

ルールベース発話生成と統計的発話生成の 融合に基づく対話システムの構築

Building a conversational system

based on the fusion of rule-based and stochastic utterance generation

目黒 豊美*¹ 杉山 弘晃*¹ 東中 竜一郎*² 南 泰浩*¹
Toyomi Meguro Hiroaki Sugiyama Ryuichiro Higashinaka Yasuhiro Minami

*¹NTT コミュニケーション科学基礎研究所, 日本電信電話株式会社
NTT Communication Science Laboratories, NTT Corporation

*²NTT メディアインテリジェンス研究所, 日本電信電話株式会社
NTT Media Intelligence Laboratories, NTT Corporation

This paper describes our rule-based chat dialogue system and its comparison to our statistical one by subjective evaluation. We found that, for expected user utterances, our rule-based system can achieve satisfactory responses; however, for unexpected ones, its responses become comparable to or often worse than those of the statistical system. Following this result, we present a method for classifying user utterances into two classes: one that can be answered by the rule-based system and the other that should be answered by the statistical system. Using features such as words, semantic categories, and perplexity of system responses, we trained an SVM classifier that achieved significantly better classification accuracy than a baseline. The result shows the possibility that a system that switches between the rule-based and statistical systems would lead to more naturalness.

1. はじめに

近年、雑談を行う対話システムに関する研究が増加している [Ritter 10, Higuchi 08, 東中 14]. 従来より研究が盛んなタスク指向型の対話システムにおいても、雑談はタスク達成率を上げる効果があることが知られている [Bickmore 01]. また基礎研究のみならず、NTT ドコモのしゃべってコンシェルヤや、iPhone の Siri の普及と共に、実ユーザが利用できる形で浸透してきている。

雑談対話システムを実現する上で最もよく用いられている手法は、ユーザの発話に応答するルールを手手で大量に記述する手法である。このようなルールベース型の対話システムは、ユーザ発話に合致するルールがある場合、適切な返答をすることができ、ルール作成者の直感や考えに沿った変更が容易であるという利点がある。しかし、任意の話題が現れる雑談に対応するためには、莫大な量のルールを手手で記述する必要があり、膨大なコストがかかる。また、刻々と変化する時事の話題に対応するためには、常に手手でメンテナンスを行う必要があり、ランニングコストも高いという欠点がある。

一方で近年、大規模なテキストデータから、統計的な手法を用いて発話を生成するシステムが出現してきている [Bessho 12, 柴田 09]. これらのシステムは、大規模なデータにより多様なユーザ発話に対応することを目的としている。このような統計的発話生成手法は、様々な話題について手手を要せずに発話を生成することができ、ルールベース型の対話システムと比較してコストは低い。しかしながら、手手でチェックされていないために、不適切な応答が生成され得るという欠点がある。

本稿では、まず、人手によりルールを大量に作成し、そのルールを用いた対話システムを評価することで、日本語ルールベース型雑談対話システムの性能を明らかにする。これは日本語の雑談対話システムにおいて、膨大なルールを手手で作成し、その性能を他のシステムと比較した例はなく、他の統計的発話生成手法との差も明らかになっていないからである。ルールベース型対話システムは、適切なルールが存在する発話には適切な返答ができるが、事前に想定されていない発話に応答する場合や、部分的なルールのマッチを許容する場合などには不適切な応答をすることがある。そのため、このようなルールが

連絡先: meguro.toyomi@lab.ntt.co.jp

表 1: ルールベース型対話システムとの対話例

話者	発話	マッチしたルール
ユーザ	こんにちは	こんにちは
システム	こんにちは	
ユーザ	お名前は何ですか	
システム	愛子といいます	*お名前*何*か*
ユーザ	愛子さんですね、こんにちは	
システム	こんにちは	こんにちは
ユーザ	ご機嫌はどうですか	
システム	ええー?どうなのでしょう…。	どう*か
ユーザ	体調はどうですか	
システム	CM でよく見るパブロンという風邪薬がすごく	体調*どう
ユーザ	よくきいてびっくりしました。	
ユーザ	風邪をひいてたんですか	
システム	そうなんです	ですか

適切な返答を返すことができない発話には、統計的発話生成手法で返答することで、より広い話題に対して頑健に応答でき、ユーザの満足度が上がると考えられる。そこで、本稿ではユーザ発話がルールを用いて返答すべき発話か否かを識別することで、ルールベース型対話システムと統計的発話生成手法との融合を検討する。

2. 先行研究

多くの雑談対話システムは手作業によるルールに従って動作する。例えば、雑談対話システムのコンテストであるローブナー賞における多くのシステムは入力と出力のペアからなるパターンを大量に保持し、これらを用いて応答する。これらのパターンは Artificial Intelligence Markup Language (AIML) と呼ばれる言語で記述されることが多い [Wallace 04]. 日本語の対話システムとして「たけまる君」があるが、このシステムは内部に質問応答ペア (QADB と呼ばれる) を多く保持しており、これらを用いてガイダンスのみならず自身のプロフィールなどについて応答する [Takeuchi 07].

統計的発話生成手法には、例えばユーザ発話をクエリとして Web 検索を行い、その結果得られるニュース記事や、ツイートを用いて答えるシステムがある [Bessho 12, 柴田 09]. 我々も、ツイッターから比較的高精度な発話のデータベースを構築したり [Higashinaka 14], ユーザ発話に含まれる単語について、その単語が係っている単語や表現を同じくツイッターから抽出し、テンプレートに当てはめて応答を生成する手法を提案

表 2: 評価項目

評価項目名	説明
(1) 応答文の自然さ	応答文として直感的に違和感を感じないか
(2) 文法の正しさ	日本語として理解できる文法か
(3) 意味の関連性	文として意味が通った発話か
(4) 発話間の矛盾	入力文に対して相反していないか
(5) 話題の連続性	話題がずれていないか
(6) 話題の広がり	話を広げる情報が含まれているか
(7) 話の続けやすさ	システム発話の後に発話を続けやすいか

表 3: 入力文が雑談の場合の各システムの評価値 (5 が最も良い, すべての項目, システムとルール間で有意差あり)

	ルール	文節ペア	単語	IR-st.	IR-re.
(1) 応答文の自然さ	4.11	3.16	3.31	2.11	2.96
(2) 文法の正しさ	4.99	4.07	4.19	3.15	4.08
(3) 意味の関連性	4.17	3.32	3.41	2.13	3.17
(4) 発話間の矛盾	4.20	3.34	3.43	2.28	3.38
(5) 話題の連続性	4.22	3.45	3.51	2.28	3.45
(6) 話題の広がり	3.99	2.94	3.08	2.07	2.82
(7) 話の続けやすさ	4.15	3.08	3.19	2.07	2.80

している [Sugiyama 13]. また, 単語だけではなく, ユーザ発話の係り受け構造に着目し, その述語の係り先の節を応答に用いている [杉山 13]. これらの検索ベースの手法は Web 検索と同様, 入力について何らかの応答を返すことが可能になる一方, ルールベースシステムのように応答の質を担保することが難しいという問題がある.

我々が本稿で扱う問題は, 対話における発話モジュール選択の問題とみなすことができる. 江頭らは雑談システムにおいて, 応答に用いるべきモジュール (例えば, 質問応答モジュールや Wikipedia の定義文を返すモジュール) を強化学習で選択する手法を提案している [江頭 12]. CONVERSE は, ユーザ入力について, 複数のモジュールに回答可能かどうかを表すスコアを計算させ, そのスコアが最も高いものを回答モジュールとして選択するという方法で雑談を行う [Batacharia 99]. 我々は, これらの研究と同様, 発話モジュール選択の問題に取り組むが, ルールベースと検索ベースの発話モジュールを選択するという点が異なる.

3. ルールベース型対話システムの構築

3.1 ルール作成方法

ルールはユーザ発話とマッチするためのパターンと, それに対応するシステム発話を持つ. 対話システムとして動作する時には, ユーザ発話を入力文として受け取り, マッチしたパターンに対応するシステム発話を応答として出力する. パターンは単語とアスタリスクから成り, アスタリスクの部分はどのような文字列ともマッチする. これらの記述形式は, AIML に準拠する. 以下に構築手順を述べる.

まず, 我々が収集した雑談コーパス約 3600 対話 (約 12 万発話) のうち一割を最終的な性能評価テスト用に除外し, 残りに含まれる連続した 2 発話の対について, 先の発話をパターン作成用の発話, 後の発話をシステム発話として抽出した. 同様に先行研究で収集した対話システムのパーソナリティを問う質問コーパス [杉山 14] 約 25000 質問回答ペアのうち, 20 代女性向けへの質問約 4200 件を, 質問をパターン作成用の発話, 回答をシステム発話として抽出した. 質問の対象パーソナリティを限定したのは, ルール中の矛盾を避けるためである.

抽出したこれらの発話ペアのうち, パターン用発話を形態素解析し, 発話内で重要と思われる自立語を中心とした単語 (名詞, 動詞, 形容詞, 終助詞「か」, 連体詞など) を残し, その他は削除した. 削除した部分には, アスタリスクを挿入しパターンとする. このように自動的に作成した発話ペアを手で取捨選択する. 加えて, 雑談コーパス中に頻出したキーワード 250 個に対して, 20 人が 25000 文の「対話を続けたいと思う

表 4: 入力文が Twitter の場合の各システムの評価値 (5 が最も良い, 有意差はルールベースとの比較. *: $p < .1$, **: $p < .05$)

	ルール	文節ペア	単語	IR-st.	IR-re.
(1) 応答文の自然さ	3.17	3.33	3.39	2.20**	2.74**
(2) 文法の正しさ	4.97	4.42**	4.42**	3.30**	4.25**
(3) 意味の関連性	3.16	3.41	3.45	2.25**	2.89
(4) 発話間の矛盾	3.19	3.46	3.48	2.26**	3.20
(5) 話題の連続性	3.20	3.56	3.51	2.30**	3.25
(6) 話題の広がり	3.51	3.06**	3.23	2.10**	2.51**
(7) 話の続けやすさ	3.81	3.12**	3.20**	2.10**	2.34**

発話」を作成し, キーワードをユーザ発話とマッチングするためのパターンとして, 作成された発話をシステム発話としてルール化した. このように既存の対話コーパスや質問コーパスを用いることで, ルール作成者が想像だけで作るよりも, 網羅性の高いルールになると期待できる.

作成されたルールの品質を向上させるため, ルール作成者とは異なる別の実験参加者が, 作成されたルールによって動作するシステムと対話を行い, 適宜ルールを追加・変更した. 具体的には, 性能評価用に除外した発話に対する応答文のうち, 90% 以上の応答文に問題がないと判断されるまでこの作業を繰り返した. 最終的に, 149300 個のルールが得られた. これらのルールを用いて構築されたシステムの対話例を表 1 に示す.

3.2 ルールベース型対話システムの評価実験

我々が構築したルールを用いたルールベース型対話システムと統計的発話生成システムを比較することで, 作成したルールベース型対話システムの有効性を検証する. まず, 著者ではないアノテータが, 「文脈を考慮せず, 一文だけ読んで理解ができる」という評価軸のもと, 雑談のデータから 149 文, Twitter コーパスから 80 文を選択し, システムへの入力とした. 各文について, 構築したルールベース型対話システムと, 後述する 4 つの比較システムからそれぞれ発話を出力させ, 著者ではないアノテータ 3 人が, 表 2 の 7 項目について評価を行った.

ルールベース型対話システム 3.1 節の手法で構築したパターンと, ユーザ発話の間で TF · IDF の重み付きコサイン類似度を測り, 最も近いパターンを持つシステム発話を出力する. AIML のデフォルトのルール選択方法は, 先頭から単語をマッチングさせる方法であるが, 日本語のように語順が入れ替わりやすい言語では, 適切なルールが選ばれない場合がある. 予備実験において, 先頭からのマッチングと, TF · IDF の重み付コサイン類似度を用いた手法を比較し, 自然性の高かったコサイン類似度の手法を採用した.

文節ペア間の係り受けを利用した発話生成 我々が先行研究で提案した手法 [杉山 13] で, Web コーパスからユーザ発話中の係り受け構造に関連する係り受け構造を抽出し, テンプレートに埋め込むことで発話を生成する. 係り受け関係のある文節ペアを利用することで, ドメイン知識を前提とせずに様々な発話に関連した発話を返答することができる. コーパスとしてツイート約 1 億 2 千万文を用いる.

単語間の係り受けを利用した発話生成 単語間の係り受け関係に基づき新しい話題を導入し, テンプレートに代入することで発話を生成する手法を用いる [Sugiyama 13]. 単語間の係り受け関係に着目している点が, 文節ペア間の係り受けを利用したシステムと異なる.

IR-status Twitter からユーザ発話文に類似したツイートを検索し, in-reply-to 機能で対応付けられた返信ツイートをシステム発話文として出力する. [Ritter 10] で用いられた手法と同一で, 入力文と似た意味をもつ発話への返答は, システム発話として適切なものを返答できると考えられる.

表 5: 生成された発話例. 入力文として, ツイートを利用している中で, ルールベースの評価が高かった場合.

入力文	生成した発話	システム	評価値
することなく テザリングを Twitterにのみ 使う。	次々と新しい機能の名前が登場して、ついていけないよ。	ルール	5
	5sでテザリングして使ってるよね	文節ペア	2
	LTEとかは?	単語	4
	わいもや、テザ代みたいの何百円か払ってるで。	IR-st.	3
	テザリングじゃなくてテザリングなデザリングするか	IR-re.	3

IR-response Twitterから何らかの投稿への返信となっているresponseを検索し, ユーザ発話文と類似したツイートを出力する. IR-statusと似た手法であるが, 返答文自体で検索することにより, 返答文の文体を持ち, 入力文とかけ離れた内容ではない発話が出力されると考えられる.

3.3 評価結果

表3に結果を示す. 雑談コーパスから抽出した発話を入力文としたときには, ルールベース型の評価が高い. これは, ルールが雑談コーパスを基に作られており, 本評価で用いたユーザ発話に合致するパターンも, ルールに含まれているためである. このように, 事前に想定することができる発話については, 高い自然性で返答することができていることが確認できた.

表4にTwitter文を入力としたときの結果を示す. 応答文としての自然さ(以下, 自然性)では, ルールベース型は文節ペア間の係り受けを利用した発話生成手法とほぼ同程度の評価であった. 表5はツイートを入力文とした場合の各システムの出力と, その評価値である. 評価値として, 7項目のうち発話全体の適切さを評価する自然性を用いている. 表5の例のような場合, ツイートの発話はルールにとって未知の発話であるが, ルールの中に「テザリング」にマッチするルールが存在し, 適切な発話を返答することができている. このように, 雑談のような既知の入力文だけでなく, Twitterのような未知の入力文についても適切な応答ができるルールが構築できていることがわかる.

しかし, 雑談コーパス中の発話に適切に返答できていない場合もある. 表6は雑談コーパス中の発話をユーザ発話とした場合の各システムの出力とその評価値である. ユーザ発話文「卵とカレーほどよく混ぜり合って美味しそう。」に対して, 最も類似度が高いルールは「雷 * 混ぜ * ました *」となった. これは, パターンとのマッチにTF・IDFを用いており, 「混ぜ」というTF・IDFの値が比較的高い単語にマッチしたため選択された. このユーザ発話に本来マッチすべきルールは「美味し」というパターンの返答文である「美味しいですよ〜!」などであるが, 「美味し」は「混ぜ」よりTF・IDFの値が低く, 本ルールは適用されなかった.

このように, 平均値として見た自然性ではルールベースと統計的発話生成手法は, 未知の発話に対してほぼ同程度の自然性ではあるが, ルールベース型のシステムを用いた場合に, 適切なルールがありながらも, うまくマッチングすることが出来なかったり, そもそもルールがないということが起こっているようである. このような場合には, 統計的発話生成手法を用いて返答することで, 自然性が向上する可能性がある.

4. ルールベース型対話システムと統計的発話生成手法の融合

前節で示されたように, ルールベース型の対話システムは, ルールとマッチするユーザ発話文には適切な返答ができるが, ルールにうまくマッチしない発話文には適切な返答ができない. そのようなときには, 広い話題に対して返答が可能な統計的発話生成手法を用いて発話を生成した方が良いと考えら

表 6: 生成された発話例. 入力文として, 雑談コーパスでの発話文を利用している中で, ルールベース以外の統計的発話生成手法の評価が高かった場合.

入力文	生成した発話	システム	評価値
卵とカレーほどよく混ぜり合って美味しそう。	雷は怖いですね	ルール	1
	いい感じに混ぜって美味しかったな	文節ペア	5
	ヤングコーンも卵だ	単語	1
	カレーだと	IR-st.	2
	混ぜり合おう	IR-re.	1

れる. 前節の評価実験で, 各入力文に対してルールベース型が出力した発話が最も自然性が高かった例は, 雑談の入力文の約70%, Twitterの入力文の50%であった. つまり, 残りの各30%, 50%では, 他のシステムを用いて発話を生成する方が良かったと言える. すなわち, ルールベース型の対話システムと, 統計的発話生成手法を切り替えることができれば, どちらか一方のみを用いる時より, 高い自然性を得られると考えられる.

そこで, 本節では前節の評価実験に用いた入力文を対象とし, ルールベース型が応答すべきか, それ以外のシステムが応答すべきかを識別する手法を検討する.

4.1 ユーザ発話文の識別手法

前節の実験でユーザ発話として用いた入力文から後述の特徴量を基に, SVMを用いて, ルールベース型対話システムで返答すべきか否かを識別する分類器を学習する. 前節の評価実験では, ユーザ発話一文につき, 5つのシステムがそれぞれ一文ずつ発話を出力した. その中で最も評価が高かったシステムがルールベース型であるか否かを識別する分類器を学習する. 同点で一位のシステムがあった場合は, ルールベース型を選ぶことを正解とした. 特徴量としては以下を用いる.

品詞, 語彙大系, 単語 unigram

ユーザ発話に含まれる品詞, 単語, 語彙大系 [池原 99] のクラスのすべての unigram を用いる.

ルールとの一致度

ルールベース型対話システムが発話を生成する際には, ルールに含まれるマッチングのためのパターンと, ユーザの発話文との一致度をTF・IDFで重み付されたコサイン類似度が最も近いものを選択する. その際に用いた類似度を特徴量の一つに用いる. これは, ルールとの一致度が高く, よくマッチしている場合は, システム発話としてルールベースの出力を採用すべきと考えられるからである.

パープレキシティ

大量のデータから学習した言語モデルにおいて, パープレキシティが高い発話は頻出しない発話であり, ルールの中に適切な返答ができるものが含まれていない可能性が高い. 逆に, パープレキシティが低い発話は頻出する発話であるため, 適切なルールが存在する可能性が高い. この仮説を反映させるための特徴量として, 言語モデルによるパープレキシティを用いる. 言語モデル作成にあたり, 2011年4月1日~2012年3月5日のブログ記事から抽出した約2億ページ, 約100万語彙のテキストデータを用いた.

4.2 実験結果

ユーザ発話内の各特徴量を用いてSVMで学習した結果は, 表7のようになった. 雑談をユーザ発話としたときには, すべての特徴量を用いたときが最も精度が良く, ランダムと比較して有意に精度が向上した(マクネマー検定: $p < .05$). これらの特徴量の中で, ルールとの一致度が最も精度向上に寄与しており, ルールとの一致度が高いときにはルールベース型で返答すべきで, ルールとの一致度が低いときには他のシステムで返答すべきであるということが分かった.

表 7: 各特徴量を用いた時の識別精度 (10 分割交差検定). unigram, 一致度, パープレキシティのカラムは各特徴量のみを用いた場合.

入力文	ランダム	unigram	一一致度	パープレキシティ	すべて
雑談	69.8%	71.3%	75.3%	58.7%	78.7%
Twitter	50.0%	53.8%	22.5%	62.5%	53.4%

表 8: ルールベース型対話システムと、他システムを融合したときの自然性. proposed は、ルールベースで返答すべきではないと判断されたときに、単語間の係り受け関係を用いた発話生成を用いた場合. (有意差は proposed との比較, **: $p < .05$)

入力文	oracle	proposed	ルールのみ	単語のみ
雑談	4.42**	4.17	4.11	3.31**
Twitter	4.02**	3.61	3.17**	3.39**

一方で、Twitter を入力文としたときには、すべての特徴量を用いた時が、ランダムの性能と変わらず、パープレキシティのみを用いた時が最も精度がよく、有意に分類精度が向上した (マクネマー検定: $p < .01$). つまり、ユーザ発話が、珍しい発話かどうかで、ルールベース型で返答すべきか否かが決定するということを表している。ただ、他の特徴量での精度が安定していないことから、学習データにオーバーフィットしている可能性もある。学習量がユーザ発話 80 文しかないため、今後はデータを増やして検討する必要がある。

次に、ルールベース型で返答すべきか否かという識別がユーザ評価を向上させるか検証する。返答すべきでないと判断されたときに、単語間の係り受けを利用した発話生成手法 (以下、単語間生成) を用いて返答する場合の自然性を表 8 に示す。単語間生成を選んだのは、3.3 節の評価実験において、統計的発話生成手法で最も評価が高かったためである。特徴量には、表 7 で入力文の種類ごとに最も精度が高かった特徴量を用いており、雑談はすべての特徴量、Twitter はパープレキシティを用いる。oracle はルールベース型と単語間生成のうち、常に自然性の高い方を選べたとき、proposed は、ルールベースで返答すべきではないと判断されたときに単語間生成を用いた場合、ルールのみ、単語のみは、それぞれ常にルールベース型、単語間生成のみで返答した場合の自然性である。

雑談コーパスの発話を入力文とした場合には、提案手法は、ルールベース型単独で動作した場合と自然性の評価値はほぼ変わらなかった。これは、単語間生成がルールベース型よりも自然性が高い場合であっても、これらシステム間の自然性の評価値の差が小さく、全体の自然性向上には寄与しなかったためと考えられる。しかし、オラクルとは有意な差があるため、識別精度を向上させることができれば、提案手法の自然性を改善できる可能性がある。Twitter を入力文とした場合には、ルールベース型と単語間生成のみをそれぞれ単独で使用した場合と比べて、自然性に大きな向上が見られた。Twitter に見られる発話はルールベース型に未知であり、適切に返答できる入力文とできない入力文の評価値の差が大きく、これらを適切に識別をすることで、ルールでは適切に返答できない文について、単語間生成で返答できるようになった。

なお、これらの実験では、入力文を雑談コーパスと Twitter 由来のものに分けて、識別モデルの学習と自然性の推定を行った。ルールベース型として、雑談コーパスは事前に想定していた発話、Twitter は想定していなかった発話という位置づけである。本システムを使用する場合には、事前にユーザ発話が想定内の発話であるか否かという情報が得られるという状況は考えにくい。今後は、想定内発話であるか否かに関する事前知識が無くとも、想定内・外発話のどちらにも高評価を得られるシステムを構築していきたい。

5. おわりに

本稿では、日本語のルールベース型対話システムを構築し、統計的発話生成手法との比較を通して性能評価を行った。その

結果、ルールベース型のシステムはルール構築に利用した発話に対してはすべての評価項目において有意に高いが、想定外の未知の発話に対しては統計的発話生成手法と同程度の自然性が得られることがわかった。しかし、ルールを適切にマッチングできなかったり、そもそもマッチするルールが存在しない場合には、統計的発話生成手法を用いた方が自然性が高くなることがわかった。

そこで、ルールベース型対話システムで返答すべき発話であるかを識別することによって、ルールベース型と、他の統計的発話生成手法を融合する手法を検討した。この識別によって、識別精度、自然性ともに向上が見られた。このことから、複数の手法を組み合わせることでより自然性の高い対話システムを構築できる可能性を示すことができた。

ルールベース型対話システムでは返答すべきではないと判断されたユーザ発話に対して、単語間係り受け関係を用いた発話生成手法を用いたが、これは、評価実験で自然性の平均値が最も高かったためである。しかし、他の統計ベースのシステムで良いものがあればそれを適切に選ぶことで、より高い自然性を得られる可能性がある。今後は、どのシステムで返すべきかという発話モジュール選択の問題としてだけでなく、実際に各システムが候補文として生成した発話の内容自体を考慮し、適切な返答文を選択する手法を考案していく必要があると考えている。

参考文献

- [Batacharia 99] Batacharia, B., Levy, D., Catizone, R., Krotov, A., and Wilks, Y.: CONVERSE: a conversational companion, in *Machine conversations*, pp. 205–215, Springer (1999)
- [Bessho 12] Bessho, F., Harada, T., and Kuniyoshi, Y.: Dialog System Using Real-Time Crowdsourcing and Twitter Large-Scale Corpus, in *Proc. SIGDIAL*, pp. 227–231 (2012)
- [Bickmore 01] Bickmore, T. and Cassell, J.: Relational Agents: A Model and Implementation of Building User Trust, in *Proc. CHI*, pp. 396–403 (2001)
- [Higashinaka 14] Higashinaka, R., Kobayashi, N., Hirano, T., Miyazaki, C., Meguro, T., Makino, T., and Matsuo, Y.: Syntactic Filtering and Content-based Retrieval of Twitter Sentences for the Generation of System Utterances in Dialogue Systems, in *Proc. IWSDS* (2014)
- [Higuchi 08] Higuchi, S., Rzepka, R., and Araki, K.: A Casual Conversation System Using Modality and Word Associations Retrieved from the Web, in *Proc. EMNLP*, pp. 382–390 (2008)
- [Ritter 10] Ritter, A., Cherry, C., and Dolan, B.: Unsupervised Modeling of Twitter Conversations, in *Proc. HLT-NAACL*, pp. 172–180 (2010)
- [Sugiyama 13] Sugiyama, H., Meguro, T., Higashinaka, R., and Minami, Y.: Open-domain Utterance Generation for Conversational Dialogue Systems using Web-scale Dependency Structures, in *Proc. SIGDIAL*, pp. 334–338 (2013)
- [Takeuchi 07] Takeuchi, S., Cincarek, T., Kawanami, H., Saruwatari, H., and Shikano, K.: Construction and Optimization of a Question and Answer Database for a Real-environment Speech-oriented Guidance System, in *Proc. COCOSDA* (2007)
- [Wallace 04] Wallace, R. S.: *The Anatomy of A.L.I.C.E.*, A.L.I.C.E. Artificial Intelligence Foundation, Inc. (2004)
- [江頭 12] 江頭 勇佑, 柴田 知秀, 黒橋 禎夫: 雑談対話システムにおける強化学習を用いた応答生成モジュールの選択, 言語処理学会第 18 回年次大会論文集, pp. 654–657 (2012)
- [柴田 09] 柴田 雅博, 富浦 洋一, 西口 友美: 雑談自由対話を実現するための WWW 上の文書からの妥当な候補文選択手法, 人工知能学会論文集, Vol. 24, No. 6, pp. 507–519 (2009)
- [杉山 13] 杉山 弘晃, 日黒 豊美, 東中 竜一郎, 南 泰浩: 任意の話題を持つユーザ発話に対する係り受けを利用した応答文の生成, 人工知能学会研究会, SIG-SLUD, pp. 55–60 (2013)
- [杉山 14] 杉山 弘晃, 日黒 豊美, 東中 竜一郎, 南 泰浩: 対話システムのパーソナリティを問う質問に対する応答生成, 人工知能学会研究会, SIG-SLUD, pp. 33–38 (2014)
- [池原 99] 池原 悟, 宮崎正弘, 白井 謙, 横尾 昭男, 中岩 浩巳, 小倉 健太郎, 大山 芳史, 林 良彦: 日本語語彙大系, 岩波書店 (1999)
- [東中 14] 東中 竜一郎: 雑談対話システムに向けた取り組み, 人工知能学会研究会, SIG-SLUD(招待講演), pp. 65–70 (2014)