雑　談　対話ロボット　対話破綻回避

気軽に雑談できるシステムの実現をめざして

ここ数年で大きく進歩し，一気に身近な存在となってきた雑談システムですが，音声認識誤りや扱える話題の範囲の狭さなど，まだまだ課題も多く残されています．本稿では，雑談システムの発話内容の改善および実ロボットへの搭載に関する取り組みと，ロボットを複数体としたときの対話破綻回避効果について紹介します．

杉山 弘晃（すぎやま ひろあき）/東中 竜一郎（ひがしなか りゅういちろう）
目黒 豊美（めぐろ とよみ）
NTTコミュニケーション科学基礎研究所

雑談システムの普及

ここ最近，さまざまなところで，人と雑談することを志向したアプリケーションやロボットを目にするようになりました．ほんの2，3年前までは，しゃべってコンシェルやSiriなどの，スマートフォン上のアプリケーションが雑談システムの主流でしたが，PepperやOHaNASの登場を皮切りに，実ロボットとの雑談が一気に身近なものになってきています．気軽に雑談できるロボットは，人と情報をつなぐ自然なインタフェースとしてだけでなく，人自身のコミュニケーションスキルの向上にも役立つと期待されています．

ところで，こうした雑談システムは，実際はどのくらい上手におしゃべりできているのでしょうか？　実際に彼らと雑談したことのある人は，正しくこちらの声を聞き取ってもらえず意味の分からないことを言われたり，少し細かい話題について話そうとしてもこたえてもらえなかったりして，がっかりした経験があると思います．現在の（特に商用の）雑談システムは，音声認識によって正しいテキストが得られることを前提として，そのテキストに応答するルールを人手で大量に記述することでつくられています．音声認識と応答ルールを切り分けて考えることで，ルールを記述するコストを低減できますし，人がルールを書くことで，よく話されるユーザ発話に対して面白い発話を仕込むこともできます．しかしこの方法では，音声認識が失敗したときに不適切な発話をしてしまいますし，ユーザが話し得るすべての話題についてルールを書ききるのは現実的ではありません．本稿では，これらの問題に対する，私たちの最新の研究成果を紹介します．

さまざまな話題に対する応答発話の自動生成

これまで，雑談システムを構築する場合，ユーザ発話に含まれ得る単語について，対応するシステム発話をあらかじめ人手で記述しておく，ルールベースのアプローチが多く用いられてきました．代表的なルールベース対話システムとしてA.L.I.C.E.などが知られており，雑談システムの国際的なコンテストであるローブナー賞においても，優秀な成績を収めています．しかし，ルールベースの対話システムの欠点として，対応可能な話題の拡充が難しいことがあげられます．雑談に現れる話題は極めて多様であり，かつ日々増えていくという特徴があります．このようなユーザ発話に対して，適切なシステム発話を出力するルールを人手で作成・維持し続けるには，莫大なコストが必要となります．

雑談システムが対応可能な話題を自動的に拡充する試みとして，マイクロブログやWebの文章からユーザ発話に類似した文を検索し，システム発話として出力する，検索ベースのアプローチが提案されています．Webでは利用者が多くの文を定期的に投稿していますので，多様かつ新しい話題を持つユーザ発話に対しても，ユーザ発話に関連した内容をシステムが発話することができます．また，検索によって得られる文は，実際に人が記述したものですので，文法的な誤りが入り込みにくく，かつ発話のバリエーションが豊富という利点があります．こうしてみると，これらの検索ベースのアプローチが持つ特徴は，雑談対話システムを構築するうえで有用にみえます．しかし，検索によって得られる文は，書かれた状況に応じた固有の文脈を持ちますので，内容の理解に必要な情報が省略されていたり，不要な情報が含

まれていたりする場合があります．これらをそのままシステム発話として出力してしまうと，ユーザにとって理解できない発話になってしまいます．

この問題に対して，私たちは，Webコーパスの持つ話題の豊富さを活かして話題を広げつつ，不要な情報が入り込むのを抑える，新しい発話生成方法を提案しています[1]（**図1**）．この方法ではまず，話題（単語や文節のペア）と話題の関係性の強さを，コーパス内での係り受け関係の出現数に基づいて自動的に計算し，データベースに保存しておきます．ユーザ発話が入力されると，その中の重要な話題（例：「お腹・空いた」）を抜き出し，その話題に関係の深い話題（例：「ご飯・食べる」）を先ほどのデータベースから取得します．こうして得られた2つの話題を，コーパス内での実際の使われ方を参考に助詞などを補って組み合わせて，「お腹が空いたらご飯を食べましょうか」のように，システムの発話を合成します．こうすることで，文法的に妥当，かつユーザ発話に関連する話題を含む，新しい発話を得ることができます．応答の適切さを測る実験を通して，従来の検索ベースおよびルールベースの方法よりも，応答文として自然な発話を生成できることが示されています．

システムパーソナリティの設計

前述の方法によって，ユーザ発話に関連する発話を，自動的に得ることができるようになりました．では，これで雑談システムは完成するでしょうか？　実は，雑談対話におけるユーザ発話の特徴として，話題の広さに加えて，対話相手の好みや経験といった，話者の詳細なパーソナリティに関する質問（「どんな映画が好き？」など）がしばしば発話されることが知られています[2]．このようなパーソナリティ質問は，商用のタスク達成機能を持つシステムに対しても同じように発話されるため[3]，雑談対話システムを実現するためには，こうした質問に対して適切に応答できる必要があります．しかし，先ほどの関連発話の生成だけでは，こうした質問に適切に答えることができません．また，Web上の知識を基にして質問応答システムをつくると，質問応答の一貫性を保てないという問題があります．

そこでNTTコミュニケーション科学基礎研究所では，多数の実験参加者から世代・性別のみを定義した特定のキャラクター（ペルソナ）に対する質問を大量に集め，このペルソナと同じ世代・性別の1人の回答作成者からそれらに対する回答を一貫して集めることで，一貫性のある質問・応答のペアを大量に収集する方法を考えました[4]．世代・性別によって収集できる質問は異なりますので，私たちは，20代・50代の男女（4種類）＋ロボット2種類の計6種類のペルソナについて，質問・応答のペアを収集し，Person DataBase（PDB）を構築しました（**図2**）．得られた質問・応答のペアは，対話開始時のパーソナリティ質問の80%以上をカバーし，対話がしばらく進んだ後に出現する細分化された話題を含めても，およそ半数の質問をカバーしています．収集した質問応答ペアを用いて質問応答システムを構築し

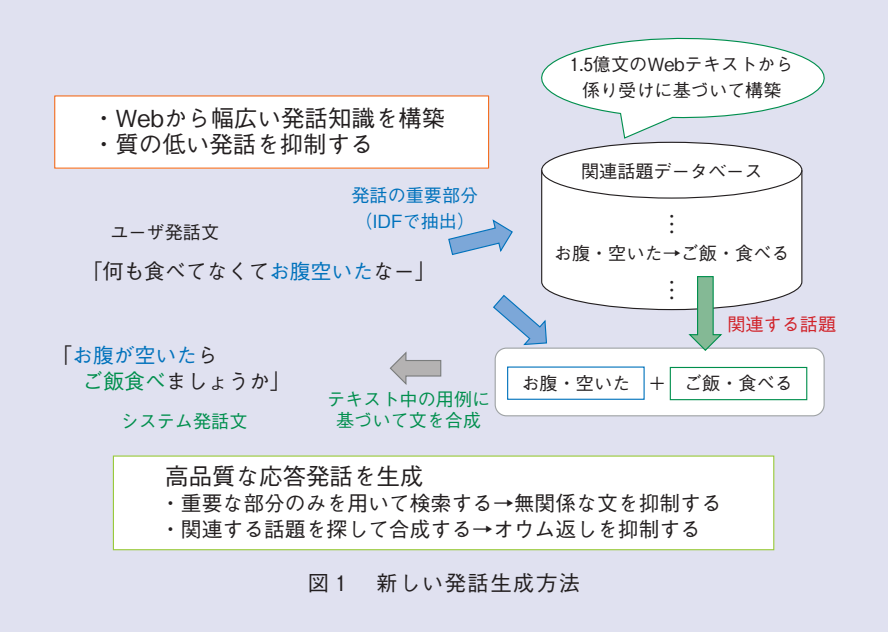


図1　新しい発話生成方法

・パーソナリティに関する質問文・応答文の組を大規模に収集
　6キャラクタに対する10082種類の質問を収集（計26595質問文・応答文）

20代男性　20代女性　50代男性　50代女性　ロボット（身体あり）　ロボット（身体なし）

・質問文を質問カテゴリ・話題へ分類しPDBを構築

質問文	質問カテゴリ	話題	キャラクタ	回答
何色が好きですか？	好きな色	好きな色	20代女性	ピンク
好きな色は？			50代男性	緑
旅行に行きたい場所は？	どこに旅行/旅したいか	行きたい場所はあるか/どこか	50代女性	ブータン
どこに旅行したいですか			50代男性	フランス
国内旅行に行くとしたらどこに行きたいですか	国内旅行するならどこへ行きたいか		20代女性	安曇野

図2　PDBの構築

たところ，およそ6割の質問に適切に応答し，発話の主観評価も有意に向上することが分かりました．

実ロボットへの雑談システムの搭載

私たちはこれまで，テキストチャットを対象とした発話内容の改善に取り組んできましたが，実際の利用シーンでは，実体を持ったロボットと音声で会話する状況も多く見られます．こうした状況でどのくらい自然に話せるかを調べるため，私たちは，大阪大学の石黒浩教授と連携して，非常にリアルなアンドロイドに私たちの雑談システムを載せてみることにしました．マイクから入力されたユーザの声をテキストに変換し，私たちの雑談システムを通してロボットの発話テキストをつくり，それを音声合成にかけてシステムの声として出力する，という流れです．これまで，TV番組『マツコとマツコ』に出演する，マツコ・デラックスさんそっくりのマツコロイドや，石黒教授そっくりのHI-4へ搭載し，それぞれTV放映やSXSW（サウス・バイ・サウスウエスト）でのデモを行っています（**図3**）．

http://mashable.com/2016/03/13/geminoid-robot-conversation/?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+Mashable+%28Mashable%29#nwn9Y3cBCSqN

図3　石黒教授そっくりのHI-4

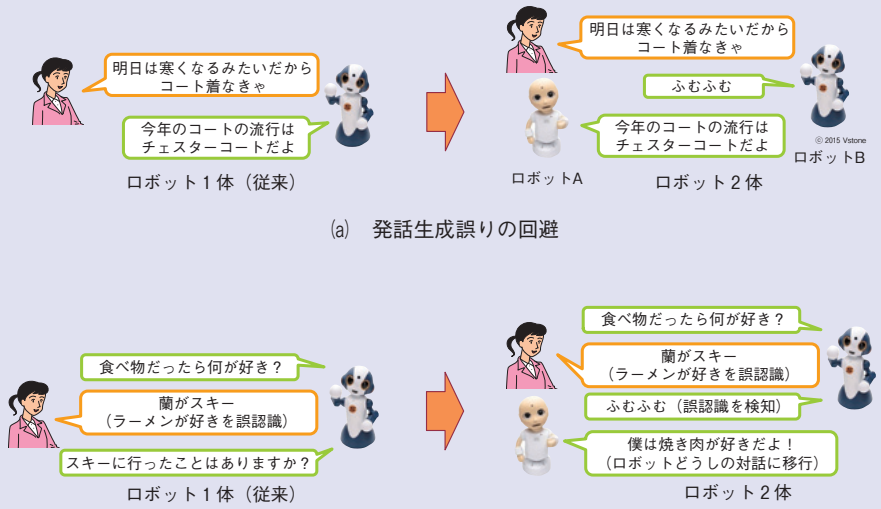
実際にロボットに載せて音声のみで対話してみると，当初の想定どおり，音声認識誤りによって，しばしば不適切な発話が出力されていました．また，ユーザが人に話しかける調子でどんどん話しかけると，処理速度が追いつかず，発話のタイミングがずれてしまうことも分かりました．これらの困難は予想どおりのものでしたが，一方で，音声のみで対話することで予想外に簡単になりそうな問題もありました．例えば，テキスト対話では違和感を覚えるような話題やロジックのずれに対しても，音声のみで（テキストを見ずに）対話をすると，ユーザが違和感を持ちにくくなる，という効果がありました．また，テキスト対話では，システムが意味のつながらない発話をしてしまうと，途端にユーザの対話へのモチベーションが下がっていましたが，実際のロボットを相手にすると，ロボットがそうした発話をしたときでも，ユーザは対話を続けようとする傾向がありました．人は対話相手との関係を維持しようとしますが，リアルなアンドロイドであれば，ロボット相手でもこうした効果が現れるようです．これらの利点は，いくつかのデモを通して私たちが得ている感触レベルのものではありますが，これから精査・検証し，「使える」技術へとつなげていきたいと考えています．

複数ロボットとの対話

これまでの研究で，ロボットやCGエージェントが主となって話す対話（美術館の紹介など）において，ロボットを複数体にしてロボットどうしが掛け合いながら話すことで，単純に1体が話し続けるよりも，スムーズに対話が進むことが知られています[5]．私たちはこの効果に着目し，複数の雑談ロボットを適切に協調させて対話を進めることで，発話生成誤りや音声認識誤りが生じた場合でも，対話そのものの破綻を回避する技術を開発しました（**図4**）．

(1)　発話生成誤り

例えば，ユーザ発話の一部（ここでは「コート」）しか理解できず，多少ずれた話題の発話（「今年のコートの流行はチェスターコートだよ」）しか生成できない場合を考えます（図4(a)）．ユーザと対話しているロボットには，ユーザの発話内容に正しく応答する義務が発生しますが，ロボットが従来どおり1体しかいない状況では，



(a)　発話生成誤りの回避

(b)　音声認識誤りの回避

図4　複数ロボットとの対話

そのロボットに応答義務が発生します．しかし，ずれた話題の発話しか生成できませんので，この発話を聞いたユーザはロボットに違和感を持ってしまいます．一方，ロボットが２体いる場合で，ユーザ発話をロボットAが相槌で受けると，ロボットAにユーザ発話の話題に対する応答義務が発生します．このとき，別のロボットBが先ほどの発話を生成すると，この場合は「ロボットAに受け止められた話題について，新しい情報を提供している」という判断が成立しますので，同じ発話でもユーザは違和感なく対話を進めることができます．加えて，対話の話題が新しい話題に移るため，ロボットAに発生していたユーザ発話の話題に対する応答義務は自然に解消されます．このように，同じ発話を使っていても，ロボットを複数体に分けて提示することで，ユーザがロボットに違和感を持たないように対話を進めることができます．

⑵　音声認識誤り

致命的な音声認識誤りが起きた場合は，システムの発話とユーザの発話は，全く意味がつながらない，不適切な発話になっていると予想されます（図４⑹）．私たちは，つながりの悪い発話を自動的に検出する対話破綻検出技術を用いて，音声認識誤りを検出する技術を開発しました．また，音声認識誤りを検出した場合は，ユーザの発話を利用せず，それまでの対話の履歴に含まれる話題を利用して，ロボットどうしで対話を進めます．ユーザからすると，自分の発話が無視されたように感じられますが，話題はユーザ発話直前のロボット発話とはつながっていますし，ロボットどうしでは対話が成り立っていますので，認識誤りに基づく理解不能な発話に比べれば，違和感は大きく軽減されます．この方法により，誤った認識結果を基に話が進む場合や，ロボット１体で話を進める場合よりも，ユーザのロボットに対する印象を有意に改善することができました．

今後の展開

本稿では，これまで私たちが取り組んできたテキストチャット上の雑談システムの改善と，それをロボットへ搭載したときの効果について紹介しました．私たちは，ここで紹介した以外にも，発話の意図を表す音声合成や，雑談システムの自動評価，話者交代タイミングの改善など，さまざまな問題への取り組みを通して，気軽に雑談できるロボットの実現をめざしています．

本稿は，大阪大学の石黒浩教授との共同研究成果，およびNTTメディアインテリジェンス研究所音声言語メディアプロジェクトとの技術協力による成果を含んでいます．

■参考文献

⑴　杉山・目黒・東中・南：“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.
⑵　L. C. Tidwell and J. B. Walther: “Computer-Mediated Communication Effects on Disclosure, Impressions, and Interpersonal Evaluations: Getting to Know One Another a Bit at a Time,” Human Communication Research, Vol.28, No.3, pp.317-348, 2002.
⑶　大西・吉村：“コンピュータとの自然な会話を実現する雑談対話技術，”NTT DOCOMOテクニカル・ジャーナル，Vol.21，No.4，pp.17-21，2014.
⑷　杉山・目黒・東中：“対話システムのパーソナリティを問う質問の大規模な収集と分析，”人工知能学会論文誌，Vol.31，No.1，p.DSF-D1-9，2015.
⑸　A. Baylor and S. J. Ebbers: “Evidence that multiple agents facilitate greater learning,” Artificial intelligence in education: Shaping the future of learning through intelligent technologies, pp.377-379, 2003.

（左から）杉山　弘晃／東中　竜一郎／目黒　豊美

人は雑談を通して，対話相手と社会的な信頼関係を構築しています．人と自然に雑談し，信頼関係を構築できるようなロボットの実現をめざして，研究を進めています．

◆問い合わせ先
NTTコミュニケーション科学基礎研究所
協創情報研究部　インタラクション対話研究グループ
TEL　0774-93-5243
FAX　0774-93-5245
E-mail　sugiyama.hiroaki@lab.ntt.co.jp