

# 誰でも手が届く AI 対話システム構築とその意義 —AI との共生社会に向けて—

企画責任者・指定討論者・進行：宇佐美まゆみ(東京外国語大学)

話題提供者：西川寛之(明海大学) 高橋哲朗(富士通)

話題提供者：白井宏美(FCL/慶應義塾大学 SFC 研究所)

## 1. このワークショップまでの経緯・趣旨・目的

対話システムとは、「人と会話を行うコンピュータ」のことである。工学（対話知能学）においては AI に人の発話の理解や発話生成を行なわせる手法の開発が進められてきたが、今日では、家庭内の情報検索、ホテルや空港などでの通訳、案内、オンライン上での人からの質問に対話システムが回答するなど、すでに生活のあらゆる場面で活用されるようになってきている。これらの例のように、状況がある程度特化されていて目的が明確なものは、内容や「やりとり」の展開の予測がしやすく、必要な語彙は限定され、文型にも一定のパターンがあるため、技術開発も急速に進み、現在ではかなりレベルが向上したと言える。

一方、このワークショップで扱う「誘いや依頼を断る」とか「借りた本をなくしてしまったことを伝えて謝罪する」というような、円滑な人間関係を確立・維持するために、様々な形の「対人配慮」が必要になってくるようなやりとりは、未だ、対話システムが最も苦手とすることの一つである。しかし、技術が向上するに伴い、工学分野においても、ポライトネス（対人配慮行動）を踏まえた「人らしい」発話ややりとりができる対話システムの構築が目指されるようになった。ここに、対話システム構築と人文・社会科学系の研究との接点が明確になる。

折しも、2018 年に対話システム研究の分野で、聴衆の前でライブでいくつかの対話システムを動作させて参加者が評価し、問題点をコミュニティで共有するという「対話システムライブコンペティション（以降、ライブコンペ）」（話題提供 1（高橋）を参照）の第 1 回が開催された。これを観戦していた企画責任者の宇佐美は、「もっと自然で人らしい会話を対話システムに行わせるために、人文系・言語研究の知見が活かせるはずだ」と確信し、対話知能学の研究者と協働を始めることになった（ライブコンペの（人文系）委員となる）。第 1 回は、オープントラックという比較的自由に対話させる形であったが、コンペという短い時間の中で話題が多岐にわたる複数の対話を比較して評価するのは難しいことや、人文系からのライブコンペ参加者を増やすには、ある程度、場面や状況を限定したほうがやりやすいと考え、委員会で検討した結果、第 2 回からは、「シチュエーショントラック」を導入した。これは、ACTFL の OPI インタビューのロールプレイを参考にしたもので、話題提供 2（白井）、3（西川）でも例にあげられるように、なるべく「シチュエーション」という文脈を与えた上で、対話システムがその文脈に応じたやりとりができるかどうかを評価するものにしたのである。ここで、まず、言語教育・研究の知見が応用されることになった。外国語学習者の口頭能力のレベルを判定する方法は、対話システムの評価にも応用できるのだ。そのため、ライブコンペの「課題」となるシチュエーションは、基本的にライブコンペの「人文系委員」が原案を作ることになった。AI は、外国語学習者に比して、共有している背景知識が少ないため、言語教育でよく用いられるロールプレイよりも、遂行させたい課題に応じたより詳しい文脈をシチュエーションに盛り込む必要がある。その文脈提示の度合いを最適にするためにも、言語・言語教育研究の知識・経験が必要だ。

工学の側からは、ライブコンペへの人文系からの参加を促進することを主たる目的に、「誰でも手が届く対話システム構築」に向けて工夫もなされた。それが、2018 年に公開された xAIML SUNABA（以降 SUNABA）<sup>1</sup>の活用である。これは自然対話作成用プラットフォームで、視覚的かつ直感的に対話のシナリオを作成できるものである。このプラットフォームを用いることで、プログラミング言語を用いることなく、対話システムを構築することができるようになった。これは、人文系の研究者にとっては、画期的なことである。

このワークショップでは、「対話システム構築」というのは、工学的な手法を駆使して構築するものだけではなく、話題提供 2（白井）で紹介するように、「システム構築者が、対話相手であるユーザ（人）の反応を予測・先読みして、それに対する応答を事前に何通りか用意しておいて、その中から、ユーザの反応に応じた発話をシステムにさせるようなやり方」（ルールベースと呼ばれる方法）で構築するものも含めている。談話研究や心理学の知見を活用すれば、ユーザの反応を予測して誘導することによって、用意しておいたフレーズ

<sup>1</sup> <https://docs.sunaba.docomo.ne.jp/>

が使えるように仕向けることも可能だ。このように、SUNABA を使えば、プログラミングをしなくても、対話のシナリオを書くことによって、人の設計通りに動く対話システムを構築することができる。本ワークショップでは、そのことを「誰でも手が届く」と表現した。

このように、SUNABA 等を用いて「誰でも手が届く対話システム」が構築できるようになったことで、人文・社会学系の研究者も、その専門知識を活かす形でコンペティションに参加しやすくなっている。対話システムが様々な場面で人との接点を増やしていく中で、人と対話システムとの共生社会を実現するためには、工学的な基準だけではなく、対話の背景、話者間の関係、ポライトネスを考慮した発話や、対話が引き起こす感情の変化などの社会・心理的観点も取り入れた対話システムの設計が益々求められるようになっていく。そもそも人と AI との共生社会の「望ましいあり方」とはどのようなものなのか。人と AI との共生社会の中で対話システムには何が期待されているのか。

この問いを考えるために、3 名が以下のような話題提供を行う。最初の話題提供者の高橋は、ライブコンペティションを簡単に紹介するとともに、工学者の観点から、対話システム開発に必要な社会言語科学の視点についてまとめる。次の白井は、ここ数年ライブコンペティションに参加し、人文系として初めて最優秀賞も受賞している経験を活かし、対話システム構築に談話研究や社会言語科学的視点や知見がいかんにか活用できるかを論じる。3 番目の西川は、言語系研究者のライブコンペへの参加を促進するために行っている講習会について紹介する。最後に、指定討論者・討論進行の宇佐美が、3 つの話題提供に対して学際的視点を踏まえながら指定討論を行う。その後は、フロアも交えて、「誰でも手が届く対話システム構築の意義」、AI との共生社会に向けて社会言語科学がなすべきこと、できることについて、対話システム構築やウェルフェア・リングイステイクスの観点から議論したい。

## 2. 話題提供

以下に 3 名の話題提供者の論考を示す。

### 話題提供 1. 工学側が期待する社会言語科学の視点

#### 対話システム開発に求められる社会言語科学の視点

高橋哲朗, 富士通

人工知能学会 言語・音声理解と対話処理研究会 (SLUD) では 2018 年から対話システムライブコンペティション(ライブコンペ)というイベントを毎年開催している(東中ら, 2023)。ライブコンペは、より高度な対話システムの実現を目的としたコンペティションであり、ユーザとシステムの対話を大勢のオーディエンスが一斉にライブで鑑賞・評価することにより、対話システムの現状の問題点が共有され、対話システム研究を加速することを目的としている。これまでの 6 回のライブコンペではそれぞれ 10~26 チームが参加し、独自の対話システムを構築してライブコンペに臨んできた。それぞれのシステムは異なる特徴を持っているため、これらの複数の対話システムを比較しつつ評価することにより対話システムの技術開発に関して幅広い議論を行ない上述した目的への貢献を目指してきた。

第 1 回から第 4 回までは、図 1 に示すインスタントメッセージシステム Telegram<sup>2</sup>上でテキスト対話を行なう対話システムを参加者が開発し、そのシステムと人間の評価者の間のテキスト対話により評価を行なった。

第 5 回および第 6 回では、テキスト対話ではなく図 2 に示すバーチャルエージェントを用いたマルチモーダル対話を行なう対話システムを用いた。このバーチャルエージェントは、アンドロイドロボット ERICA(Glasら, 2016) の 3 次元 CG モデルである。ここでは音声での入出力に加えて、バーチャルエージェントの表情やジェスチャーもコントロールできるようにした。また第 6 回からは評価者である人間の映像も入力し、評価者の表情から推定される感情や、顔の向き、および年齢と性別の推定値も含めて入力として用いることができるようにした。

対話システムを開発するためには、形態素解析や構文解析などの基礎的な解析に加えて、話者の意図推定や照応解析などの高次の解析を含む多様な自然言語処理の技術が必要となる。さらにマルチモーダル対話を実現するためには、自然言語処理に加えて音声認識や表情認識、音声合成、表情合成などの技術も必要となるが、これらすべての機能を参加者全員が独自に開発することになればそのハードルは非常に高く、ライブコンペの参加の障壁となってしまふ。そこでライブコンペでは、Telegram との通信のためのサンプルコードやマルチモ

<sup>2</sup> <https://telegram.org/>.

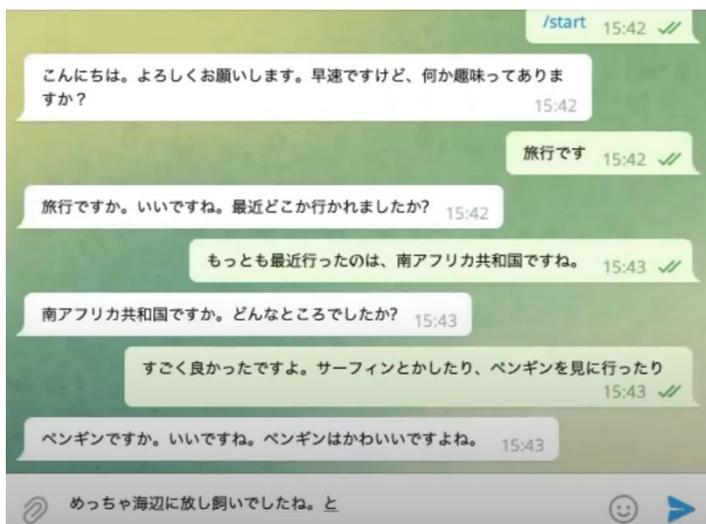


図 1. Telegram を用いた対話例

一ダルの制御を行なう共通モジュールを参加者へ提供することにより、ライブコンペの参加者は対話システムの開発のみに集中できるようにした(図 3)。

対話システムの開発のみに集中できるとは言え、その対話システムの開発には一般的に情報システムの構築スキルや経験が求められる。そうなる参加者が工学系に偏ってしまい、その結果としてライブコンペを通して得られる知見も偏ったものになってしまう懸念がある。そこでライブコンペでは、プログラミングをすることなく対話システムを構築することができる SUNABA というサービスと連携し、SUNABA 上で対話システムを作ることによって上述した Telegram 上での対話やバーチャルエージェントを介した対話を実現できる開発ツールを構築し配布した。SUNABA にはボットエディタと呼ばれる GUI のエディタが用意されており、図 4 に示すようにシステムの発話内容と想定されるユーザ発話とをブロックでつないでいくことにより対話のルールを記述できる。つまり図 3 中の「対話制御」の部分を図 4 に示す SUNABA で作成することができ、それによりバーチャルエージェントとマルチモーダル対話を行なえるシステムを構築できる。

SUNABA 上では、想定されるユーザ発話は正規表現などを用いて柔軟に記述することができ、様々な可能性に基づいて対話のシナリオを分岐させることができる。またマルチモーダルの入出力も SUNABA で扱えるようにしており、たとえば「おはよう:fp(fullsmile)」という発話を SUNABA で設定することにより「おはよう」という発話をさせつつ表情を笑顔(fullsmile)にしたり、「p(80)本当に w(0.3)ごめんなさい」という発話を設定することにより、発話のピッチを 80%にしつつ(p(80))、「本当に」の後に 0.3 秒のポーズ(w(0.3))を入れるといった発話のコントロールができる。また、ユーザ発話には推定された感情が「元気です em:happy」のような制御文字付きで入力されるため、SUNABA 上ではこの情報を用いることで、感情が happy の場合と angry の場合で条件分けすることも可能である。

ライブコンペで提供した開発環境により、マルチモーダルまで含めた複雑な対話システムを図 4 に示したユーザインターフェースを用いた簡単な操作で構築できるようになった。対話システムの構築が誰でも手が届くようになったことにより SUNABA を導入した第 2 回以降(第 2 回時点では SUNABA の前身 Rep1-AI を用いた)は、工学系だけではなく人文系からもライブコンペに参加してもらえるようになり、これにより、ライブコンペは人文系の対話研究者と工学系の対話システム研究者が共に参加し、対話システムで用いる理論や技術について議論できる場となった。

対話システムの開発において従来は、質問に回答できるか、特定のタスクを実行できるか、破綻なく対話を継続できるか、といったプリミティブな評価基準が定められ、それらの評価基準に基づき開発が行われてきた。ここには主に工学的な観点から「使えるシステム」を目指してきた背景がある。また、これらの評価基準を満たすことは容易ではなかったため、まずこれらの評価基準を満たすことが優先されて長い期間研究開発が続けられてきた。

システムと対話をする人間の評価においても、システムがこれらの評価基準を満たすことができないうちは、発話における命題の正誤が主に注目されてきた。しかし、昨今のディープラーニングや言語モデルに関する技術の進歩により、プリミティブな評価基準を高いレベルで満たす対話を実現できるようになってきており、そうなる人間側の期待も上がってくることは必至である。対話システムを次のレベルに推し進めるためには、発話の命題の正誤だけではなく、



図 2. CGErica のジェスチャと表情の例。上段左から、ノーマル、おじぎ、相手への促しを表現したジェスチャ。下段左から、ノーマル、怒り、困惑、喜びの表情

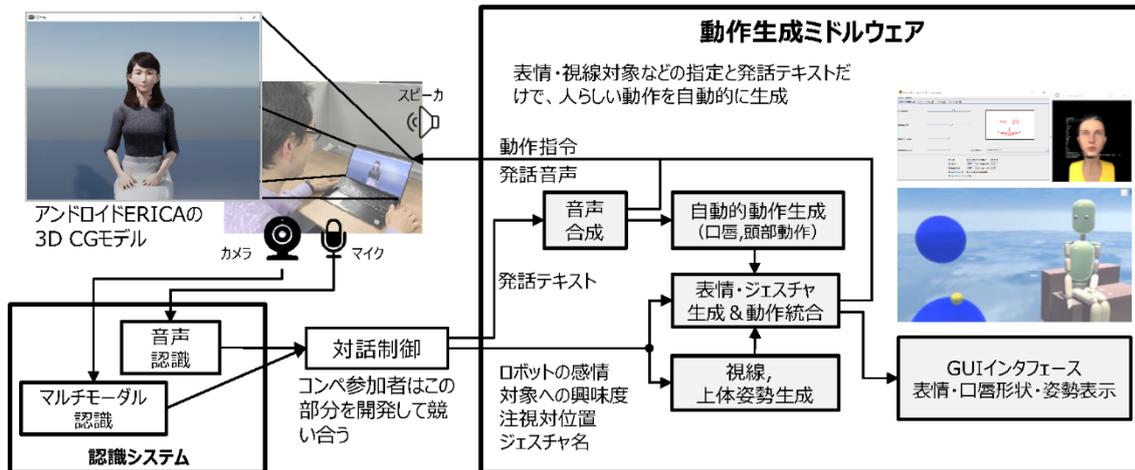


図 3. マルチモーダル対話システムの共通モジュール

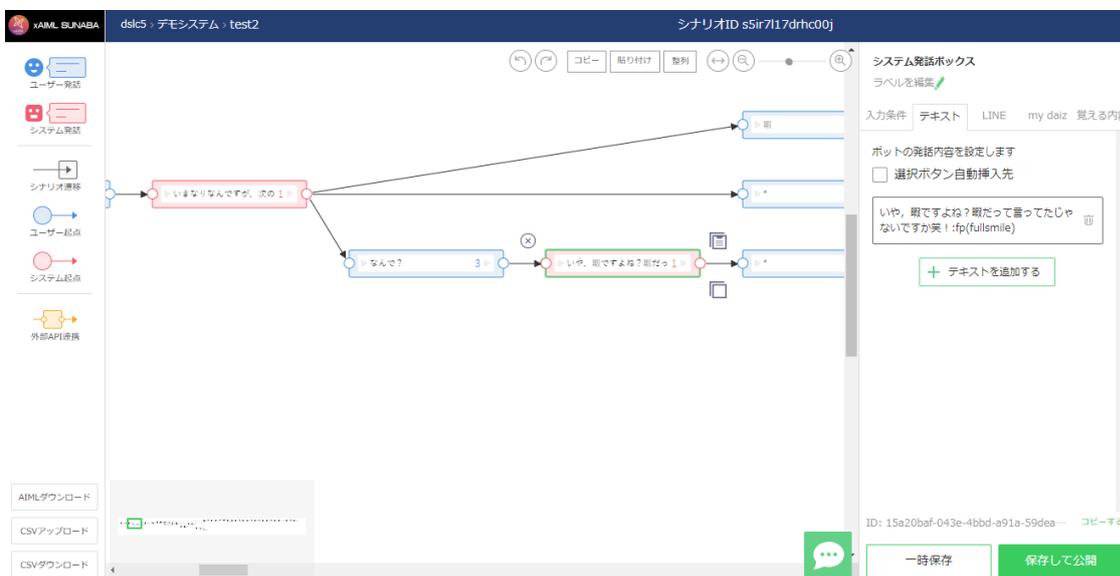


図 4. SUNABA を用いた対話システムの開発

- ・ 談話構造の解釈
- ・ 対話の背景や話者間の関係の考慮
- ・ 対話を引き起こす感情の変化
- ・ 相槌や間などのパラ言語の利用

などが求められるが、工学における対話システムの開発では、これらの実装はまだまできていないのが現状である。一方で社会言語科学の分野においてはこれまでに研究され蓄積されてきた知見があるため、その知見を用いていくことが求められる。これを実現するための、工学系と社会言語科学との共同の場として、上述したライブコンペの場や SUNABA のような誰にでも対話システムの開発を可能とするツールが貢献できると考えられる。これまでのライブコンペを通して、Telegram との接続や SUNABA によるマルチモーダル対話システムの基本的な制御は実現できているが、即応性の高いバーチャルエージェントの制御や ChatGPT を始めとした言語モデルの活用ができるツールはまだ用意できていないため、それらの開発が工学系の責務の一つと言える。

Siri や Amazon Echo などの対話システムの登場から約 10 年が経ち、さらには ChatGPT の登場により対話システムが今後さらに身近になることが予想される。対話システムが様々な場面で人との接点を増やしていく中で、対話システムと人との共生社会を実現するためには、上述した工学的なアプローチに基づく対話システムの開発だけでは不十分であり、社会言語科学からの視点や知見を取り入れた対話システムの設計や評価が求められる。本ワークショップではこのような対話システム開発にとっての新たな視点について議論したい。

## 参考文献

- 東中竜一郎・高橋哲朗・稲葉通将・斉志揚・佐々木裕多・船越孝太郎・守屋彰二・佐藤志貴・港隆史・境くりま・船山智・小室允人・西川寛之・牧野遼作・菊池浩史・宇佐美まゆみ(2023). 対話システムライブコンペティション 6, 第 99 回人工知能学会 言語・音声理解と対話処理研究会(第 14 回対話システムシンポジウム).
- Dylan F. Glas・Takashi Minato・Carlos Toshinori Ishi・Tatsuya Kawahara・Hiroshi Ishiguro (2016). ERICA: The ERATO Intelligent Conversational Android, In Proc. ROMAN, pp. 2229.

## 話題提供 2. 談話研究の観点から見た対話システム構築

### 人文科学と対話システムの接点

白井宏美 (FCL/慶應義塾大学 SFC 研究所)

チャットボットが普及するなか、その応答に違和感を覚えたり、やりとりに不満を募らせたりしている人文・社会科学系研究者は多いのではないだろうか。筆者もその一人であったが、プログラミングができるわけでもなく手をこまねいていた。しかし、「対話システムのコミュニティは主に工学系の研究者から構成されている。そのため、対話システムをひいき目に見ていたり注目する点が固定化されていたりして重要な問題点を見逃しているかもしれない。」(宇佐美・東中・杉山・角森, 2019) という問題意識から人文科学系研究に向けて「対話システム講習会」が開催されることを知り、2020 年に参加することで一歩を踏み出した。この講習会では、プログラミングができなくても対話システムを構築することができる SUNABA の使い方の教示や演習などが行われる。SUNABA を使えるようになった筆者は、2020 年に「人工知能学会 言語・音声理解と対話処理研究会の対話システムシンポジウム内」で開催された第 3 回「対話システムライブコンペティション (以降、ライブコンペと呼ぶ)」に初参戦した。ライブコンペは 2018 年度開始以来、主に工学系の研究者や企業の研究チームが参戦し、好成績を収めてきた。そのなかで筆者は、談話研究の知見を活かし、場面に適した人らしい会話のチャットボットを作成することを目指した。

2020 年の第 3 回ライブコンペにおいて与えられた設定は、「システムがユーザからの依頼を断る」というものである。ユーザ側である大学の先輩(同性)から同窓会の幹事をしてほしいと「依頼」され、それを断るシステムを作り、15 発話ずつ展開したところで対話を終了しなければならない。筆者は、対話の序盤で「談話の展開メカニズム」の知見を取り入れ、中盤では「テーマのある順番交替」を活用することにより、対話の主導権を握ることに成功した。終盤では「ポライトネス・ストラテジー」(ポライトネス理論)や、「談話標識」、「会話の終結」を活用し、対話をうまく終わらせる工夫をした。その結果、初挑戦にも関わらず予選を 1 位で通過することができたのである。さらにライブコンペ本選でも最優秀賞(1 位)を獲得した。人文系研究者が最優秀賞(1 位)を獲得したのは初めてのことである。

2021 年の第 4 回ライブコンペにおいて与えられた設定は、「ユーザがシステムからの誘いを断る」というものである。図 1・2 のとおり、非常に具体的なシチュエーションが用意されていた。そこで、語用論の知見を活かし、相手を気遣ったり、相手の気持ちを察したりできる対話戦略を考え、それを駆使したシナリオ構成を試みた。対話戦略には、主に「発語内行為」「含意」「結束性」を取り入れた。その結果、2 年連続で予選を 1 位で通過することができ、本選では優秀賞(2 位)を獲得した。

清水 (システム) は、次の週末に、会社の仲の良い同期とオンライン飲み会を企画している。今のところ、声がけしたメンバーの佐藤、鈴木、高橋、渡辺、小林は、全員参加の予定だ。コロナ禍の昨今、居酒屋で酒を飲むことはできないため、以前に上司の湯川 (ユーザ) がオンライン飲み会を企画したことがあり、これが結構盛り上がった。前回とは違って今回は、オンライン飲み会のために酒やおつまみを、当日に各自の自宅に届けてくれるサービスも使うので、更に楽しくなりそうだ。ところで、同期のメンバーらが、上司である湯川も誘ってはどうかと言い出した。確かに湯川は、年も近いし、話が合うこともあって、私たちとプライベートでも仲が良い。会社では上司であるが、同時に公私にわたって良き先輩であるともいえる。週末に特に予定が無いことを「暇だ」とぼやいていたし、もともと飲み会が好きで人なので、誘えばきっと参加してくれるだろう。オンラインサービスの予約期限が迫っており、メンバーを確定しなくてはならないので、早速連絡しよう。

図 1. システム側の背景

部下の清水（システム）から、オンライン飲み会の誘いがきた。私（湯川＝ユーザ）は飲み会は好きだし、以前に自分が企画したオンライン飲み会は結構楽しかったので、参加したい気持ちは正直ある。しかし、清水の同期ということは、小林も参加するというのだ。ここだけの話だが、実は、小林のことがあまり好きではない。というより、はっきり言って嫌い。前は部署全体の飲み会だったので仕方がなかったが、今回は内輪での集まりのようだし、無理してまで参加する必要もない。小林がいないのであれば参加したいのだけれども、それが我儘なのもよくわかっている。我慢して参加しても楽しくないので、今回は断ろう。

図 2. ユーザ側の背景

2022 年の第 5 回ライブコンペにおいて与えられた設定は、「システムがユーザに謝罪する」というものである。システムとユーザは、同じ大学のゼミの友人同士で、システムがユーザから借りていた本を紛失してしまったことを謝罪する場面である。対話戦略として、今回は特に友人間での謝罪時に用いられる語用論的方策(ボイクマン・宇佐美, 2005)を参考にした。表 1 に示した対話例のとおり、言い淀み、話題の前触れ、責任の承認、事情説明、謝罪、申し訳なさの表明、人間関係への言及などの語用論的方策を用いた。

表 1. 対話例(原文ママ)と語用論的方策

| 話者   | 発話   | 語用論的方策   |
|------|--|----------|
| 01 U | 話って何   |          |
| 01 S | いやー、あの一、えーっとねー   | 言い淀み     |
| 02 U | しずかどうしたの   |          |
| 02 S | えーっと、ほんとに、あの、何から話したらいいか  | 言い淀み     |
| 03 U | えなにになにどうしたの  |          |
| 03 S | ユウキがめちゃくちゃ頑張っ、て、お金ためて買った専門書、それ、私も興味があるって言って、ほんとにお願いして、貸してもらったじゃない                  | 話題の前触れ   |
| 04 U | あーそうだねあの本大切な本なんだけどどう読んでみて役に立った   |          |
| 04 S | それ、きょう返そうと思ってたのに、私、バカなこと、しちゃって   | 責任の承認    |
| 05 U | 何したのしずか  |          |
| 05 S | 昨日、本、カバンに入れて電車に乗ってたんだけど、すごく混んで、手に持ってるのと周りの人に迷惑になりそうだったから、網棚に乗せてしまったの               | 事情説明     |
| 06 U | それでどうしたの   |          |
| 06 S | それで普段、網棚にのせることないのと、ほんと混んで降りられるかなあ、早く降りなきゃってあせっちゃって、私、ほんとバカなこと、網棚の荷物、持っておりの、忘れちゃったの | 事情説明     |
| 07 U | ていうことは本をおきっぱなしにしちゃったってこと   |          |
| 07 S | 降りたあと気づいて、あっしまったって思って、電車へ戻ろうとしたんだけど、目の前でドアしまっちゃって、間に合わなかったの                        | 事情説明     |
| 08 U | じゃあもう戻ってきてないってこと気づか探した   |          |
| 08 S | 本当に、ごめんなさい   | 謝罪       |
| 09 U | それで駅とか駅には聞いたの  |          |
| 09 S | もちろん鉄道会社と警察に、連絡して探してもらってるんだけど、まだ連絡きてなくて  | 対処       |
| 10 U | そうなんだあの本すぐバイトして予約買った本なんだないと困るなあ  |          |
| 10 S | 見つかりますように一って、ずうっと祈ってるんだけど  | 願い       |
| 11 U | そうだよ鉄道会社から連絡はまだ来ないんだ   |          |
| 11 S | いや、ユウキが、めちゃくちゃ頑張っ、てバイトしてさ、すごいお金いっぱいためて、あの本買ったこと、知ってるから                             | 申し訳なさの表明 |
| 12 U | そうなんだあの本買うために毎日のようにバイトしてすごい大変だったんだよね   |          |
| 12 S | 高いだけじゃなくて貴重な本なんだよね、ユウキ、やっと見つけたって言って、手に入れた時、すごく喜んでたから                               | 申し訳なさの表明 |
| 13 U | そうなのあの本一般的には売ってないからかなり手に入れるのも大変だったんだよねなくなっちゃったらどうしよう                               |          |
| 13 S | それに私も、無理言って、貸してもらったじゃない  | 人間関係への言及 |
| 14 U | しずかがどうしても見たいって言うから先に貸したんだよね私まだ読んでないし困るなあ   |          |
| 14 S | あー私、なんてひどいことしちゃったんだろう、ゆうきからの信頼を失ってしまっ  | 人間関係への言及 |

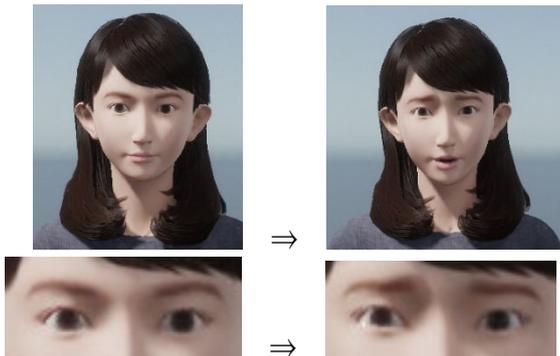
U:ユーザ(ゆうき) S:システム(しずか)

これまでと異なる点は、テキストチャットシステムから CG キャラクターが話すというマルチモーダル対話システムを構築することになったことである。そこで SUNABA 上で記述する発話文の中に、音声、表情、姿勢を制御する命令を埋め込んだ。音声の制御については、音声合成のパラメータとして、下記のとおり 3 種類の値を与えることが可能であった。

p: pitch (50~200)  
v: volume (50~200)  
s: speed (50~400)  
規定値: 100

例えば、冒頭の言い淀む箇所で「えーっと」のあと 0.3 秒、「ほんとに」のあと 0.2 秒、「あの」のあと 0.2 秒の間を与え、眉をひそめて辛そうな表情に見える fp (eye-down) を与えた(例 1)。

例 1



p(90)s(90)えーっと w(0.3)ほんとに w(0.2)あの w(0.2)何から話したらいいか:fp(eye-down):bm(0.0,1.2,1.5)

例 2 は人間関係への言及の箇所で、少し頭を下げて軽く上体を前に倒す姿勢を bp(greeting) で与えている。

例 2



p(90)s(90)あー私、なんてひどいことしちゃったんだろう w(1.0)ゆうきからの信頼を失ってしまって:bp(greeting)

例 3 は、改めてしっかり謝罪する箇所で、さらに頭を下げて深く上体を前に倒す姿勢を bp(greeting\_deep) で与えている。

例 3



p(80)s(80)ほんとにごめんなさい:bp(greeting\_deep)

その結果、3 年連続で予選を 1 位で通過することができ、本選では 2 年連続で優秀賞 (2 位) を獲得した。

このように順調に好成績を収めてきたが、2023 年の第 6 回ライブコンペでは、予選を通過することすらできなかった。ChatGPT と SUNABA を使ったのであるが、SUNABA が ChatGPT に対応しきれず、応答反応が著しく遅くなったのが敗因であった。

以上の結果から、「技術的サポートがあれば、談話や会話などを研究している人文系研究者がその知見や経験を活かし、対話戦略の面で力を発揮できる」(白井, 2022)が、工学技術の進展が不均衡になると難しくなると考えられる。今、その時期にさしかかっている。しかし、技術は進歩の波を繰り返して高くなっていくものであるため、次に来る技術進歩の波に向けて人文・社会科学系研究者が準備しておく必要がある。工学技術が発展していく過程において釣り合いがとれた時、人文・社会科学の知見を活かせることができれば、対話システムを進化させ、AI との共生社会をよりよい方向へ導ける可能性があるのではないか。

## 参考文献

- ボイクマン総子・宇佐美 洋 (2005) . 友人間での謝罪時に用いられる語用論的方策—日本語母語話者と中国語母語話者の比較—語用論研究, 7, 31-44.
- 白井宏美 (2022). 談話研究の知見に基づく対話システムの現状と課題 小特集—音声対話システムにおける“不気味の谷”を超えるには—日本音響学会誌, 78(5), 277-282.
- 宇佐美まゆみ・東中竜一郎・杉山弘晃・角森唯子(2019). 対話システム構築と談話研究・日本語教育の接点 日本語教育学会秋季大会予稿集, 26-35.

## 話題提供 3. 応用言語学・日本語教育学の観点から見た対話システム構築

### 誰でも手が届く対話システム構築のために

西川寛之 (明海大学)

インターネットインフラの充実と AI 技術の発展によって、ヒトがコンピュータやロボットを操作する時代から自律的な AI システムとの共生社会へと変化してきた。はじめに AI 対話システムについて述べた上で、言語系研究者のライブコンペへの参加を促進するために行っている講習会について紹介する。

AI 対話システムについて、AI という語と対話システムに分けてここでの定義について述べる。AI とは人工知能とも呼ばれ、その定義は明確な定義があるわけでない。「知能」とは何かという観点からの定義も考えられるが、ここでは自然知能と対になる語として扱うこととする。

- ・人工知能 (AI : Artificial Intelligence)
- ・自然知能 (NI : Natural Intelligence)

NI があらず具体的なものは、人間の知能である。ヒトやイヌなどの生物が持つ知能を NI とし、これに対して、ヒトが作った知能を AI とする。つまり、AI はヒトが作り出した知能であるというのが AI の大前提として考えられる。この AI が作った知能の存在も考えられる。これに別の名称を用いるならば、AII (AI-generated Intelligence), RI (Regenerated Intelligence), GI (Generated Intelligence) 等と新たなカテゴリーとして名称を作ることも可能だが、ここでは AI に含め、AI とはヒトが作り出したもので、それが自律的に何らかの選択を行うことができるものとする。

次に対話システムについて述べる。今回紹介する講習会で扱った対話とは、ヒト (以下、ユーザ) がコンピュータやロボットに言葉を投げかけ、それにコンピュータやロボットが応答する一連のやり取りを指す。そして、その対話を実現させるもの全体を対話システムとする。具体的にはユーザは含まないが、文字や映像を表示するモニタ、キーボードなどのハードから対話のやり取りを創り出すプログラムまでを指すものとする。

\*提喻で以下に示すものを対話システムと称することもある。

この対話システムを模式図にしたものが図 1 である。ユーザのメッセージを受け取るマイクやカメラ、その他キーボードやモニタが「インターフェイス」に該当する、このメッセージを受け取ってから応答のフレーズを創り出す作業をするハードウェアがコンピュータで、その中で働くのが対話プログラム、対話プログラムの中で 1 つ 1 つの応答に対してどう応答するかを選択・決定する条件が記述されたものが応答選択コードである。

実際にチャットボットなどをはじめ、実用化されている対話システムには、あたかもロボットが自発的に発話しているように見える Woz 法 (Wizard of Oz 法) もある。実際にヒトが対話の応答をしているのにロボットやキャラクタ (エージェント) を映したモニタ画面を介す理由の一つは、将来的にはヒトが介在せずに応答することを目指すというものが考えられる。近年注目されている生成 AI を仕事に用いる例は増えているようである。

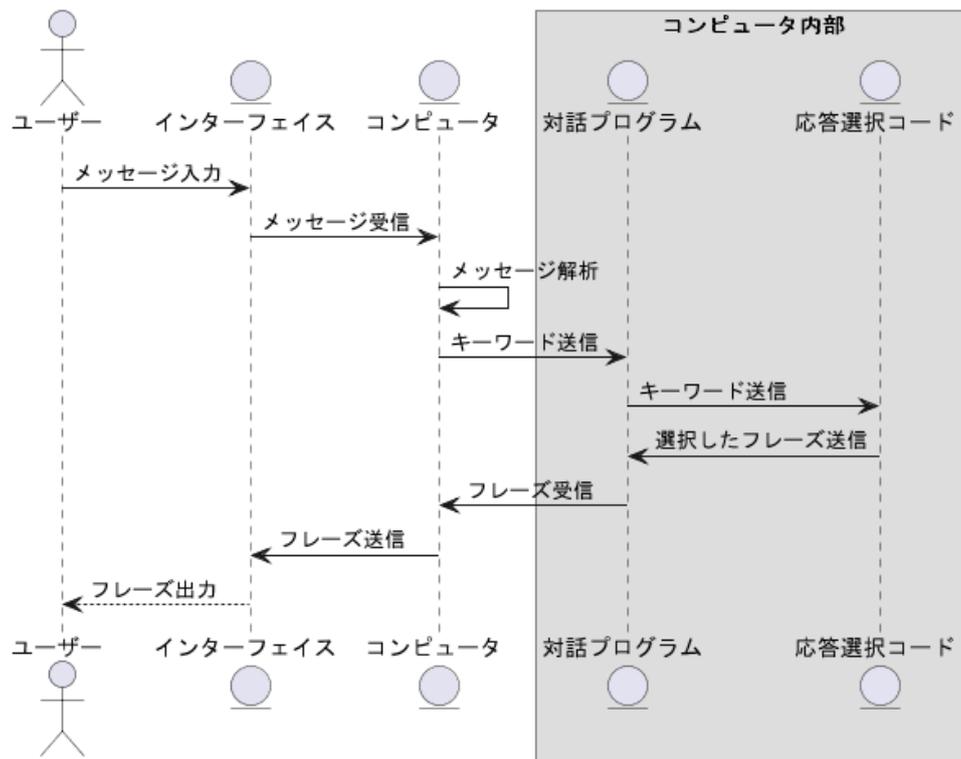


図 1. 対話システム内の処理イメージ

るが、傾聴などを除きチャットボットなどで責任ある発言をさせる例が一般化している例は管見の限りない。

羽田空港の案内を行なうロボットには Pepper (ペッパー) をはじめ、さまざまなロボットが使われているが、その中の一つのロボット「mini MORK」(ミニモーク)は人間が遠隔操作し、情報の伝達の中にロボットを介しているが応答はヒトが行う形、いわゆる Woz 法の形で運用されている。ユーザとシステムの応答にヒトが介入していないものとしては Apple 社のタブレット端末等で使うことができる Siri などがある。このほか、銀行の窓口の案内やインターネットを介した問い合わせなどにすでに広く使われはじめている。Siri の例のようにユーザが文字や音声で問いかけることから対話を開始するシステムもあれば、ユーザがシステムからの質問で提示される選択肢をクリックするなどして選ぶことで対話を進めるものもある。

いずれも対話システムと呼ばれるものであるが、ライブコンペではヒトが介在して話すことなく対話が進められる対話プログラムを対象としている。

#### 講習会について

対話システムライブコンペティション (以下、ライブコンペ) の講習会は、ライブコンペの中でもシチュエーショントラックへの出場者をサポートすることを第一の目的としている。対象者としては特に文系の研究者を取り込むことを意識した対応を行っている。講習会の内容はライブコンペ紹介と対話システム作成に分けられる。ライブコンペの紹介では、対話システムの概要と各回シチュエーショントラックで課される課題の説明が行われる。第 4 回の課題は図 2 のとおりである。

対話システム作成では、システム全体の構築を行うのではなく、図 1 にあるコンピュータ内部の処理である対話プログラム作りを行う。プログラムを作成するが、講習会ではいわゆるコンピュータ言語を用いたプログラミングではない。視覚的に対話システムを構築することができる SUNABA<sup>3</sup>というサービスの操作を学ぶ。このサービスは Web ブラウザ上で発話やその条件を設定したブロックを並べることによって図 1 の対話プログラムと応答選択コードが自動的に生成されるものである。特別なプログラミングの知識を用いることなく対話システムを完成させることができる。

講習会ではこの SUNABA を紹介したうえで、「シナリオの作成」という表現でタスクにあった会話の流れを考

<sup>3</sup> <https://docs.sunaba.docomo.ne.jp/>

える作業を行う。ユーザの発話と応答という組み合わせを規定するのが図1の応答選択コードに当たる作業で、正規表現を用いて記述した条件と直前のユーザの発話を照合し、次の応答フレーズの選択（もしくは生成）を行う作業の繰り返しとなる。ユーザが「今何時ですか」と尋ねてくるのであれば時間を答える応答フレーズを用意することが必要となる。「日本の首都はどこですか」という問いがあると考えた場合にも同様にその答えを準備することとなる。しかし、ユーザの発話に合せた応答を繰り返すだけではタスクを達成することはできない。タスクを達成するためにシステムの応答フレーズを使ってユーザの発話をコントロールすることも必要となる。さらに自然な流れを作ることも重要である。ライブコン

清水（システム）と湯川（ユーザ）は、学生時代、同じサークルの先輩・後輩の関係であった。清水にとって湯川は、いろんなことについて相談しやすい先輩であり、湯川にとって清水は自分を頼ってくるかわいい後輩であった。ただ、湯川が大学を卒業して会社員になってからは、会う機会も連絡をとることもなくなっていた。清水も就職が決まり、自宅できつろいでいた春休みのある日の夜、久しぶりに湯川からテキストメッセージが送られて来た。卒業式の時期を迎え、清水も卒業の年になったということを出し出してくれたようで、連絡をくれたらしい。久しぶりにお互いの近況報告をしたあと、湯川がおもむろに頼み事をしたいと言ってきた。どうやらサークルの同窓会の幹事を、特に清水に任せたいらしい。清水としては決して暇がないわけではないのだけれど、日程調整やら出欠の確認やらは非常に面倒だ。とはいえ湯川は非常にお世話になった先輩であり、失礼な態度もとりにづらい。なんとか頼みを断りたいけれど、さてどうやって断ろうか。（東中ら 2020）

## 図2. 第4回シチュエーショントラック課題

ライブコンペでは対話システムが人と比べて自然であることも評価の対象とされているからである。この評価は複数の評価者によって行われる。そこでは西川(2022)などで指摘されているようにユーザの発話から応答までの時間や視覚的な情報も対象とされる。形態素だけが対象となるのではなく表情や手の動きや話者の姿勢なども重要な役割を果たす。この点が、特に文系の研究者の参加を促す理由である。さらにこの点に関しては主催者側の設定にも表れている。技術の発展やハードウェアの進歩もあり、2023年の対話ライブシステムコンペティション6（東中ら 2023）では対話システムはユーザからの入力機器にマイクとカメラが使われ、音声入力と顔の向きなどの情報をシステムに入力し、システムはこれに対する応答として、音声とモニタ上のキャラクタ（エージェント）の表情や手の動きを含む姿勢を組み合わせた映像の情報を出力するものが用いられた。

ライブコンペおよび講習会の対話システムの変化は文系研究者の参加者数に影響を及ぼす可能性がある。

講習会第1回から第4回までの対話システムはキーボードからの文字入力に文字出力で応答するシステムであった。参加者数に関して講習会の中で参加者から聞かれる感想には、「対話システムと聞いて期待したのは音声によるやり取りで文字によるやり取りに魅力を感じない」という声もあったが、徐々に文系の参加者は増え全6回の中では文字による対話システムの最終回である第4回が最多であった。第5回からは対話システムの入力出力が音声となった。対話システムとして文系の参加者の声に応えたものとも考えられるが文系の参加者は減少した。この原因の可能性として、対話システム自体の変化について紹介する。

第1回との比較では、第2回からSUNABAの使用が可能となり、文系の参加者の参加が容易になった。しかしこの段階ではSUNABAによって対話プログラム部分を作ることができたが、図1の「コンピュータ内部」以外に、出力のためにtelegramと連携させること、さらにハードウェアの条件として予選期間中24時間自分のコンピュータをネットワークに接続した状態を継続する必要がある。図1のインターフェイスの部分だけはユーザが用意するもののそれ以外は参加者が用意する必要があった。文系の参加者が最も多くなった第4回ではこの点が改善され、SUNABAで作成した対話プログラムがあれば参加者自身のコンピュータの電源を切っただけでも参加が可能となった。第5回からは対話プログラムが音声及び表情のコントロールまで発展した。結果として第4回までにはなかった表情や姿勢の制御をするためのコマンド、エージェントを対話プログラムと連携させることなど、文系の研究者の日常的な活動にない作業が必要となった。これが文系の参加者を増やすことに意味があるのであれば、負担の大きさの影響について考える価値はあろう。

## 参考文献

- 西川寛之(2022)「マルチメディア型データベースと評価 -リアリティのある評価を求めて」『日本語プロフィール シンシー研究の広がり』
- 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・ドルサ・テヨルス(2020). 対話システムライブコンペティション3, 第96回人工知能学会 言語・音声理解と対話処理研究会(第11回対話システムシンポジウム).
- 東中竜一郎・高橋哲朗・稲葉通将・斉志揚・佐々木裕多・船越孝太郎・守屋彰二・佐藤志貴・港隆史・境くりま・船山智・小室允人・西川寛之・牧野遼作・菊池浩史・宇佐美まゆみ(2023). 対話システムライブコンペティション6, 第99回人工知能学会 言語・音声理解と対話処理研究会(第14回対話システムシンポジウム).