



NTT人間情報研究所
客員上席特別研究員

東 中 竜 一 郎 Ryuichiro Higashinaka

古くて新しい課題「共通基盤」を、最先端技術を活用して解明し、コンピュータと人のコミュニケーションに挑む

人と人のコミュニケーションにおいて、共通の話題・前提条件・認識等の「共通基盤」が一致しているときにお互いに理解し合えてコミュニケーションが成立します。共通基盤を構築することができれば、コンピュータと人のコミュニケーションもより円滑になります。共通基盤の概念は古くからありましたが、人の認知と深く関連することもあり、いまだに解明されていないことも多々あります。それを1つひとつ解明していくことで、共通基盤構築のプロセスを実現することが可能となります。こうした共通基盤を、生成AI（人工知能）等を活用して解明し、コンピュータと人のコミュニケーション実現に挑む、NTT人間情報研究所 東中竜一郎客員上席特別研究員に、古くて新しい課題に取り組む姿勢と「失敗ではなく成功しなかっただけ」という考え方を伺いました。



「共通基盤」構築のメカニズムを解明し、コンピュータと人のコミュニケーションの実現をめざす

現在、手掛けていらっしゃる研究について教えていただけますでしょうか。

人と人のコミュニケーションにおいて、共通の話題・前提条件・認識等を理解し合う、「共通基盤」が一致しているときにお互いに理解し合えてコミュニケーションが成立します。私は、この共通基盤の研究をテーマとして、人の共通基盤構築のメカニズム解明、モデル化を行い、それをコンピュータと人の共通基盤に適用することで、コンピュータと人の円滑なコミュニケーションの実現をめざしています。

前回（2022年5月号）では、電気通信大学、静岡大学、慶應義塾大学との共同研究に関連し、①2名の作業者が、配置の異なる共通の図形集合を見ながら、テキストチャットにより図形配置を共同で決定していく「共同図形配置課題」を設定し（図1）、図形集合どうしの距離を測ることで、対話の各時点での共通基盤の構築度合いを可視化可能とした、②共同図形配置課題について、図形配置方法を一般的な物等に例えて説明する（例：自動車の車

輪のように配置する）「名付け」の有用性・重要性を確認した、③テキストチャットで行われた共同図形配置課題の研究を拡張し、音声や映像を用いたコーパス（自然言語等のデータを大規模に集積したもの）を構築・分析し、映像が存在するほうが共通基盤構築が行いやすいことを確認した、④初対面条件よりも知り合い条件のほうが共通基盤構築が行いやすいことが分かった、ことを紹介しました。

その後、研究を継続していく中で、以下の進展がありました。
(a) 2人で話し合いの目標を決める「目標名付け」が非常に重要であるということが分かった、
(b) タングラム（正方形を分割したピースを用いるパズル）で作成した図形について対話をしながら名前を付けていくタングラム命名課題（図2）において、形全体に名前を付けるような発話（Holisticな発話）と図形の詳細を述べる発話（Analyticな発話）を人は使い分けていることが分かり、それをどのように使い分けているのかについて実験的に明らかにした、
(c) 相手と同じタングラムを見ていることを伝える際の、うまくいく場合といかない場合の認知モデルに関する実験を行った。
(d) 共通基盤を測るための新しい課題を考案してデータを収集し、マルチモーダル情報がどのくらい共通基盤・相互理解に影響を与えるかの調査を行った。

「目標名付け」については、名付けの効果を検証する中で、ま

だ存在していないものに対して名付けする。例えばある図形集合に対して「運転席と荷台があるように見えるので、これをトラックにしましょう」といったように最初に名付けて（目標名付け）、それをベースとしてどの位置を示すのか（例：荷台の位置）、どのように配置するのか（例：荷台を上に向ける）等のやり取りをすると、2人が同じ図形集合を認識して同じ配置にする（成功する対話）目的に対して有効であることが分かってきました。そして、目標名付けが登場したタイミングを分析し、成功する対話は目標名付けが対話の流れの序盤に出現する、つまり2人の意識を最初に合わせることで成功対話が多くなることが分かりました（図3）。

Holisticな発話とAnalyticな発話については、実際人間の活動をみると、相対的にHolisticな発話の頻度が高い傾向にあり、最

初はHolisticな発話で、それで問題が生じたらAnalyticな発話になるということが分かりました。さらに、それを生成AI（人工知能）を用い、まずHolisticで話してみて反応に応じてAnalytic発話を行うようなプロンプトに基づくシステムを試作し、システムの精度（タスクの達成度合い）を分析しました（図4）。HolisticとAnalyticを意識して使い分けのP条件では安定した発話数でタスクが達成できていますが、Holisticのみを用いるH条件ではたまたに極端に少ない発話数でタスク達成できるものの分散が大きく、Analyticのみを用いるA条件ではタスクの達成に非常に多くの発話が必要になる（手間がかかる）ことが分かりました。これは、人の対話において、最初はHolisticな発話で、それが分からないとAnalyticな発話になるという順番が安定したやり取りを実現で

オブジェクトを共通の配置に移動するデザインタスク

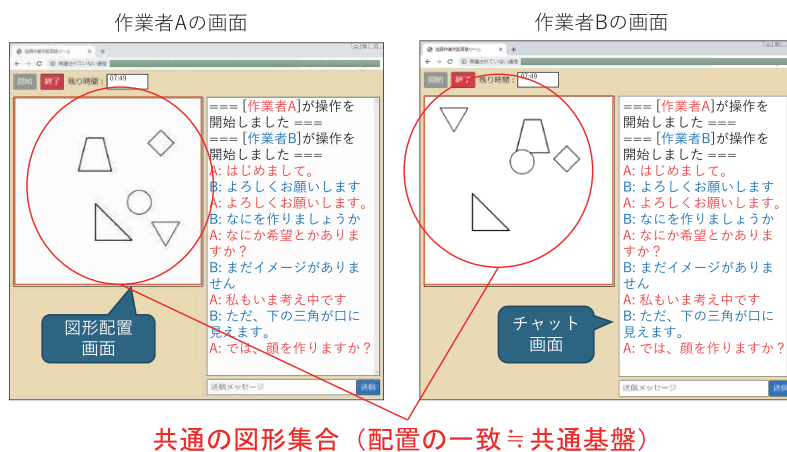


図1 共同図形配置課題

オブジェクトに共通の名前を付与する命名タスク

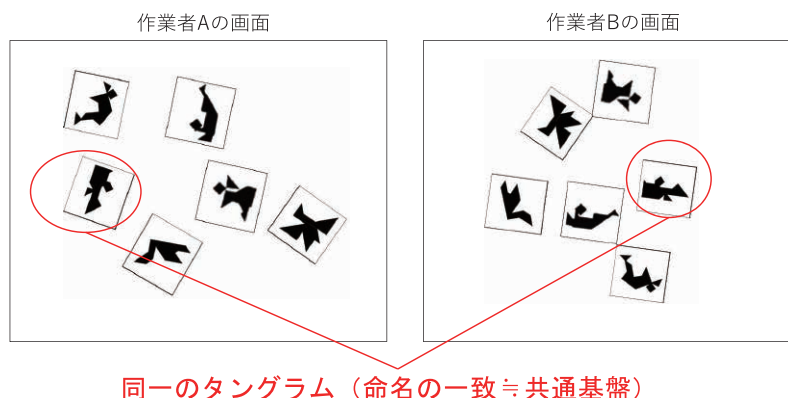


図2 タングラム命名課題

きることを示していると考えられます。

言葉以外ではどのようなことが分かってきたのでしょうか。

タングラム命名課題で人間が行っている処理を認知モデルの視点から説明すると、まず送り手側で伝えたい画像をイメージし、そのイメージを言語化して受け手に伝えます。そして、受け手側は、その言葉からイメージをつくり、そのイメージを手元の図形とマッチングをして選択するという過程を辿ります。そこで、認知モデルのプロセスをコンピュータ上でシミュレータとして実現し、相手がもし送り手側の図形を理解できなかったときに、どのような調整をすると理解できるようになるのかという点について実験を行いました。具体的には、イメージ化と言語化の両方を調整対象とし、それらのどこを調整すると精度が上がるのかを調査しまし

た（図5）。その結果、言語化の調整だけでも精度はよくなるのですが、イメージ化を含めて調整することで多様な表現を生み出し、人間らしいやり取りに近づいている、つまりイメージ化と言語化、両方の調整が重要であることが確認されました。この結果については、2025年のThe 47th Annual Meeting of the Cognitive Science Society (CogSci 2025)という国際会議で発表されます。

新しい課題とデータ収集については、「共同物語並びかえ課題」を新たに提案してデータ収集を行いました。この課題は、2人と同じ動画から切り出された画像集合（動画を10分割し、それぞれの分割から1枚ずつ切り出したもの。話者によって画像集合は異なり、ランダム順で提示される）が与えられ、身体を使いながら対話をし、それぞれの画像に対してどの順番が正しいかを予測して並び変える作業を行います。そして、画像の並びが一致してい

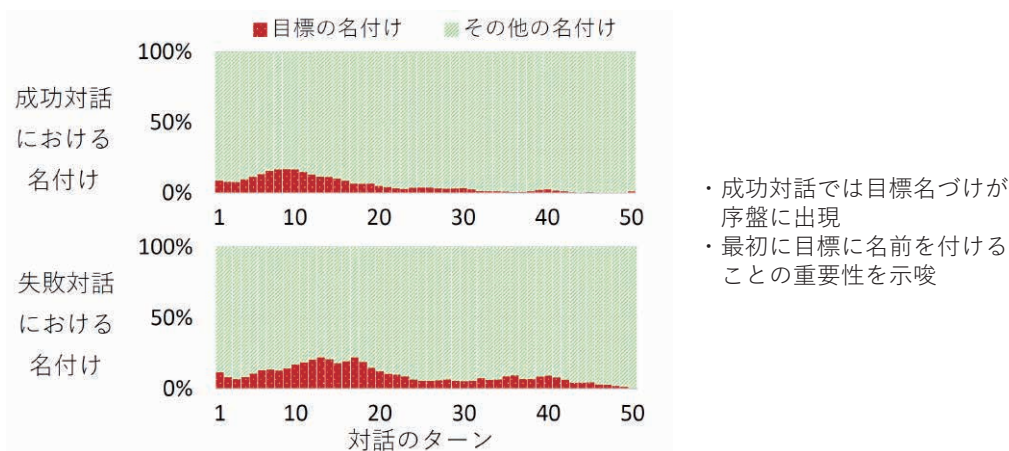


図3 目標名付け

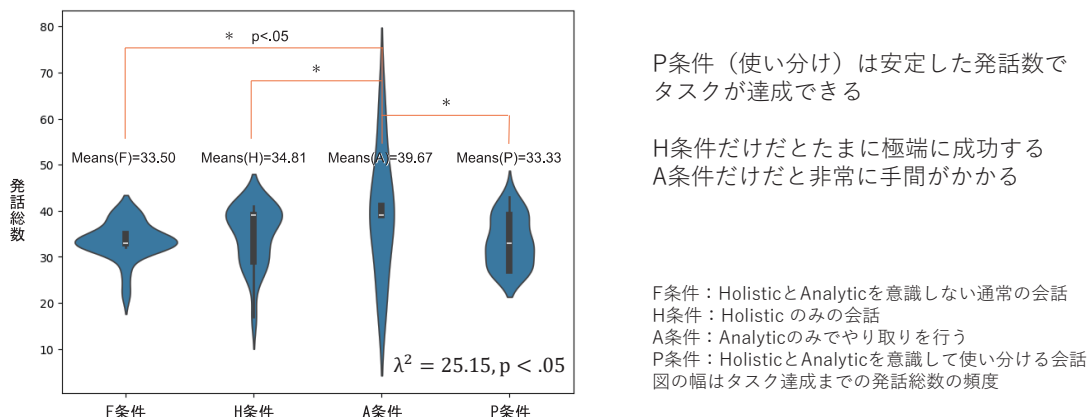


図4 実験条件ごとの発話総数とシステムの精度

る度合いを共通基盤の一致とみなすことで、表情やジェスチャと共通基盤の関係を分析します。これまでの共同図形配置課題に比べて、時間という軸が増えており、それを身体動作等により表現するので、共同図形配置課題のマルチモーダル版のような位置付けです。動画の並べ替えの時間と動画の説明の時間を分離することで、表情やジェスチャの共通基盤への影響のみを計測できるようにしました。この実験では、セッションが繰り返されるごとに2人の間の一致率が上がっており、セッションごとに共通基盤が構築されている様子が確認されました(図6)。この中でどのような表情やジェスチャが共通基盤構築と関連して出現したのかを調査することで、モダリティとの共通基盤の関係が分析できます。

ロボットどうしの対話、人の会話におけるターンテイキングの研究に取り組む

共通基盤以外ではどのような研究をしているのでしょうか。

共通基盤以外では、大阪大学、国際電気通信基礎技術研究所

(ATR) と共同で、対話ロボットの研究を行い、複数体ロボットとのインタラクションが話者の意見に与える影響を検証しました。2台の対話ロボットを配置して、それぞれが異なる意見を持って議論する場面で、2台とも同意してくれるパターン、片方が不同意を言い続けるパターン等、いくつかのパターンをつくり、議論している途中で対話ロボットが人に「あなたはどう思うの?」と聞き、その意見にロボットたちが同意するか同意しないかによって、人の意見がどのように変容するかを調べました。一番相手の行動が変わったパターンは、例えば旅行に行くなら北海道か沖縄かという議論をロボットがしているときに、「あなたはどう思うの?」と聞かれ「北海道がいい」と言ったときに、沖縄がいいと言っていたロボットも翻って「確かにそうだ(北海道がいい)」と収束する同意パターンにすると、納得感が増して意見の確信度が高まったと感じることを確認しました。これは、2022年3~8月に日本科学未来館で、「きみとロボット」展としてデモを展示しました。また、『International Journal of Social Robotics』にも採録されました。

このほか、私が教員をしている名古屋大学とNTTコミュニケーション科学基礎研究所の共同研究で、会話の流れによって主導権



図5 認知モデルをもとにしたモデル化

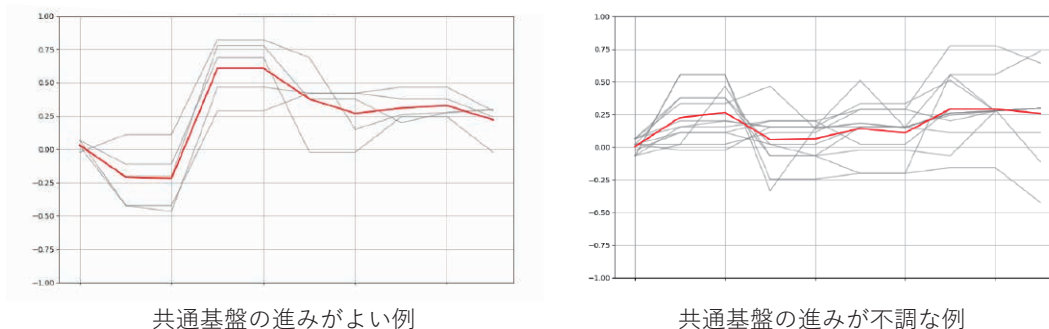


図6 共通基盤の構築の可視化

が入れ替わる、話者交替のメカニズムである「ターンテイキング」の研究をしています。ターンテイキングでは相手の発話終了を知ることが重要であるため、日本語で相手の発話がいつ終わるかを予測するモデル（Voice Activity Projectionモデル：VAPモデル）の構築に取り組んでいます。深層学習モデルのTransformerをベースとしたVAPモデルについて、日本語、英語、英語+日本語のデータで学習させ、それぞれのパターンでどうVAPモデルの性能が変わるかという基礎的な評価実験を行いました。また、そのVAPモデルを実際の対話システムに組み込んで性能が上がるのかどうかといった影響を検証しました。

そして、これまでの活動をベースとして、2023年2月に分野全体の教科書となる『対話システムの作り方』（近代科学社刊）、2024年6月に『Pythonと大規模言語モデルで作る リアルタイムマルチモーダル対話システム』（共著、科学情報出版株式会社刊）を上梓しました。

チャレンジングな古くて新しい課題に、最先端技術を活用して挑む

研究者として心掛けていることを教えてください。

どの分野の研究においてもそうだと思うのですが、古くて新しい課題というものが多いです。それにしっかりと取り組むことが重要ではないかと考えています。私のテーマである「共通基盤」は概念や古典的な理論そのものは以前からありましたが、分からないことがあまりに多い、あるいは研究が進むにつれて次々と新しい課題が出てきているため、まさに、古くて新しい課題なのです。

こうした課題をテーマとすると、データ収集やシミュレーションそしてそのための環境づくり等、地味で根気のいる部分もあり、さらにはそれが問題解決につながるかどうか分からない、やってみなければ分からない場合もあります。こういったことを繰り返しているうちに、何かを発見できてくるのですが、それはある現象のほんの一部でしかない、ということもよくあります。このため、古い課題ではありながら、長期にわたって手が付けられてこなかった課題ということもあり、チャレンジングな課題でもあるのです。

最近は大規模言語モデルのような新しい技術、最先端の技術も多々登場してきており、これを活用することでこれまでできなかったようなシミュレーション等を行うことも可能となりました。最先端の技術をウォッチ・キャッチアップし、それにより研究の新たなステップを踏み出すことができれば、それを実用レベルにす

ることができるのではないかと思います。古くから手が付けられずにいた課題であるからこそ、そのインパクトは大きなものでもあります。だからこそ、古くて新しい課題にしっかりと取り組むことが重要なのです。

 **結果が失敗でも何かを得ることができれば、それは成功しなかっただけ。そのために本質をとらえた研究が大切**

後進の研究者へのメッセージをお願いします。

課題・テーマの本質をとらえて研究を進めてもらえればいいと思います。そのために、課題の基礎的な部分をしっかりと押さえることが大切です。前述のように、古くて新しい課題に取り組むときに、最先端の技術をキャッチアップしてそれを活用して研究に取り組むと、新たなステップに踏み出すことができるのですが、最先端の技術を追いかけしているとそれが華やかに見えて、ついそちらに目がいくようになり本質から遠ざかる可能性があります。これは、古くて新しい課題に限った話ではなく、基礎的なことがしっかりと押さえることができていれば、最先端の技術との位置関係も理解できるので、逆に効果的な活用方法のアイデアも出てきて、研究を前進させることができます。

そして課題・テーマの本質をとらえたうえで、失敗をおそれずにチャレンジしてもらいたいと思います。本質をとらえることができていれば、たとえ結果が失敗だったと思えても、その状況、原因等を理解することで、必ず何かを得ることができます。つまり、それは失敗ではなく、単にその取り組みは成功しなかっただけであり、得られたものを糧にさらなる前進につながるのです。