

# 〈wis-Bones〉 先走りしながら共話的インタラクションを生み出す おしゃべりエージェント

石山 智也<sup>†1</sup> 長谷川 孔明<sup>†1</sup> 大島 直樹<sup>†2</sup> 岡田 美智男<sup>†1</sup>

**概要:** 「授業って...」とつぶやくと「13 時だと思うよ」と返してくれる。友人同士の何気ない会話では全てを言い切らずに情報としては不十分な会話が行われている。お互いに不足している会話片を補う中で自己非分離の関係が生まれると心が通じたようで心地いい。本発表では GPT-2 を用いて人の言葉の先を補いつつ共話的なコミュニケーションを生み出す〈ウィズボーン〉(wis-Bones) を提案する。

## 1. はじめに

「えっと、今日やることって」と言えば、「ミーティングだったような」と誰かが続ける。また、「確か時間が」と言えば「13 時でしたっけ」と続ける。このようなありふれた日常会話のコミュニケーションのなかではお互いに情報を提供しあうことで疎通感や楽しさも生まれてくる。

人同士のコミュニケーションのなかで生まれる楽しさや疎通感は一切どこから生まれてくるのだろうか？本研究ではこの問題を水谷ら [1] の提唱する「共話」が大きく関わっていると考え、その共話的な会話の成り立ちをロボットによる構成論的に研究している。

本研究では日常的に行われる共話的なコミュニケーションを人とロボットの間で実現することを狙いとして共話の生成モードを選択できる会話システムとして、人の発話を足がかりに相手の言葉を先走るおしゃべりエージェント〈wis-Bones〉を構築し、その動作例を示すことにする。

## 2. 研究背景

### 2.1 〈対峙しあう関係〉から〈並ぶ関係〉へ

これまで、人とシステムとの関わりの多くは〈対峙しあう関係〉を前提としていた。人がロボットに命令して、ロボットはそれに対して応答を表示する。そのため「使う側-使われる側」と明確に役割が線引きされてきた。

このように〈対峙しあう関係〉は「主導する側と従属する側」が固定された関係にある。しかし、母親と子どもと一緒に絵本を眺めるときや家族と一緒にテレビを見ると

き、一方的に感想を言い続けるようなことはなく、お互いに少しずつ言葉を交わしながら感想を共有しあっている。どちらかが主導権をもつわけでもなく、主導する側も従属する側も存在しない。この関係は〈対峙しあう関係〉ではなく〈並ぶ関係〉 [2] といえる。

テニスのラリー、散歩中の手、話したいことなど二者の間に「第三項」や「媒介物」のようなものが存在する時、それらを介して相手の考えていることがなんとなく伝わる三項関係と呼ばれる状態がしばしば生まれる。このような状態では相互に相手はどう考えているのか相手の身体に入り込んで思っていることや今の気持ちを推測する〈なり込み〉 [3] によるものだろうと考えている。

### 2.2 共同想起対話

お互いに経験した出来事や過去に体験したことについて誰かと一緒に思い出すことを「共同想起 (joint remembering)」と呼ぶ。共通の知識や経験などをともに思い出しながら紡ぎ出される会話であり、自分自身を相手になり込ませるようにして志向を調整しあうコミュニケーションである。

忘れてしまったことについて人同士でああでもないこうでもない対話を行う。小野田ら [4] は〈Talking-Bones〉を用いて共同想起対話を行う様子を観察し子どもたちが忘却した言葉に志向を向ける共同想起対話が行われていること。ロボットが忘れてしまった単語を教えてあげる協力行為、共通の目的に向けた子供たちのコミュニケーションなどを引き出す場を形成していたことを確認している。

<sup>†1</sup> 豊橋技術科学大学 情報・知能工学系

<sup>†2</sup> 豊橋技術科学大学 エレクトロニクス先端融合研究所



図 1 時々もの忘れしてしまう 〈Talking-Bones〉

### 2.3 投機的な会話

これまで筆者らは、聞き手からの「あなたの話を聞いていますよ。」という聞き手性をリソースとし、発話タイミングや内容を相手の様子を見ながら発話文を生成するロボット〈Talking-Ally〉の研究を行ってきた [5].

この研究では日常的な会話の発話連鎖を「投機的な行為」と「グラウンディング」の組み合わせで行われていると説明してきた。この2つの要素を欠いたロボットとのコミュニケーションでは、「相手に発話を一方的に任せて、ロボットが喋ったことをただただ受け入れることしかできない」、「ロボットの生成した発話文を人がどう受け取るかロボット自体が評価することができない」などの問題があった。



図 2 相手の聞き手性を参照し発話を調整する 〈Talking-Ally〉

### 2.4 共話

水谷らは、欧米のように相手が話し終わるまで話さず発話が完結するのを黙って待つ「対話」と比較して、日本語の話し方は「共話」と呼ぶべき発話形態ではないかと指摘している [1].

対話と共話を比較すると以下の図 3,4 のように表すことができる。対話の場合、相手が話し終わるまで待つため発話タイミングは重なり合うことはなく2つの独立した線になる。

一方で共話的な会話では会話の途中で相槌をうったり相手の発話を引き取って発話を続けたり、相手の発話で不足していた部分を補ったり、その相手の補完に対して補完したり、といったようにするうちに相手のお互いの線が重な

り合い、まるで一つの線になる。

一つの発話を必ずしも一人の話し手が完結させるのではなく、話し手と聞き手のふたりで作っていくという考えで、お互いに相手が次に何を話そうとしているか想像し、相手を補足したり情報を提供しあう発話モードが共話である。

例として、「昨日ちょっと用があって鎌倉へ…」と言いかけた話を引き取ってもらい「おいでになったんですか」と完結させる機会を与える。

すべてを言い切らずに相手に続きを話してもらうことで話し手と聞き手の間に共通の理解があることを認識することができ、安心感や心がつながっているように感じられるような話し方となる。



図 3 対話的な会話

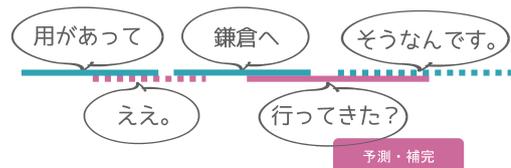


図 4 共話的な会話

## 3. 〈wis-Bones〉

〈wis-Bones〉とは、人の発話を足がかりにその言葉を先走りすることで共話的なコミュニケーションを生みだすおしゃべりエージェントである。

### 3.1 ハードウェア

本研究のプラットフォーム〈Talking-Bones〉は以下のような、骨を模した外観となっている。図5に〈Talking-Bones〉の外観を示す。

〈Talking-Bones〉は図6に示すように、センサには人の状態を把握するためのヒューマンビジョンコンポ (HVC-P2) と発話用のスピーカ、音声認識用のマイクから構成されている。

内部にはサーボモータを3つ搭載しており、頭部と胴体の振る舞いを制御している。また、〈Talking-Bones〉のハードウェア構成の1つの特徴として胴体下部の接合にバネを用いていることが挙げられる。サーボモータでの制御では機械的な印象を与えてしまうため、バネによる不規則な揺れを追加して生き物らしい印象を与え、振る舞いとの適応ギャップを回避することを目的としている。



図 5 <Talking-Bones> の外見

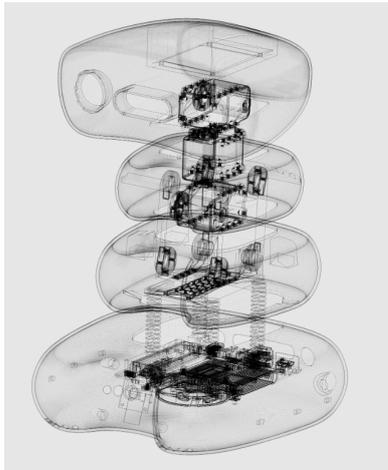


図 6 <Talking-Bones> のハードウェア構成

### 3.2 ソフトウェア

3章で示したインタラクション例を実現するために以下のようにシステムを構築した。

まずはじめに音声認識エンジンを用いて人の発話を認識し、音声から文字情報に変換する。次にその人の発話をもとに次に話そうとすること深層学習の推論モデルを用いて、与えられた文字列の続きを生成する。文章生成モデルはGPT-2にCC100[6]のデータセットを元に事前学習を行い日本語に特化されたモデル rinna[7]を用いた。生成された文章を形態素解析器[8]によって整理し、項構造をから「会話片」を作り合成音声によって発声することで最終的に相手の会話の「ひきとり」を達成する。

図7にシステムの構成を示す。

#### 3.2.1 深層学習によるひきとり

本システムでは先走りのための手法として深層学習(GPT-2)による文章生成を用いた。現時点での深層学習による文章生成では相手の意図してなかった先走りが行われる場合がある一方で、人が意図していなかった先走りが行われた場合に相手がそれに触発されてより会話が発展したり、間違えた出力を訂正しながら教えてあげる有能感を感じられることができると期待したためである。

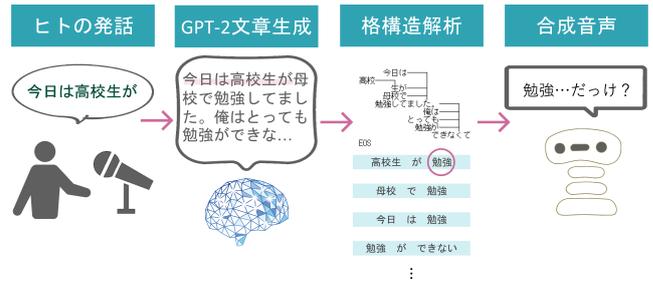


図 7 本システムの構成

#### 3.2.2 格構造生成部

本研究ではPythonからKNPを用いるためにpyknpを介して構文解析を行った。rinnaの推論によって得られた出力を係り受け解析すると図8に示すようにどの単語がどの単語に対応しているか解析できていることがわかる。

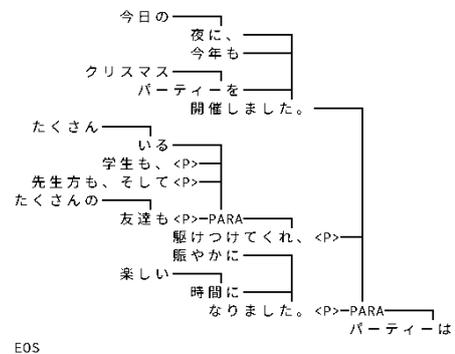


図 8 KNPによる係り受け解析結果

出力から格構造を抽出し、名詞句と動詞のペアを生成する。pyknpによって分析を行った出力に「項構造」が含まれている行を抜き出し、動詞とそれに係っている名詞句が含まれるデータフレームを取得することができる。それをもとにより扱いやすい形に整形したものを表1に示す。

表 1 得られた項構造の例
('クリスマス', [( '今年', 'の' )])
('開催', [( 'クリスマスパーティー', 'を' )])
('要る', [( '友達', 'が' )])
('駆け付ける', [( '友達', 'が' )])
('楽しい', [( '時間', 'が' )])

## 4. インタラクション

インタラクション例を以下に示す。表2のように「高校生と一緒にね」という発話を引き取って「作った」と発話を行う。

## 5. 終わりに

本研究では共話的なコミュニケーションの成立について構成論的に探るために人の発話を足がかりに相手の言葉を

表 2 <wis-Bones> のインタラクション

01:	Robot	今日何があったの？
02:	Human	今日はお客さんがね
03:	Robot	父ちゃん
04:	-	かな
05:	Human	そうじゃなくってあの...
06:	Robot	そうなんだね
07:	-	それじゃあなに？
08:	Human	来たんだよ
09:	Robot	そうなんだ
10:	-	そうしたらどうなったの？
11:	Human	高校生と一緒にね
12:	Robot	作った
13:	-	かな
14:	Human	そうそう
15:	-	作ったんだ

先走るおしゃべりエージェント <wis-Bones> を実装した。<wis-Bones> は深層学習を用いた予測による「ひきとり」など共話的な会話を指向するおしゃべりエージェントである。今後は共話的なコミュニケーション形態を生み出していることを実証実験を行って確認していきたい。

## 参考文献

- [1] 水谷信子. 「共話」から「対話」へ. 日本語学, Vol. 12, pp. 4-10, 1993.
- [2] 長谷川孔明, 林直樹, 岡田美智男. マコにて：並ぶ関係に基づく原初的コミュニケーションの研究. ヒューマンインタフェース学会論文誌, Vol. 21, No. 3, pp. 279-292, 2019.
- [3] 鯨岡峻. 原初的コミュニケーションの諸相. ミネルヴァ書房, 1997.
- [4] 小野田慎平, 西脇裕作, 窪田裕大, 大島直樹, 岡田美智男. 子どもたちはときどきモノ忘れするロボット<Talking-Bones>とどのように関わるのか？ ヒューマンインタフェース学会論文誌, Vol. 23, No. 2, pp. 213-226, 2021.
- [5] 松下仁美, 香川真人, 山村祐之, 岡田美智男. 非流暢性を伴うロボット <Talking-Ally> の発話調整方略とその聞き手に対する適応に関する研究. ヒューマンインタフェース学会論文誌, Vol. 20, No. 2, pp. 255-268, 2018.
- [6] Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 8440-8451, Online, July 2020. Association for Computational Linguistics.
- [7] 趙天雨, 沢田慶. 日本語自然言語処理における事前学習モデルの公開. 人工知能学会研究会資料 言語・音声理解と対話処理研究会, Vol. 93, pp. 169-170, 2021.
- [8] 黒橋禎夫, 河原大輔. 京都大学自然言語処理ツール, 2020.