

NII Today

National Institute of Informatics News

FEATURED TOPIC

ロボットは井戸端会議に入れるか

NII Interview
「井戸ロボ」を通じて、コミュニケーションの本質に迫りたい

NII Special 1
「会話分析」から探る、人とロボットとのインタラクション

NII Special 2
遠隔地間でも自然な会話を実現する、未来の井戸端会議とは

That's Collaboration 1
「井戸ロボ」と「東ロボ」、自然言語理解へのそれぞれの挑戦

That's Collaboration 2
「身振り、手振り」が会話で果たす役割をロボットに理解させる



INTERVIEW WITH

坊農真弓 Mayumi Bono

国立情報学研究所 コンテンツ科学研究系 助教
総合研究大学院大学 複合科学研究科 情報学専攻 助教
「井戸ロボ」代表研究者



「井戸ロボ」を通じて、 コミュニケーションの本質に迫りたい

ロボットは井戸端会議に入れるだろうか？ それは、ロボットが東大に入ることと同じくらい、いや、それ以上に難しいだろう。そもそも人間がどうやって井戸端会議に加わり、自然な会話をしているのかということ自体、解き明かされていない。そうした人間のインタラクション理解を深めるべく、NII グランドチャレンジの1つとして、2012年度から「ロボットは井戸端会議に入れるか」（通称：井戸ロボ）がスタートした。その概要と方向性について、代表研究者の坊農真弓助教に、会話情報学の提唱者であり、AI 研究者の西田豊明教授が話を聞いた。

西田 お久しぶりです。最近取り組んでおられる「井戸ロボ」プロジェクトはとても面白そうですね。どういうきっかけで始められたのですか？

坊農 本日はお忙しいところ、ありがとうございます。井戸ロボの端緒は、2009年10月に私がさががけ^{*1}の研究者に選ばれたことにあります。さががけでは、手話を使う聴覚障害者の遠隔コミュニケーションの研究と同時に、人間のインタラクション理解について手掛けていました。そうしたなか、2011年11月にNII グランドチャレンジとして「東ロボ」^{*}2がスタートしたんですね。そこで人間のコミュニケーションの研究をより深めたい、と「井戸ロボ」を提案したのです。東ロボが、人間の能力・頭脳・論理といったものから知能や知性にアプローチしようとしているのに対して、井戸ロボは会話を通じて人間の社会性

にアプローチしようという試みです。これまで、インタラクション理解というと、私の出身でもある人文科学の研究対象で、会話分析など各研究者の名人芸とも言える独自の手法が使われていました。そこに情報学の汎用性の高い手法を取り入れることで、インタラクション理解をサイエンスとして捉えるプラットフォームを構築できないか、と考えています。

西田 会話するロボットを作るのですか？

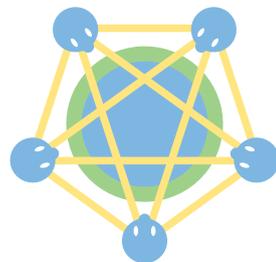
坊農 私自身が実体としてのロボットを作るわけではありませんが、実際にものを作ることで対象を理解しようという、構成論的なアプローチをとるロボット工学者にも議論に参加いただいています。インタラクション研究には、ロボット工学者にも面白いと思ってもらえるような要素がたくさんあると思っています。それらを共有したいな、という思いです。

今後は、共同研究という形で私もロボットを作ることに直接関わる可能性があるかと思っています。

西田 井戸端会議には単なる対話を超えた面白さがありますね。

坊農 2人の対話から3人以上の多人数インタラクションになるとどうなるか、ということがポイントです。従来のロボット研究では、人とロボットとの対話を想定することでロボットデザインに反映させてきたわけですが、インタラクションに重きを置くと、もう少し表面的な現象を見る必要があります。例えば、発話のタイミングやジェスチャーなど、会話の中で産出されたものを受けて人間がどう反応し、行動するかということが重要になる。ロボットデザインの仕方も変わってくると思うのです。

西田 なるほど。ところで、井戸端会議って、



英語でなんて言うのですか？

坊農 英語で Multi-Party (マルチパーティ) という講演のような大人数の集まりのことを指すようですし、Small-Party では何かネガティブな感じがして……。井戸端会議のほうがアジア的でいいという意見もあって、現状はそのまま「Ido-bata kaigi」としています。こうした学際的な研究ができるのも、日本ならではでしょう。

西田 アジア的な会話に焦点を当てたのは面白いと思います。井戸ロボの具体的な内容を教えてください。

坊農 まず大きな柱となっているのが、ロボット演劇をフィールドとする研究です。これは、ロボット研究者である石黒浩先生が手掛けたアンドロイドが登場する、演出家・平田オリザさんによる演劇を研究対象としています。この発端は、ある研究会で石黒先生から、「あなたの研究はつまらない。あなたのやりたいことは、全部オリザさんの頭の中にある！」と言われたことでした。つまり、オリザさんの頭の中には、「人間の会話のモデルがある」と言うのです。そこで発奮して、ロボット演劇の制作現場を見せてもらうことにしました。数ヶ月間にわたる稽古の様子をビデオに収めて分析を進めています。もう1つは、日本科学未来館のサイエンスコミュニケーターと来館者との会話分析。どうすれば、効果的に科学の知識を伝達できるのか、というところに着目しています。さらに、実際の井戸端会議を見てみよう、野沢温泉でのフィールドワークを始めたところです。最近では井戸端での会話はあまり見られなくて、実際には火祭りの準備のための寄り合いを研究対象としています。

西田 じつにチャレンジングな取り組みで

すね。東ロボが目標達成を第一に考える関東的なアプローチであるのに対して、井戸ロボはフィールドワークで思ってもみなかった面白いものを見つけ出そうとする関西型のアプローチであるとも言えるでしょうか。科学に対するチャレンジでもありますね。会話を科学的に分析しようとすればするほど抽象化が必要になり、切り捨てなければならぬ部分が出てきてしまいます。しかし、切り捨てられたところにこそ、コミュニケーションの本質がかなり含まれているのではないのでしょうか。

坊農 視線や身体の動きを取り入れたマルチモーダルインタラクション研究の第一人者の1人に、チャールズ・グッドウィンがいます。彼は、定量的な分析は通常しないのですが、ある論文の中で、冒頭であえて定量的な分析をして見せて、さらにそこから外れたものを詳細に分析することによって、その重要性を示すという試みをしています。

西田 それってとても重要なことですね。科学/非科学の二項対立ではなくて、双方が歩み寄ることで新たな知見を得るという方向性は、現在の成果主義・効率主義の弊害の解決にも通じる話だと思います。井戸ロボでは、最終的には何を目標しているのですか？

坊農 東ロボが1人のロボットの頭脳に知能を詰め込んで賢くしていくプロジェクトだとすると、井戸ロボでは対話やインタラクションが人間の社会を支える重要な役割を担っていることを示したいと思っています。また、従来研究において、発話、表情、ジェスチャーなどとして分断されてきたモダリティを再び統合し、まとめて見ることで人間の本質に迫りたい。今後は、それこそ「おもてなし」と

いうか、よりよいサービスのためのインタラクションにも着目していきたいと考えています。

西田 統合することで人間の本質に迫る、という部分は東ロボとも共通するわけですね。世の中には問題解決型のエキスパートも必要だけれど、吉本新喜劇のようなお笑いやお

もてなしがないとつまらないですからね。後者をサイエンスとして取り組むことは大変難しいでしょうが、そこにどう切り込むかが腕の見せ所ですね。井戸ロボがどう発展していくか、大変楽しみにしています。

(構成・文=田井中麻都佳)



インタビュアーの一言

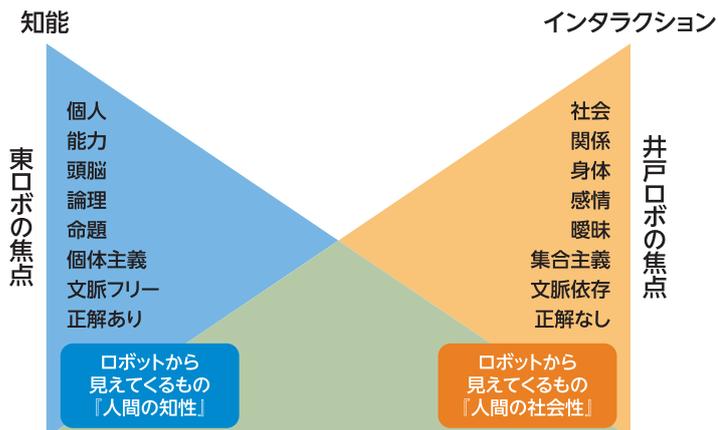
会話情報学は会話という現象の理解とその工学的应用を軸として、個体の知と集合の知の相互作用を通して人の心の過去と未来を探ることをねらっています。坊農さんの井戸ロボは合理性だけでは割り切れない人の心の深部にある光と影を浮き彫りにする可能性を秘めています。東ロボといろいろな面に対照的ですが、統合することで人間の本質に迫るという点が共通しているところが面白いと思います。これからのチャレンジの進展が大変楽しみです。

西田豊明 Toyooki Nishida

京都大学大学院 情報学研究科
知能情報学専攻 教授

1977年京都大学工学部卒業。1979年同大学院修士課程修。1993年奈良先端科学技術大学院大学教授、1999年東京大学大学院工学系研究科教授、2001年東京大学大学院情報理工学系研究科教授を経て、2004年4月京都大学大学院情報学研究科教授、現在に至る。会話情報学、社会知デザイン、原初知識モデルの研究に従事。

※1 さきがけ
科学技術振興機構の個人型研究プロジェクト (PRESTO)。戦略目標に基づいて未来のイノベーションの芽を育むことを目的としている。
※2 [東ロボ]
NII グランドチャレンジ・人工頭脳プロジェクト「ロボットは東大に入れるか」<http://21robot.org>
本誌60号 http://www.nii.ac.jp/muom2c5rm-4542/#_4542



図：東ロボと井戸ロボの対比

「会話分析」から探る、 人とロボットとのインタラクション

「井戸ロボ」は、会話や身振りなど、人間の日常のインタラクションを理解するというアプローチを、情報学の中に位置づけるためのプロジェクトだ。学際的な取り組みにより、今後のロボットデザインに生かせるような知見を探りたいという。その意味で、社会学者の西阪仰教授が専門とする「会話分析」とも深く繋がっている。同じく人文科学出身で、会話分析も手掛ける坊農助教が、西阪仰教授に「人とロボットのインタラクション」に対して社会学が果たす役割について、話を聞いた。

発話の構造を探る

坊農 従来、会話ロボットの研究は、人間の言語を認識する装置を作り、それをロボットに組み込めばロボットがしゃべるようになるだろうというフレームワークの中で進められてきたように思います。それに対して「井戸ロボ」では、二者以上の人のインタラクションの中にある「構造」のようなものを追究しています。つまり、人間のソーシャルなやりとりの構造を探ることで、ロボット研究に別の視点を提供できればと思っています。

その研究の柱の1つが、「ロボット演劇」へのアプローチです。演出家の平田オリザさんの演出を見ながらいろいろデータを取って、細かな動作や視線の動きと会話の関係性を探っています。なかでも、会話の開始とコミュニケーションに着目しています。稽古中、演出家の指示によって繰り返されるセリフの流れを分析する目的で、「会話分析」という手法が使えらると思うのです。

西阪 面白いですね。会話分析では、たとえば、会話で次に誰が話すかという順番交替のルールが、一見、そのままコンピュータでプログ

ラミングできるような形で定式化されています。そのため、一時、人工知能では、それを使って会話するロボットを作れるのではないかと考えられていました。でも、うまくいかなかった。なぜなら、ルールの定式化は可能であっても、実際の発話はあまりにも文脈に依存しているからです。

順番交替のルールというのは、次の話者を選択するための2つのテクニックの間に、優先順位をつけるというもの。それは、現在の話者が次の話者を選ぶか、次の話者が自ら話し始めるか、という2つです。前者のテクニックは、2つの要素から成立しています。1つは、現在の話者が発言を特定の人に向けてことで、もう1つは、特定の行為を強く促すような発話をすることです。例えば、ある特定の人に質問が向けられるならば、その質問をされた人が次の話者となり返答することが強く促される、という具合です。

坊農 でも、実際の場面では、理屈通りにはいかないわけですね。

西阪 はい。「質問」といってもさまざまで、「～か?」のように明確な終助詞が常に用いられるわけではありません。さらに、質問が誰に向けられたものか一義的には決まらない。「坊農さんのご出身は?」という問いかけは、坊

農さんがその場にいるときは坊農さんへの質問ですが、坊農さんがいない場面でも、坊農さんを知っている人に対して使うことができます。つまりそれは、文脈次第です。そういう「発話のデザイン」の構造を定式化してロボットに会話させるのは、非常に困難なのです。逆に言えばこれは、文脈に対する人間の適応性がいかに柔軟かということを示しています。「井戸ロボ」を大変興味深いと思ったのは、人間のように会話するロボットを作るのではなく、人間みたいな格好の物体を人間のなかに置くことによって、会話する人間のほうを見ようという、発想の転換が感じられるからです。

ロボットらしさ、 人間らしさ

坊農 そもそも、工学者があまりにも簡単に会話を文脈から切り離してデザインしてしまうので、もう少し会話の複雑性を取り入れたという思いから「井戸ロボ」をスタートしました。ロボット演劇の演出を見ても、会話の始まりは何気ない会釈やアイコンタクト、歩み寄ることだったり、実にさまざまです。例えば実験室でデータ収録を行った、ロボットが施設案内をするという想定の場合。これは案内ロボットと人が会話しているところに、もう1人が通りかかるという設定です。その通行人の視線を受けて、3人に複雑な視線交錯が生まれます。さらに、見るタイミングや視線の返し方、ロボットの通行人に対する会釈など、オリザさんの細かな演出指示が何度も加えられていく。すると、通行人は無言の



坊農真弓 Mayumi Bono

国立情報学研究所 コンテンツ科学研究系 助教
総合研究大学院大学 複合科学研究科 情報学専攻 助教
「井戸ロボ」代表研究者

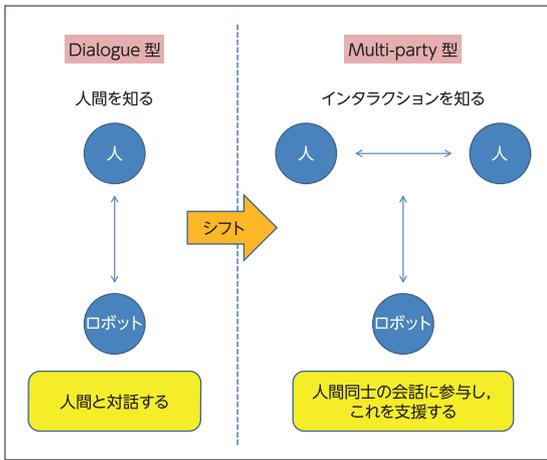


図1：本プロジェクトの核

第三者でありながらもインタラクションへの関与を高め、3人の会話の関係性を明らかに変えていくのです。

西阪 ロボットらしさがどう演出されるのか、興味がわきますね。やはりロボットは人間とは違う存在ですから。まさに、人間的にふるまうロボットが人間とどこがどう違うのか、その差を明らかにしようという試みにも見えます。

「会話分析」の知見を 社会に還元する

西阪 会話分析とはそもそも社会学の研究領域の1つで、会話という人間の相互行為を観察・分析する手法です。会話を分析するという発想は、1960年代初め、民生用の録音機の登場により、容易に音声の記録、再生が可能になったことがきっかけになりました。その後、小型ビデオで映像も手軽に記録できるようになると、視線やジェスチャーなどの視覚的リソースも、会話を運営していくための仕組みとして注目されるようになります。ただ、言葉のやりとりを動かしていく要素はある程度体系的に抽出されているのに対し、音声以外のリソースはアナログ的すぎて、定式化するのは難しいのです。

坊農 音声以外はまだまだ難しいですね。でも、音声会話を対象にした会話分析の長年の成果が社会に役立ちつつもある。応用例として、警察や消防への通話記録の分析が裁判での資料として役立つことがあると聞きました。西阪先生は最近、東日本大震災後の福島の避難所での「足湯」の会話分析をされて、本に

図2：未来の人間とロボットの関係性を考えるために、ショッピングモールのロボットと客のやり取りを平田オリザ氏が演出。50パターン近く演出し分ける。ここではロボット役も人間の役者が演じている(2013年5月14-15日撮影、ATR社会メディア総合研究所協力)。



まとめられましたね。

西阪 まだ混乱状態にある避難所では、被災者自身も何を必要としているのかよくわからないことが多い。そんなとき、阪神淡路大震災、中越地震の際の経験から、足湯が導入されました。足湯につかっていると、いま困っていることとか自然にいろいろな話が出てくるというのです。そこで、会話分析によって足湯のコミュニケーションの構造を見出せないだろうか、と考えました。ボランティアの活動の一助にもなれば、と考えたのです。

坊農 何が見えてきたのでしょうか？

西阪 マッサージにより話がしやすくなることもあるかもしれませんが、それ以上に、会話自体は止まってもインタラクションは継続するという構造の重要さがわかってきました。つまり、視線の動きやマッサージの手の動きが、インタラクションを支えているベースになっている。一方で、それはあくまでも土台

であって、その上に構築される会話の妨げにはならないという仕組みがあるんですね。だから、足湯につかりマッサージを受けているうちに、通常、ボランティアの側からは尋ねることをしない被災経験についても、被災者のほうから自然と語り出すことが少なくないのです。同じような構造があれば、足湯ではなくても、避難所や仮設住宅の特徴にあったコミュニケーションのあり方を提示できるかもしれません。

坊農 何かをしながら会話をすることで、新たなインタラクションが生まれる——「井戸ロボ」につながる話ですね。

西阪 インタラクションの中にロボットを置くことで、逆に人間のインタラクションの構造が見えてくる。それこそが「井戸ロボ」の醍醐味なのでしょうね。研究の進展に期待しています。

(取材・文＝小原誠之)

西阪 仰 Aug Nishizaka
明治学院大学 社会学部 教授

遠隔地間でも自然な会話を実現する、 未来の井戸端会議とは

日常会話においては音声による言語情報だけでなく、視線やジェスチャー、身振り・手振りといった非言語情報が頻繁にやり取りされている。人間とロボットとの自然な会話を実現するには、そうした非言語情報の伝達も重要なテーマだ。その前駆的な研究とも言えるのが、テレビ会議システムの取り組みである。NIIと共同研究を進めるNTTコミュニケーション科学基礎研究所の大塚和弘主幹研究員に、研究の最新動向と成果を伺った。

日常に溢れる 非言語情報のやり取り

日常会話や会議等のコミュニケーションの場では、音声による言語情報だけでなく、視線や表情、ジェスチャー、身振り・手振り、声の調子といった非言語情報がやり取りされている。これらの情報が総合的に作用することで、さまざまなコミュニケーションが行われ、会話や話題も変化していく。したがって、人間との共生を目指す井戸ロボの実現には、言語情報だけでなく、非言語情報も含めたコミュニケーションプロセスを明らかにすることが重要だ。

このようなテーマを情報工学の立場から追究するため、NTTコミュニケーション科学基礎研究所（以下CS研）によって開発されたのが「実時間マルチモーダル会話分析システム」である。

同システムは、複数の人が会話する場面を自動的かつリアルタイムで分析するもの。具体的には、8人程度の小規模会議において、テーブル上に設置された全方位カメラ・マイク統合システムが画像情報や音声情報を収集。得られた情報を処理・統合することで会話の状況を分析し、その結果をディスプレイ上に表示する。これにより、「誰がいつ話しているか」「誰が誰を向いているか」「誰が誰に注目しているか」といった会話の状態をリアルタイムにモニタリングできるようになる。

CS研の大塚和弘主幹研究員は、同システムの開発に至った背景について、「遠隔地間における人と人とのコミュニケーションを可能にする手段として、テレビ会議システムに期待が寄せられていました。しかし、実際の対面会話ほどスムーズな会話を実現するには至っていません」と説明する。

場の空気を伝えきれない 従来のテレビ会議システム

従来の平面上で展開されるテレビ会議システムは、「誰が誰を見ているのか」わかりにくいという視線不一致の問題をはじめ、場の空気を伝えることが困難だった。そこで、よりリアルな会話の場をテレビ会議で再現するためには、場の空気の醸成をはじめ、人どうしのコミュニケーションがどのようにして成立するのか、その過程を明らかにすることが必要との結論に至ったという。

大塚主幹研究員は、「実時間マルチモーダル会話分析システム」では、CS研が蓄積してきた画像技術、音声技術を活用し、人の行動を画像・音声情報として観測。そこで得られた



古山宣洋 Nobuhiro Furuyama

国立情報学研究所 情報社会相関研究系 准教授
総合研究大学院大学 複合科学研究科 情報学専攻 准教授
東京工業大学 大学院総合理工学研究科
知能システム科学専攻 連携大学院講座 連携准教授

情報を計算機で分析することで、コミュニケーションのメカニズムを解明しようと考えたのです」と話す。

同システムの開発にあたり、1つの示唆となったのが、坊農真弓助教の研究だったと大塚主幹研究員。

「これまでのコミュニケーション研究は、言語が先にありましたが、会話の場における言語的な内容だけではなく、人間の行動も研究対象になるという気づきを、坊農先生の研究から受けたのです」

一方、坊農助教も、「コミュニケーションの研究を人間の行動から探究していくという発想をもっている研究者は情報学には少なく、お互いに協力しあえると思っていました。とくにこのシステムの優れた点は、複数の会話映像をディスプレイ上で一覧できるよう、シリンダー式に提示する方法です（写真1）。これにより、複数のビデオカメラによる従来の撮影手法では落ちてしまっていた、会話場の



大塚和弘 Kazuhiro Otsuka

日本電信電話株式会社
NTTコミュニケーション科学基礎研究所
人間情報研究部 感覚共鳴研究グループ
主幹研究員（特別研究員）博士（情報科学）

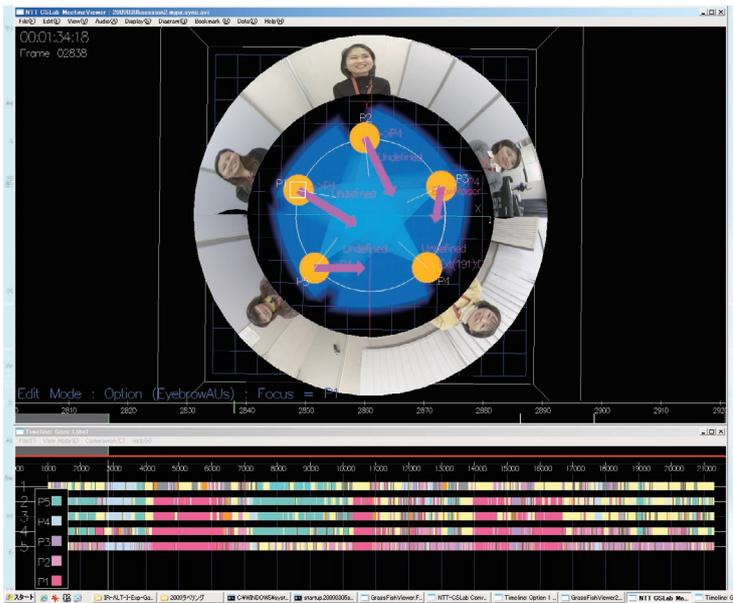


写真1：実時間マルチモーダル会話分析システム

空間情報が再現され、視線や身体方向の分析がとてもしやすくなりました」と話す。

また、人間の知覚、行動、身体性コミュニケーションを研究する立場から井戸ロボのプロジェクトに参加する古山宣洋准教授も、「人々がコミュニケーションのためのリソースとして、曖昧な部分が多いジェスチャーをどのようにして利用しているのか、それを追究するためにも、実時間マルチモーダル会話分析システムは非常に有効なものであると実感しました」と語る。

2009年には、同システムと大塚主幹研究員の研究室に設置された映像収録装置を用いた「音声4者会話・手話4者会話コーパス」を共同で開発。さらに同年10月のNTTと国立情報学研究所の組織対応型（包括的）連携



坊農真弓 Mayumi Bono

国立情報学研究所 コンテンツ科学研究系 助教
総合研究大学院大学 複合科学研究科 情報学専攻 助教
「井戸ロボ」代表研究者

契約の締結も契機となり、以来、両者による井戸ロボ、および「未来の井戸端会議」の実現に向けた共同研究を進めているところだ。

次世代のテレビ会議システム 「t-Room」「MM-Space」とは

遠隔地間における円滑なコミュニケーションの実現を目指してきたNTT・CS研。その研究成果の一例が、次世代テレビ会議システムの「t-Room」、[MM-space]だ。

t-Roomは、遠隔地間においても会話者どうしが“同じ部屋にいる”と思わせるようなコミュニケーションの実現を目的に開発されたもの。具体的には、壁面に複数のディスプレイを配置した多角形の空間を用意。各ディスプレイの上側には、ビデオカメラが設置されており、撮影された参加者の映像が他の場所にあるt-Roomのディスプレイ上に投影される（写真2）。参加者が移動すれば、映像も合わせて隣のディスプレイに移動。さらに、参加者が特定の方向を指させば、全員がその方向を見るなど、t-Room内にいる参加者は同じ空間を共有しているかのようなコミュニケーションが可能だ。

一方、「MM-Space」は、遠隔地にいる複数の人の会話を、あたかもその場で聞いているかのように再現可能なシステム（写真3）。実際に会話を行う空間において会話者の顔画像と音声を取得し、会話を再現する空間に伝達する。伝達先となる再現空間には、会話者の



写真2：t-Room



写真3：MM-Space

配置に合わせて、複数のプロジェクタや透過型スクリーン、アクチュエータ、スピーカーを配置。会話者の顔画像は背景を除去した上で透過型スクリーンに映し出されるが、そのスクリーンは会話者の頭の動きと同期して制御されるようになっている。つまり、実際の会話で行われている視線の動きや首振り、うなずきといった非言語情報を解析し、スクリーンの物理的な動きとして反映させているのだ。

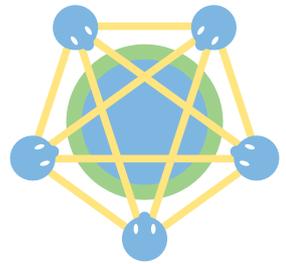
大塚主幹研究員は、「これにより、『誰が誰に注目しているか』『話題に共感しているのか』といった、いわばその場の空気までも、よりはっきり伝えることができると期待しています」と強調する。

現在、会話における非言語情報のより詳細な伝達を実現するための機能強化を進めている。その一例がアクチュエータを前後左右に稼働可能にしたことで、「興味を引いた話題には前のめりになる」といった動作も、反映できるようにしている。

古山准教授は、「複数人の会話でやりとりされる多様な情報を解析していくことで、よりコミュニケーションの本質に近づけると考えていますが、MM-Spaceによって得られる知見は、今後の井戸ロボ研究の1つの指針になるものです。MM-Spaceは単なる分析ツールの域に留まらない、研究のパラダイムを変えてくれるものと考えています」と、大きな期待を寄せる。

テレビ会議システムの進化によって、未来の井戸端会議の姿が見え始めた。今後の進展に、さまざまな夢が膨らむ。

（取材・文＝伊藤秀樹）



「井戸ロボ」と「東ロボ」、 自然言語理解へのそれぞれの挑戦

NII では、「ロボットは井戸端会議に入れるか」と「ロボットは東大に入れるか」という2つのテーマでそれぞれグランドチャレンジが進行中だ。どちらも人工知能の研究領域でありながら、研究戦略は正反対に見える。言葉やジェスチャー、表情などの多様な「モダリティ」の総合で行う複数人の自然な会話を研究する伝康晴教授と、文字や記号で記述された論理的で正解のある命題を解く人工頭脳を研究する宮尾祐介教授。自然言語処理研究のバックグラウンドが共通する両氏の研究アプローチの違いと共通点はどこにあるのだろうか。

自然言語処理の基盤になるのは 「コーパス」づくり

宮尾 私たちが取り組んでいるいわゆる「東ロボ」では、自然言語で記述された入試問題を解くために、言語の電子化資料である「コーパス」を活用しています。コーパスは1980年代から充実するようになり、機械翻訳などはこれをベースに大きく発展しました。伝先生は言語学の立場からコーパスの研究をされておられますが、言語学ではどのような位置づけなのですか？

伝 マイナーな領域かもしれませんが。しかしコーパスは機械翻訳をはじめ、自動応答システムなどさまざまなシステムに応用されて

います。機械翻訳の場合、1980年代以前は学者がアタマで考えた手法でシステム化されたので、現実にはあまり役に立ちませんでした。コーパスとして新聞記事や会話の書き起こしなど実世界の言語資料がまとめられることにより、実用可能なレベルに大きく前進しました。自然言語を最小の単位に分割して品詞を判別する「形態素解析」という基礎技術の精度は、すでに新聞記事で98.5%、ブログでも95%以上に達しています。これは現実の言葉によるコーパスがあればこそでしょう。

宮尾 自然言語処理にコーパスは不可欠ですね。「東ロボ」は2021年の東大入試突破を目標にしていますが、本当は「思考するプロセス」の研究です。研究をできるだけシンプルにするために、明快に記述された文章に焦点を絞り、曖昧な言語要素を省くようにした結果、入試試験問題への解答という目標ができたのです。これは「井戸ロボ」のアプローチとはだいぶ違うようですね。

単一モードのコーパスから マルチモーダルなコーパスへ

伝 「井戸ロボ」は非言語要素も含めた現実の会話に入り込むのが目標ですから、ジェスチャーや表情、視線の変化、会話中に行われる直接内容と関連しない動作も研究対象となります。そうした要素は「モダリティ」と言い、会話は「マルチモーダル」に行われるのが普通です。

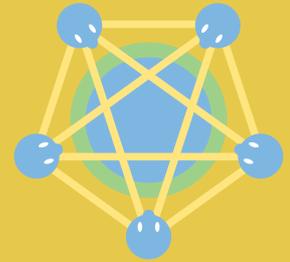
マルチモーダルな会話から1つのモダリティだけを取り出してきたのが過去のコーパスです。たとえば1990年代末には「道順を教える対話」などの課題指向型対話のコーパスづくりが行われました。やがて「雑談」や

伝 康晴 Yasuharu Den
千葉大学文学部 行動科学科 教授



宮尾祐介 Yusuke Miyao
国立情報学研究所 コンテンツ科学研究系 准教授
総合研究大学院大学 複合科学研究科 情報学専攻 准教授
「東ロボ」 サブプロジェクトディレクター





食べ物を口に入れる					食べ物をつかむ		
P	H	P	H	P	S	P	S

A: クジラ漁ってさー、絶滅する
からーだめなのそれとも可哀そうだからだめなの

ふーん

B: 絶滅ー [でしよ

C: [絶滅ーじゃない

事例の説明

- Aは発話を終えてから食べ物を口に運ぶのではなく、発話途中から箸を上げ始める (P: 準備)。
 - Aは上げた箸を空中で止め (H: 保持)、発話が終わるタイミングを見計らう。
 - Aは発話を終えるや否や、食べ物を口に入れ (S: ストローク)、他者が反応を返している間に咀嚼を終え、次発話が行なえる状態にいち早く到達する。
- ⇒ 保持は次発話に備えるための迅速な摂食につながる

イラスト制作者 (人物) NII 坊農研究室 特任研究員 城 綾実

3人以上の人の会話などのコーパスも出してきました。しかし会話を本当に理解するには、会話をマルチモーダルなままでも収録して分析する必要があります。

これに取り組んだのが2003年の千葉大学による「3人会話コーパス」です。これは友人同士12組がサイコロを振って初期の話題を決めて、その後は自由に話す様子をヘッドセットマイクとビデオで収録し、一部をアノテーション (付加情報の付与) しました。しかし、この事例を含め、コーパスづくりは実験室に人を集めて、特殊な環境の中での会話収録にとどまっておらず、日常場面の中での会話とは違います。

円卓での会食風景を収録してアノテーションを行う

伝 そこで、私たちは3人が円卓を囲んで会食しながら会話するという日常的なシチュエーションでのコーパスづくりに取り組みました (映像データは東京電機大学・武川直樹教授より提供)。ここでは口は発話と食べることの両方に使いますし、手はジェスチャーにも使えば、食器や箸を使うためにも使います。食事と会話をどう両立するかなんて、普段は

意識することはありませんが、人工知能で模倣しようとするの大問題です。これはマルチモーダルな会話分析の新しい視点だと考えました。

「食べ物を口に入れる」のような食事動作を「準備」「保持」「ストローク」「復帰」といったサイクルに分割して時間軸で記録し、会話の内容との突き合わせを行いました。するといろいろなことがわかりました。たとえば自分が話しているときは、しゃべりながら箸で食べ物を持ち上げて空中で保持しており、話し終わったら約0.2秒で口に入れていることです。これは話を終えるタイミングを図って、食事動作を事前に開始しているわけですね。また他人が話した直後に自分が話し出すときの時間差も0.2秒ほどでした。相手の話が終わる手がかりを事前につかんでいるからできることです。

分析前には「発話時間と食事時間は交互にとる」とか「発話量が多い話者ほど口に入れる動作が短い」といった仮説がありましたが、実際には裏付けられず、発話量の多い人はいったん開始した「口に入れる」動作を一時停止しながら発話して、食事と会話を両立させていたのです。まだ話の内容にまで踏み込んで分析していませんが、体験談を話している状況では途中で食事動作をしても、他の人が口をはさまない傾向なども見えてきました。

宮尾 おもしろいですね。英語の入試問題では2人の会話の穴埋め問題があります。人間が取り組みばなんでもない問題でも、機械ではこれがなかなか解けません。人間なら意識

せずに自然に理解できるような会話の流れを、機械が常識として備えていないからではないかと思います。そうした部分を機械に教えてくれるようなデータがあると助かります。

「火祭り」準備のシチュエーションでさらに多様な状況も対象に

伝 「井戸ロボ」のためには井戸端会議のコーパスが欲しいところですが、プライバシーなど制約条件が厳しくて収録ができずに困りました。そこで昨年からは野沢温泉 (長野県) で毎年1月に行われる火祭り (道祖神祭り) の準備現場に向き、三夜講と呼ばれる実行チームの行動記録をとっています。三夜講は後見人、世話人、見習いの年齢別3グループからなっていて、ご神木の切り出しや社殿の組み立てなどの作業を役割分担しながら多人数が連携して仕事をします。言葉に加えて仕事を見せることで、古くからの知識を伝承していくわけです。

これは目的指向のインタラクションの典型的なシーンといえます。他人をモニタして自分の行動を判断する場面が多いこと、また日常的な会話がくり広げられ、多人数が関わるという点で、「井戸ロボ」に通じるデータが得られると思っています。

宮尾 人間が頭の中で考えていることを研究している「東ロボ」プロジェクトに、この研究を今すぐに適用できるわけではないですが、私たちの研究が進めば、やがて足りないところが見えてくるでしょう。そこに伝先生たちの成果が生かせそうです。20年から30年といった将来には、研究が統合できる日が来るかもしれません。ぜひこの分野の開拓を進めて、情報科学の研究者が利用できるデータにまとめていただければ幸いです。

(取材・文=土肥正弘)

「身振り、手振り」が会話で果たす役割をロボットに理解させる

東京工業大学の岡田将吾助教は、人々の会話での「ジェスチャー」を機械学習の手法で分析する研究を続けてきた。「井戸ロボ」では、井戸端会議のような日常会話でのジェスチャーの役割を、コミュニケーション科学の知見を取り入れつつ分析していく方針だ。一方、人工知能、HAI (Human-Agent Interaction)、ISS (Intelligent Interactive Systems) を研究してきたNIIの山田誠二教授は、「井戸ロボ」について、外部の立場から今回の研究の工学的な意味を問いかける。

前例が乏しい研究に、 コミュニケーション科学の 知見を活用

「井戸ロボ」では、音声だけでなくジェスチャー（身体の動き、身振り）を含めたマルチモーダル（多手段）なインタラクション（会話）の理解を狙う。また、会話の理解には、話者以外の人々のジェスチャーも重要だと考える。そこで、人々の会話を、言葉だけでなく居合わせた人々のジェスチャーまでを含めて記録、分析し、それらを「会話するロボット」の設計に結びつけようとしている。

このようなジェスチャー分析に関わっているのが、東京工業大学の岡田将吾助教だ。岡田助教は、以前から人間のジェスチャーをコンピュータに取り込み、機械学習の手法により分析する研究を続けてきた。

ジェスチャー認識は、すでに一部では実用

化されている。

「ゲーム機の任天堂 Wii のコントローラや、マイクロソフト Xbox の Kinect では、ゲームのコントロールにジェスチャーを使っています。例えば釣りのゲームで、“竿を振る”動きをしたら、ゲーム内でも竿を振るようにする。このように、約束事（プロトコル）が決まっているものでは、実用化されている例や研究例がたくさんあります」と岡田助教は言う。

その一方で、日常会話に伴うジェスチャーはより難しい研究対象で、研究例は乏しい。

「日常会話において表出される非言語のうち、“視線の動き”などは話し相手・興味対象の特定に有用であることから盛んに研究が行われています。しかし、手の動きのようなジェスチャーは“ノイジー”で意味づけが難しい。例えば、会話の中で、言葉が出ない“言いよどみ”の最中に手が動いていることがありますが、これも非言語コミュニケーションの一種です。こうした今までノイズとして捨てていた情報の中から、意味がある情報を認識できればと考えています」（岡田助教）。

そのため、「井戸ロボ」では、「ジェスチャー

は喋っている言葉と共起性がある（同時に出現しやすい）」といったコミュニケーション科学の知見を取り入れ、より深いレベルで会話を理解することを狙う。

会話中のジェスチャーを 機械に認識させることに成功

岡田助教の研究成果として、日常会話の中のジェスチャーのパターンの発見がある。

例えば、Aさん、Bさんが会話をしていると。ここで、「Aさんが話している」「AさんがBさんを見ている」「Bさんがうなずいている」といった一連の会話やジェスチャーを含むデータを、「マルチモーダル時系列データ」と呼ぶ。

岡田助教は、このような「マルチモーダル時系列データ」の中で、会話の進行に伴うジェスチャーの順番（時系列）には意味があるとの仮説を立てた。そこで、マイクや、人の動きを検出するモーションキャプチャーなど複数の手法で人々の会話中のジェスチャー、つまり「非言語行動」を検出し、記録した。話者の顔が向いている方向（視線方向）や頭部などの身体の動き、それに手の動きのパターンを集め、これらを「教師なし学習」*と呼ぶ機械学習のアプローチでパターンを自動的に抽出した。

その結果、「話者が説明時に使うジェスチャーは聞き手の視線を伴う」「説明時のジェスチャーと“うなずき”が共起しやすい」「聞



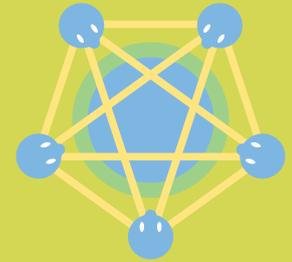
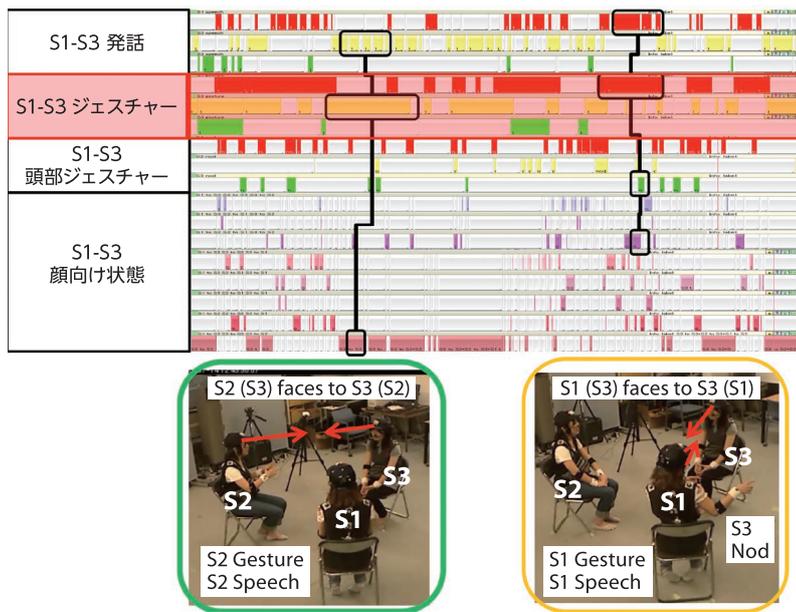
岡田 将吾 Shogo Okada

東京工業大学大学院 総合理工学研究科
知能システム科学専攻 助教

山田 誠二 Seiji Yamada

国立情報学研究所 コンテンツ科学研究系 教授
総合研究大学院大学 複合科学研究科 情報学専攻 教授
東京工業大学 大学院総合理工学研究科 知能システム科学専攻
連携大学院講座 連携教授

会話者3名の発話・頭部動作・顔向け状態



会話中に用いられたジェスチャーの役割・機能を認識するためには、会話参加者の非言語行為（発話・頭部動作・顔向け状態）を利用することが有効である。左下、右下の2つの例のように、説明の補助的役割を担うジェスチャーを行っている人を聞き手が見ていたり、うなずいていることがわかる。ジェスチャーと共起する非言語行為を同時に観測することで、ジェスチャーの役割認識精度を向上させることが可能となる。

「私たちは社会性をもっていて、例えば『視線を向けられたら、喋ってもいい』といった非言語コミュニケーションを理解できる。ところが、こんな簡単なことがロボットにとっては難しい。『場の空気を読む』じゃないが、その場の人々の役割、社会性を機械に理解させることは難しいのです」（岡田助教）。

そして、「井戸ロボ」での研究は、こうしたロボットにとって難しいテーマを、工学、計算機科学のような「理系」だけでなく、コミュニケーション科学など「文系」の知見を取り入れて研究しようとしている。例えば、コミュニケーション科学の知見を取り入れ、会話中のジェスチャーが「説明時のジェスチャーか、否か」を機械学習により判定する際に、ジェスチャーを行っている人以外の動き（非言語情報）を取り入れることで、認識精度を2～16%向上できるという成果が得られた。

このようなコンピュータ処理によるジェスチャーの理解が進めば、コミュニケーション科学の研究手法にも革新がもたらされる可能性がある。従来はビデオを見ながら人間が分析していた作業を、機械で自動化でき、より大量のデータを蓄積できるようになるからだ。

「学際的な取り組みゆえの成果が出せると考えています」と、岡田助教。一方で山田教授は、「ぜひ、科学としての成果だけでなく、工学的なアプローチから実用への道筋をつけてほしい」と鼓舞する。今後の研究の進展に大いに期待したい。

（取材・文＝星 暁雄）

き手の質問が出た直前・直後でジェスチャーを用いた説明が行われる」などのパターンが検出できた。

これらのパターンは、私たち人間にとっては当たり前にも感じられるかもしれないが、コンピュータが認識できるようになったことは大きな進化だ。機械的な手法により、大量の人間の会話を、より深く理解できるようになり、ロボットにも応用可能となることを意味するからだ。

工学的にも挑戦的な 課題が山積み

一方、ロボットと人間の対話に関して長く研究を続けてきた山田誠二教授は、「井戸ロボ」の外部という立場で次のようにコメントする。

「『井戸ロボ』では、現実のロボットは作らずに会話の研究をする、ということになっているが、私としてはロボットまで作ってほしいですね。人工物であるロボットが、おばあちゃんたちの世間話に加わるところまでいけば、それは本当に素晴らしいことです」

現実には、人間そっくりなロボットを作ったとしても、人間同士の会話に加わることは難しい。ロボットに可能な動きは、人間の自然な動きとは明らかに違うために、人間から

見るとかえって不自然に感じてしまうからだ。そこで、「人間とは違う外見をもつモノの、人間からみて親しみやすい動き、ジェスチャーとはどのようなものか、これは研究対象としては興味深いものになるはず」と山田教授は指摘する。例えば、アニメーション映画『天空の城ラピュタ』（宮崎駿監督、1986年）に登場するロボットは人間とは外見が異なるが、手足の動きやライトの反応などにより、人間とのコミュニケーションがある程度成立しているように表現されている。ここにもロボット研究のヒントが含まれている、と山田教授は言う。

また、今回の研究で扱うマルチモーダル時系列データを機械学習により処理することは、そもそも難しい。

「どのような特徴量に注目するのか、学習アルゴリズムをどうするか、そのような機械学習の本質に関わる貢献ができる可能性がある」と期待しています」（山田教授）。

学際的な研究経験を 積むことに意義がある

岡田助教は、山田教授の指摘を受け、「井戸ロボ」と現実のロボットのギャップについて、次のように話す。

※ 教師なし学習

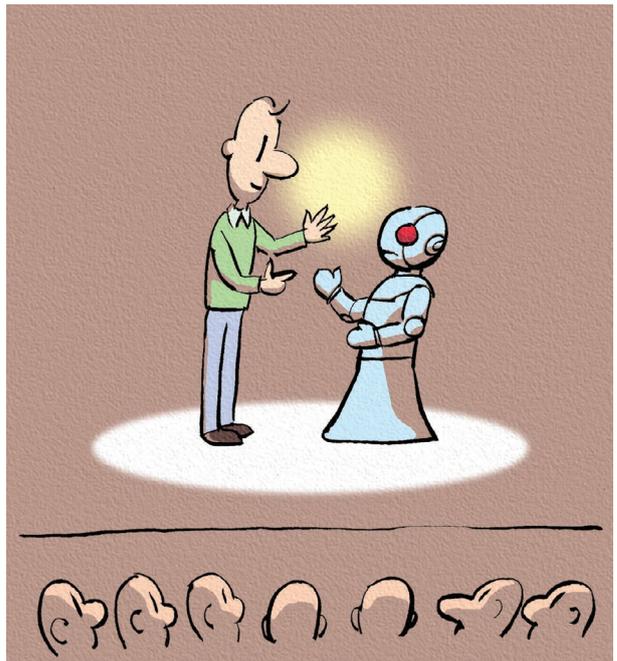
正解がわかっていない問題に対する機械学習の手法。大量の入力データの背後に隠れた本質的な構造を抽出するために行われる。

ロボット演劇事始め

平田オリザ Oriza Hirata

劇作家、演出家

大阪大学 コミュニケーションデザイン・センター
教授



大阪大学の石黒浩教授とロボット演劇のプロジェクトを始め、すでに6年になる。この間、15カ国33都市（国内12都市を含む）で上演を行い、大きな反響を得てきた。科学研究費を使っただけのプロジェクトで、これほど短期間に目に見える成果をあげた研究も珍しいだろう。

私が大阪大学に移籍して1年ほど経ったある日、大学のPRイベントの控え室で、当時阪大総長だった鷺田清一氏と世間話をしてた。鷺田氏の総指揮で始まった大阪大学コミュニケーションデザイン・センターは徐々に軌道に乗り、私もいくつか演劇の授業を開講した時期だった。

「他に、何かやりたいことはありますか？」と聞かれて、かねがね心に抱いていた懸案を相談した。

「ロボットを使って演劇がやりたいんですが？」

鷺田総長はすぐに、石黒先生と、当時の上司の浅田稔先生に連絡を取ってくださった。1週間後、浅田先生の研究室をはじめて訪ねた際のことは、いまも、よく覚えている。最初に私は、以下のことを伺った。

「私がプロジェクトに入ると、ロボットが、いま持っている技術以上のものを、あたかもそれを持っているかのように見せる

ことができますが、それをやってもいいですか？」

これは、通常の学術の世界では禁じ手であろう。学会発表でそのようなことをしたら、ねつ造のそしりさえ免れない。しかし浅田先生は、開口一番、「望むところだ」と仰った。そして、浅田先生のお墨付きを得て、私と石黒先生のプロジェクトが始まった。二人は、ほぼ同世代であるだけでなく、人間に対する見方、コミュニケーションに関する考え方が、おもしろいほど一致しており、研究は最初からフルスロットルで加速していった。石黒先生がかつて画家を志しており、私も科学技術を題材にした演劇を多く書いてきたといったお互いの来歴も、プロジェクトをスムーズに進行させる要因となった。

どのようなロボットを作るかだけでなく、それをどう見せるか、そのために「演出」を行うという発想は、石黒先生だけが持っていた。そこに私が大阪大学に赴任し、このプロジェクトが実現した。日本の大学が、まだ大学としての価値を多少は有しているということの証左であろう。

私はこのロボット演劇プロジェクトの誕生の経緯を、世界中で話してきた。この話のオチは、そうはいつでも二人を出会わせた鷺田清一氏、すなわち哲学者が一番偉いという結論である。

情報から知を紡ぎだす。

NII

表紙イラスト

かつて共同井戸は生活用水として利用され、そこに集まった人々は水くみや洗濯をしながら、世間話や噂話に興じた。それは地域を支える重要な情報交換の場でもあった。はたして、そんな井戸端会議にロボットは入れるだろうか？ それは人間と猫が会話するくらい難しいことかもしれないが、その実現に夢が膨らむ。

国立情報学研究所ニュース [NII Today] 第62号 平成25年12月

発行：大学共同利用機関法人 情報・システム研究機構 国立情報学研究所 <http://www.nii.ac.jp/>

〒101-8430 東京都千代田区一ツ橋2丁目1番2号 学術総合センター

編集長：東倉洋一 表紙画：小森誠 写真撮影：川本聖哉 / 佐藤祐介 デスク：田井中麻都佳 制作：クディラ アンド アソシエイト株式会社
本誌についてのお問い合わせ：総務部企画課 広報チーム TEL：03-4212-2164 FAX：03-4212-2150 e-mail：kouhou@nii.ac.jp