

共感表出と発話促進のための 聞き手応答を生成する傾聴対話システム

An Attentive Listening System

Generating Empathetic Responses and Promoting User Utterances

石田真也^{1*} 井上昂治¹ 中村静¹ 高梨克也¹ 河原達也¹
Ishida Masanari¹ Inoue Koji¹ Nakamura Shizuka¹
Takanashi Katsuya¹ Kawahara Tatsuya¹

¹ 京都大学 大学院情報学研究科

¹ Graduate School of Informatics, Kyoto University

Abstract: We develop an attentive listening system generating various responses and selecting an appropriate one among them. First, the proposed system generates five kinds of responses: formulaic responses, sentiment responses, repeated responses, self-disclosure, and questions for elaboration. Next, the system statistically predicts the appropriateness of each response with regard to attentive listening. Finally, the system selects among the filtered responses. To generate these kinds of responses, we use focus detection, sentiment analysis, and language model evaluation. For training of the model for appropriateness prediction, acoustic features, linguistic features, and contextual features are used. We have collected the dialogue sessions between senior subjects and the android ERICA in a WoZ setting. We have conducted an additional annotation in terms of appropriateness. Using the data set, we have conducted evaluation of the accuracy of the prediction model, and also subjective evaluation of the generated responses.

1 はじめに

近年、人と対話を行うシステムが日常生活においても盛んに利用されるようになってきている。対話の種類として、タスク遂行型の対話だけでなく雑談対話の研究も行われている [1]。雑談対話の一つとして、傾聴対話 [2] がある。システムが傾聴対話を行うことで、ユーザの話したいという欲求の充足や高齢者の認知機能の維持などが期待される [3][4]。

傾聴対話を実現するためには、ユーザが発話しやすいように、システムのふるまいを設計する必要がある。従来研究では、相槌の韻律 [5] や形態 [6] が扱われている。また、相槌以外の聞き手応答として、我々は、ユーザの発話内容に応じて、慣用的表現の語彙的応答、一部を繰り返す繰り返し応答、詳細を問う掘り下げ質問の選択と生成に取り組んできた [7]。[8] では聞き手応答を評価応答を含む 6 種に分類している。これらの知見を踏まえ、本研究では、より自然な傾聴対話を行うために、掘り下げ質問、繰り返し応答、語彙的応答、評価応答という聞き手応答に加えて、システムが自分自身に関

する話をする自己開示の生成を行う。上記の各応答を提案手法で生成したのち、ユーザ発話の音声認識、焦点解析の結果や、各応答モジュールによって生成された応答などの情報を特徴量として、それぞれの応答候補の傾聴的な妥当性の判定を行い、妥当と判定されたもののうちの一つを出力する。本稿ではシステム構成の概要、システム応答の精度評価、被験者による印象評定を行う。

2 関連研究

2.1 ルールベースの傾聴対話システム

下岡らはルールベースの傾聴対話システムを開発している [2]。このシステムでは、ユーザの発話を音声認識し、認識結果の信頼度をもとにシステムの応答の種類を決定し、各応答生成アルゴリズムに従って応答内容を生成、出力するという過程を繰り返す。応答の種類として問い返し応答、評価応答などの 4 種類が採用されているが、応答選択の基準としては音声認識の信頼度のみを考慮している。ただし、予期しないユーザ発話がなされたときに対話が破綻する可能性があり、頑

*連絡先：京都大学 大学院情報学研究科
〒606-8501 京都市左京区吉田本町 36-1
E-mail:ishida@sap.ist.i.kyoto-u.ac.jp

健性に欠ける。また、ルールベースのシステムであるため、ルールの適用基準や優先順位などに関する修正や保守が難しい。これに対して、本研究では、生成した応答内容の妥当性を3種類の基準から判断することとし、応答の種類を選択を対話コーパスに基づく統計的学習によって行い、予期しない発話に対しては無難な応答（語彙的応答）を行うことで頑健性を保つようになっている。

2.2 応答種別の選択

太田 [9] は高齢者を対象とした雑談対話における応答種別選択に取り組んでいる。この研究では先行発話の音響特徴量、言語特徴量を用いて9種類の応答種別の選択を行っている。学習や評価のデータセットとして、高齢者が語り手となって自由に語るという1対1の雑談対話を収録している。実験の結果、音響特徴量、言語特徴量を組み合わせたパターンにおいて、それぞれの応答種別のF値の平均がおよそ0.54となっている。当該手法は統計的に応答選択を行っているが、応答の生成に関しては具体的に検討されていない。また、9種類の応答がそれぞれ排他的に分類されている。これに対して、本研究では、具体的な応答生成を個別に述べた上で、収録した対話データに対して実際の応答種類以外で出力可能な種類の追加アノテーションを行うことで、応答選択の任意性を実現する。

3 傾聴システムの構成

提案するシステムにおける処理の流れを図1に示す。はじめに、ユーザ発話に対して音声認識を行う。構文解析は形態素解析をJuman¹で、係り受け解析をKNP²で、焦点解析を吉野らの手法 [10] で行う [7]。その解析結果を用いて各モジュールが独立に生成した応答文のそれぞれについて傾聴的な妥当性を判定し、妥当なものの中から一つを出力する。

3.1 システム応答の妥当性

本研究では、出力するシステム応答の妥当性を言語的、文脈的、傾聴としてという3種類の観点から考察する。

言語的妥当性は、生成した文が日本語として意味をなすかという観点から見た妥当性である。例えば「どんなアメリカですか」というシステム応答は、「アメリカ」という固有名詞に対して、その種類を問う「どん



図 1: 傾聴システムの全体像

な」という疑問詞を用いているため、日本語文として破綻している。

文脈的妥当性は、生成した応答が先行するユーザ発話の内容を考慮しているかという観点から見た妥当性である。例えば「沖縄の海に行きました」の発話に対して「どこ海ですか?」という応答は、先行発話内に特定の場所を表す地名「沖縄」があるので、文脈的に不適切である。

傾聴的妥当性は、生成した応答が傾聴対話の聞き手応答としてそのタイミングで出力すべきかという観点から見た妥当性である。傾聴的に妥当でないものの例を以下に示す。

ユーザ「初めての海外旅行はポルトガルで、」
システム「ポルトガルのどこですか?」← ユーザの続く発話を遮っている

具体例では、ユーザは「ポルトガルで。」の後に、さらなる発話をする意思があるのに、そのタイミングで掘り下げ質問を出力することで、ユーザの話を遮ってしまう可能性がある。

本研究では、言語的妥当性と文脈的妥当性は既存の辞書などを用いることで考慮し、傾聴的妥当性は3.6節で述べる応答選択器において、コーパスからの学習に基づいて、判定する。

3.2 応答文生成モジュール

本節では、本研究が提案する、語彙的応答、繰り返し応答、評価応答、自己開示、掘り下げ質問という5種類の聞き手応答の生成方法について述べる。ただし、語彙的応答は「そうですか」のみをそのまま出力し、繰り返し応答は以下の例のように、焦点単語と文末表現を用いて繰り返すものであるため、説明を省略する。

ユーザ「先週友人と沖縄に行ってきました」
システム「沖縄ですか」

¹<http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN>

²<http://nlp.ist.i.kyoto-u.ac.jp/index.php?KNP>

表 1: ラベルに対応する評価応答

	ポジティブ	ネガティブ
客観的	素敵ですね	大変ですね
主観的	いいですね	残念ですね

3.3 掘り下げ質問

掘り下げ質問は先行するユーザ発話の詳細を問うものである。本研究では、焦点単語と疑問詞との相性を単語 N-gram 確率を用いて算出し、その値に応じて応答文を生成する [7]。以下に例を示す。

ユーザ「昨日友達と海に行ってきました」
システム「どこの海ですか？」

ユーザ「昨日兄と夕食を食べました」
システム「どこで食べましたか？」

3.4 評価応答

評価応答は先行するユーザ発話に対してシステムが評価を示す応答である。本研究では小林らの日本語評価極性辞書 [11] を用いて生成する。当該辞書内には、用言・名詞の両者の各単語について、「ポジティブかネガティブか」と「客観的か主観的か」の2つのラベルが付与してある。提案手法では、ユーザ発話を文節毎に区切り、各文節にポジティブ/ネガティブ、主観/客観のラベルを当該辞書を参照しながら付与し、各発話内の多数決で発話文全体のラベルを決定する。そして、そのラベルに応じて表 1 のようなシステム発話を生成する。

3.5 自己開示

自己開示は先行するユーザ発話に関連した事象に対して、聞き手自身の考えや情報を提示する短い応答である。この自己開示は対話をより自然にする目的に加えて、直接掘り下げ質問を生成できないようなユーザ発話に対しての緩衝となり、その直後に掘り下げ質問が続くことの自然さを実現する効果が期待される。本研究では、先行発話の焦点単語と述語をテンプレートに当てはめるスロット形式を採用する。テンプレートを表 2 に示す。これらの応答の中から言語モデル的により妥当なものを選択するために、単語 N-gram 確率を用いて、その値が最大になるものを応答の候補とする。

表 2: スロット形式の自己開示文の候補

私も (焦点) が好きです
私も (焦点) なんです
私は (焦点)(述語) したことはありません
私も (焦点)(述語) してみたいです
私は (焦点) に関しては知りません

3.6 応答選択器

応答生成モジュールによって生成された 5 つの応答から妥当なものを選択する。生成された 5 種類の応答に関して、それぞれ独立に、その文脈における傾聴行動としての妥当性をコーパスに基づいて算出し、妥当と判定されたもののうちの一つを出力する。詳細は次節で述べる。本研究では、先行ユーザ発話はユーザの当該ターン内でなされた発話全てをその範囲とする。

4 応答選択器の性能評価

応答選択器の学習や評価に用いるデータセットを作成するために、対話データへのアノテーションを行った。使用した対話データは、語り手役の被験者の語りをアンドロイド ERICA [12] が傾聴する 1 対 1 の WOZ 対話である。ERICA の音声発話と非言語行動の制御は別室の操作者が行っている。今回は 1 セッション 10 分程度の対話データを 13 セッション使用する。

4.1 アノテーション

2.2 で述べた通り、聞き手応答の選択には任意性があり、実際に出力された応答のみを正解とするのは不自然であると考えられる。そこで、本研究では、コーパス内の ERICA の発話ターンにおいて提案する 5 種類の聞き手応答が傾聴対話の聞き手応答として適切 (O)、不適切 (X)、不適切とまでは言えないが妥当とは言えない (T) かの 3 段階評価値を付与した。これにより、一つの応答箇所について、各種類の応答の中から一つのみを排他的に正解とするのではなく、複数の種類の妥当な応答の情報がより利用できるようになる。アノテーションを行う発話単位は全セッション合わせて 561 箇所であった。今回のアノテーションは著者と研究室所属の学生の計 2 名が行った。2 名のアノテーションの一致率を考慮して、今回は評価値の統合方法として、両者が O のもの以外にも、片方が O でもう一方が T の場合も正解として許容することとした。

表 3: 傾聴的な妥当性の予測結果

		適合率		再現率		F 値
語彙的応答	O	382/398	95%	382/505	71%	0.85
	X	40/163	25%	40/56	71%	0.37
評価応答	O	133/182	73%	133/322	41%	0.53
	X	190/379	50%	190/239	79%	0.61
繰り返し応答	O	246/298	83%	246/360	68%	0.75
	X	149/263	57%	149/201	74%	0.64
自己開示	O	38/131	29%	38/110	35%	0.32
	X	358/430	83%	358/451	79%	0.81
掘り下げ質問	O	73/184	40%	73/174	42%	0.41
	X	276/377	73%	276/387	71%	0.72

4.2 特徴量と評価

各応答種類ごとに傾聴的な妥当性の予測モデルを作成する。学習、予測、評価は各応答種類ごとに独立に行われる。入力には以下の特徴量とする。直前のシステム応答の種類を素性の一つとする理由は、一つの種類の応答を連続して出力することの不自然さや自己開示のあとには掘り下げ質問が起こりやすいといった応答の連続性を考慮するためである。

1. ユーザ発話文の末尾の表現 (文末表現・終助詞・終助詞の有無)
2. ユーザ発話全体に含まれる内容語の数
3. 先行発話の極性値 (ポジティブ/ネガティブ, 主観/客観)
4. 一人称単語の有無
5. 焦点解析結果
6. 固有表現の有無
7. 応答生成の際の N-gram 確率
8. ユーザ発話の末尾 100ms の韻律情報
 - ・パワーの最大値, 平均
 - ・F0 の最大値, 最小値, 平均, レンジ
9. 先行ユーザ発話の全部分の韻律情報
 - ・F0 の最大値と平均値の差
 - ・パワーの最大値と平均値の差
10. ユーザ発話後からのポーズ長
11. ユーザ発話のターン継続時間
12. 直前のシステム応答の種類

出力は傾聴対話の聞き手応答として妥当かどうかを表す O/X とした。また、学習と評価は 13 セッション中の 561 発話を 5 分割した交差検定によって行う。ただし、自己開示や掘り下げ質問はアノテーション後も正解が少ないため、各クラス間に重みを付与して、実験を行った。結果を表 3 に示す。実験の結果、言語的かつ文脈的に妥当と判断され生成された応答は平均で 3.55 個であり、そのうち平均で 1.45 個が当該傾聴妥当性判定により除外されていることを確認した。

4.3 考察

表 3 より、F 値はおよそ 0.65 から 0.75 の間におさまっており、比較的精度が高く予測を行えている。しかしながら、語彙的応答の X, 自己開示の O, 掘り下げ質問の O の F 値の値が低かった。これらのうち、語彙

的応答に関しては大半の箇所です容されると考えられるため、対話が大きく破綻する可能性は低いと考えられる。

自己開示の F 値が低い要因の一つとして、傾聴対話の文脈で自己開示を出力するのが妥当かどうかを判定するための特徴量の抽出が十分に取込みできていないという点が考えられる。自己開示は、ユーザが自身の経験について述べている箇所においてユーザへの共感を示す形で行われることが多いため、先行ユーザ発話の内容上の主体がユーザ自身であるかを検知する必要があると考えられるが、現在の「一人称単語の有無」という表層的な方法だけでは不十分であると考えられる。

掘り下げ質問の O の適合率の低さは、応答内容の文脈的な妥当性を強く表す素性を使用していないためであると考えられる。現段階では、掘り下げ質問の文脈的な妥当性を推定する素性として Named Entity を用いているが、この情報だけでは不十分である可能性が高い。

5 印象評定実験

5.1 方法

各モジュールを統合したシステムの印象評定を行った。今回比較に用いたベースラインは、提案システムの処理の流れのうちから、傾聴妥当性の判定モジュールのみを取り除いたものとした。また、今回の実験で用いる音声データは、コーパス内の対話の一部 (1 分から 3 分程度) を切り取り、操作者の各発話箇所を、提案システムとベースラインシステムのそれぞれで生成した応答文を ERICA 用の TTS で音声合成したものに置き換えたものである。これらの音声対話データのペアを 8 組準備した。本実験の被験者は 8 人で、各被験者に対して以下の手順で 2 種類の評価をしてもらった。

- 評価 1: 各音声対話データの全体の印象評定 (下記の 9 項目に関して)
 1. データ A の音声を聴き、印象評定を行う
 2. データ B の音声を聴き、印象評定を行う
- 評価 2: 音声対話データ内での 2 つのシステム応答の対比較
 1. 評価 1 と同じ対話部分について、語り手の音声のみを聴く
 2. それぞれの応答箇所での応答としてシステム A と B のどちらによるものの方がより適切かを選択する

表 4: 評価 1:各項目ごとの両システムの比較

	質問 1	質問 2	質問 3	質問 4	質問 5	質問 6	質問 7	質問 8	質問 9	合計
提案システムの評価値の平均	-0.1875	0.0781	-0.1562	0.2500	0.0469	-0.4062	0.0156	-0.0625	-0.2344	-0.0729
ベースラインの評価値の平均	-0.5781	-0.5312	-0.4062	-0.3906	-0.3750	-0.8750	-0.5156	-0.2031	-0.7812	-0.5174
p 値	0.0701	0.0102	0.1687	0.0080	0.1522	0.0275	0.0535	0.6422	0.0272	0.1288

表 5: 評価 1:各対話ごとの両システムの比較

	対話 1	対話 2	対話 3	対話 4	対話 5	対話 6	対話 7	対話 8	合計
提案システムの評価値の平均	-0.1389	-0.3611	-0.4444	0.9861	0.7639	0.3333	-1.0278	-0.6944	-0.0729
ベースラインシステムの評価値の平均	0.0556	-0.0694	-1.0139	-1.3611	-0.6667	-0.0417	-0.5556	-0.4861	-0.5174
p 値	0.4706	0.1151	0.0163	0.0000	0.0000	0.0464	0.0141	0.1988	0.1077

以上の流れを対話 1 から対話 8 まで行ってもらい、各対話データで提案法とベースラインのどちらが A,B になるかをランダムに入れ替えており、さらに、半数の 4 人の被験者については、この A と B を全て逆にしたものを用いた。評価 1 の印象評価では、[13] の実験を参考に作成した次の 9 つの評価項目についてそれぞれ 7 段階 (-3 から 3) で評価してもらった。

1. 全体を通して聞き手の応答は自然でしたか
2. 全体を通して聞き手は真面目に聞いていると感じましたか
3. 全体を通して聞き手は集中して聞いていると感じましたか
4. 全体を通して聞き手は積極的に聞いていると感じましたか
5. 全体を通して聞き手は親身に聞いていると感じましたか
6. 全体を通して聞き手は理解していると感じましたか
7. 全体を通して聞き手は関心を持っていると感じましたか
8. 全体を通して聞き手は話し手に共感していると感じましたか
9. 全体を通して聞き手は語りを促していると感じましたか

5.2 結果と分析

まず、評価 1 の対話全体の印象について評価結果を提案システムとベースラインシステムで比較する。評価は、各評価項目ごとの値による比較と、各対話セッションごとの全項目の値による比較を t 検定で行う。結果をそれぞれ表 4 と表 5 に示す。

表 4 より、全ての質問項目において提案システムの平均値の方が高い値となった。そのうち、質問 2,4,6,9 では p 値が 0.05 を下回り有意差が見られたため、ベースラインシステムに比べて提案システムはこれらの項目においては印象が良いことが確認できた。また、表 5 より、対話ごとに関する評価でも、対話 3,4,5,6,7 において有意差が見られ、対話 7 以外で提案システムの評価値平均がベースラインシステムを上回った。

以上の結果より、全体的に見て提案システムの方がより高い印象評価を得たが、評価値平均が 0 を下回ることが多く、性能が十分高いとは言えない。

次に、評価 2 のシステムの応答の対比較の結果を見る。表 6 では、両システムの応答の種類全ての組み合わせについて、提案システムが被験者に選択され

た場合をカウントした結果を示している。ただし、両手法が同じ種類の応答を選択した箇所は今回の実験の評価対象から予め除外されている。

表 6 より、まず、応答の種類を区別しない場合の合計では両手法に差がなかった (183/360)。次に、提案システムとベースラインシステムがそれぞれ掘り下げ質問、繰り返し応答を選択したとき、もしくはその逆である場合の数値を見ると、それぞれ 18/24 と 19/24 となっており、どちらの場合も提案システムの印象評価の方が高いことが確認できた。両手法の差が生まれなかったのは、主に、提案システム、ベースラインシステムに関わらず、自己開示や掘り下げ質問をシステムが生成した時に被験者に選択される場合が非常に少なかったことによるものと考えられる。その原因として、当該二種類の応答が生成される際に、言語的もしくは文脈的に不適切なものが出力されていることがあげられる。したがって、より無難な種類の応答を選択したもう一方の手法による出力の方が適切であると判断されたと推測できる。

そこで、自己開示や掘り下げ質問のうちで生成された表現が妥当でないものを言語的もしくは文脈的に妥当性に欠けるかという基準で取り除き、再び印象評価の比較を行った結果が表 7 になる。表 7 から分かるように、提案システムが掘り下げ質問を行った場合の選択割合は改善された。しかし、他の箇所については、内容が破綻しているものを除いても選択割合はほとんど変化しなかった。したがって、当該 2 種類の応答が被験者に選択されない理由が個々の応答の内容を超えたところにもあると考えられる。その一つの要因が、今回の印象評定で用いたデータは話し手の発話と聞き手の応答でのみ構成されており、相槌が一切含まれていなかったため、相槌に近い語彙的応答や評価応答もしくは繰り返し応答といった軽い応答の必要性を被験者が感じていたということが考えられる。ただし、以下の例に見られるように、提案システムの掘り下げ質問が肯定的に評価されている場合もある。

表 6: 評価 2: 両システムの応答の一対比較

		提案システム (分子は提案法が選択された度数)					平均
		語彙的応答	評価応答	繰り返し応答	自己開示	掘り下げ質問	
ベースライン	語彙的応答	0/0	32/40	16/40	0/8	3/16	51/104
	評価応答	1/32	0/0	8/24	2/8	3/24	14/88
	繰り返し応答	7/8	0/0	0/0	3/24	18/24	28/56
	自己開示	28/32	0/0	8/8	0/0	5/8	41/48
	掘り下げ質問	12/16	15/16	19/24	3/8	0/0	49/64
	合計	48/88	47/56	51/96	8/48	29/72	183/360

表 7: 評価 2: 両システムの応答の一対比較 (破綻応答を除いた場合)

		提案システム					平均
		語彙的応答	評価応答	繰り返し応答	自己開示	掘り下げ質問	
ベースライン	語彙的応答	0/0	32/40	16/40	0/8	3/16	51/104
	評価応答	1/32	0/0	8/24	2/8	0/8	11/72
	繰り返し応答	7/8	0/0	0/0	2/16	18/24	27/48
	自己開示	13/16	0/0	0/0	0/0	5/8	18/24
	掘り下げ質問	6/8	8/8	0/0	0/0	0/0	14/16
	合計	27/64	40/48	24/64	4/32	26/56	121/264

話し手「タイトルをつけるとするならば
異文化交流ですか」
提案システム「どんな異文化
交流ですか？」(6人選択)
ベースラインシステム「異文化
交流ですか」(2人選択)

6 おわりに

本稿では、傾聴対話システムの構築について述べた。具体的には、傾聴対話の聞き手応答として必要と考えられる語彙的応答、評価応答、繰り返し応答、自己開示、掘り下げ質問の5種類の各応答の生成方法を述べ、また、それらの各応答の言語的、文脈的妥当性を判定し、傾聴的な妥当性を考慮した応答選択を行う手法を提案した。また、この手法の性能評価、選択された応答について被験者による印象評価を行った。その結果、性能評価では、傾聴として妥当な応答の予測が比較的精度よく行えることが確かめられた。また、印象評価の結果、提案システムがベースラインの印象評価を全体的に上回り、一部の応答種類において適切に傾聴妥当性の判定を行えていることが確認された。今後は、自己開示や掘り下げ質問の応答文生成の精度の向上に取り組むことに加え、対話システムとして振る舞う場面では、相槌生成システム [13] などと組み合わせ、より統合的な対話制御、応答生成手法を構築していく必要がある。

謝辞

本研究は、JST ERATO 石黒共生ヒューマンロボットインタラクションプロジェクトの支援を受けて実施されたものである。

参考文献

- [1] 河原達也. 音声対話システムの進化と淘汰: 歴史と最近の技術動向. 人工知能学会誌, Vol. 28, pp. 45–51, 2013.
- [2] 下岡和也ほか. 音声対話ロボットのための傾聴システムの開発. 言語・音声理解と対話処理研究会, Vol. 58, pp. 61–66, 2010.
- [3] R. Butler. The life review: An interpretation of reminiscence in the aged. *Psychiatry*, Vol. 26, pp. 65–76, 1963.
- [4] 大武美保子ほか. 回想法から見た共想法の考察と連携の可能性. 人工知能学会全国大会, 6 2010.
- [5] 上里美樹ほか. 傾聴対話における相槌の韻律的特徴の同調傾向の分析. 言語・音声理解と対話処理研究会, Vol. 70, pp. 7–13, 2014.
- [6] 山口貴史ほか. 傾聴対話システムのための言語情報と韻律情報に基づく多様な形態の相槌の生成. 人工知能学会研究会資料, Vol. 76, pp. 49–54, 02 2016.
- [7] 石田真也ほか. 傾聴対話システムのための発話を促す聞き手応答の生成. 人工知能学会研究会資料, Vol. 77, No. 01, pp. 1–6, 08 2016.
- [8] 伝康晴. 対話への情報付与. 小磯花絵 (編), 日本語コーパス 3: 話し言葉コーパス-設計と構築, pp. 101–130. 朝倉書店, 東京, 2015.
- [9] 太田健吾, 西村良太, 北岡教英. 発話の分散表現に基づく雑談音声対話システムの応答種別選択. Technical report, 電子情報通信学会, 2017.
- [10] K. Yoshino, et al. Conversational system for information navigation based on pomdp with user focus tracking. *Computer Speech Language*, Vol. 34, pp. 275–291, 2015.
- [11] 小林のぞみほか. 意見抽出のための評価表現の収集. 自然言語処理, pp. 1483–1486, 2005.
- [12] K. Inoue, P. Mihorat, D. Lala, T. Zhao, and T. Kawahara. Talking with erica, an autonomous android. *17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pp. 212–215, 2016.
- [13] 山口貴史, 井上昂治, 吉野幸一郎, 高梨克也, G. Nigel, 河原達也. 傾聴対話システムのための言語情報と韻律情報に基づく多様な形態の相槌の生成. 人工知能学会論文誌, Vol. 31, No. 4, pp. 1–10, 2016.