The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

複数ロボット協調による一問一答型雑談対話からの脱却

Continuous conversation with two-robot coordination

杉山 弘晃 *1　水上 雅博 *1　成松 宏美 *1
Hiroaki Sugiyama　Masahiro Mizukami　Hiromi Narimatsu

*1NTT コミュニケーション科学基礎研究所
NTT Communication Science Laboratories

We have been working on a conversational dialogue system that raises user satisfaction and draws interests through dialogues, unlike a task-oriented dialogue that answers simple instructions and questions. To raise user satisfaction, systems should precisely respond to a user utterance and continue dialogue with topics that strongly related to the user utterance. However, most conversational systems, which are usually utilize one-turn query-response pairs, are difficult to naturally continue dialogue. In this study, by controlling dialogue topics with multiple robot coordination, we propose a novel conversational system with two-turn query-response pairs that can naturally continue consistent dialogue. Our field experiment at Kyoto city zoo shows that our proposed system achieves very high user satisfaction.

1. はじめに

近年，従来のタスク指向の対話システムとは異なる，雑談を行う対話システムに注目が集まっている [Ritter 11, Higashinaka 14]．雑談対話は，単純な命令や質問応答に答えるタスク対話とは異なり、対話そのものへのユーザの満足度を高めたり，対話を通して話している話題への興味や思考を引き出すことを目的とする対話である．ユーザの対話への満足度を高めるためには，対話相手に発話を受け止めてもらったという印象をユーザに感じさせることが重要である．この印象は，単に一問一答的によい応答をすれば充足されるわけではない．ユーザ発話に対してロボットが適切な応答をしたとしても，そこから話題が発展しない場合，ユーザは自身の発話が表層的にしか受け止められなかった，もしくはその話題に興味がないと表明されたように感じてしまう．そのため，受け止めたところから話題が発展していくことも，満足度向上には重要な要素である．

一方，現在の雑談対話システムでは，対応可能な話題の広さを優先した，一問一答的なアプローチが主に用いられている [Vinyals 15, Ritter 11, Higashinaka 14]．応答の複雑さを単純な一問一答に限定することで，雑談中の話題に応答できる幅広さを実現している．しかしながら，一問一答では対話が細切れとなり，まとまってしっかりと対話できたという満足を得られにくい．この課題に対し，ユーザ発話による話題遷移を許容しない，もしくはごく少数の分岐を用意しておく前提で，複数ターンのシナリオとして構築する場合もある [Iio 16, 渡辺 16]．話題遷移を許容しない場合の一般的な方策は，ロボットからユーザに質問し，ユーザの答えによらず,「そっか」などの相槌で受け止め,「僕は○○だよ」と切り返すという流れを繰り返すものである．このアプローチの問題点として，展開される話題がユーザ発話と直接対応するものではないため，十分な満足を与えることは難しい点がある．また，ユーザ発話に応じてシナリオを分岐させていく場合も，ある程度の話の展開に影響を与えているものの，複雑な内容は伝わらないため，理解されている感は少ない問題がある．

こうした課題に対し，杉山らは，ユーザ発話に一問一答の知識に基づいて応答するとともに，その内容に関連する別の一問一答知識を利用してロボット間で対話を行うアプローチを提案しており，通常の 1 対 1 の対話よりも対話の継続感が向上することを報告している [杉山 17]．しかしながら，彼らの構築した発話知識は，特定の話題に特化するよりもむしろ一般的な内容で構築されているため，対話の個別の話題・文脈とはやや乖離した（ユーザ発話の詳細とは関連しない）内容になることが多い．

本研究では，一問一答とシナリオの組み合わせとして，ミニシナリオと呼ぶ 2 ターン分の発話知識に基づく複数ロボット雑談対話システムを提案する．ミニシナリオは，ユーザが発話しそうな文とそれに後続する 3 発話から構成される．2 体のロボットと 1 人のユーザとの対話を前提とし，ユーザ発話へロボットが応答したあと，その内容を反映した追加の一問一答をロボット間で行うことで，ユーザ発話を起点として，詳細に話題が繋がる自然な対話を実現する．ユーザへの応答および追加の一問一答は全てロボットが発話するため，対話として自然につながるようにあらかじめ作成しておくことができることがポイントである．

また，ロボット間の対話を利用して，自然に話題を誘導することも可能である．そのため，システムが限られたドメインの発話知識しか保有しない場合でも，ユーザは違和感を感じることなく雑談を継続できる．本研究ではこの特性を活かし，狭いドメインに特化して質問応答と同程度に詳細な雑談用の発話知識を構築することで，雑談と質問応答を相互に行き来しながら知識を伝達するシステムの実現を目指す．また，本システムを京都市動物園に設置し，来場者（主に子ども）と対話させた結果について報告する．

2. 発話知識に関する関連研究

雑談対話システムが発話するための知識として主に用いられるものとして，一問一答形式の発話対がある．ユーザ発話に類似する入力文を選び，それに紐付いた出力文をシステム発話として用いる形式である．この方法の場合，人手で想定される応答対を記述するもの（ルール [Wallace 04, Sugiyama 14]）や，実際の人の対話例を利用するもの（用例 [水上 16, Ritter 11]），ある単語や述語項に対する応答を自動生成するもの [杉山 15]，ニューラルネットを用いて End-to-end で生成するもの [Vinyals 15] などがある．

ルールで記述する場合は，想定されるユーザ発話に対しては質の高い応答が期待できる一方，ユーザ発話に対するカバー

連絡先: 杉山 弘晃，NTT コミュニケーション科学基礎研究所，京都府相楽郡精華町光台 2 － 4，TEL: 0774-93-5243，FAX: 0774-93-5245，sugiyama.hiroaki@lab.ntt.co.jp

発話文	応答文	後続発話文	後続発話応答文
お鼻が長いところが好き	ゾウさんのお鼻は筋肉でできてて小さいものもつかめるんだよ	【質問】鼻で吸ってるんじゃないの？	鼻の先が指のように分かれてて、その部分でつかんでるんだよ
ゾウさんすごくお鼻長いね！		【平叙】すごく器用なんだね	鼻の動きを観察していると、ゾウの気持ちが分かるらしいよ
ゾウさんのお鼻はとっても長い		【継続】しかも鼻の動きを観察していると、ゾウの気持ちが分かるんだって	鼻で気持ちが分かるってすごいね。
ゾウさん大きくてかっこいい	肩までの高さは2.5〜3mくらいあるんだよ	【質問】鼻の長さはどれくらいあるの？	1.5〜2mくらいの長さがあります
体が大きい		【平叙】そんなに大きいんだ	近くで見ると迫力があるよ
超でかい！		【継続】近くで見ると迫力があるよ	近くで見てみたいな

図 1: ミニシナリオの例（ゾウのいいところ）

率が低いことが課題である．用例を利用する場合は，マイクロブログ等の実際の対話を利用するか，実験用に集められた対話を利用するかで特性が異なる．実際の対話を利用する場合は，その話者間でのみ成立する前提知識なしには一問一答としても成立していないように見える用例が問題となる．また，対話の方向性は状況，時刻や期間，目的によって異なるが，そうした違いを取り扱わず一括りに扱ってしまうため，文脈と矛盾した発話が生成されうるという問題がある．さらに，対話中の文そのものを利用する場合には，著作権的な問題も存在する．実験用に集めた対話を利用する場合はこれらの問題は比較的緩和されるものの，用例の量が大幅に少なくなり，ユーザ発話に対するカバー率が低下する問題がある．

自動的に文を生成する場合は，カバー率は非常に高く，また著作権的な問題も発生しないという利点がある．しかしながら，文の生成自体が非常に難しいタスクであるため，一文として意味の通る文を作成することが難しい．また，複雑な内容の発話を作成できないという問題点がある．さらに，これらの方法のいずれにも当てはまる問題として，一問一答を基本として知識が作られているため，対話が細切れとなりしっかり対話できたという満足が得られにくいという点がある．

3. システムの構成

本研究では，2 体のロボットと 1 人のユーザとの対話を想定し，2 ターンの発話対から成るミニシナリオ（図 1）に基づく対話を提案する．さらに，ロボット間対話や話者切り替えにより応答可能な範囲へ話題の流れを誘導することで，狭いドメインに集中的に構築した発話知識を基に自然な対話を実現する．本章では，話題が連続的につながっているように感じられる対話を実現するための，発話知識の構築・運用方法および全体の実装について説明する．

3.1 ミニシナリオ：2 ターン単位の発話知識

ミニシナリオは，発話文，応答文，それらに対する後続の発話文とその応答文の 4 文を単位として構成される．ここでは，動物に関する対話を想定した例に基づいて，各文の説明および構築方法を以下に示す．

発話文は，ユーザが発話すると想定される文であり，ユーザの発話する範囲を詳細にカバーできるよう多数作成する．本研究では，対象とする動物について，いいところ，質問，トリビアのいずれかの発話種類ごとに，発話文を 50 文ずつ（5 文 x 10 名の発話作成者）作成した．例えば，ゾウのいいところについて，「お鼻が長いところが好き」や「ゾウさん大きくてかっこいい」などとなる．合わせて，ユーザ発話の表現の揺れを吸収できるよう，それぞれの発話文と異なる表現で同じ意味となる文を 5 文ずつ作成する．なお，同じ意味の発話文をまとめたところ，意味の異なり数は概ね 25〜30 種類程度となっていた．

次に，このように作成した発話文について，ロボットが発話する応答文を作成する．ロボットの発話に矛盾が生じないよう，応答文は動物の種類ごとに 1 名の発話作成者が作成するものとし，同じ意味の入力文に対しては，同じ応答文となるよう作成する．また，応答文に質問を入れると，後述する後続発話文との整合が取りにくくなるため，応答文は平叙文で作成することとする．ゾウの「お鼻が長いところが好き」という発話文に対する応答文として，「ゾウさんのお鼻は筋肉でできてて小さいものもつかめるんだよ」が作成された．

後続発話文は，それに紐づく発話文と応答文のペアに対して，対話として自然につながるよう作成された発話である．ここでは，質問，平叙，継続の 3 つのタイプの発話を作成している．質問と平叙は応答文の発話者に対して別の話者が発話するものとして作成し，継続は応答文の発話者自身が継続して発話するものとして作成する．例えば，応答文「ゾウさんのお鼻は筋肉でできてて小さいものもつかめるんだよ」の後続の質問には「鼻で吸ってるんじゃないの？」が作成された．

後続発話応答文は，後続発話文に対する自然な応答になるよう作成された発話であり，応答文と同様の方法で作成する．

以上のように発話知識を構成することで，後続発話は先行する発話・応答文に密接につながる発話となるため，一問一答をつなげて複数ターンとするよりも自然な対話を実現できる．

3.2 対話の構成

構築したミニシナリオに基づき，ユーザとの対話を実現する方法を説明する．通常のシナリオを用いた対話では，対話の内容と進行が一体となって構築されていることが多い．しかしこの方法では，ミニシナリオのように大量のシナリオにスケールさせることは難しい．本研究では，ミニシナリオとして構築した多数の知識を効率よく利用するため，対話の進行と内容を分離して設計する．具体的には，対話の進行については，図 2 のように yaml 形式でスクリプトを記述し，その中の一部分をミニシナリオを利用してテンプレート的に補完する方法である．図 2 はユーザから話しかけられた直後の応答部分について示したものである．図中，

```
- ["R1", "H", 1.0, "/応答文/", "/SPEAK"]
```

の 1 つ目の項目は話者（H：ユーザ，R1, R2：ロボット 1，2），2 つ目の項目は話しかける相手話者，3 つ目の項目は発話終了後の待機時間，4 つ目の項目は発話内容，5 つ目の項目はジェスチャを表す．<ANIMAL>のように<>で囲われた部分は，話題としている動物やユーザの名前など，対話を通して固定される内容が代入される．また，//で囲われた部分には，BC（BackChannel）以外は，現在の話題で選択されているミニシ

```
INITIAL_FAMOUS_YES:
- ["H", "", 0.0, "FAVORITE:BEGIN_WITH_USER_UTT", ""]
- ["R2", "H", 100.0, "<TARGET_ANIMAL>さんのどんなところが好きなの？", "/ASK"]
...

BEGIN_WITH_USER_UTT:
- ["R1", "H", 0.5, "/BC/", "/NOD"]
- ["R2", "H", 1.0, "/発話文/", "/SPEAK"]
- RANDOM:
  - TRANSIT:
      BODY_QA
  - TRANSIT:
      BODY_DEC
  - TRANSIT:
      BODY_SELF

BODY_QA:
- ["R1", "H", 1.0, "/応答文/", "/SPEAK"]
- RANDOM:
  - ["R2", "R1", 1.0, "へえー", "/NOD"]
  - ["R2", "R1", 1.0, "ふむ", "/NOD"]
- ["R2", "R1", 1.0, "/後続質問文/", "/SPEAK"]
- ["R1", "H", 1.0, "/後続質問文-応答文/", "/NOD"]
...
```

図 2: yaml で記述したスクリプト

ナリオの内容が代入される．話者が H の行は，ユーザ発話を待ち受ける状態へ移行するコマンドを表し，発話内容の項目にどのような発話を待ち受けるかと，その発話が入力された後の遷移先等の処理項目が記述されている．ユーザ発話を待ち受けている状態以外のタイミングでユーザ発話が入力された場合は，その発話の種類に応じた行動を行う．具体的には，動物のいいところについての発話であれば共感を，質問であればその回答を発話し，「へえ」などのフィラーであれば相槌を打った後，元の発話に戻る処理を行う．

図 2 の流れに沿って，基本的な対話の流れを説明する．ここでは TARGET_ANIMAL がゾウであるとする．まず 2 行目で，ユーザ発話が動物についてのいいところ（FAVORITE）である場合を対象とする待ち受け状態へ移行する．ここでは，ユーザ発話が FAVORITE であった場合，BEGIN_WITH_USER_UTT へ遷移するものとする．

次に，3 行目で「ゾウさんのどんなところが好きなの？」を発話し，ユーザ発話を待ち受ける．それに対し，例えばユーザが「鼻が長いところ」と発話した場合，いいところについての発話と認識され，BEGIN_WITH_USER_UTT に遷移する，同時にいいところに関するミニシナリオ中で，類似する発話文を検索する．

BEGIN_WITH_USER_UTT では，7 行目で R1 が相槌を打ち，8 行目で R2 がユーザ発話と同じ内容の発話をすることで，共感を表現する．その後，後続発話文のどれを選択するかをランダムに選択する．BODY_QA（質問）が選ばれた場合，18 行目に遷移し，/応答文/に図 1 の応答文を代入して「ゾウさんのお鼻は筋肉でできていて小さいものもつかめるんだよ」と R1 が発話する．ユーザ発話を一旦受け止めてから応答文を発話することで，ユーザ入力と多少異なる意味の発話文が検索され応答がやや不自然になった場合でも，違和感を軽減して対話を続けることができる．

R1 の応答文に対し，20，21 行目を利用して R2 が R1 に相槌を打ちつつ，22 行目の後続質問文「鼻で吸ってるんじゃないの？」を発話し，R1 は H に向き直りながら，23 行目の後続質問文-応答文の「鼻の先が指のように別れてて，その部分でつかんでるんだよ」を発話する．このようにロボット間での対話を見せることで，ユーザの発話によって対話の内容が変動し，かつそこから対話がスムーズにつながっている印象を与えることができる．

ミニシナリオ 1 つ分の対話が終了した後は，類似した意味のミニシナリオを利用してロボット間で対話を継続したり，ユーザに質問を出して新たな発話を引き出すなどして対話を継続する．

3.3 システムの実装

対話ロボットには，CommU[*1] を用いる．視線や首を動かすことができ，話者交代をスムーズに表現することができる．音声を受け取るマイクは，指向性・ノイズ抑圧性能が高い，インテリジェントマイク [*2] を用いる．音声認識エンジンには NTT-TX 製の SpeechRec[*3] を用いる．今回，話題が動物に特化していること，また話者が小学生程度の子どもを想定していることから，通常のドメインの音声認識エンジンでは，認識率の低下が予想される．そのため，こうした発声を正しく認識できるようにするため，事前に収録した子どもの対話音声を用いて音響モデルのドメイン適応を，ミニシナリオ中の文を利用して言語モデルのドメイン適応を行った．さらに，ユーザが言いよどんだ場合に発話区間を誤って切り出さないようにするため，ボタンを押しながら話す Push-to-talk 方式を採用する．音声合成は，CommU の外見に合わせた子供の声として，NTT-TX 社製の音声合成 [*4] を用いる．

4. 京都市動物園における実証実験

4.1 実験設定

提案システムを京都市動物園に設置し，2/1 から 2/28 までの 1ヶ月間来場者と対話する実証実験を行った．実施場所は，京都市動物園の無料エリアにある，図書館カフェと呼ばれるエリアである．主に親子で本を読みながら食事や休憩を取るスペースとなっており，特に休日は多数の来場者が訪れる場所である．本実験では，提案するシステムがどの程度満足度の高い対話を実現できるかを，実ユーザを対象に評価することを目的とする．合わせて，適切な発話タイミングやユーザの対話への興味を推定する元データとして，対話中のユーザの表情や音声の収録を行う．対話の様子を図 3 に示す．

図 3: 動物園における対話の様子

対話の対象とする動物は，京都市動物園で飼育されている動物の中で人気の高い，ゾウ，キリン，カバ，レッサーパンダ，ツシマヤマネコ，トラ，フクロウ，ゴリラ，ペンギン，バクの 10 種類である．

来場者への案内は園内の看板や Web 等を通して行った．対話に参加する場合には，対話の方法について説明するとともに，タブレット PC を用いて対話中のユーザの呼び名や年齢・性別の設定，対象動物の選択，および本人が 18 歳以上もしくは保護者がいる場合のみ動画等のデータ取得に関する説明および同意取得を行った．

上記準備の後，実際に来場者とロボットとの間で対話を行った．なお，デモ時間や対話安定性の制約上，ユーザが 6 回発話した段階で，ミニシナリオの切れ目で終了モードに移行し，

∗1 https://www.vstone.co.jp/products/sota/index.html
∗2 http://www.ntt.co.jp/md/products/product_29.html
∗3 http://www.v-series.jp/speechrec/
∗4 http://www.v-series.jp/futurevoice/

「そろそろ時間みたい」と対話の終了を促す形で対話の終了処理を行った．また対話終了後，ユーザ評価を 5 段階（1：そう思わない，5：そう思う）で入力した．対話の楽しさや話題の対象への興味が対話の満足度を表すと考え，評価項目には，1. ロボットと話すのは楽しかったですか？(楽しさ)，2. 選んだ動物に興味を持てましたか？(興味)，3. 選んだ動物に詳しくなれましたか？(知識) の 3 項目を設定した．

4.2 結果と分析

実験に参加した延べ人数は，付き添う保護者を含め，概ね 400-600 人程度であった．そのうちデータ取得の同意を取れた人数は 238 名であった．本研究では，有効な同意を取得できた体験者のデータのみを用いて分析を行う．

まず，参加者全体の評価値は，1. 楽しさ：4.52，2. 興味：4.28，3. 知識：4.04 であった．5 段階評価で 4.5 以上は極めて高い値であり，ほとんどの体験者が楽しいと感じたことがわかる．一方，3 の知識については，4.0 は超えているものの楽しさ・興味に比べるとやや低い評価値となっていた．

次に，年齢の分布，および年齢ごとの評価値を図 4 に示す．来場者として，当初小学生低学年くらいを想定していたものの，実際には未就学児が非常に多く体験していた．一方，小学生中学年以上および中高生はほとんど来園していないことがわかる．評価値で見ていくと，1. 楽しさと 2. 興味は年齢に依らず概ね横ばいであった．3. 知識については，有意差も出ていないものの，6-8, 13-19, 20-39 歳の評価が高い一方，9-12 歳の落ち込みが大きい．実際に体験者の様子を観察していると，6-8 歳は知識のレベルが程よく合致しており，知識の満足度向上につながったものと考えられる．しかしながら，9-12 歳程度で動物園に来場する子どもはもともと非常に動物に興味があり知識も極めて豊富な子が多く，小さい子どもに合わせた知識では十分な満足を与えられなかったものと考えられる．一方，それより大きい 13 歳以上，特に 20 歳以上になると，普通程度の知識の来場者が再び増加し，かつ一般的な対話システムやロボットの対話レベルとの比較で評価するようになるため，評価値が向上したものと考えられる．

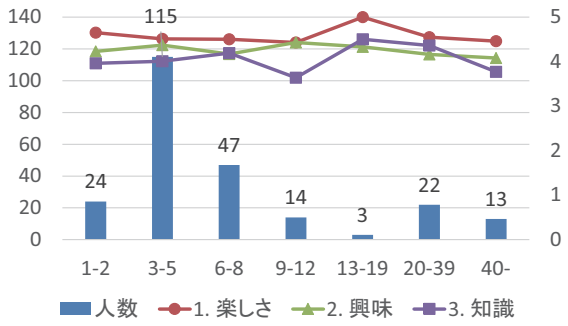

図 4: 年齢に対する評価値

男女の体験者数はそれぞれ 116 名，119 名（回答なし 3 名）であり，評価値は男性は 4.47, 4,32, 3.95，女性は 4.56, 4.23, 4.11 で有意差はなかった．

また，観察に基づく定性的な分析として，4 歳以下はロボットの発話を正しく理解すること自体が難しい（オープンな質問に的確に答えられないなど）場合が多く，論理的に見れば破綻している状態がほとんどであった．しかしながら，その状態であっても，図 4 の結果からも，楽しく対話していた子が多いことがわかる．内容のやりとり以外の観点での対話の楽しさを解き明かす手がかりになると考えられる．

加えて，対話後に感想を尋ねたところ，今回の対話の仕方（ロボット発話→人発話→ロボット間で対話 の繰り返し）でも，しっかりつながった対話と感じたという意見が多かった．ロボット間で話すところまでを応答と見れば，構造的には一問一答と類似しているものの，つながった対話と感じられていたという結果は，今後の対話ロボット研究を進めていく上で非常に有用な知見である．一方，ロボットが話しすぎている，という意見も多くあった．スクリプトでは頻繁に人に話を振るように設計していたが，それでもなお不足と感じられていたため，話を振るタイミングやユーザが割り込みやすい隙をうまく制御する必要があると考えられる．特に今回，対話の安定性を志向して Push-to-talk 式のターンテイクを採用していたものの，これにより，話を振られるまで割り込みにくいという印象を強めていた可能性があるため，ターンテイクの制御と合わせたデザインが必要である．

5. まとめ

本稿では，2 体のロボット間の対話や話者交代による話題制御を利用し，特定の動物の話題についてユーザ発話に的確に応じながら複数ターンに渡る自然な対話を実現するシステムを提案した．また，京都市動物園に本システムを設置し来場者と対話する実証実験を通して，非常に高いレベルで対話が楽しいと感じられていたことを示した．

今後の展開として，ユーザが話に入り込みやすいタイミングの推定や，内容以外の部分でユーザが心地よく感じるための要素について検討を行う．

謝辞

本実証実験は，京都市動物園および NTT 西日本の協力により実現されたものである．両者の多大なる支援に感謝する．

参考文献

[Higashinaka 14] Higashinaka, R., Imamura, K., Meguro, T., Miyazaki, C., Kobayashi, N., Sugiyama, H., Hirano, T., Makino, T., and Matsuo, Y.: Towards an open-domain conversational system fully based on natural language processing, in *Proc. COLING*, pp. 928–939 (2014)

[Iio 16] Iio, T., Yoshikawa, Y., and Ishiguro, H.: Pre-scheduled Turn-Taking between Robots to Make Conversation Coherent, in *Proc. HAI*, pp. 19–25 (2016)

[Ritter 11] Ritter, A., Cherry, C., and Dolan, W. B.: Data-Driven Response Generation in Social Media, in *Proc. EMNLP*, pp. 583–593 (2011)

[Sugiyama 14] Sugiyama, H., Meguro, T., and Higashinaka, R.: Large-scale Collection and Analysis of Personal Question-answer Pairs for Conversational Agents, in *Proc. IVA*, pp. 420–433 (2014)

[Vinyals 15] Vinyals, O. and Le, Q.: A Neural Conversational Model, in *ICML Deep Learning Workshop* (2015)

[Wallace 04] Wallace, R. S.: The Anatomy of A.L.I.C.E., *ALICE Artificial Intelligence Foundation, Inc.* (2004)

[水上 16] 水上雅博, Nio, L., 木付英士, 野村敏男, Neubig, G., 吉野幸一郎, Sakti, S., 戸田智基, 中村哲：快適度推定に基づく用例ベース対話システム, 人工知能学会論文誌, Vol. 31, No. 1, pp. DSF–C_1–12 (2016)

[杉山 15] 杉山弘晃, 目黒豊美, 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と用例を利用した応答文の生成, 人工知能学会論文誌, Vol. 30, No. 1, pp. 183–194 (2015)

[杉山 17] 杉山弘晃, 目黒豊美, 吉川雄一郎, 大和淳司：複数ロボット間連携による対話破綻回避効果の分析, 人工知能学会全国大会, pp. 1B2–OS–25b–2 (2017)

[渡辺 16] 渡辺美紀, 小川浩平, 石黒浩：タッチディスプレイを通じて誘導的な対話を行う販売アンドロイド, 人工知能学会全国大会, pp. 2O3–2 (2016)