** での分析において示唆された．

今後の方針としては，**3.7** で挙げられた問題点の改善に加え，過去研究でも触れられている駄洒落の面白さ (適切性) を文の内容語から推定することのできる手法の開発が考えられる．有効性の評価方法として，**3.5** で述べた手法との比較が妥当である．そのためには，**3.2.1** で述べた重畳型・併置型，perfect・imperfect の駄洒落の類型の正確な分類が必要になる．これは多クラス分類になるため，新たに適切な素性群を設定し学習用のデータセットを準備する必要がある．

3.1 において，駄洒落文の意味理解の前段階として駄洒落を含む文の精度の高い検出が必要であることを述べた．駄洒落文の意味理解の実際に関しては，駄洒落文の理解を語義曖昧性解消と捉えた研究がある．英語を対象にした研究では Miller らの手法 [26]，日本語対象では前述の Yokogawa [14] の手法があるが，いずれも **3.2** で述べた駄洒落の定義における一部の類型にしか対応していない．具体的には，Yokogawa の手法は形態素境界や語音の変化を伴う imperfect な駄洒落に対象を限定している．本章の研究の手法を応用した意味理解手法を

perfect な併置型駄洒落にも適用すれば，自然言語意味理解の対象範囲がより拡大すると考えられる．

参考文献

[1] 山田浩平, 朝野聡, 物部博文. 対人葛藤場面での断り行動に対する自己効力感と社会的スキル及びアサーティブな態度，ユーモア対処との関わり. 学校保健研究, Vol. 54, No. 3, pp. 203–210, 2012.

[2] 塚脇涼太, 深田博己, 樋口匡貴. ユーモア表出が表出者自身の不安および抑うつに及ぼす影響過程. 実験社会心理学研究, Vol. 51, No. 1, pp. 43–51, 2011.

[3] 椙本知子, 山崎勝之. 対人ストレスユーモアコーピングが敵意，意識的防衛性と抑うつに及ぼす影響. 心理学研究, Vol. 82, No. 1, pp. 9–15, 2011.

[4] Pawel Dybala, Michal Ptaszynski, Jacek Maciejewski, Mizuki Takahashi, Rafal Rzepka, and Kenji Araki. Multiagent system for joke generation: Humor and emotions combined in human-agent conversation. *Journal of Ambient Intelligence and Smart Environments*, Vol. 2, No. 1, pp. 31–48, 2012.

[5] 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 A, Vol. 95, No. 1, pp. 27–36, 2012.

[6] 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*, pp. 171–173, 2015.

[7] 吉田裕介, 萩原将文. 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: 日本知能情報ファジィ学会誌, Vol. 26, No. 2, pp. 627–636, 2014.

[8] 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, Vol. 14, No. 1, pp. 9–16, 2015.

[9] 滝澤修. 記述された「併置型駄洒落」の音素上の性質. 自然言語処理, Vol. 2, No. 2, pp. 3–22, 1995.

[10] Shigeto Kawahara. Probing knowledge of similarity through puns. *Proceedings of Sophia University Linguistic Society*, Vol. 23, pp. 110–137, 2009.

[11] 永田靖. サンプルサイズの決め方. 朝倉書店, 2003.

[12] 木村洋二. 笑いを科学するユーモアサイエンスへの招待. 新曜社, 2010.

[13] Gary Zalewski. Laugh detector and system and method for tracking an emotional response to a media presentation. US Patent US20090195392, 2009.

[14] Toshihiko Yokogawa. Japanese Pun Analyzer Using Articulation Similaries. In *Proceedings of the 2002 IEEE International Conference on Fuzzy Systems*, pp. 1114–1119, 2002.

[15] 天谷祐介, ラファウジェプカ, 荒木健治. 単語間類似度を用いた物語ユーモア認識手法の性能評価. 人工知能学会第 2 種研究会 ことば工学研究会資料, pp. 63–69, 2013.

[16] 北垣郁雄. おかしみの誘発モデルに基づいた笑́うコンピュータ́の開発事例. 知能と情報: 日本知能情報ファジィ学会誌, Vol. 15, No. 5, pp. 577–583, 2003.

[17] 山根宏彰, 萩原将文. 笑いを生むことわざすかしの自動生成システム. 知能と情報:日本知能情報ファジィ学会誌, Vol. 24, No. 2, pp. 671–679, 2012.

[18] Michal Ptaszynski, Pawel Dybala, Rafal Rzepka, Kenji Araki, and Yoshio Momouchi. YACIS: A Five-Billion-Word Corpus of Japanese Blogs Fully Annotated with Syntactic and Affective Information. In *Proceedings of The AISB/IACAP World Congress in Honour of Alan Turing – LaCATODA workshop*, pp. 40–49, 2012.

[19] Samuel S. Shapiro and Martin B. Wilk. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, Vol. 52, No. 3/4, pp. 591–611, 1965.

[20] Corinna Cortes and Vladimir Vapnik. Support–vector networks. *Machine Learning*, Vol. 20, No. 3, pp. 273–297, 1995.

[21] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, Vol. 12, pp. 2825–2830, 2011.

[22] Shigeto Kawahara and Kazuko Shinohara. The role of psychoacoustic similarity in Japanese puns: A corpus study. *Journal of Linguistics*, Vol. 45, pp. 111–138, 2009.

[23] 注連隆夫, 土屋雅稔, 松吉俊, 宇津呂武仁, 佐藤理史. 日本語機能表現の自動検出と統計的係り受け解析への応用. 自然言語処理, Vol. 14, No. 5, pp. 167–197, 2007.

[24] 葦原史敏, 木村泰知, 荒木健治. 地方議会会議録における節単位による議員の要望抽出. 電子情報通信学会論文誌 D, Vol. J98–D, No. 11, pp. 1390–1401, 2015.

[25] Shigeto Kawahara and Kazuko Shinohara. Calculating vocalic similarity through puns. *Journal of the Phonetic Society of Japan*, Vol. 13, pp. 101–110, 2009.

[26] Tristan Miller and Iryna Gurevych. Automatic disambiguation of English puns. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pp. 719–729, 2015.

第 4 章

対話における駄洒落ユーモアの認識/生成と話題適応

4.1 はじめに

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．

図 4.1 Siri との対話例

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検出・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，駄洒落の検出・生成は，文脈とは関係なく無秩序に行う方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推定される．

これらの仮説を検証するため，この駄洒落ユーモアの検出・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．本章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 **4.2** 節で関連する研究を挙げる．**4.3** 節では，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検出・生成能力及び話題への適応能力について述べる．**4.4** 節では提案手法に対して行った実験について，**4.5** 節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．

4.2　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のほかに，吉田ら [5]，ディバワら [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検出

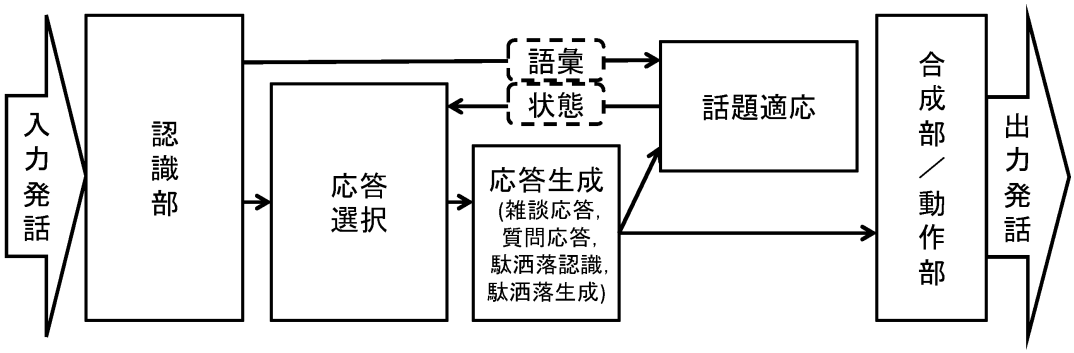

図 4.2　ユーモア機能付き統合型対話システムの構成

表 4.1　システムの状態及び 2 種の駄洒落動作の発生確率

状態	駄洒落検出	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

した場合に応答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．ディバワら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，著者ら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，本章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検出及び生成の頻度を調整し，実験を行う．

4.3　提案手法及び実験システム

4.1 節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検出・生成機能を統合し，小型ロボット RAPIRO[*1]への実装を行った．音声認識には Google Speech API[*2]を，音声合成には Open JTalk[*3]を用いた．システム全体の構成を図 2 に示す．

4.3.1　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，**4.3.2** 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，**4.3.3** に述べる手法により駄洒落の検出及び生成を行う．

4.3.2　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 Leacock-Chodorow 法 [11] により計測する．このようにして累積した値が閾値を超えているか否かを，発話の入力毎に確認する．閾値を超えている状態においては，式 (4.1) で表される減衰関数を用いてターンの経過毎に累積値を減少させる．

[*1] http://www.rapiro.com/ja/

[*2] https://cloud.google.com/speech/

[*3] http://open-jtalk.sourceforge.net/

$$f(t) = e^{-at}\cos bt \tag{4.1}$$

式 (4.1) における t は，累積値が閾値を超えた後で経過したターン数であり，$t \geq 0$ である．パラメタ a 及び b は，予備実験を行い最適な値を実験的に求める．後述する駄洒落の検出及び生成において，累積値が閾値を超えている状態とそうでない状態とで検出及び生成を行う確率を変化させる．この変化の手法については，**4.3.3** において述べる．

入力発話との関連性の計測対象を「食事」トピックに定めた理由は以下に述べる通りである．まず，**4.3.3** で述べる駄洒落ユーモアの生成手法が食べ物の語彙を多く利用するものであることが主な理由である．そして，「食べることが好き」というキャラクター付けは対話ロボットに親近感を持ち接するために妥当であると考えられ，多数のトピックを用いる場合と比較して **4.4.1** において述べる雑談対話への導入も容易になると考えられるためである．

4.3.3　駄洒落の検出及び生成

駄洒落ユーモアの検出は，発話の入力がある毎に検出モジュールを用い駄洒落を検出することにより行う．駄洒落の検出モジュールは，子音間の音韻類似度及び bag-of-words を素性とする教師あり機械学習による手法 [12] を用いる．駄洒落の検出が行われた場合，「それは駄洒落」という命題文にランダムにモダリティ表現を付加した応答発話を出力する．

駄洒落ユーモアの生成は，(1) 入力文の改変（漫才のボケ形式），(2) なぞなぞ，(3) 駄洒落データベースの，いずれかのタイプの駄洒落を生成するシステム [13] を用いる．それぞれのタイプの生成手法は下記の通りである．

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する．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．

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、天丼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 (w_1 とする) について，音韻の類似した食べ物・飲み物・菓子の名称 (w_2) をシソーラスを用いて抽出する．抽出に成功した場合，”「$< w_1 >$ は $< w_1 >$ でも，＜修飾子＞ $< w_1 >$ はなーんだ？」「何？」「$< w_2 >$」”（＜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．w_2 の類似部分を w_1 に置換して出力する．w_1 が「職人」，w_2 が「冷凍食品」である例を以下に示す．

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．

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

表 4.2　実験条件の一覧

条件	話題適応	駄洒落検出	駄洒落生成
c_0	-	-	-
c_1	-	-	○
c_2	-	○	-
c_3	-	○	○
c_4	○	-	-
c_5	○	-	○
c_6	○	○	-
c_7	○	○	○

い...」という発話を応答とする.

駄洒落の検出及び生成は，一定の確率により確率的に行われる．その確率は，**4.3.2** 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 4.1 に状態に応じた各動作の発生確率を示す.

4.4　印象評価実験

本節では，**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.

4.4.1　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 4.2 に示す条件 c_0〜c_7 のうち，条件 c_4 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 c_0 と同一の出力となるため，c_4 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

べることが大好きである」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．(駄洒落検出の適切さ)
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 及び Q5 は，駄洒落検出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．

4.4.2　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, 駄洒落検出, 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，話題適応の有る場合の条件と無い条件との評価値の比較を行った．この結果を表 4.3 に示す．

また，設問 Q1,Q2 及び 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 4.4 に示す．

表 4.3　実験条件間の評価値の比較結果

比較	話題適応 無	話題適応 有	Q1 (自然性)	Q2 (文法性)	Q3 (知識量)	Q4 (検出)	Q5 (生成)	平均点
生成のみ	c_1	c_5	**無**	**＝**	**無**	N/A	N/A	**無**
検出のみ	c_2	c_6	**＝**	**有**	**＝**	N/A	N/A	**有**
生成＋認識	c_3	c_7	**有**	**有**	**有**	N/A	**有**	**有**

凡例 ［**無**, **有**］ 話題適応を (行わない, 行う) 条件における評価値がより高い
［**＝**］ 話題適応の有無にかかわらず，評価値が等しい
［N/A］ いずれかの評価値が得られなかったため，比較不可

4.4.3　対話例

本実験において行われた対話の一部を表 4.5 及び表 4.6 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．

4.5　考察

4.4 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 4.3 より，駄洒落発話の生成のみを行う場合，話題適応を行う条件の応答の自然性及び知識量の印象評価値は行わない条件の値以下のものとなっている．一方，駄洒落発話の検出のみを行う条件及び検出・生成の両方を行う条件のもとでは，Q1, Q2, Q3 及び全項目の平均点は話題適応の有る条件は無い条件に比べ同等以上の評価値となり，駄洒落発話の検出を行う条件と行わない条件とで相反する傾向が確認された．

このことから，駄洒落を生成することを主とする従来のユーモア機能付き対話システムにおいては，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 4.3 の「駄洒落検出」に対し Q1 が B を示していることから，駄洒落検出発話

表 4.4　実験で用いた 7 つの条件及び設問の得点平均値（降順に並び替え）

条件	Q1(自然性)	Q2(文法性)	Q3(知識量)	Q4(認識)	Q5(生成)	平均値
c_0	3.33	4.33	4.00	(値なし)	(値なし)	3.89
c_7	3.00	4.00	4.00	3.67	3.33	3.60
c_1	3.33	3.33	3.67	(値なし)	(値なし)	3.44
c_6	2.33	4.33	3.33	(値なし)	(値なし)	3.33
c_2	2.33	3.67	3.33	(値なし)	(値なし)	3.11
c_3	2.67	3.00	3.00	(値なし)	3.00	2.92
c_5	2.00	3.33	2.67	(値なし)	(値なし)	2.67
c_4	(値なし)					—

注 (値なし) は，実験の条件設定 (c_4) 及び被験者へのアンケート設問の条件 (Q4,Q5) により回答がなされかった項目である (**4.4.1** 参照).

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，システムによる駄洒落の生成においては，不適合理論から推測される，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 4.4 において，話題適応及び駄洒落の両機能を一切用いない条件 c_0 が設問 Q1, Q2 及び Q3 の平均値で最上位となっている．このことから，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．

対話例（表 4.6）における駄洒落検出（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．

4.6　まとめ

本章では，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．

表 4.5　システムとの対話例（駄洒落生成のある例）

ターン	話者 (モジュール名)	発話
1	ユーザ	遊びたい
	システム (雑談応答)	しりとりやりましょうよ。
2	ユーザ	しりとり
	システム (雑談応答)	しりとりが好きですね。
3	ユーザ	猫
	システム (雑談応答)	いぬ。
4	ユーザ	動物は好きですか
	システム (雑談応答)	嫌いじゃないですよ。
5	ユーザ	猫は好きですか
	システム (雑談応答)	はい、好きです。
6	ユーザ	猫は怖いです
	システム (駄洒落生成)	猫と言えば、猫がロンドンで寝ころんどん?って言いたくなるよね。
7	ユーザ	犬はいいですか
	システム (雑談応答)	はい、犬は好きですよ。

実験の結果より，駄洒落検出が有効な条件において，話題適応による機能変化により性能が向上することが示唆された．ただし，被験者数の少なさにより，統計的な検定による結論を得るには至っていない．予備実験によるパラメータの調整を経た上で，20 名以上の規模の評価実験により有効性を詳細に渡り再評価することが必要と考えられる．

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検出・生成モジュールの統合の方法を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 回平均で 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．

表 4.6　システムとの対話例（駄洒落検出のある例）

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

† このユーザ発話の入力時に音声認識誤りが発生した．被験者は「飛行機と電車はどちらがいいですか」という入力を意図していた．

参考文献

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

[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, 萩原将文. 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] 亀山恵祐, 荒木健治, 木村泰知. 質問応答・システムにおける知識源選択規則の自動獲得の有効性について. 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, pp. 875–886, 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*, Vol. 22, No. 140, pp. 5–55, 1932.

第 5 章

結論

本学位論文では，雑談システムをベースとする統合型対話システムにおいて，話題の変遷への適応に基づく対話戦略とユーモア理解・生成能力との間の適切な相互作用を実現することを目的に，ベースとなる統合型対話システム及び話題適応手法の提案，駄洒落ユーモア発話の検出手法の性能評価，そして駄洒落ユーモアの認識及び生成機能を有するユーモア機能付き統合型対話システムにおける話題適応の有効性の検討を行った．

第 **1** 章では，研究の背景となる問題及び関連を行った研究について，文献を示しながら考察を行った．

第 **2** 章では，話題適応発話の生成機能をもつ統合型対話システムを提案し，印象評価実験による性能評価を行った．初めに対話ドメインの遷移を検出する話題適応の手法について説明し，対話システムの構成を説明した後に応答精度評価実験及び印象評価実験の結果について考察を行った．実験の結果，提案手法が既存の非タスク指向型対話システムの対話の自然性を低下させずに，話題適応発話の生成が可能であることが示された．課題として，各種パラメータの最適化のための対話ログに対する客観的自動評価手法の確立と，再現性の高いユーザモデルの採用によるシステムの基本動作フローの見直しが挙げられた．

第 **3** 章では，駄洒落ユーモアの検出手法の性能評価を行った．初めに，併置型

駄洒落のルールベースでの検出を行った．文に含まれる語彙情報に加え, 子音音韻類似度に基づく検出ルールの結果を素性とする SVM を用いた併置型駄洒落の検出手法を提案し, 提案手法と前述のルールベースの検出手法との比較およびパラメータの変化に対する性能の特性の調査を行った．この結果, SVM 及び子音音韻類似度を用いた検出手法がルールベース検出手法を上回る検出性能を有することに加え, データ量の現象に対を行った頑健性においては, 子音音韻類似度によるルールベース素性を含む全素性を用いる場合が最も優れていることが確認された．今後は母音の音韻類似度を利用することによる検出性能の向上や, 語彙の変換に基づく種・変形表現ペアの検出による重畳型駄洒落への対応を考えている．

第 **4** 章では, 話題遷移に対応したユーモア機能付き対話システムの性能評価について現時点での結果を述べた．ユーモア機能付き統合型対話システムの内部状態の変動 (話題適応) に対する印象評価値の変化の調査を行った．実験結果より, 話題変遷への適応による印象評価値の増減が, 用いるユーモア関連機能の組合せにより異なることが示された．対話システムの応答の自然性の向上等の課題は残るが, 話題変遷への適応がユーモア機能付き対話システムの印象評価に正の影響をもたらしうることが実験により示唆された．これは，話題適応に基づくユーモア機能の動作の変更が適切であればあるほど，統合型対話システムの印象評価に正の影響が生じるという仮説を支持するものである．

本学位論文で述べた研究を通じて, ユーモアの理解・生成の手段を話題適応の形で変化させることが可能な統合型対話システムを構築し, その印象評価を行った．実験の結果, ユーモアの認識を行う場合における話題適応に基づくユーモア機能の動作の変更の有効性が確認された．

今後の課題として初めに挙げられるのは, 第 **4** 章における印象評価実験において大きく影響していたと考えられる統合型対話システムの対話の自然性の向上である．ユーモア機能における課題としては, ユーモア発話の頻度及びユーモア認識の確率といった量的な調整に加え, 駄洒落ユーモアの面白さを考慮した質的な幅のある応答を行う能力を付与を行うことが挙げられる．以上の課題の解決

が, 人間により近いユーモアセンスを持ち合わせた対話機能を実現可能とすることに寄与すると考えられる．今後は，ユーモア機能付き統合型対話システムが人間により近い印象を与えることのできるよう，上記の課題の解決手法を主な研究対象とする予定である．

謝辞

本学位論文は，著者が北海道大学大学院情報科学研究科メディアネットワーク専攻情報メディア学講座言語メディア学研究室において同専攻博士課程に在籍した 2010 年 4 月から 2017 年 3 月までに行った研究をまとめたものである．執筆にあたり，多くの方々からご支援をいただいた．

本学位論文を取りまとめるにあたり，研究の遂行のみならず，終始，子細なるご指導とご高配を賜りました北海道大学大学院情報科学研究科メディアネットワーク専攻情報メディア学講座言語メディア学研究室，荒木健治教授に深く感謝の意を表します．国際会議及び諸学会での発表に際して多大なるご支援をいただきましたことにつきましても，この場を借りまして厚くお礼申し上げます．

ご多忙のなか，快く副査をお引き受けくださいました北海道大学大学院情報科学研究科メディアネットワーク専攻情報メディア学講座情報メディア環境学研究室，山本強教授に深く感謝の意を表します．

ご多忙のなか，快く副査をお引き受けくださいました北海道大学大学院情報科学研究科メディアネットワーク専攻情報メディア学講座メディアダイナミクス研究室，長谷山美紀教授に深く感謝の意を表します．

ご多忙のなか，快く副査をお引き受けくださいました北海道大学大学院情報科学研究科メディアネットワーク専攻情報メディア学講座メディア創生学研究室，坂本雄児教授に深く感謝の意を表します．

本研究を進めるにあたり，有益なご指導，ご討論をいただいた北海道大学大学院情報科学研究科メディアネットワーク専攻情報メディア学講座言語メディア学研究室，伊藤敏彦准教授に深く感謝の意を表します．

本研究を進めるにあたり，数多くのご指導，ご討論，そしてユーモアに関する有益かつ面白い示唆に留まらず様々な面でのご協力をいただいた北海道大学大学院情報科学研究科メディアネットワーク専攻情報メディア学講座言語メディア学研究室，Rafał Rzepka 助教に深く感謝の意を表します．

著者の北海道大学文学部言語情報学講座への在籍当時より，数多くのインスピレーションを齎すご教示を賜りました北海道大学大学院文学研究科言語文学専攻言語情報学講座，小野芳彦特任教授に深く感謝の意を表します．言語情報学の世界の奥行きの深さを先生のもとで垣間見たことが，自然言語処理ないし計算言語学のフロンティアへと一歩を踏み出すきっかけとなりました．

お忙しい中，被験者として実験にご協力いただきました皆様に厚くお礼申し上げます．

著者の研究期間中，公私にわたりご協力いただいた北海道大学大学院情報科学研究科メディアネットワーク専攻情報メディア学講座言語メディア学研究室の学生・OB・OG 諸氏に厚くお礼申し上げます．

最後に，著者が北海道大学において学ぶチャンスを与え，そして支えてくれた家族や友人達に感謝します．

研究業績

査読付学会誌等

(1) 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), 2016 年 28 巻 5 号, pp. 875–886 (2016-10).

査読付国際会議プロシーディングス

(1) M. Yatsu, R. Rzepka, and K. Araki: 「A Modules-Based, Task-Navigational Dialogue System」, Proc. of the Pacific Association For Computational Linguistics, Kuala Lumpur, Malaysia. July 19-21, paper no. 44 (2011-7).

(2) M. Yatsu, R. Rzepka, and K. Araki: 「A Domain Analytic Method in Modular-Designed Dialogue System: Application to a System for Japanese」, Proc. of Linguistic and Cognitive Approaches To Dialogue Agents (AISB/IACAP Symposium), Birmingham, UK, July 2-6, pp. 25-30 (2012-7).

学会発表等

(1) 谷津 元樹, ジェプカ ラファウ, 荒木 健治: 「汎用的な対話システムにおけるドメイン検出のための複合的手法」, 第 28 回ファジィシステムシンポジウム講演論文集, pp. 420-425 (2012-9).

(2) 谷津 元樹, ジェプカ ラファウ, 荒木 健治: 「トピック推定を用いたタスクドメインを選択するための発話生成」, 言語処理学会第 19 回年次大会予稿集, pp. 142-145 (2013-3).

(3) 谷津 元樹, ジェプカ ラファウ, 荒木 健治: 「自由対話からの話題遷移検出のためのタグ付け調査」, 言語処理学会第 20 回年次大会, pp. 236-239 (2014-3).

(4) 谷津 元樹, 荒木 健治: 「子音の音韻類似度を用いた併置型駄洒落検出手法の有効性について」, 言語処理学会第 22 回年次大会, pp. 31-34 (2016-3).

(5) 谷津 元樹, 荒木 健治: 「駄洒落の面白さにおける要因の分析」, 第 32 回ファジィシステムシンポジウム, pp. 237-242 (2016-9).

(6) 谷津 元樹, 荒木 健治: 「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」, 人工知能学会第 2 種研究会 第 52 回ことば工学研究会資料, pp. 23-27 (2016-10).