

修　士　論　文

題　目

人型ロボットを用いた対話 AI の構築

指導教員

竹内 孔一

報　告　者

山崎 瑶

岡山大学大学院自然科学研究科電子情報システム工学専攻

令和 4 年 2 月 3 日提出

修　士　論　文

題　目

人型ロボットを用いた対話 AI の構築

指導教員

報　告　者

山崎 瑶

岡山大学大学院自然科学研究科電子情報システム工学専攻

令和 4 年 2 月 3 日提出

要約

目次

1	はじめに	1
1.1	研究背景	1
1.2	研究の位置付け	1
1.3	研究目的	2
1.4	本論文の構成	2
2	関連研究	3
2.1	音声対話システムに関する先行研究	3
2.2	対話システムの類型	3
3	対話システムの構築	4
3.1	ロボット管理	4
3.1.1	音声認識	4
3.1.2	音声合成	5
3.1.3	口形状生成	6
3.1.4	首動作・状態姿勢生成	7
3.1.5	表情生成・動作統合	7
3.2	対話管理	7
3.3	言語理解	9
3.3.1	フレーム表現を用いた発話理解	9
3.3.2	行動選択	10
3.4	対話を円滑に進める機構	10
3.4.1	対話破綻の抑止	10
3.4.2	観光地に関する情報の拡充	12
4	評価実験	13
4.1	実験設定	13
4.1.1	旅行代理店対話タスク	13
4.1.2	体験やへの事前の指示	14
4.1.3	レギュレーション	14
4.2	対話の評価手法	17
4.2.1	対話者の評価手法	17
4.2.2	ビデオ評価の評価手法	17
4.3	実験結果	18
4.3.1	対話の結果	18

4.3.2 アンケート評価の結果	21
5 考察	23
5.1 対話ロボットに関する考察	23
5.2 対話制御に関する考察	23
6 まとめ	24
謝辞	25
使用したツール	25
参考文献	25

1 はじめに

1.1 研究背景

近年，Apple 社の Siri などのスマートフォン上で動作する音声エージェントサービスや，Amazon 社の Alexa などのスマートスピーカーに見られるように，ユーザの発言に対して正しい応答や人間らしい応答を返すための音声対話技術がめざましく進歩している。今後，このような対話システムは，様々な形で，私達の日常活動を支援するものと期待される。^[1] ^[2]しかし，日常活動での対話は，スマートスピーカーとの対話よりも複雑であり，現在の音声対話技術でも，様々な状況においてうまく対話を継続して目的を達成するのは困難である^[3]。例えば対話を通した接客業務では，話し方や要望の出し方など対話相手によって様々であり，それに適切に対応する必要がある。このとき，私達人間であれば，対話相手のタイプによって話し方を切り替えて対応したり，音声だけでなく視線や表情などもうまく使って対話を継続することができるが，音声対話システムではこのような対応を取ることができない。人型ロボットは，様々なセンサを用いてユーザの音声だけでなく表情やしぐさなどを認識できたり，体を用いてジェスチャや表情など様々な表現ができる。音声対話システムよりも複雑である反面，多くの情報や多くの表現手段を用いることで，従来の対話システムとは異なる方法で対話をうまく継続を実現できる可能性がある。本研究では音声，表情，身振り，手振りなどのモダリティを用いて円滑に対話を進めるマルチモーダル音声対話システムの構築を目指す。

1.2 研究の位置付け

情報化社会の次に来るであろう人間と知能ロボットや情報メディアが共生する社会を実現するため，「人間機械共生社会を目指した対話知能システム学（対話知能学）」^{*1}が文部科学省科学研究費助成事業「新学術領域研究」として2019年に創生された。対話内容を完全に理解できずとも違和感なく対話を継続できる能力を実現する「対話継続関係維持」，特定の状況において特定の目的に関して対話理解と対話生成を組み合わせた対話能力を実現する「対話理解生成」，システムが自らの行動決定モデルを構築したり相手の行動決定モデルを推定する機能を実現する「行動決定モデル推定」，実証実験を通して意図や欲求を持つロボットの人々への影響を研究，ロボット共生社会における社会規範を提案する「人間機械社会規範」の4つの軸を元に，人間と機械や情報メディアが互いの意図や欲求を推定し合いながら関わり合う社会の実現を目指す。複数の情報を統合し状況に適応的にふるまうマルチモーダル音声対話システムの構築は「対話理解生成」に相当する。

^{*1} <https://www.commu-ai.org/>

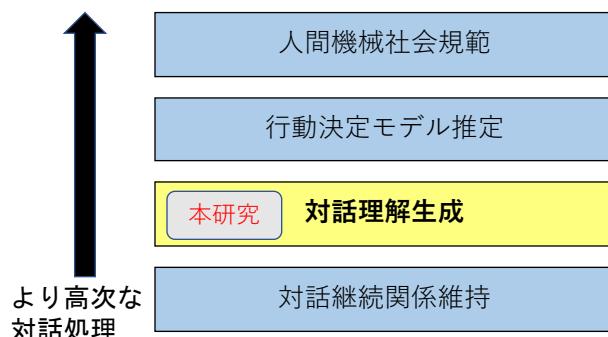


図 1 対話システム構築の位置付け

1.3 研究目的

本研究では、旅行代理店業務において、カウンターセールスを行う音声対話システムの構築を目指す。対話システムは、音声、表情、身振り、手振りなどのモダリティをもち、身体性も有する。複数のモジュールを用いてロボットの身体管理、対話管理を行うことでロボットと音声で対話実現をすることで、テキストだけの対話システムや音声だけの対話システムに比べ、ホスピタリティの高い対話を実現することを目標とする。

1.4 本論文の構成

以下に本論文の構成を記述する。2章では、音声対話システムの関連研究について説明する。3章では提案手法の概要とシステムのモジュールについて説明する。4章では対話ロボットコンペティションにおいて実験について述べる。5章で構築した対話システムに関する考察を行い、6章でまとめについて述べる。

2 関連研究

本章では、音声対話システムに関する先行研究について述べる。

2.1 音声対話システムに関する先行研究

対話システムは、数度のフェーズを経て、古いものでは 1950 年代後半から開発が行われている。^[4] 初期の対話システムとして Weizenbaum の ELIZA^[5] や Winograd の SHRDLU^[6] が挙げられる。ELIZA は手作業で作成した If-Then ルールに基づき挙動するシステムで、簡単な単語の一致によるパターンマッチングにより動作する。SHRDLU は対話により積み木を動かすというシステムで、「赤いブロックを左に移動して」という発言に対し、赤いブロックが 2 つある場合「どちらの赤いブロックですか？」等の問い合わせができる。これらの 2 つのシステムは限定的な状況でしか問題を解決することができず、現実の問題を解決するには至っていない。1980 年代になると、データベースをもとにしたエキスパートシステムが台頭する。Buchanan らの MYCIN^[7] は医療診断システムとして専門医に若干劣るほどの診断を可能とした。Bobrow らの GUS^[8] はフレーム表現と呼ばれる知識構造を利用し、ユーザの発話から発話理解を行う。フレーム表現は今日の対話システムでも利用されることが多く、本実験でもフレーム表現を利用して構築を行う。エキスパートシステムのデータベース作成は全て手作業で行われるため、専門知識を体系的にシステムに落とし込む作業がボトルネックとなった。2000 年代になると機械学習の発展により、事例から対話を生成することが可能となる。Wu らの TOD-BERT^[9] は 10 万件を超える対話事例から対話を自動生成する。機械学習により音声認識や音声合成などの精度も飛躍的に向上し、音声認識や音声合成を用いた対話システムの構築が積極的に行われている。^[10]

2.2 対話システムの類型

対話システムは以下のような観点で分類することができる。

- タスクの有無
- ユーザの人数
- モダリティ
- 主導権
- 身体性の有無

3 対話システムの構築

本章では、対話システムの構築方法について述べる。対話システムは、ロボットのハードウェアを制御するロボット管理と対話の制御を行う対話管理に分けることができる。対話システムの全体像を図 2 に示す。ロボット管理と対話管理の通信は全て TCP で行う。ロボット管理の各モジュールではそれぞれの部位に対応するロボットの制御を行う。3.1 節で各モジュールの詳しい説明を行う。対話管理ではユーザの発話を理解し、システムの発話や行為を決定する。3.2 節で対話管理について詳しく述べる。

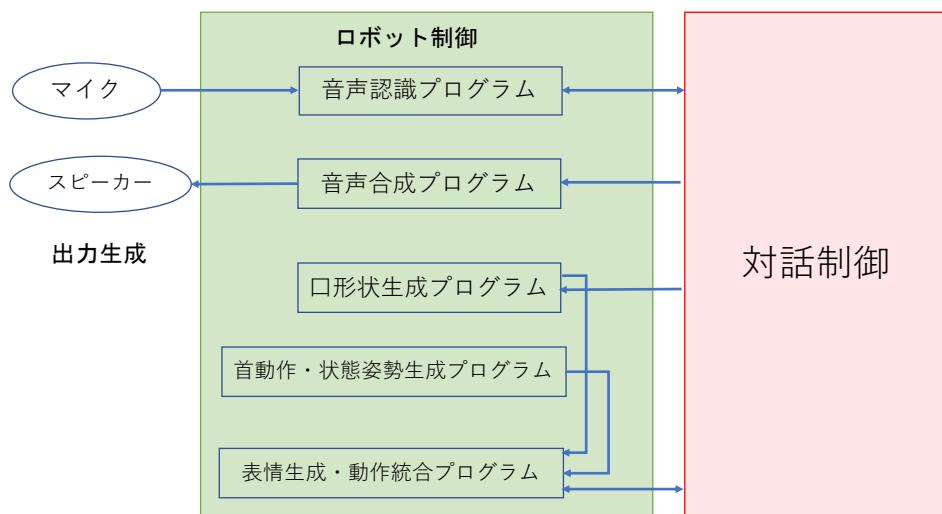


図 2 対話システムの全体像

3.1 ロボット管理

対話ロボットとして、国際電気通信基礎技術研究所 (ATR)^{*2}が管理するアンドロイドを使用する。身長 165 センチ、体重 38 キログラムであり、空気圧アクチュエータで駆動する。本節でロボットを制御するモジュールを詳しく述べる。

3.1.1 音声認識

音声認識には Google の Speech-to-Text を利用する。文字コードは utf-8 で、音声認識を開始してからマイク入力があると、途中結果を返しながら最終結果と信頼度を返すストリーミング形式の通信を行う。コマンドのフッターは改行コードである。送信プロトコル、受信プロトコルを表 3 に示す。

^{*2} <https://www.atr.jp/>



図3 アンドロイドの近景

表1 送受信プロトコル

送信プロトコル	受信プロトコル
音声認識スタート	start\n
音声認識ストップ	stop\n
	音声認識途中結果 interimresult:
	音声認識最終結果 result:
	最終結果の信頼度 confidence:

一度音声認識を開始すると音声認識を終了するまでずっと音声認識をする。受信サンプルを時系列順に並べると以下のようになる。

3.1.2 音声合成

音声合成には Amazon の Amazon Polly を使用した。通信方式は同期通信で、全て JSON 形式でやりとりが行われる。文字コードは utf-8 で、コマンドのフッターは改行コードである。再生に成功した場合、のように、音声の再生開始、終了と音声の再生時間を受信する。失敗した場合は、となる。また、

で現在発話中かを確認することができ、

で再生中に音声を停止することができる。音声合成に使用するパラメータを表に示す。

システムに標準のピッチ、声量、大きさで「コマンドサンプル」と発話させるためのコマンドの例を以下に示す。

```

startrecog:
interimresult:今日は
interimresult:今日は
interimresult:今日は
interimresult:今日は週
interimresult:今日は中止
interimresult:今日は中止度
interimresult:今日は修士論文
nterimresult:今日は修士論文発
interimresult:今日は修士論文発表
interimresult:今日は修士論文発表会です
result:今日は修士論文発表会です
confidence:0.9313719153404236

```

{(送信) 再生コマンド →
(受信) {"result":"success-start", "duration":12491}
→ (受信){"result":"success-end"}

{(送信) 再生コマンド → (受信){"result":"failed"}

(送信){"engine":"ISSPEAKING"} →(受信){"isSpeaking":true/false}

(送信){"engine":"STOP"} → (受信){"result":"success-end"}

{"engine":"POLLY", "speaker": "Mizuki", "pitch": 100, "volume":100,
"speed":100,"vocal-tract-length":0, "duration-information":false,
"speechmark":false, "text":"コマンドサンプル"}\n

3.1.3 口形状生成

口形状生成には，Oculus の Oculus Lipsync Unity を利用する。Oculus Lipsync Unity はマイク入力やオーディオファイルからのオーディオインプットストリームを分析し，口形素と呼ばれる唇や顔の表情の一連の値を予測するソフトウェアである。プログラムにアクセスする必要はなく，合成音声を再生すると，それに同期するようにアンドロイドの口形状を制御する。

3.1.4 首動作・状態姿勢生成

首動作・状態姿勢生成は非同期通信で行い、コマンドのフッターは改行コード、文字コードは utf-8 である。視線、上体姿勢、感情状態を任意のタイミングで指令することができる。ロボットの正面方向を z 軸、右を x 軸、上を y 軸、ロボットの腰の下を原点とした座標系で、x,y,z で視線、顔、体を向ける座標をメートル単位で指定することができる。ロボットの視線を、正面の 1.5 メートル先、高さ 1.2 メートルの位置に向ける際のコマンドを以下に示す。“translateSpeed”で表される移動速度の単位はメートル毎秒である。また、座

```
EyeController={"id": "EyeController","motionTowardObject": "",  
"targetMotionMode": 2,"targetPoint": {"x": 0.0,"y": 1.2,"z": 1.5},  
"translateSpeed": 2.0}\n
```

標指定で動作をせず、事前に定義されたジェスチャを指示することもできる。定義されたジェスチャでは両手の指の開閉や、頭のみ、首のみ、全身を使ったお辞儀、首や視線の縦振り、横振りを指示することができる。

3.1.5 表情生成・動作統合

表情生成・対話生成は非同期通信で行い、コマンドのフッターは改行コード、文字コードは utf-8 である。事前に定義された表情のラベルにより、頬、口角、眉毛などの位置を制御し、表情を変化させる。表情のラベルを図 2 に示す。瞬きは支持する必要がなく、定感覚で自動で行われる。

表 2 定義された表情ラベル

表情名	説明
MoodBasedFACE	基本の表情
fullsumile	笑ったような表情
angry	怒ったような表情
bad	印象の良くない表情
mouth-a/i/u/e/o	それぞれの母音を発する時の表情

3.2 対話管理

対話管理では、音声認識した発話を文字列で受け取り、発話理解をし、対話のフローとフレーム表現を含めた内部状態の更新、参照を繰り返しながら行動選択、発話生成を行う。本研究で構築する対話システムは「旅行先の決定」というゴールを持ったタスク指向型対話システムである。対話管理で構成する対話の全体の流れを図 4 に示し、本システムの想定する対話の流れを詳しく述べる。

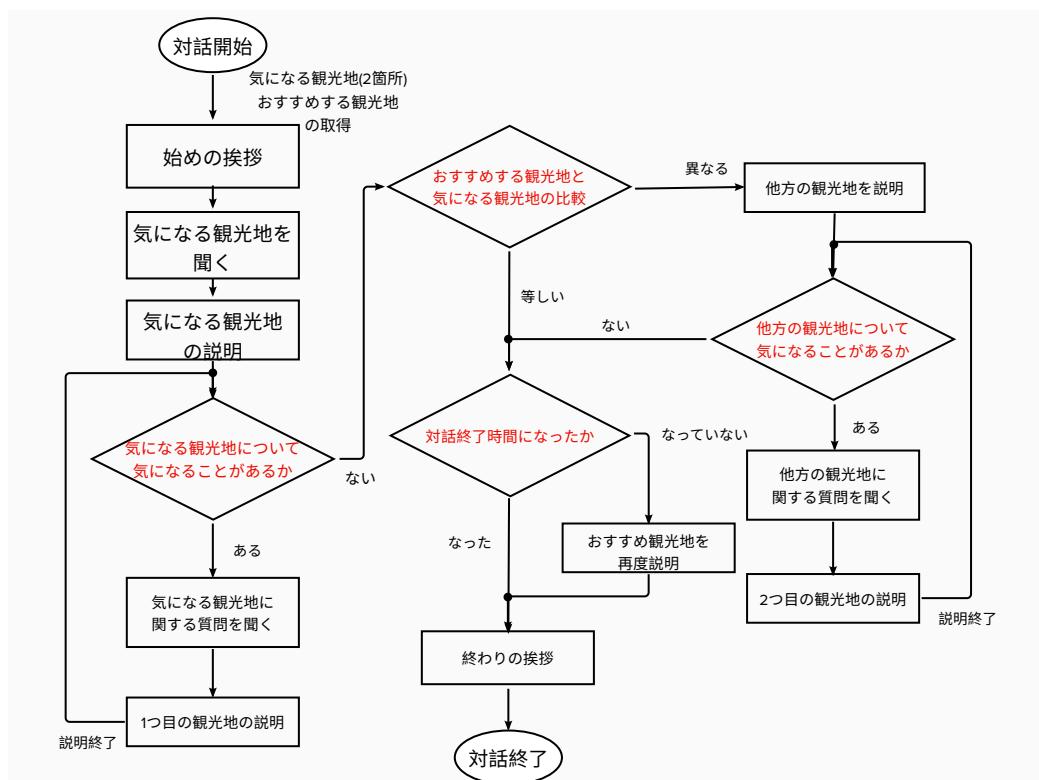


図4 対話の流れ

1. 対話開始

対話を開始すると、事前にID管理サーバへアクセスし、対話者の選択した2つの観光地と、おすすめ観光地（体験者に選ばせればコンペティションの評価の上がる運営がランダムに選定した勧めるべき観光地。）を取得する。選択した観光地についてどちらの観光地が気になるかを尋ね、気になる観光地の基本情報を説明する。

2. 気になる観光地に関する質疑応答

基本情報の説明ののち、「観光地に関して何か気になることはないか」と尋ね、質問がなくなるまで対話者主導で複数回の質疑応答を行う。

3. 気になる観光地とおすすめ観光地の比較

質問がない場合、気になる観光地とおすすめ観光地の比較を行う。気になる観光地とおすすめ観光地が等しい場合、他方の観光地の説明は行わず、再度気になる観光地についての簡単な説明を行い対話を終了する。気になる観光地とおすすめ観光地が異なる場合は他方の観光地の基本情報を追加で説明する。

4. 他方の観光地に関する質疑応答

基本情報の説明ののち、「観光地に関して何か気になることはないか」と尋ね、質問がなくなるまで対話者主導で複数回の質疑応答を行う。

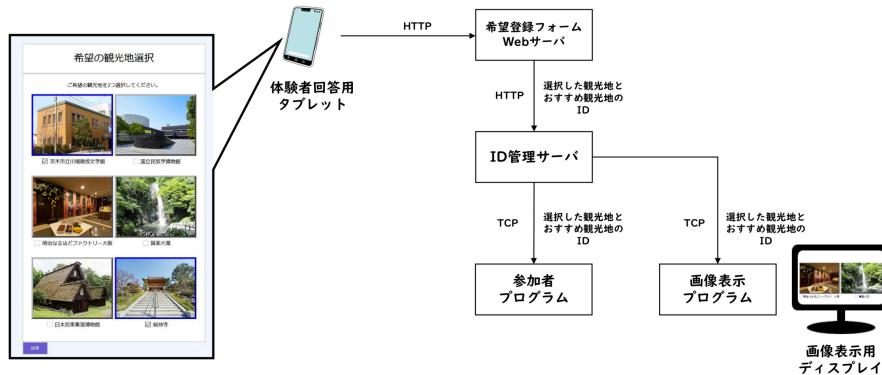


図5 観光地の選択システム

5. 対話終了

他方の観光地について質問が無くなった場合，対話の終了フェーズに入る．定められた対話時間がまだ残っている場合，再度気になる観光地についての簡単な説明を行い対話を終了する．対話時間が来た際は質疑応答の途中でも終わりの挨拶をして対話を終了する．

3.3 言語理解

言語理解では対話者の発言から対話行為を推定する．推定には単純な文字列マッチングを用いる．MeCabを用いた形態素解析を行い，ユーザの発言と対話行為タイプの一致判定により対話行為タイプを決める．

3.3.1 フレーム表現を用いた発話理解

本システムは「フレーム」と呼ばれるデータ構造を用いて対話を進める．図6に示すように，対話者が事前に選択した2つの観光地，おすすめする観光地，現在話題になっている観光地，対話履歴（観光地とその観光地の何について対話したかの組）を保持する．この例では，現在箕面大滝について話しており，すでに観光地の説明，アクセス方法，車での行き方，駐車場について対話が行われたことがわかる．このようなフレーム表現を用いることで「どのようにすれば行けますか？」という質問に対して，現在話題に上がっている箕面大滝への行き方を尋ねられたとして発話の理解を行うことができる．話題が国立民族学博物

館は話題の項目を国立民族学博物館へと変更し対話を進める。この時、「どのようにすれば行けますか?」と同じ質問をされても、話題に上がっている国立民族学博物館への行き方を尋ねられたと理解することができるため、適切な返答をすることができる。

```
[選択された観光地 : 国立民族学博物館, 箕面大滝
 おすすめ観光地   : 国立民族学博物館
 現在の話題       : 箕面大滝
 対話履歴         : [箕面大滝, 説明], [箕面大滝, アクセス], [箕面大滝, 車],
                      [箕面大滝, 駐車場].....
]
```

図 6 フレーム表現の例

私たちは対話の中で主語の省略や指示代名詞をよく使う。例えば、「私は箕面大滝へ行きたいです。そこへはどうやっていきますか。駐車場はありますか。」という文書の場合、「そこ」は直前に述べたを指し、駐車場はありますかという文書には箕面大滝にという語が隠れている。人間はその場の状況を参照し文書の理解を試みるが、システムがなんの情報もなくこの文書を理解することは難しい。このような照応表現の解析は自然言語処理における重要な課題として挙げられ[11]、対話システムの構築においても対話の中での照応解析は大きな問題である[12]。フレーム表現を利用してことで、旅行代理店におけるカウンターセルバースという限られた条件の中で、照応解析を実現し、主語の省略や代名詞の使用に対応した人間らしい会話の実現を可能にする。

3.3.2 行動選択

内部状態を参照しながら発話理解を行ったのち、行動選択に移る。

3.4 対話を円滑に進める機構

3.4.1 対話破綻の抑止

対話システムがユーザに対して不適切な返答をしてしまう「対話破綻」と呼ばれる現象が起こることがある。対話破綻は大きく4つに分類することができる[13]。

1. 構文などの崩れにより、日本語として成立せず発話そのものが破綻している場合
2. 日本語としては正しいが、相手の発言に対する「応答」として成立せず破綻している場合
3. 単発のやりとりとしては成立しているが、既に話した内容と異なる発話をするため、「文脈」が破綻している場合
4. 社会通念や倫理的におかしな発言をしてしまう場合

体験者の発話から、返答を生成して応答せず、あらかじめ用意された返答を返すため、1つめと4つめの対話破綻は未然に防ぐことができる。また、フレーム表現を用いて対話履

表 3 発話行為タイプ

発話行為タイプ	説明	発話行為タイプ	説明
explanation	概要の説明	parking	駐車場の有無
address	住所	child	子供も楽しめるか
time	営業時間	near	近隣の飲食店情報
open	営業開始時間	event	直近のイベント情報
close	営業終了時間	covid	covid-19 情報
close_day	定休日	thanks	お礼
tell	電話番号	yes	はいと言われた時
access	どうやっていくか	no	いいえと言われた時
by_train	電車でのアクセス方法	stay	観光地を迷ってる時
by_car	車でのアクセス方法	decision	観光地を決める時
photo	写真撮影の可否	error	対話理解ができなかった時
height	観光地の高さ	benefit	ご利益 (総持寺)
width	観光地の広さ	goshuin	御朱印 (総持寺)
seasons	訪問におすすめの季節	swim	遊泳の可否 (箕面大滝)
tourists	年間観光客数	choco	チョコレートの飲食
price	発生する料金		(明治なるほどファクトリー大阪)

歴を参照しながら対話を進めるため、発話の整合性を保ちながら対話を進め、3つめの対話破綻を防ぎながら対話を進める。

2つめの、日本語としては正しいが、相手の発言に対する「応答」として成立せず破綻している場合の典型的な例として、音声認識や発話認識、行動選択をしている間に対話者が新たな発話をすることで、発話認識、行動選択を再度開始してしまうといったケースが挙げられる。この場合、システムは発話すべきテキストを複数抱えてしまい、円滑に対話が進まなくなる。発話内容とアンドロイドの姿勢・表情の2つの面からこの破綻を防ぐ。

発話内容 システムが発話を終了する際は全て疑問文で終わる。システムの発言する順番と対話者が発言する順番を明確にすることで、発話認識、行動選択をしている間に対話者が新たな発言をすることを防ぐことができる。

アンドロイドの姿勢・表情 システムが発話を終了した際に音声認識を完了した際に笑顔で頷く。対話者にシステムの発話を終了したことや、音声が正常に認識されたことを表情により示すことで、対話者が不要な発言をすることを防ぐことができる。

3.4.2 観光地に関する情報の拡充

6箇所の観光地に関する基本情報として、運営から、るるぶ DATA^{*3} ^{*4}を使用した観光案内情報を付与されたが、情報量が不十分であったため、Google Maps API と手作業による観光地情報の拡充を行なう。

事前に Google Maps API により観光地周辺の飲食店について情報を収集し、データベースを作成することで、観光地の周辺情報について聞かれた際の返答を用意した。また、「箕面大滝で遊泳することは可能か」、「明治なるほどファクトリーでチョコレートを試食することは可能か」など、各観光地に対して事前に想定される質問への回答を手作業でいくつか用意することで対話の成功確率の向上を試みた。

^{*3} 旅行ガイドブック「るるぶ情報版」掲載の観光情報コンテンツのデータベース。

^{*4} <https://solution.jtbpublishing.co.jp/service/domestic/>

4 評価実験

本章では3章で説明した対話システムについて、大阪府吹田市にある大型複合施設EXPOCITY内にあるショッピングモール「らぼーとEXPOCITY」で行われた対話ロボットコンペティション^{*5}に参加し、評価実験を行う。

4.1 実験設定

4.1.1 旅行代理店対話タスク

対話ロボットコンペティションでは、旅行代理店における対話タスクとして、カウンターセールス役となったロボットが、対話を通してお客様役である対話者の要望に応える。体験者は図7に示す通り、「日本民家集落博物館」、「茨木市立川端康成文学館」、「総持寺」、「日本民族学博物館」、「箕面大滝」、「明治なるほどファクトリー大阪」の6箇所の観光地候補から行きたい2箇所決め、そのどちらに行きたいをシステムとの対話を通して決める。



図7 観光地の候補

^{*5} <https://sites.google.com/view/crobotcompetition>

4.1.2 体験やへの事前の指示

ロボットと対話する体験者は、当日らぼーと EXPOCITY を訪れた買い物客であるため、実際に対話をするにあたり、以下のような事前の指示を行う。

- 日本語でロボットと対話して頂きます。
- お客様役として振る舞って頂きます。
 - Expo City に休暇で訪れる予定を持っており、その近辺で 1 日遊びに行く観光地を決める目的を持っているつもりになって、カウンターセールス役のロボットに相談してください。
 - 対話をを行う前に、近辺の観光地候補の中から行ってみたいと思う観光地を 2 箇所選んでください。カウンターセールス役のロボットと相談して、体験者自身がお金を払ってでも遊びに行きたいと思える観光地をその 2 箇所の中から 1 箇所を決めてください。
 - 2 つの観光地に関する情報をまんべんなく確認して、行きたい観光地を決めてください。
- 対話中の注意事項
 - カウンターの椅子に座ってから対話を始めてください。
 - ロボットと相談する時間は最大 5 分間です。5 分経過すると対話を終えて頂きます。5 分経過する前に行きたい場所が決まった場合は、カウンター上のタブレットの「行きたい場所を決めた」ボタンにタッチして対話を終えてください。
 - 相談が終わった後、カウンター上のタブレット上に、ただいまの対話についての質問が表示されますので、それについて回答して頂いた後、行きたい観光地（強いて行きたいならどこか）を選んでください（対話を通して観光地の情報がうまく聞き出せなかった場合などで選ぶのが難しい場合は「選べない」を選択して頂くことも可能です）。
 - 最後に、体験についてのアンケートについてタブレットで回答して頂きます。

4.1.3 レギュレーション

対話ロボットコンペティションではレギュレーションが定められており、このレギュレーションに従って開発、実装を行なう。

1. 対話状況
 - 体験者とロボットは 1 対 1 で対話する。
 - カウンターテーブルの大きさ、ロボットと体験者の椅子の位置は固定。
 - テーブル上に体験者とロボットが同時に見ることができるようにモニタを設置（位置と向きは固定）。
2. ロボットに行わせてもよいこと

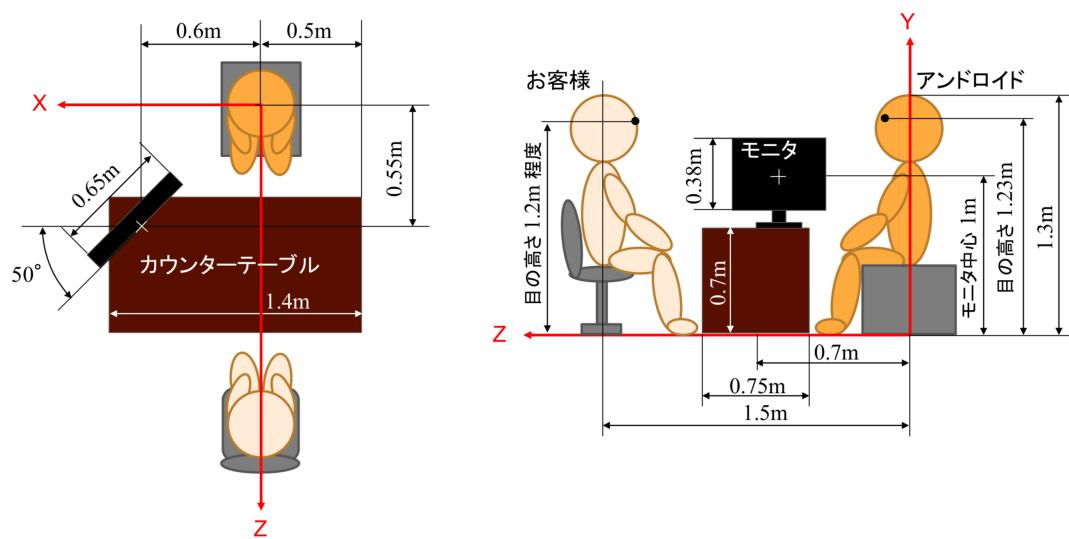


図 8 体験者とロボットの位置関係

- 任意のタイミングで発話させること .
- 任意のタイミングで , 視線 , 表情 , 頭部 , 上体等を動かすこと .
- 貸与された観光地情報を用いて説明すること .
- カウンター上のモニタに表示された観光地の写真について説明すること .
- 説明している観光地についての感想や意見を言うこと . ただし , 開発指針として , 本番で未知の観光地情報が与えられたとしても対応できるようにシステムを開発すること .

3. ロボットが対話中に使える情報

- 参加登録時に貸与される 12箇所の観光地に関する観光案内情報 .
- マイクとカメラで認識された , 体験者の音声 , 表情 , 性別 , 年齢 .
- モニタの位置 (置いてある場所) の情報 .

4. 観光地情報の扱い方

- 観光地情報データベースのレコードをプログラムで自由に操作して利用してよいものとする .
- 参加登録時に , 本番で使用する観光地情報を渡すが , 開発指針として , 本番で未知の観光地情報が与えられたとしても対応できるようにシステムを開発する

こと。

5. 提案する観光地とお薦めの観光地

- 提案する観光地:対話開始前に、体験者は日本科学未来館近辺の 6箇所、あるいは Expo City 近辺の 6箇所の観光地の中から、行ってみたいと思う観光地を 2つ選ぶ。
- お薦めの観光地:2箇所の中からランダムに決定。

6. 対話タスクの開始・終了

- タスク開始前に、体験者から聞き出した 2か所の観光地 (A, B)、その中からランダムに決定したお薦め観光地を参加者のプログラムに入力。
- テーブルのモニタ上に、観光地 A と B の写真を並べて出力 (A が左, B が右) モニタへの出力は主催者側で実行
- 体験者が椅子に着席した状態で、参加者のプログラムを開始。
- 対話開始から 5 分経過した時点、あるいは体験者がテーブル上のタブレットで観光地を決めたことを知らせてきた時点で、参加者のプログラムに対話終了。命令を入力し、対話タスクを終了。対話開始から 5 分経過しても対話中である場合は、体験者にタブレットで対話を終えるように伝える(主催者側が用意)。
- 対話の始まりは、ロボットから話始めても、お客様の話始めを待っても、どちらでもよいとする。
- 対話終了後に、体験者にタブレットで観光地を選んでもらう。

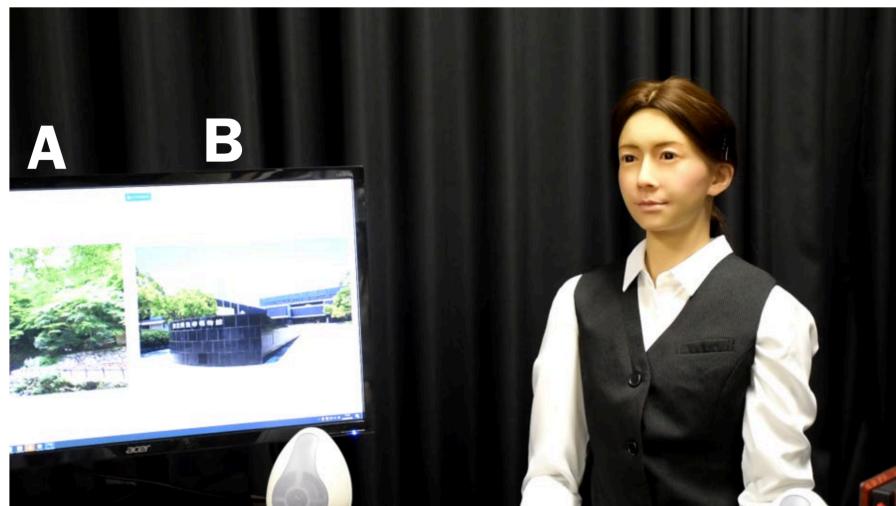


図 9 観光地の出力

7. ロボットに行わせてはいけないこと

- 対話以外の方法で、お薦めの観光地を体験者に選んでもらうこと（例：ロボットが、体験者に、お薦めの観光地を選んでもらったら賞品をあげるなどと言う）。

4.2 対話の評価手法

下記の 2 つの観点から、システムの総合的な評価を行う。

- カウンターセールスとしてお薦めの観光地を選んでもらえたかどうか
- 体験者の満足度（対話後のアンケートによる評価）

アンケート評価の評価項目は、下記の 8 つの観点である。1 点の「そう思わない」7 点の「そう思う」までの 7 段階で評価を行う。

4.2.1 対話者の評価手法

1. 満足して遊びに行く観光地を選ぶことができましたか？（選択の満足度）
2. 観光地の情報を十分に聞くことができましたか？（情報の十分さ）
3. ロボットとは自然に対話できましたか？（対話の自然さ）
4. ロボットの対応は適切でしたか？（対話の適切さ）
5. ロボットとの対話に満足しましたか？（対話の満足度）
6. ロボットの対応は好ましいものでしたか？（対応の好ましさ）
7. 観光地を選ぶのにロボットから得られた情報を参考にしましたか？（情報の参考度）

4.2.2 ビデオ評価の評価手法

COVID-19 の影響で対話者が想定より集まらず、チーム間での対話者の人数にばらつきがあるため、参加チームごと予選会場での対話を記録した映像を用いて、クラウドで追加の印象評価を実施する。評価項目は、下記の 8 つの観点である。1 点の「そう思わない」7 点の「そう思う」までの 7 段階で評価を行う。システムが正常に動かなかった場合などを評価から排除できるよう、各チームが 3 つの対話映像を選択し、それらを第三者に評価させる。

- お客様とロボットの対話を第三者視点でどう思うかお答えください。
 1. お客様はロボットから観光地の情報を十分に聞けていましたか？（情報の十分さ（客視点））
 2. お客様はロボットと自然に対話できていましたか？（対話の自然さ（客視点））
 3. お客様はロボットの対応を好ましく思っていましたか？（対応の好ましさ（客視点））
 4. お客様はロボットの対応に満足していましたか？（対応の満足度（客視点））
- あなたがこのロボットと対話したお客様だったらという視点でどう思うかお答えください。

1. あなたはロボットから観光地の情報を十分に聞けたと思いますか？（情報の十分さ（評価者視点））
2. あなたはロボットと自然に対話できたと思いますか？（対話の自然さ（評価者視点））
3. あなたはロボットの対応を好ましかったと思いますか？（対応の好ましさ（評価者視点））
4. あなたはロボットの対応に満足したと思いますか？
5. ロボットの話を聞いてあなたならどちらの観光地に行きたいと思いましたか？（どちらとも言えない場合は、強いて決めるしたらどちらかで決めてください）
6. このロボットは旅行代理店で実際にサービスできると思いますか？

4.3 実験結果

2021年8月23日，9月4日の2日間にかけて，10代から50代の男女21人の対話を通じてシステムの評価を行なった。コンペティション全体として，大学8，高専1，企業2の11チームがコンペティションに参加し，それぞれ別日に同様の評価を行う。ビデオ評価は参加11チームのシステムと，実行委員の用意したベースラインを含めた12のシステムについて評価を行う。

4.3.1 対話の結果

21件の対話のうち，システムの想定した対話フローに沿って旅行先の選定を行い終了するか，定められた対話時間いっぱい対話をし終了することで，最後まで旅行代理店のカスタマーサービス業務を遂行できた対話は全体の71%当たる15件であった。残りの6件はシステムが生成した音声を再度拾いループ的に対話を生成したり，想定外の質問応答により音声生成が行えず対話終了となったりし，対話を中止した。また，ユーザとシステムの対話が成立した20件の対話における体験者とシステムの平均発話数は14.3件で，対話前に選択した2箇所の観光地からシステムのおすすめする観光地を選んだ割合は，76%当たる16件であった。

本項では実際の対話事例をいくつか示す。なお，事例中に登場する発話者のUはユーザ（体験者），Sはシステムとする。

対話が円滑に進んだ事例

事前の想定通り，システムのフローに沿って進んだ対話の例を図10から図に示す。

図10は対話の冒頭部分である。最初に質疑応答する観光地を決めるための対話で，体験者の発話を受けてシステムが国立民族学博物館の概要情報の説明を行う。ここでは，システムのお勧めする観光地と気になる観光地が一致したため，国立民族学博物館をはじめに説明しようとしていたとすることでお勧めする観光地を旅行先に決定することを誘発している。図11は，図10の対話の続きを示す。住所，アクセス方法などに対して一連の

発話者 発話内容

- S : いらっしゃいませ。私は旅行代理店スタッフのあと申します。本日はご来店ありがとうございます。お客様はで国立民族学博物館と明治なるほどファクトリー大阪で旅行の行き先を迷っているのですね。どちらの方が気になりますか？
- U : 国立民族学博物館の方が気になります。
- S : 奇遇ですね。以前訪れたことがあるので、私も国立民族学博物館について説明しようと思ってました。国立民族学博物館は文化人類学と民族学をテーマにした世界最大級の民族学博物館です・・・

図 10 対話が円滑に進んだ事例 1

会話で適切な返答をすることができている。

発話者 発話内容

- S : 国立民族学博物館について何か気になることはありますか？
- U : 国立民族学博物館はどこにありますか？
- S : 大阪府吹田市千里万博公園10-1です。
- U : どうやったら行けますか？
- S : 電車を利用しますか？それともお車で行かれますか？
- U : 車で行きます。
- S : 名神高速吹田ICから4.6kmのところにあり、20分ほどかかります。駐車場はありません。他に気になることはありますか？

図 11 対話が円滑に進んだ事例 2

図 12 は各観光地の質疑応答における対話の例である。体験者の質問に対し、システムが適切な返答をできていることが分かる。

図 13 は話題となる観光地の変更をする対話の例である。限られた対話時間の中でシステムのお勧めする観光地を旅行先に決定に決定してもらうため、システムのお勧めする観光地と気になる観光地が一致しない場合、このような対話を通じて他方の観光地へと話題を移す。

図 14 は対話の終了フェーズである。上が観光地に対しての質疑応答が終わり体験者の質問がなくなった場合の対話終了で、下が対話の途中で対話終了時刻を過ぎてしまった場合の対話終了である。

質疑のズレが生じつつも対話自体は進んだ事例

次に、体験者の発言意図とシステムの返答にズレが生じつつもシステム上では正常な対話として対話が続行下例を示す。

図??は、「近くの駅が気になります」という発言に対して、システムは近隣の施設の案内を返答している。体験者は質問を変え再度駅について尋ねるが、「電車を利用しますか？それとも送るまで行かれますか？」という少し的外れな返答を再度している。この後の対話では電車でのアクセスについて適切な対話が行われた。

発話者	発話内容
S :	国立民族学博物館について何か気になることはありますか？
U :	入場料金はいくらですか？
S :	観覧420円、特別展観覧料は別途料金です。ふと気になったのですが、私も入場料金で発生するんですかね？他に気になることはありますか？
S :	箕面大滝について何か気になることはありますか？
U :	電車で行けますか？
S :	阪急箕面駅から徒歩40分です。
S :	他に気になることはありますか？
U :	駐車場はありますか？
S :	ありません。他に気になることはありますか？
S :	他に気になることはありますか？
U :	どの季節に行くのがいいですか？
S :	自然の豊かな場所にありますので季節ごとに違った楽しみ方ができると思いますよ。アンドロイドの私は熱に弱いので出来れば夏以外に行きたいですね。他に気になることはありますか？

図 12 対話が円滑に進んだ事例 3

発話者	発話内容
S :	他に気になることはありますか？
U :	ありません。
S :	私のおすすめは箕面大滝で、お時間もまだあるのでぜひこちらの説明もさせてください。箕面大滝は日本の滝百選のひとつです。高さ33mの滝の周辺一帯は、年間100万人以上の観光客が訪れます。

図 13 対話が円滑に進んだ事例 4

図 16 は時間に関する対話である。体験者は観光にかかる時間を知るために、「所要時間は何分くらいですか?」、「全部の展示物を見て回るのにどれくらいの時間がかかりますか?」と言い換えを用いて質問しているがシステムは2回とも営業時間を答えていた。この後、体験者は観光にかかる時間を知ることを諦め、別の質問を行なった。

知りたい情報に対してシステムが適切な返答を変えさないとき、体験者が同じ質問を別の表現で行うという場面が何度か見られた。図 15 のように言い換えによりシステムが適切な返答を返す場合と、図 16 のように適切な返答を返すことができない場合があるが、このような煩雑なやりとりは体験者のシステムに対する満足度を下げ、好印象で対話を終了するための足枷となる。

対話が続かない事例

最後に対話を続けることができない事例について示す。図 17 は対話が続かない典型的な例である。箕面大滝の観光において「猿」という予見不可能な単語が体験者の発話に含まれているため、適切な返答を行うことができない。この他にも、「もみじ天ぷらはあります

発話者	発話内容
S :	他に気になることはありますか？
U :	特くないです。
S :	箕面大滝はとても良い場所だと思うのですが、旅行先に決定するのはいかがですか？
U :	そうします。
S :	旅行先を選ぶためのお力添えができてよかったです。楽しい旅行になることを祈っています。改めまして、私のお勧めは箕面大滝です。旅行先を決めるお手伝いができたなら嬉しい限りです。本日はどうもありがとうございました。
S :	他に気になることはありますか？
U :	ありません。
S :	すみません、お時間になってしまいました。私のバッテリーが切れてしまいそうです。改めまして、私のお勧めは日本民家集落博物館です。旅行先を決めるお手伝いができたなら嬉しい限りです。本日はどうもありがとうございました。

図 14 対話が円滑に進んだ事例 5

発話者	発話内容
S :	他に気になることはありますか？
U :	近くの駅が気になります。
S :	近隣には徒歩圏内にカフェやレストランなどが点在しております。お客様の好みに合うお店や思わぬ出会いをするお店などあると思いますので、ぜひご自身の足でお確かめください。他に気になることはありますか？
U :	駅はどこですか？
S :	電車を利用しますか？それともお車で行かれますか？

図 15 対話が進んだ事例 1

か？」、「どういった御朱印がありますか？」など観光地に対してひどく限定的な質問であったり、「見所はなんですか？」、「何があるんですか？」など抽象度の高い質問が行われた際には対話を続けることができない。

4.3.2 アンケート評価の結果

本項ではアンケート評価の結果を示す。体験者 21 人のアンケート結果から表 4 のような評価が得られ、ビデオ評価者 50 人のアンケート結果から表 5 から以下のような評価が得られた。

表 4 体験者によるアンケート結果

	選択の満足度	情報の十分さ	対話の自然さ	対話の適切さ	対話の満足度	対応の好ましさ	情
評価の平均	4.05	4.10	3.14	3.62	3.86	4.00	4.2
順位	12	12	9	10	8	8	12

発話者 発話内容	
S :	他に気になることはありますか？
U :	所要時間は何分くらいですか？
S :	10~17時で、入館は16時30分までです。
	他に気になることはありますか？
U :	全部の展示物を見て回るのにどれくらいの時間がかかりますか？
S :	10~17時で、入館は16時30分までです。
	他に気になることはありますか？

図 16 対話が進んだ事例 2

発話者 発話内容	
S :	他に気になることはありますか？
U :	猿に出会えますか？
S :	すみません。もう一度お願いします。
U :	猿に出会うことはできますか？
S :	すみません。もう一度お願いします。もしくは言い方や質問を変えていただけると助かります。アンドロイドですが新人なものでご迷惑おかけします。

図 17 対話が続かない事例

表 5 ビデオ評価によるアンケート結果

客観視点					
	情報の十分さ	対話の自然さ	対話の適切さ	対話の満足度	
評価	4.10	3.76	3.62	3.86	
順位	11	9	10	8	
評価者視点					
	情報の十分さ	対話の自然さ	対話の好ましさ	対話の満足度	実用性
評価	3.96	3.90	3.81	3.51	3.74
順位	10	11	11	11	11

5 考察

5.1 対話ロボットに関する考察

5.2 対話制御に関する考察

6 まとめ

謝辞

本研究を進めるに当たり，指導して下さいました竹内孔一先生に心より感謝致します。また，議論に参加して下さいました竹内研究室の諸氏に心より感謝致します。

使用したツール

- 工藤拓 . 形態素解析 MeCab . <http://taku910.github.io/mecab/>
- Google . 音声文字変換 Speech-to-Text . <https://cloud.google.com/speech-to-text/>
- Amazon . テキスト読み上げ Amazon Polly . <https://aws.amazon.com/jp/polly/>
- Oculus . 口形状生成 . <https://developer.oculus.com/downloads/package/oculus-lipsync-unity/>
- Google . Google Maps API . <https://developers.google.com/maps/>

参考文献

- [1] 目黒豊美, 東中竜一郎, 堂坂浩二, 南泰浩ほか. 聞き役対話の分析および分析に基づいた対話制御部の構築. 情報処理学会論文誌, Vol. 53, No. 12, pp. 2787–2801, 2012.
- [2] 山本浩司, 水谷研治, 岩野賢二. 対話ロボットを用いた高齢者コミュニケーション支援システムの開発. 映像情報メディア学会誌, Vol. 54, No. 6, pp. 798–801, 2000.
- [3] 小林峻也, 萩原将文. ユーザの嗜好や人間関係を考慮する非タスク指向型対話システム. 人工知能学会論文誌, pp. DSF-502, 2016.
- [4] 東中竜一郎, 稲葉通将, 水上雅博. Python でつくる対話システム. 株式会社 オーム社, 2020.
- [5] Joseph Weizenbaum. Eliza – a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, Vol. 9, No. 1, pp. 36–45, 1966.
- [6] Terry Winograd. Procedures as a representation for data in computer program for understanding natural language. 1971.
- [7] Bruce G Buchanan and Edward H Shortliffe. Rule-based expert systems: the mycin experiments of the stanford heuristic programming project. 1984.
- [8] Daniel G Bobrow, Ronald M Kaplan, Martin Kay, Donald A Norman, Henry Thompson, and Terry Winograd. Gus, a frame-driven dialog system. *Artificial intelligence*, Vol. 8, No. 2, pp. 155–173, 1977.
- [9] Chien-Sheng Wu, Steven Hoi, Richard Socher, and Caiming Xiong. Tod-bert: pre-

- trained natural language understanding for task-oriented dialogue. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pp. 917–929, 2020.
- [10] 東中竜一郎, 船越孝太郎, 高橋哲朗, 稲葉通将, 角森唯子, 赤間怜奈, 宇佐美まゆみ, 川端良子, 水上雅博, 小室允人ほか. 対話システムライブコンペティション 3. 人工知能学会研究会資料言語・音声理解と対話処理研究会 90 回, p. 23. 一般社団法人 人工知能学会, 2020.
- [11] 中岩浩巳, 池原悟. 語用論的・意味論的制約を用いた日本語ゼロ代名詞の文内照応解析. *自然言語処理*, Vol. 3, No. 4, pp. 49–65, 1996.
- [12] 徳永健伸, 関谷幸恵, 田中穂積ほか. 対話システムにおけるプランベースの照応解析. 第 65 回全国大会講演論文集, Vol. 2003, No. 1, pp. 221–222, 2003.
- [13] 東中竜一郎, 船越孝太郎, 小林優佳, 稲葉通将ほか. 対話破綻検出チャレンジ. *SIG-SLUD*, Vol. 5, No. 02, pp. 27–32, 2015.
- [14] 竹内孔一. 語彙概念と語彙概念構造. 松本裕治, 奥村学(編), コーパスと自然言語処理, pp. 94–113. 朝倉書店, 2017.