

特別研究報告書

傾聴対話における関心・共感表出のための 焦点と感情語に基づく応答生成

指導教員：河原 達也 教授

京都大学工学部情報学科

辻 爽汰

2023年1月31日

傾聴対話における関心・共感表出のための 焦点と感情語に基づく応答生成

辻 爽汰

内容梗概

近年日本の高齢化が進む中で、高齢者の社会的孤立を軽減するために傾聴ボランティアの需要が高まってきている。傾聴対話とは、相手の立場や意見を理解し、それに対して適切な反応をすることを意味する。これは相手との関係を深め、信頼関係を築くために必要であると考えられる。そして相手の話をさらに聞き出すことで、話を聴いてもらいたいという相手の欲求を満たすことが傾聴対話の目的である。しかし、傾聴ボランティアを務めるために必要なスキルや知識が多く、ボランティア活動に取り組む上での負担が大きいため傾聴ボランティアは不足している。

一方、音声対話システムの研究は近年活発に進められており、スマートフォンやスマートスピーカなどの普及により日常生活においても利用されるようになってきている。音声対話システムを用いた傾聴対話システムの研究もなされており、アンドロイドを用いた音声対話システムが傾聴対話を行うことで傾聴ボランティアを代替することが期待できる。しかし、現在の傾聴対話システムでは話し手が話している内容に関わらず、ポジティブな発話内容なら「いいですね」や「素敵ですね」と応答するなど固定的な応答が多い。そこで、本研究では、話題や焦点語、そして感情語に応じた応答を生成することで話し手への理解や共感、関心を示し、話し手がより多く話したいと思うような傾聴対話システムの構築を目指す。

話し手の発話からポジティブな感情を検出したタイミングで ~~OpenAI 社の GPT-3~~ を用いたプロンプティングにより応答文を2文連続で生成するシステムを構築した。1文目は、話題語と感情語を組み合わせた「簡単応答」または焦点語と感情語を用いた「要約応答」である。2文目は、1文目に続く「質問」または「自己開示」である。「それが美味しかったです」や「とても綺麗でした」とユーザが発話した場合の「簡単応答」とは、それぞれ「お弁当が美味しかったですね」、「富士山がきれいだったんですね」などであり、「要約応答」とは、「お弁当の卵焼きがふわふわで美味しかったですね」、「富士山の山頂からの景色が別世界のようで綺麗だったんですね」などである。また、「質問」とは、

「どんなお弁当が好きなんですか」, 「富士山には何度も行っただことがあるんですか」などであり, 「自己開示」とは, 「私は焼肉弁当が好きで毎週食べるんです」, 「私は毎年富士山を登って山頂からの景色を楽しんでいます」などである. 1文目を生成する際に使用するプロンプトに含める情報として, 話題, 焦点語の履歴, 感情語, 応答例を用いた. また, 2文目を生成する際に使用するプロンプトに含める情報として, 1文目の応答文, 2文目の応答例を用いた.

提案システムの応答を人間の応答, 従来システムの応答と比較する評価実験を実施した. 参加者には日本語としての自然さ, 文脈整合性, 理解の表出, 関心の表出, 共感の表出, システムの印象, の6項目について7段階で評価してもらった. この評価結果に対して多重分析を行ったところ, 日本語としての自然さの項目, 文脈整合性の項目, 理解の表出の項目においては提案システムの評価値が従来システムの評価値よりも有意に低い結果となったが, 関心の表出に関しては提案システムは従来システムを有意に上回り, 人間と同程度の評価値となった. また, 提案システムの応答のうち, 日本語として自然なものだけを取り出して多重分析を行った. その結果, 提案システムは, 文脈整合性の項目, 理解の表出の項目において従来システムより有意に低い評価結果となることなく, 関心の項目については従来システムより有意に高い評価結果を得ていることが分かった.

Response Generation Based on Focus and Emotional Words to Express Interest and Empathy in Attentive Listening

Sota Tsuji

Abstract

As Japan's population ages in recent years, there has been an increasing demand for attentive listening volunteers to help reduce the social isolation of the elderly. Attentive listening means paying close attention to the speaker and focusing on their words, thoughts, and emotions. This is considered necessary to deepen the relationship with the speaker and to build trust. The purpose of attentive listening is to satisfy the speaker's desire to be listened. However, there is a shortage of attentive listening volunteers because of the large number of skills and knowledge required to be volunteers and the heavy workload involved in volunteer activities.

On the other hand, research on spoken dialogue systems has been active in recent years, and with the spread of smartphones and smart speakers, these systems have come to be used in our daily lives. Research has also been conducted on an attentive listening system using a spoken dialogue system and it is expected that an android-based spoken dialogue system can replace listening volunteers by conducting attentive listening. However, current attentive listening dialogue systems often return similar responses regardless of what the speaker is saying. Therefore, in this study, the aim is to construct an attentive listening system that expresses understanding, empathy, and interest towards the speaker through the generation of responses based on the topic, focus words, and emotional words, so that the speaker is encouraged to continue speaking.

The proposed system generates two consecutive response sentences by a prompt with OpenAI's GPT-3 at the timing when positive emotion is detected from the speaker's utterance. The first sentence is a *simple* response containing a topic word and an emotion word, or a *summary* response using a focus and an emotional word. The second sentence is a *question* or *self-disclosure* following the first sentence. When a user says, "That was delicious", a *simple* response is, for example, "I see your lunch was delicious" etc. A *summary* response

is, for example, "So the egg rolls in your lunch were fluffy and delicious" etc. A *question* is "What kind of bento do you like?" etc. A *self-disclosure* is "I like yakiniku bento, I eat it every week" etc. The information to be included in the prompt used to generate the first sentence is the topic, focus words, emotional words, and the response example. As information to be included in the prompt used to generate the second sentence, the response sentence of the first sentence and the example of the response of the second sentence were used.

A comparison experiment was conducted to evaluate the proposed system against human responses and a conventional system. Participants rated six elements (naturalness, contextual consistency, expression of understanding, expression of interest, expression of empathy, and system impression) on the 7-point scales. Multiple comparisons were performed on the evaluation results. The proposed system received lower ratings than the conventional system in naturalness, contextual consistency, and expression of understanding, but received higher ratings in expression of interest and comparable ratings against the human responses. When unnatural responses by the proposed system were filtered out in this analysis, it was found that it performed similarly to the conventional system in contextual consistency and expression of understanding and outperformed it in expression of interest.

傾聴対話における関心・共感表出のための 焦点と感情語に基づく応答生成

目次

| | | |
|-------|-------------------------------|----|
| 1 | はじめに | 1 |
| 2 | 関連研究 | 3 |
| 2.1 | 話題推定 | 4 |
| 2.2 | 要約 | 4 |
| 2.3 | 傾聴対話システムにおける応答生成 | 4 |
| 2.4 | GPT-3 | 5 |
| 3 | 話題と焦点語・感情語に基づくプロンプティングによる応答生成 | 6 |
| 3.1 | 応答生成タイミング | 8 |
| 3.2 | 応答内容 | 9 |
| 3.2.1 | 1文目 | 9 |
| 3.2.2 | 2文目 | 9 |
| 3.3 | 応答文候補 | 10 |
| 3.4 | 応答文選択 | 11 |
| 4 | 対話データ | 12 |
| 5 | システム実装 | 13 |
| 5.1 | 焦点語の抽出 | 13 |
| 5.2 | 感情分析 | 15 |
| 5.3 | 話題の決定 | 15 |
| 5.3.1 | 旅行の話 | 15 |
| 5.3.2 | 食べ物の話 | 17 |
| 5.4 | プロンプト | 17 |
| 5.4.1 | 簡単応答用のプロンプト | 17 |
| 5.4.2 | 要約応答用のプロンプト | 17 |
| 5.4.3 | 質問用のプロンプト | 18 |
| 5.4.4 | 自己開示用のプロンプト | 19 |
| 5.5 | 応答選択システム | 19 |

| | | |
|----------|-----------------|-----------|
| 5.5.1 | 感情スコアの算出 | 19 |
| 5.5.2 | 焦点スコアの算出 | 20 |
| 5.5.3 | 応答文選択 | 20 |
| 6 | 評価 | 20 |
| 6.1 | 評価用データセット | 20 |
| 6.2 | 条件 | 21 |
| 6.3 | 評価結果 | 23 |
| 6.4 | 考察 | 26 |
| 7 | おわりに | 28 |
| | 謝辞 | 28 |
| | 参考文献 | 29 |
| | 付録 | A-1 |
| A.1 | 簡単応答の生成文 | A-1 |
| A.2 | 要約応答の生成文 | A-3 |

1 はじめに

近年日本の高齢化が進む中で、高齢者の社会的孤立を軽減するために傾聴ボランティアの需要が高まってきている。傾聴対話とは、相手の立場や意見を理解し、それに対して適切な反応をすることを意味する。これは相手との関係を深め、信頼関係を築くために必要であると考えられる。そして相手の話をさらに聞き出すことで、相手の話を聴いてもらいたいという欲求を満たすことが傾聴対話の目的である [1]。しかし、傾聴ボランティアを務めるために必要なスキルや知識が多く、ボランティア活動に取り組む上での負担が大きいため傾聴ボランティアは不足している。

一方、近年音声対話システムの研究は活発に進められており、スマートフォンやスマートスピーカなどの普及により日常生活においても利用されるようになってきている。米国では 2011 年に、Apple の Siri がスマートフォン、コンピュータ上の音声対話システムで初めて広く使用されるようになり、日本では 2012 年に Siri だけでなく、NTT ドコモの「しゃべってコンシェル」のサービスが開始され、今では広く認知されている [2]。しかし、このような音声対話システムや既存のチャット bot では、長い対話における文脈を考慮した応答が十分にできていない。音声対話システムを用いた傾聴対話システムの研究もなされており、自律型アンドロイド ERICA [3] [4] を用いた傾聴対話システムの研究もなされている。アンドロイド ERICA は人間に近い見かけを持ち、自律的に対話を行うことができるように開発されているため、人間のような存在感を対話相手に示すことができる。そこで、アンドロイドを用いた音声対話システムが傾聴対話を行うことで傾聴ボランティアを代替することが期待できる。しかし、現在の傾聴対話システムでは話し手が話している内容に関わらず、ポジティブな発話内容なら「いいですね」や「素敵ですね」と応答するなど類似の応答が多いため、長時間の対話では話し手に飽きられるということが考えられる。また、話し手の直前の発話内容のみを考慮し応答を生成するため、それより前に聴いた内容を用いた応答は生成しない。例えば、たこ焼きの話をしている場面で、話し手が「それがおいしかったです」と発話した場合、現在の傾聴対話システムでは「いいですね」などと応答し、「たこ焼きがおいしかったですね」などとは応答しない。

そこで、本研究では、話題や焦点語、そして感情語に応じた応答を生成することで話し手への理解や共感、関心を示し、話し手がより多く話したいと思うよう

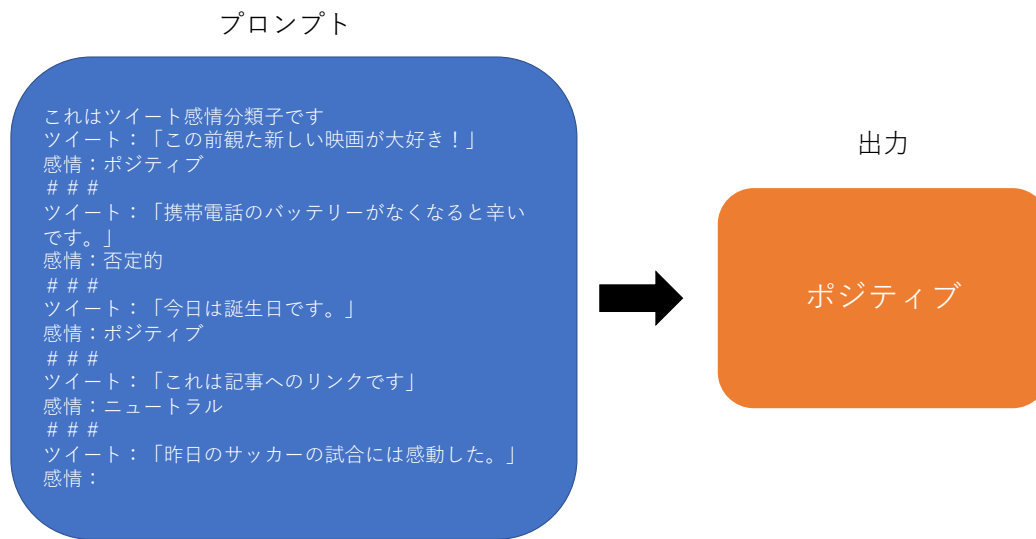


図 1: tweet 文から感情を推測するタスク

な傾聴対話システムを提案する。このシステムは既存の傾聴対話システムに実装し、ある条件下で話題、焦点語、感情語を用いた応答生成を行うことを想定している。つまり、傾聴対話システムが聴き手となり話し手と1対1の対話を行う場面を想定し、既存のシステムを用いて相槌生成や語彙的応答生成などにより傾聴対話を行うと同時に、話し手の発話内容を解析することで話題を推測し、特定の条件下で把握した話題や焦点語、そして感情語に応じた応答を生成するシステムである。

提案システムでは、汎用言語モデルである、OpenAI 社の GPT-3 (Generative Pre-trained Transformer 3) [5] を用いて応答文を生成する。GPT-3 は教師ありデータを使って特定のタスクに特化したモデルを構築するのではなく、事前学習のみで様々なタスクに対応できる汎用的なモデルを構築することを目標に開発された。GPT-3 では、パラメータの更新をすることなく、いくつかの例をプロンプトに入力するだけでタスクを解くことができる。これは決められたタスクだけでなく、初めて見るようなタスクに対しても当てはまる。tweet 文から感情を推測するタスクの場合のプロンプトと、それを GPT-3 に入力した場合の出力の例を図 1 に示す。このように、いくつかの例を与えることにより、様々なタスクに対応することができる。そこで、話題や焦点語、感情語などを入力として、それらに応じた文を生成し、関心や共感を示すことができると考えた。提案システムは

2文連続で応答を生成する。1文目は、話題語と感情語を組み合わせた「簡単応答」または焦点語と感情語を用いた「要約応答」である。2文目は、1文目に続く「質問」または「自己開示」である。「それが美味しかったです」や「とても綺麗でした」とユーザが発話した場合の「簡単応答」とは、それぞれ「お弁当が美味しかったですね」、「富士山がきれいだったんですね」などであり、「要約応答」とは、「お弁当の卵焼きがふわふわで美味しかったですね」、「富士山の山頂からの景色が別世界のように綺麗だったんですね」などである。また、「質問」とは、「どんなお弁当が好きなんですか」、「富士山には何度も行っただことがあるんですか」などであり、「自己開示」とは「私は焼肉弁当が好きで毎週食べるんです」、「私は毎年富士山を登って山頂からの景色を楽しんでいます」などである。1文目を生成する際に使用するプロンプトに含める情報として、話題、焦点語の履歴、感情語、応答例を用いた。また、2文目を生成する際に使用するプロンプトに含める情報として、1文目の応答文、2文目の応答例を用いた。

提案システムを用いることで話し手への理解や共感、関心を示すことができるのか、またそれによって提案システムと もっと話したいと思うのかを検証する評価実験を行った。また、提案システムの応答が日本語として自然であるのか、文脈に合っているのかを ~~検証する評価実験も行った。~~

本論文の構成を以下に示す。2章では、関連研究について述べる。3章では、GPT-3と提案システムの機能について述べる。4章では、本研究で使用する傾聴対話データについて述べる。5章では、システム実装について述べる。6章では、評価について述べる。7章では、研究のまとめについて述べる。

2 関連研究

本章では、はじめに、先行研究で提案されている話題推定について述べる。本研究では、話題を推定し、その話題に応じた応答生成を行っている。次に、要約について述べる。本研究では、焦点語を用いた要約応答の生成を行う。次に、傾聴対話システムにおける応答生成について述べる。本研究では、既存の傾聴対話システムに追加する形で応答生成システムを構築する。最後に本研究で使用するGPT-3について述べる。本研究における応答生成はGPT-3を用いて行う。

2.1 話題推定

丸太 [6] は複数人での雑談対話において, Sidner の焦点モデル [7] を適応し, 抽出された各話者の発話の焦点を統合したものをその時点での話題とすることで局所的な話題推移推定を行った. これは各話者が聞き手と話し手に区別されない雑談対話におけるものである. 雑談対話ではそれぞれの話者が話題となっている単語を発する頻度が高いが, 傾聴対話においては話し手が常に会話の主導権を持ち, ほとんど一方的に話す場面が多いため, その頻度が比較的低くなる. そのため, この手法を用いて傾聴対話における話題を推定するのは難しい. 金子ら [8] は話題を捉えた自然な応答ができる会話システムを作成した. システムの応答によって話題を推移させることで話題を決定している. また, 入力の手言からも話題を決定している. これらの研究では, 雑談対話での話題推定を行っている. 傾聴対話での話題推定は十分に行われていない. 本研究では, 焦点語に対する解析を行うことで傾聴対話における話題を推定する.

2.2 要約

新聞記事や論文の要約から会話履歴の要約まで幅広く要約の研究が行われている. 星野ら [9] は会話文を構造解析し, 会話集合をクラスタリングすることにより会話の流れを保持した形で大量の会話履歴に対する要約を行なった. 松尾ら [10] は大量の投稿を含む掲示板から, どのような話題がトピックになっているのか, どのような会話が行われているのかを提示する手法を提案した. これらの研究では, 対話での応答を生成することを目的とした要約は行っていない. また, 傾聴対話に特化した要約の研究は十分になされていない. 本研究では, 要約を用いた応答を生成する手法を提案する.

2.3 傾聴対話システムにおける応答生成

井上ら [11] はアンドロイド ERICA [3] [4] を用いた傾聴対話システムにおける様々な聞き手応答を実装した. ユーザへの理解, 共感の表出のために聞き手応答として, 相槌や語彙的応答に加えて, ユーザ発話の焦点語に基づく繰り返しや掘り下げ質問, ユーザ発話の極性に基づく評価応答を生成した. 相槌とは「うん」, 「はい」, 「ふん」, 「へー」といった短い応答である. これは傾聴対話における基本的な聞き手応答である. 繰り返しとは, 音声認識結果からユーザの発

話内の焦点である単語を抽出し、その焦点語を繰り返す応答である。掘り下げ質問とは、「どんな」、「どの」、「なんの」、「なにの」、「どこの」、「いつの」、「だれの」といった疑問詞を用いて焦点語に関して掘り下げていく質問である。評価応答とは、ユーザ発話の極性(ポジティブまたはネガティブ)に応じた応答を生成することで、話し手に共感しているようにみせる応答である。この研究ではポジティブの場合には「いいですね」または「素敵ですね」、ネガティブの場合には「大変ですね」または「残念でしたね」という応答を生成している。語彙的応答とは「そうですか」、「そうなんですね」、「なるほど」といった定型表現の応答である。この研究では実装した傾聴対話システムの評価を行っており、「話しやすい」、「真面目に話を聞いていた」、「集中して話を聞いていた」、「積極的に話を聞いていた」といった基本的な傾聴スキルに関する評価では人間と比較しても大きな差は見られなかった。しかし、「話を理解していた」、「関心を示していた」、「共感を示していた」といったより高度な傾聴スキルに関しては人間と比較すると有意な差が見られた。この研究では、話し手の直前の発話内容のみを考慮し応答を生成するため、それより前に聴いた内容を用いた応答は生成しない。本研究では、話題や焦点語を用いて応答を生成することにより、話し手の直前の発話以外も考慮した応答を生成する手法を提案する。

2.4 GPT-3

本研究で用いる GPT-3 について説明する。~~GPT-3 は OpenAI 社から [5] で発表された、GPT、GPT-2 に続く三番目のモデルである。~~モデルの仕組みは GPT-2 と同様に Transformer ベースのモデルとなっている。~~GPT-2 との主な違いは、学習に使用するパラメータの個数である。GPT-2 が 15 億個であるのに対し、GPT-3 は 1750 億個である。~~GPT-3 は教師ありデータを使って特定のタスクに特化したモデルを構築するのではなく、事前学習のみで様々なタスクに対応できる汎用的なモデルを構築することを目標に開発された。Transformer ベースのアーキテクチャでは、ファインチューニングをすることで文章読解や質問回答などといった多くのタスクで優れた性能を得ることができている。しかし、こうしたアプローチはタスクを既知として学習しており、目標タスクの学習データが必要になるという制約がある。タスクの種類によっては十分な量の学習データが用意できない場合もある。一方、GPT-3 は教師ありデータを使って特定のタスクに特化したモデルを構築するのではなく、事前学習のみで様々なタスクに対応できる汎用

的なモデルを構築することを目標に開発された。GPT-3では、パラメータの更新をすることなく、いくつかの例をプロンプトに入力するだけでタスクを解くことができる。本研究では、傾聴対話における応答生成を、焦点語や感情語から文を生成するタスクと考え、GPT-3を用いて応答を生成する。

3 話題と焦点語・感情語に基づくプロンプティングによる応答生成

本研究ではOpenAI社のGPT-3 [5]を用いてプロンプティングにより応答文を2文連続で生成する。入力文を受け取ってから1文目の応答文を生成するまでの過程と、1文目の応答文を受け取ってから2文目の応答文を生成するまでの過程をそれぞれ図2と図3に示す。1文目の応答文の生成は、まず入力文を受け取り、それに対して焦点解析と形態素解析を行うことで入力文の焦点語と形容詞を抽出する。焦点語は焦点語履歴に追加するとともに形態素解析を行うことで、特定の場合にその焦点語を話題とする。また、抽出した形容詞には感情分析を行い、ポジティブだと判定されればその形容詞を感情語とする。ただし、感情語には「感動」など、形容詞以外のものもいくつか含めており、形態素解析を行なった時点でそのような語が入力文に含まれる場合は、その語を感情語とする。次に、入力文に感情語が存在する場合は話題、焦点語履歴、感情語の全てと1文目の応答文生成用プロンプトをGPT-3に入力し、プロンプティングによって1文目の応答文を生成する。2文目の応答文生成は、まず1文目の応答文を質問生成用のプロンプトとともにGPT-3に入力すると同時に、1文目の応答文を自己開示生成用のプロンプトとともにGPT-3に入力する。そうすることでプロンプティングにより、質問を2文と自己開示を2文生成する。これら4文を2文目の応答の候補とする。これらの4文を応答選択システムに入力し、それぞれの感情スコア(ポジティブ度合い)と焦点スコア(文中に含まれる、焦点語履歴に存在する焦点語の個数)によって実際に2文目の応答文として出力するものを決定する。

プロンプティングとは、自然言語生成の形にしてタスクを解く手法である。例として、文書のカテゴリ分類について説明する。通常のカテゴリ分類器を用いると、「○○選手のサーブは200km/hより速い」という文に対しては「スポーツ」、「その教授はノーベル賞を受賞した」という文に対しては「学問」、「内閣総理大臣が辞任した」という文に対しては「政治」と推定するなど、いくつかの決められた

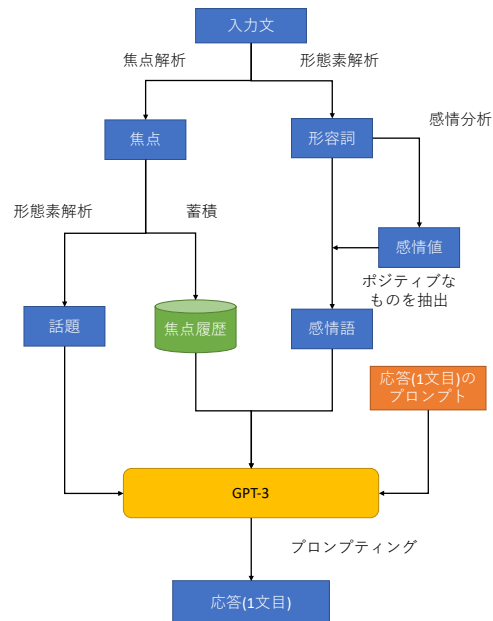


図 2: 1 文目の応答文生成

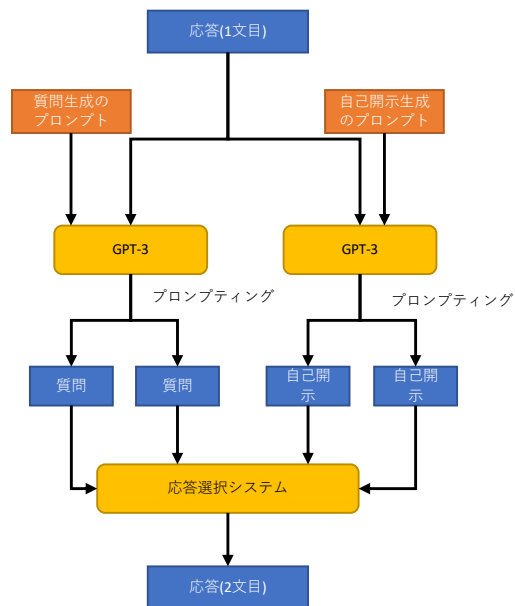


図 3: 2 文目の応答文生成

カテゴリの中から相当するものを推定する. プロンプティングでは, 言語モデルを用いた推定を行う. 「〇〇選手のサーブは 200km/h より速い」という文のカテゴリを推定する場合, 「〇〇選手のサーブは 200km/h より速い。この文のカ

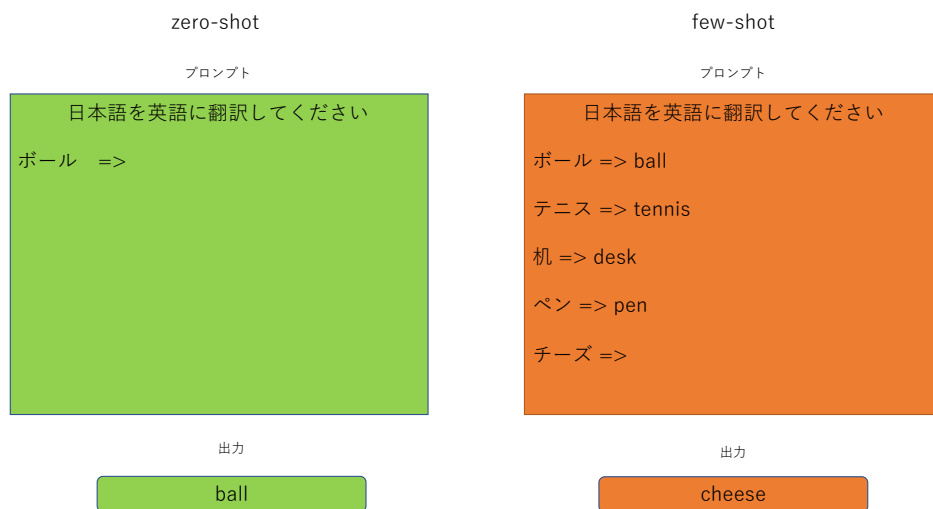


図 4: few-shot learning と zero-shot learning の例

カテゴリは」などと入力することで「スポーツです」などと出力する。つまり通常の分類器は推定の候補がラベルだけであるのに対し、プロンプティングでは推定の候補が単語や文となる。GPT-3 によるプロンプティングでは少ない学習データを用いる few-shot learning や、学習データを用いない zero-shot learning において、BERT [12] を用いた分類よりも高い精度を記録した [5]。few-shot learning と zero-shot learning の例を図 4 に示す。本研究では few-shot learning によるプロンプティングを用いて応答生成を行う。

3.1 応答生成タイミング

本研究において、傾聴対話におけるどのような入力文に対して応答を行うかを説明する。傾聴対話においては話し手への理解や共感を示し、もっと話したいと思わせることが重要である。そのため、相手の感情がポジティブになった場合に、その感情への理解を示したり、その感情に共感したりすることが有効的だと考えた。まず、入力文に対し形態素解析を行う。その結果、「感動」、「感激」などのポジティブな感情を伴う語が含まれている場合に、その語を感情語とする。また、入力文に形容詞が含まれる場合は、その形容詞に対して感情分析を行う。ただし、形態素解析は `juman` [13] を用いて行い、感情分析は¹⁾を用いて行う。感情

¹⁾ <https://taku910.github.io/mecab/>

分析の詳細は5章で述べる。感情語が存在する場合に本研究で提案する手法を用いた応答を生成する。

3.2 応答内容

本研究において、どのような応答を生成するかを説明する。入力文を受け取り上述の応答生成判断によって応答を生成すると判断した場合、その入力文に対する応答を2文連続で生成する。

3.2.1 1文目

1文目の応答内容は2種類ある。一つ目は「(話題) + (感情語 or 焦点語)」という簡単な応答である。以後これを簡単応答と呼ぶ。二つ目は焦点語を用いた要約応答である。「(話題) + (感情語 or 焦点語)」の応答については、話題の理解を示すことと共感を示すことが目的である。話題と感情語または焦点語を含む簡単な応答文を生成することにより話題の理解と共感を、話し手に分かりやすく示すことができるのではないかと考えた。焦点語を用いた要約応答についても話題と感情語、焦点語を含む応答文を生成するが、この応答ではより長い応答文になる。また、話題だけでなく、話し手の発話に含まれていた、その話題に関する情報も応答文に含むため、理解や共感だけでなく、話し手の発話内容に対するより強い関心を示すことができると考えた。簡単応答と要約応答の例をそれぞれ図5と図6に示す。

3.2.2 2文目

2文目の応答文は、1文目の応答に関連する内容の応答文とする。それによって、より話し手への関心を示すことができると考えられる。2文目の応答内容も2種類ある。1つ目は質問である。これは1つ目の応答文に続く形で質問文を生成する。例えば、1文目の応答文が「コーヒーが好きなんですね」だった場合、「どんなコーヒーが好きですか?」や「コーヒー以外にも好きなものはありますか?」、「コーヒーはよく飲まれるんですか?」などという質問が生成される。もう一つの応答内容は自己開示である。質問と同様に1つ目の応答文に続く形で自己開示文を生成する。例えば、1文目の応答文が「コーヒーが好きなんですね」だった場合、「私もコーヒーが好きです」や「私も毎日コーヒーを飲みます」、「〇〇というコーヒーが飲んでみたいんですよ」などという自己開示文が生成される。これら2種類の応答により、関心を示すだけでなく、話し手との関係構築が期待できる。質問生成例と自己開示生成例をそれぞれ図7と図8に示す。

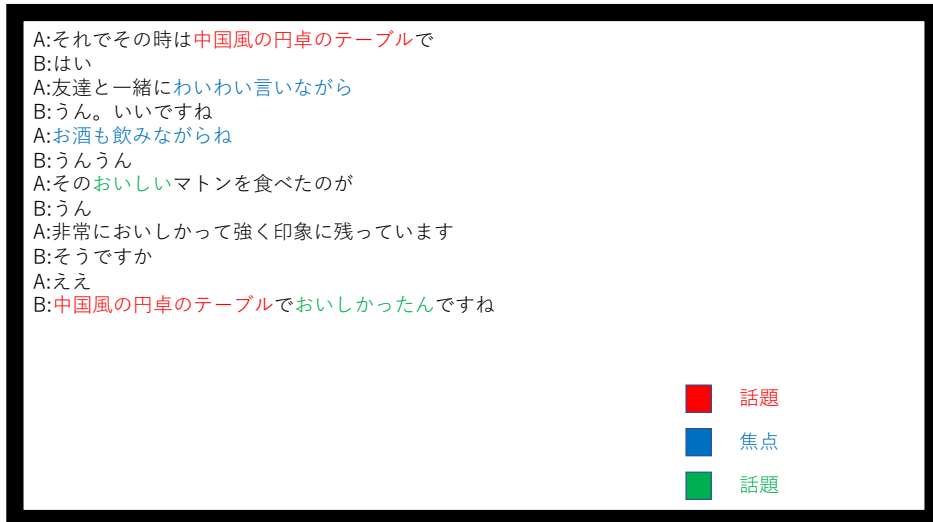


図 5: 「簡単応答例

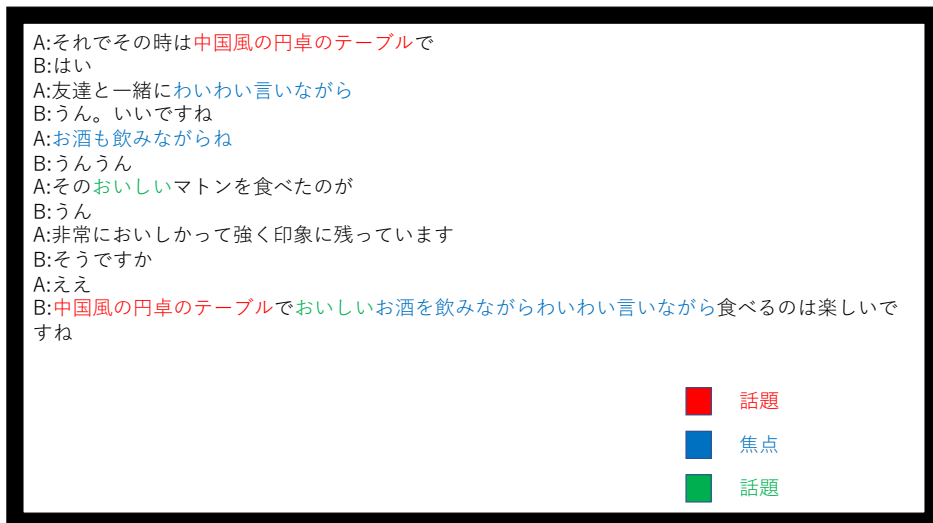


図 6: 焦点語を用いた要約応答例

3.3 応答文候補

提案システムによって生成する応答文候補について述べる。前述の通り、応答文は2文セットである。このうち、1文目は共通とし、それに続く2文目を候補として4種類生成する。その4種類は質問と自己開示がそれぞれ2種類ずつである。GPT-3からの文生成はランダムであるため、同じ入力に対する文生成を2回ずつ実行することで2種類の応答文を生成する。4種類の候補文全てを GPT-3

A:それでその時は中国風の円卓のテーブルで
B:はい
A:友達と一緒にわいわい言いながら
B:うん。いいですね
A:お酒も飲みながらね
B:うんうん
A:そのおいしいマトンを食べたのが
B:うん
A:非常においしくて強く印象に残っています
B:そうですか
A:ええ
B:中国風の円卓のテーブルでおいしいお酒を飲みながらわいわい言いながら食べるのは楽しいですね。いつも友達とわいわい言いながら食べていますか?(質問文)

図 7: 質問生成例

A:それでその時は中国風の円卓のテーブルで
B:はい
A:友達と一緒にわいわい言いながら
B:うん。いいですね
A:お酒も飲みながらね
B:うんうん
A:そのおいしいマトンを食べたのが
B:うん
A:非常においしくて強く印象に残っています
B:そうですか
A:ええ
B:中国風の円卓のテーブルでおいしいお酒を飲みながらわいわい言いながら食べるのは楽しいですね。私も友達とわいわい言いながら食べるのが好きなんです(自己開示文)

図 8: 自己開示例

を用いて生成する。

3.4 応答文選択

提案システムにおける応答文候補から、実際に応答文として出力するものを選択する。まず、4つの候補文に対し、感情分析ツールを用いて感情スコアを算出する。感情分析ツールの詳細は5章で述べる。感情スコアとはポジティブ度合い

である。これは、文中にポジティブな語が多いほど大きくなるスコアである。次に、焦点スコアを算出する。焦点スコアとは、焦点語履歴に含まれる語が文中にも含まれている個数である。これらを算出した後、感情スコアの大きさを比較する。感情スコアが最も大きい候補文を応答文として選択する。感情スコアが最も大きい候補文が複数ある場合、それらの中で焦点スコアを比較する。そして、焦点スコアが最も大きいものを応答文として選択する。焦点スコアが最も大きいものが複数ある場合、それらの中からランダムに選択する。

感情スコアと焦点スコアに注目して応答文を選択する理由について述べる。本研究の提案する応答を生成するのは前述の通り、話者の発話からポジティブな語が検出された場合である。また、本研究の目的は共感を示すことである。それゆえ、ポジティブな応答文を出力することが有効であると考えた。したがって、感情スコアが大きいものを選択することは有効であると考えた。次に、焦点スコアに関して述べる。焦点スコアが大きいものを選択することは、話し手の発話に出現した焦点語が最も多い候補文を選択することを意味する。したがって、話し手の発話内容からの逸脱が小さいものを選択する確率が大きくなる。本研究の目的は理解を示すことでもあるため、これは重要になる。よって、焦点スコアが大きいものを選択するのは妥当であると考えられる。

次に、感情スコアの優先度が焦点スコアの優先度より高い理由について述べる。共感するということは、多くの場合において理解を含んでいる。つまり、共感を示すことができているなら、理解も示せている場合が多い。逆に、理解を示している場合でも共感は示せていない場合は少なくない。したがって、感情スコアを焦点スコアよりも優先することは理解、共感の両方を示すことにおいて有効であると考えられる。

4 対話データ

本研究において使用した対話データについて述べる。本研究では、対話データとして京都大学情報学研究科音声メディア研究室で行った傾聴対話システムの被験者実験での対話データを使用する。この実験は、アンドロイド ERICA [3] [4] を用いた、被験者と傾聴対話システムの対話データである。

この被験者実験では、被験者に対して対話前に指示がされている。「最近行った旅行について話してください」または「最近食べておいしかったものについ

て話してください」という指示である。そのため、全ての対話データにおいて、被験者は旅行または食べ物について話している。

また、本研究では、提案システムを従来システムと比較するだけでなく人間とも比較するために、これらの対話データの内、ERICA を通して人間のオペレーターが被験者と対話を行ったもののみを使用する。

このデータには被験者の発話内容に対し、フィラー、相槌、言い淀み、笑いながらの発話、焦点語、咳、単独の笑いがアノテーションされている。フィラーとは、「ええと」、「あの」、「まあ」など、発話の合間に挟み込む言葉である。相槌とは、「はい」、「うん」、「うーん」などである。言い淀みとは、言葉がすらすらと出ず、滞った発話のうち意味を持たない部分である。笑いながらの発話とは、意味を持った内容の発話のうち、被験者が笑いながら発したものである。焦点語は春日ら [14] が定めた基準に基づいてアノテーションがされている。咳については、被験者実験での対話中に被験者が席をした箇所を示す。単独の笑いとは、発話をしながらの笑いではなく、ただ笑っているものである。

これらのアノテーションがされているデータに対する前処理について述べる。このデータの内、フィラーや言い淀み、咳と単独の笑いの部分の発話は消去した。また、相槌と笑いながらの発話に関してはアノテーションを消去した。焦点語については、提案システムにおいて焦点語を抽出する際に使用する。そして、対話データを発話開始時間の早いもの順に並び替えた。そのため、連続する被験者の発話とオペレータの発話の一部は被っている可能性がある。

5 システム実装

本研究における提案システムの実装について述べる。実装するシステムの内、話題の更新、焦点語履歴の更新を行う様子を、食べ物の話、旅行の話についてそれぞれ図 9、図 10 に示す。ただし、これらの例では話者は一名である。

5.1 焦点語の抽出

まず、入力文を受け取ると、焦点語が存在するかどうかを調べる。含まれる場合、それらを焦点語履歴に入れる。焦点語履歴には、抽出された焦点語を順に記録する。juman を用いて焦点語履歴内の全ての要素に形態素解析を行う。その結果、形態素の個数が 1 のものに対し品詞を調べる。品詞が「感動詞」または「指

U:一万年前のわりと氷河期の箱寿司と言いましてお寿司ですかね
 話題:(箱寿司) 焦点:(['箱寿司'])

U:氷の平原がね箱に入れて作るとおいしいですけども
 話題:(箱寿司) 焦点:(['箱寿司', '氷の平原', '箱に入れて'])

U:ずっと続いておるんですけども
 話題:(箱寿司) 焦点:(['箱寿司', '氷の平原', '箱に入れて'])

U:あれはもう寿司ご飯とそれもそれに観光いたしました
 話題:(寿司ご飯) 焦点:(['寿司ご飯', '観光'])

U:薄い卵焼きを卵焼きをのせたもうそれももうお寿司ですけども
 話題:(薄い卵焼き) 焦点:(['薄い卵焼き'])

U:もうその酢見たことのない酢の入ったご飯と卵焼きの景色ですから、大変感動いたしまして
 味の妙味が何とも言えないもう別世界にいるような気がいたしました
 話題:(薄い卵焼き) 焦点:(['薄い卵焼き', '感動', '妙味', '別世界'])

図 9: 食べ物の話における話題の更新, 焦点語履歴の更新例

U:この前大阪に行きました
 話題:(大阪,) 焦点:(['大阪'])

U:大きな公園でバーベキューをしました
 話題:(大阪, 大きな公園, バーベキュー) 焦点:(['大きな公園', 'バーベキュー'])

U:それから河川敷に行きました
 話題:(大阪, 河川敷,) 焦点:(['河川敷'])

U:そこで友達の野球の試合を観戦しました
 話題:(大阪, 河川敷,) 焦点:(['河川敷', '野球の試合'])

U:試合が終わった後にみんなでレストランに行きました
 話題:(大阪, レストラン,) 焦点:(['レストラン'])

U:とてもおいしかったです
 話題:(大阪, レストラン,) 焦点:(['レストラン', 'とてもおいしかった', '妙味'])

図 10: 旅行の話における話題の更新, 焦点語履歴の更新例

示詞」の場合はそれを焦点語履歴から除く。また要素が「も」の場合も焦点語履歴から除く。また、焦点語履歴内の要素の内、最後の形態素の品詞が「動詞」または「接尾辞」で品詞細分類が「形容詞性述語接尾辞」または「動詞性述語接尾辞」の場合にその要素を焦点語履歴から削除する。こうして残った焦点語履歴の全要素を焦点語として抽出する。

感動詞, 指示詞, 「も」を焦点語から除く理由は、それらはそれ単体で意味を持

たず、後に焦点語を用いて応答を生成する際に精度の低下につながるためである。語尾が動詞や形容詞性述語接尾辞、動詞性述語接尾辞であるものを焦点語から除く理由は、それらは応答を生成する際に使用する際にプロンプトに入力する内容をできる限り単純にするためである。

5.2 感情分析

本研究の提案システムでは感情分析は2箇所で行われている。1つ目は応答生成判断であり、2つ目は応答選択システムである。~~いずれにおいても MeCab を用いて行っている。~~

応答生成判断で行われる感情分析は、抽出された形容詞に対し MeCab で形態素解析を行い、その結果に snow 辞書¹⁾内に含まれる形態素が存在するかを確認する。存在する場合、ポジティブならスコアを「1.0」、ネガティブなら「-1.0」とする。全ての形態素に対するスコアの合計が正であればその形容詞をポジティブと判定し、スコアの合計が負であればネガティブと判定する。

応答選択システムで行われる感情分析は、候補文となる文そのものを入力とし、MeCab を用いて形態素解析を行う。その結果得られた全形態素に対し、日本語評価極性辞書 [15] [16] 内に含まれる形態素が存在するかを確認する。存在する場合、それらの形態素について、感情がポジティブならスコアを「1.0」、ネガティブなら「-1.0」とする。

5.3 話題の決定

提案システムにおける話題の決定方法について述べる。前章で述べた通り、被験者実験において被験者に旅行の話または食べ物の話をしてもらった際の対話データを扱うため、旅行の話を対話データと食べ物の話を対話データの場合で分けて考える。

5.3.1 旅行の話

旅行の話の場合、話題は「地名」、「地名以外の場所」、「食べ物」の3つのセットとする。「地名」とは「日本」、「アメリカ」、「大阪」などである。「地名以外の場所」とは「川」、「山」、「レストラン」などである。「食べ物」は「みかん」、「寿司」、「たこ焼き」などである。

¹⁾ <https://www.jnlp.org/GengoHouse/snow/d18>

次に, どのようにしてこのような話題を決定, 更新するのかを説明する. まず, 被験者の発話を受け取るたびに, その入力文に対して焦点語抽出を行う. 抽出した焦点語に対し juman を用いて形態素解析を行う. ~~そして形態素解析の結果を見ていく.~~

形態素解析の結果,

- 最後の形態素の品詞が「名詞」である
- カテゴリが「抽象物」でない
- 「方」という文字を含まない

の全てを満たす焦点語の内, ある形態素について品詞細分類が「地名」または「組織名」であるものを含むものはその形態素を固有名詞として抽出する. ただし, 品詞細分類が「地名」または「組織名」である形態素よりも後ろの形態素について, カテゴリが「場所」または「自然物」であるものが存在する場合は, その1つの焦点語全体を「地名」とする. そしてこれを話題の「地名」として更新する.

また, 「地名」でない焦点語の内,

- 最後の形態素の品詞が「名詞」である
- 最後の形態素のカテゴリが「場所」である

の両方を満たすものを「地名以外の場所」として話題を更新する.

また, 形態素解析の結果, 最後の形態素の品詞が「名詞」である焦点語の内, ある形態素について

- ドメインが「料理・食事」である
- カテゴリが「抽象物」でない
- カテゴリが「人工物-その他」でない
- 品詞が「名詞」である

の全てを満たす場合にその焦点語を食べ物とし, 話題の「食べ物」を更新する.

次に話題を更新する際の規則を述べる. 話題を更新する際の規則は以下の通りである.

- 「地名」, 「地名以外の場所」, 「食べ物」のいずれかが焦点語に出現するたびに対応する話題を更新する.
- 「地名」が更新された場合, 残りの話題を空にする.
- 「地名」が更新された場合, 「食べ物」を空にする.
- 話題を更新した場合, 焦点語履歴を空にする

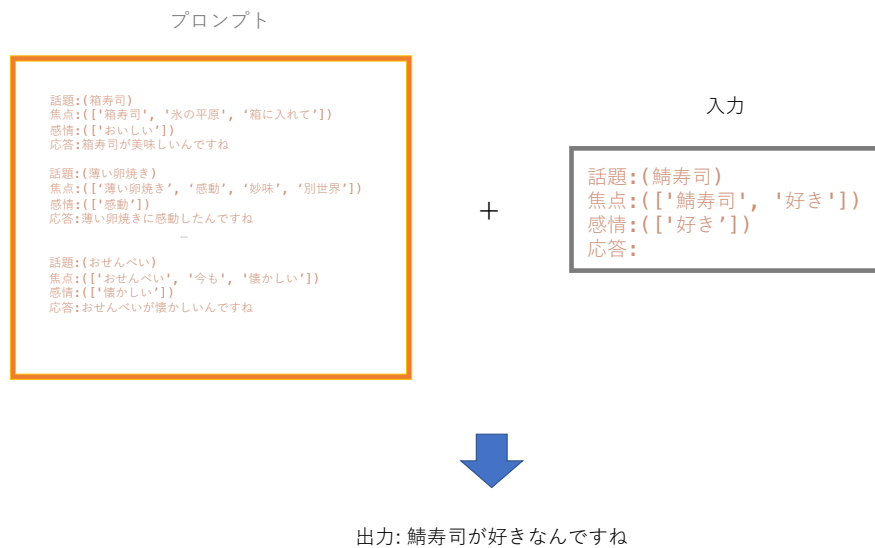


図 11: 簡単応答生成例

5.3.2 食べ物の話

食べ物の話の場合, 話題は「食べ物」のみとする. 食べ物の抽出方法は旅行の話の場合と同様である. 焦点語に「食べ物」が出現するたびに話題を更新する.

5.4 プロンプト

応答文生成の際に使用するプロンプトについて説明する. プロンプトとは, 図 13 と図 14 に示すような, 解きたいタスクの形式とその回答例を示すものである.

5.4.1 簡単応答用のプロンプト

プロンプトに記述する内容は話題, 焦点語履歴, 感情語と応答例である. そのプロンプトに話題, 焦点語履歴, 感情語の単語列をリストで入力し, GPT-3 によって応答を生成する. プロンプトと入力から応答が生成される例を図 11 に示す. GPT-3 のパラメータについては, 「engine」が「davinci」, 「max_tokens」は「70」である. 「engine」は, 応答を生成するのに用いるエンジンであり, 「max_tokens」は応答文の最大文字数である.

5.4.2 要約応答用のプロンプト

プロンプトに記述する内容は直前の入力文, 話題, 焦点語履歴, 感情語と応答例である. そのプロンプトに直前の入力文, 話題, 焦点語履歴, 感情語を入力し, GPT-3 によって応答を生成する. プロンプトと入力から応答が生成される

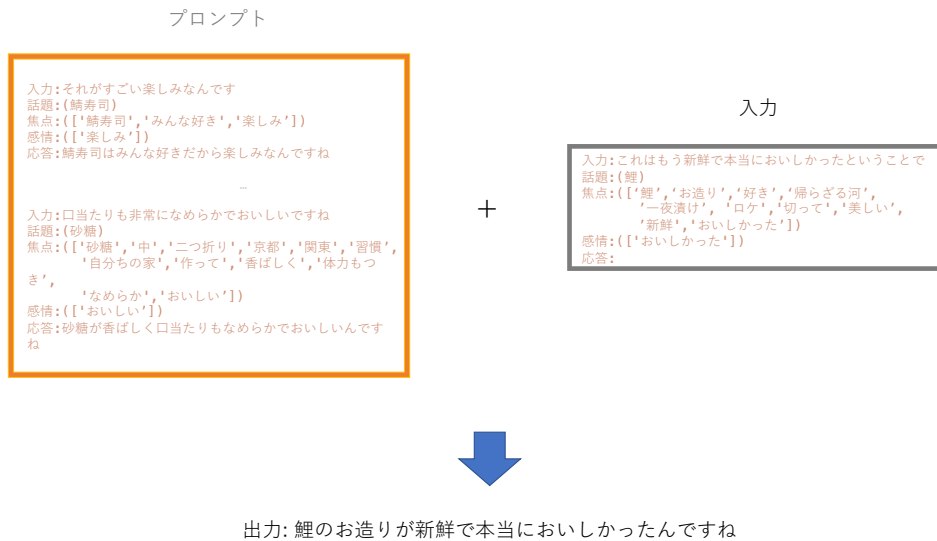


図 12: 要約応答生成例

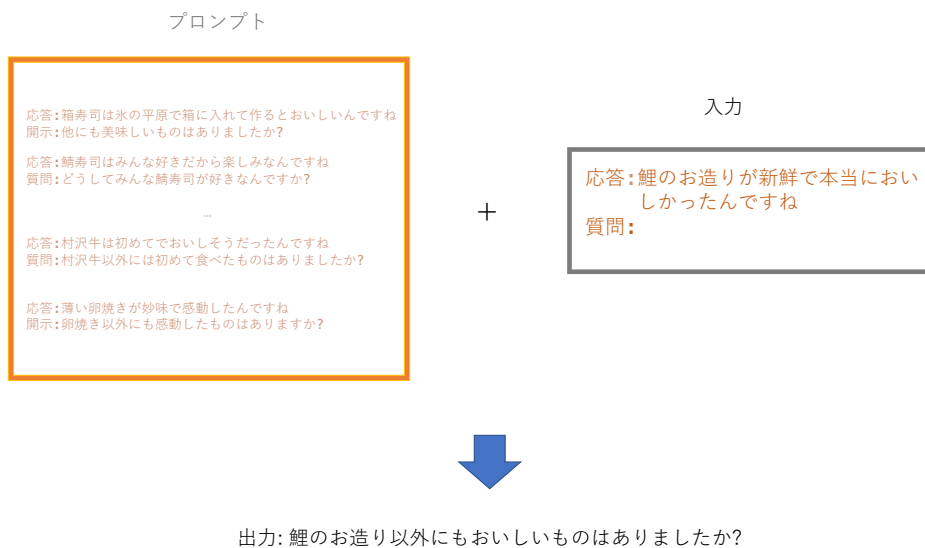


図 13: 質問文生成例

例を図 12 に示す. GPT-3 のパラメータについては, 「engine」 が 「davinci」, 「max_tokens」 は 「100」 である.

5.4.3 質問用のプロンプト

プロンプトに記述する内容は一文目の応答文とそれに続く質問例である. そのプロンプトに一文目の応答文を入力し, GPT-3 によって応答を生成する. プロ

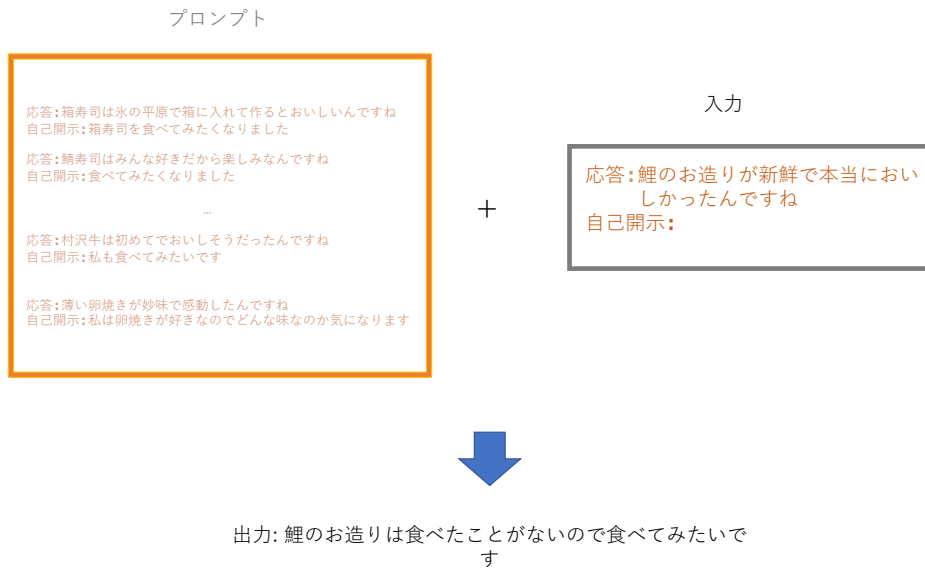


図 14: 自己開示文生成例

ンプトと入力から応答が生成される例を図 13 に示す. GPT-3 のパラメータについては, 「engine」が「davinci」, 「max_tokens」は「50」である.

5.4.4 自己開示用のプロンプト

プロンプトに記述する内容は一文目の応答文とそれに続く質問例である. そのプロンプトに一文目の応答文を入力し, GPT-3 によって応答を生成する. プロンプトと入力から応答が生成される例を図 14 に示す. GPT-3 のパラメータについては, 「engine」が「davinci」, 「max_tokens」は「50」である.

5.5 ~~応答選択システム~~

~~提案システムにおける, 4 つの候補文から実際に応答する 1 文を選択するシステムについて説明する. 応答文の選択には感情スコアと焦点スコアを用いる.~~

5.5.1 ~~感情スコアの算出~~

~~候補文に対して MeCab を用いて形態素解析を行う. その結果得られる全ての形態素に対し, 感情分析を行う. 感情分析の詳細は 5.3 節で述べた通りである. 感情分析の結果得られるそれぞれの形態素のスコアの合計を計算し, その合計の値を候補文の感情スコアとする.~~

5.5.2 焦点スコアの算出

焦点語履歴中の全ての焦点語について、その焦点語が候補文中に含まれているかを調べる。そして、含まれている焦点語の個数をその候補文の焦点スコアとする。

5.5.3 応答文選択

まず、4つの候補文の中で感情スコアが最大のものを取り出す。感情スコアが最大のものが1つの場合、それを応答文とする。感情スコアが最大のものが複数ある場合、それらの中で焦点スコアが最大のものを取り出す。焦点スコアが最大のものが1つの場合、それを応答文として選択する。焦点スコアが最大のものが複数ある場合、それらの中からランダムに1つ選択する。その取り出されたものを応答文とする。

6 評価

6.1 評価用データセット

被験者実験の対話データはオペレータが応答したものである。オペレータが応答した箇所のうち、提案システムが応答を生成する箇所と同じ箇所のオペレータの応答を提案システムの応答と置き換えたものを用意した。その箇所の応答を従来システムの応答と置き換えたものも用意した。提案システムの応答は、1文目が要約応答であるものと1文目が簡単応答であるものの2種類用意した。文脈ごとにこれら4種類の応答を用意した。評価用データは、応答を生成する前の文脈を、話題が応答を生成する時点の話題と等しい範囲で抜き出したものを使用した。ただし、最後の文は応答文としている。そして、同じ文脈における異なる4つの応答を評価した。従来システムは、アンドロイドERICAの傾聴対話システム[11]で用いられている応答生成手法である。実験に使用した4つの応答例を図15、図16、図17、図18に示す。対話データにおける被験者を話者A、オペレータを話者Bとした。対話のBの発話内容がそれぞれのシステム(オペレータも含む)の応答文となっている。対話の最後の応答のみ、それぞれの条件で、オペレータの応答が従来システムの応答、提案システムによる応答に置き換えられている。4章で述べた通り、被験者実験では、被験者に旅行の話または食べ物の話をするように伝えてある。評価用データには、旅行の話を9対話、食べ物の話を10対話使用した。文脈はそれらの対話から、旅行の話を18文脈、食べ物の話を

人間の応答

A もう〇〇ホテルも何度が寄せてもらってるんですけど
B はい
A いつも目の前のカウンターで
B うん
A 今年はサーロインステーキを
B サーロインステーキですか
A はいそのコースをいただきまして
B わあいいですね
A で、もうもちろんお肉はおいしかったんですけども
B いいですね

図 15: オペレータの応答の評価データ例

従来システムの応答

A もう〇〇ホテルも何度が寄せてもらってるんですけど
B はい
A いつも目の前のカウンターで
B うん
A 今年はサーロインステーキを
B サーロインステーキですか
A はいそのコースをいただきまして
B わあいいですね
A で、もうもちろんお肉はおいしかったんですけども
B はい

図 16: 従来システムの応答の評価データ例

を 32 文脈の合計 50 文脈を用意し、合計 200 個の評価用データを用意した。

6.2 条件

100 人に 10 個の評価用対話データに対して 6 つの質問項目について 7 段階評価をしてもらった。質問項目の内容を表 1 に示す。それぞれ、応答文の自然さ、文脈的整合性、理解の表出度、関心の表出度、共感の表出度、聞き手に対する印象の

提案システムの応答(1文目が簡単応答)

A もう〇〇ホテルも何度が寄せてもらってるんですけど
B はい
A いつも目の前のカウンターで
B うん
A 今年はサーロインステーキを
B サーロインステーキですか
A はいそのコースをいただきまして
B わあいいですね
A で、もうもちろんお肉はおいしかったですけれども
B サーロインステーキはおいしかったですね。サーロインステーキ以外のものはおいしかったですか?

図 17: 提案システムの応答 (1 文目が簡単応答) の評価データ例

提案システムの応答(1文目が要約応答)

A もう〇〇ホテルも何度が寄せてもらってるんですけど
B はい
A いつも目の前のカウンターで
B うん
A 今年はサーロインステーキを
B サーロインステーキですか
A はいそのコースをいただきまして
B わあいいですね
A で、もうもちろんお肉はおいしかったですけれども
B サーロインステーキのコースがおいしかったですね。私もサーロインステーキが食べたくなくなってきました。

図 18: 提案システムの応答 (1 文目が要約応答) の評価データ例

6つの観点に対応している。各被験者は200個の対話データからランダムにサンプリングした10対話について評価した。各対話データは5名の被験者によって評価されている。

表 1: 評価結果 (平均と標準偏差, 7 段階: 1~7)

| 評価項目 | 人間の応答 | 従来システムの応答 | 要約応答 | 簡単応答 |
|--|-----------|-----------|-----------|-----------|
| (日本語としての自然さ) | | | | |
| Q1. 最後の B の発言は、日本語として自然だと思いますか。 | 5.89(1.3) | 5.51(1.5) | 4.4(1.7) | 4.25(1.8) |
| (文脈整合性) | | | | |
| Q2. 最後の B の発言は、会話の文脈に合っていると思いますか。 | 5.51(1.4) | 5.2(1.6) | 4.28(1.9) | 3.81(1.9) |
| (理解の表出) | | | | |
| Q3. 最後の B の発言から、B は A の話を理解しているように感じますか。 | 5.16(1.5) | 4.89(1.5) | 4.34(1.9) | 3.88(1.8) |
| (関心の表出) | | | | |
| Q4. 最後の B の発言から、B は A に関心を持っているように感じますか。 | 4.66(1.6) | 4.27(1.6) | 4.71(1.6) | 4.22(1.6) |
| (共感の表出) | | | | |
| Q5. 最後の B の発言から、B は A に共感しているように感じますか。 | 4.71(1.5) | 4.64(1.6) | 4.44(1.7) | 4.26(1.6) |
| (印象) | | | | |
| Q6. もしあなたが A だとすると、最後の B の発言を聞いて B ともっと話したいと思いますか。 | 4.36(1.7) | 4.08(1.6) | 3.88(1.8) | 3.45(1.7) |

表 2: 提案法の評価用データ数

| 1 文目 \ 2 文目 | 自己開示 | 質問 |
|-------------|------|------|
| | 要約応答 | 簡単応答 |
| 要約応答 | 11 | 39 |
| 簡単応答 | 24 | 26 |

表 3: (要約+自己開示) における Tukey 検定結果 (平均と標準偏差, 7 段階: 1~7, 評価データ数: 11)

| 項目 | 要約+自己開示 | 従来システム | 人間 | 結果 |
|----|-----------|-----------|-----------|--|
| Q1 | 4.53(1.8) | 5.25(1.8) | 6.18(1.8) | (要約+自己開示)<従来システム*, (要約+自己開示)<人間**, 従来システム<人間** |
| Q2 | 4.44(1.6) | 4.98(1.9) | 5.18(1.3) | (要約+自己開示)<人間**, 従来システム<人間* |
| Q3 | 4.42(1.6) | 4.82(1.7) | 5.16(1.4) | (要約+自己開示)<人間* |
| Q4 | 4.87(1.5) | 4.36(1.8) | 4.58(1.7) | |
| Q5 | 4.78(1.7) | 4.8(1.7) | 4.64(1.6) | |
| Q6 | 4.04(1.6) | 4.16(1.8) | 4.2(1.9) | |

(*<0.05, **<0.01)

6.3 評価結果

評価結果を表 1 に示す。提案システムの簡単応答では、全ての項目で要約応答、従来システム、人間より低い評価となった。提案システムの要約応答は関心の表出以外の項目については従来システムの応答、人間の応答の両方よりも低い評価となったが、関心の表出の項目については従来システムの応答と人間の応答の両方を上回る評価となった。以下、提案システムの応答のうち、1 文目が要約応答かつ 2 文目が自己開示のもの、1 文目が要約応答かつ 2 文目が質問のもの、1 文目が簡単応答かつ 2 文目が自己開示のもの、1 文目が簡単応答かつ 2 文目が質問のものをそれぞれ (要約+自己開示), (要約+質問), (簡単+自己開示), (簡単+質問) とする。提案システムの応答の評価用データを、(要約+自己開示), (要約+質問),

表 4: (要約+質問) における ~~Tukey~~ 検定結果 (平均と標準偏差, 7 段階: 1~7, 評価データ数:39)

| 項目 | 要約+質問 | 従来システム | 人間 | 結果 |
|----|-----------|-----------|-----------|--------------------------------|
| Q1 | 4.36(1.7) | 5.58(1.4) | 5.81(1.3) | (要約+質問)<従来システム**, (要約+質問)<人間** |
| Q2 | 4.24(1.9) | 5.28(1.5) | 5.43(1.5) | (要約+質問)<従来システム**, (要約+質問)<人間** |
| Q3 | 4.32(1.9) | 4.91(1.5) | 5.15(1.5) | (要約+質問)<従来システム**, (要約+質問)<人間** |
| Q4 | 4.67(1.7) | 4.24(1.6) | 4.68(1.6) | 従来システム<(要約+質問)*, 従来システム<人間* |
| Q5 | 4.34(1.6) | 4.59(1.6) | 4.73(1.5) | (要約+質問)<人間* |
| Q6 | 3.83(1.9) | 4.06(1.6) | 4.41(1.6) | (要約+質問)<人間** |

(*<0.05, **<0.01)

表 5: (簡単+自己開示) における ~~Tukey~~ 検定結果 (平均と標準偏差, 7 段階: 1~7, 評価データ数:24)

| 項目 | 簡単+自己開示 | 従来システム | 人間 | 結果 |
|----|-----------|-----------|-----------|------------------------------------|
| Q1 | 4.31(1.8) | 5.48(1.4) | 5.92(1.2) | (簡単+自己開示)<従来システム**, (簡単+自己開示)<人間** |
| Q2 | 4.05(1.9) | 5.1(1.6) | 5.5(1.3) | (簡単+自己開示)<従来システム**, (簡単+自己開示)<人間** |
| Q3 | 4.15(1.8) | 4.79(1.5) | 5.2(1.4) | (簡単+自己開示)<従来システム**, (簡単+自己開示)<人間** |
| Q4 | 4.43(1.5) | 4.03(1.6) | 4.67(1.6) | 従来システム<人間** |
| Q5 | 4.53(1.7) | 4.43(1.5) | 4.82(1.4) | |
| Q6 | 3.68(1.7) | 3.95(1.7) | 4.38(1.7) | (簡単+自己開示)<人間** |

(*<0.05, **<0.01)

表 6: (簡単+質問) における ~~Tukey~~ 検定結果 (平均と標準偏差, 7 段階: 1~7, 評価データ数:26)

| 項目 | 簡単+自己開示 | 従来システム | 人間 | 結果 |
|----|-----------|-----------|-----------|----------------------------------|
| Q1 | 4.18(1.8) | 5.51(1.6) | 5.89(1.3) | (簡単+質問)<従来システム**, (簡単+質問)<人間** |
| Q2 | 3.6(1.9) | 5.27(1.7) | 5.53(1.5) | (簡単+質問)<従来システム**, (簡単+質問)<人間** |
| Q3 | 3.64(1.8) | 4.97(1.6) | 5.12(1.5) | (簡単+質問)<従来システム**, (簡単+質問)<人間** |
| Q4 | 4.08(1.6) | 4.54(1.6) | 4.68(1.6) | (簡単+質問)<人間* |
| Q5 | 4.03(1.6) | 4.85(1.7) | 4.63(1.6) | (簡単+質問)<従来システム**, (簡単+質問)<人間* |
| Q6 | 3.26(1.8) | 4.24(1.6) | 4.3(1.7) | (簡単+質問)<従来システム**, (簡単+自己開示)<人間** |

(*<0.05, **<0.01)

(簡単+自己開示), (簡単+質問) の 4 つに分類し, それぞれと対応する従来システムの応答の評価用データと人間の応答の評価用データを用意した. それぞれの評価用データの個数を表 2 に示す. そして, それぞれの提案システムの応答に対する評価結果と, 同じ文脈に対する従来システムの応答の評価用データと人間

表 7: 自然な要約応答における ~~Tukey~~ 検定結果 (平均と標準偏差, 7 段階: 1~7, 評価データ数:25)

| 項目 | 要約応答 | 従来システム | 人間 | 結果 |
|----|-----------|-----------|-----------|-------------------------------------|
| Q1 | 5.24(1.3) | 5.63(1.4) | 6.02(1.1) | 要約応答<従来システム*, 要約応答<人間**, 従来システム<人間* |
| Q2 | 5.16(1.5) | 5.36(1.4) | 5.64(1.3) | 要約応答<人間* |
| Q3 | 5.14(1.6) | 5.06(1.4) | 5.26(1.4) | |
| Q4 | 5.14(1.6) | 4.52(1.5) | 4.92(1.4) | 従来システム<要約応答** |
| Q5 | 4.91(1.5) | 4.91(1.5) | 4.94(1.4) | |
| Q6 | 4.54(1.6) | 4.26(1.5) | 4.65(1.6) | |

(*<0.05, **<0.01)

表 8: 自然な簡単応答における ~~Tukey~~ 検定結果 (平均と標準偏差, 7 段階: 1~7, 評価データ数:26)

| 項目 | 要約応答 | 従来システム | 人間 | 結果 |
|----|-----------|-----------|-----------|--------------------------|
| Q1 | 4.98(1.5) | 5.73(1.2) | 5.78(1.2) | 簡単応答<従来システム**, 簡単応答<人間** |
| Q2 | 4.55(1.7) | 5.38(1.4) | 5.32(1.5) | 簡単応答<従来システム**, 簡単応答<人間** |
| Q3 | 4.55(1.7) | 4.97(1.5) | 5.01(1.4) | 簡単応答<人間* |
| Q4 | 4.86(1.4) | 4.32(1.5) | 4.59(1.5) | 従来システム<簡単応答** |
| Q5 | 4.9(1.4) | 4.74(1.5) | 4.65(1.4) | |
| Q6 | 4.17(1.7) | 4.21(1.6) | 4.32(1.6) | |

(*<0.05, **<0.01)

の応答の評価用データとの間で ~~Tukey~~ 検定により多重分析を行った。その結果を表 3~表 6 に示す。(要約+自己開示) に関しては, 日本語としての自然さの項目において, 提案システムと従来システムの間で有意水準 5% の有意差が見られ, 従来システムの方が高い評価となっていた。文脈整合性の項目, 理解の表出の項目, 関心の表出の項目, 共感の表出の項目, 印象の項目ににおいては, 提案システムと従来システムとの間には有意差は見られなかった。次に, (要約+質問) に関しては, 日本語としての自然さの項目, 文脈整合性の項目, 理解の表出の項目において, 提案システムと従来システムの間で有意水準 1% の有意差が見られ, 従来システムの方が高い評価となっていた。関心の表出の項目に関しては, 提案システムと従来システムの間で有意水準 5% の有意差が見られ, 提案システムの方が高い評価となっていた。共感の表出の項目, 印象の項目については, 提案システムと従来システムとの間には有意差は見られなかった。(簡単+自己開示) に関しては, 日本語としての自然さの項目, 文脈整合性の項目, 理解の表出の項目に

において、提案システムと従来システムの間で有意水準 1%の有意差が見られ、従来システムの方が高い評価となっていた。関心の表出の項目、共感の表出の項目、印象の項目においては、提案システムと従来システムとの間には有意差は見られなかった。次に、(簡単+質問)に関しては、日本語としての自然さの項目、文脈整合性の項目、理解の表出の項目、共感の表出の項目、印象の項目において、提案システムと従来システムの間で有意水準 1%の有意差が見られ、従来システムの方が高い評価となっていた。関心の項目においては、提案システムと従来システムとの間には有意差は見られなかった。

6.4 考察

日本語としての自然さの項目 (Q1) について考察する。従来システムが「なるほど」、「いいですね」、「はい」、「素敵ですね」などのいくつかの決められた文の中から応答文を生成するのに対し、~~提案システムでは GPT3 を用いて応答文を生成するため生成される文の形式はランダムとなる。そのため、従来システムにおいては日本語として成り立たない文は生成されないのに対し、提案システムにおいては日本語として成り立たない文、または日本語として不自然な文が生成されることがある。それにより、全ての提案システムが従来システムよりも有意に低い評価結果となったのだと考えられる。~~

次に、文脈整合性の項目 (Q2) について考察する。(要約+自己開示) の提案システム以外については、提案システムの評価が従来システムの評価より有意に低くなっている。提案システムにおいて、感情語を用いて応答する際、否定語の検出ができていない。例えば、「おいしくない」を「おいしい」と判定してしまっている。そのため、「〇〇がおいしくなかった」という被験者の発話に対し、「〇〇がおいしかったんですね」というような応答文を生成してしまう。それにより、提案システムにおける文脈整合性の評価が低くなっていると考えられる。また、質問については、話し手が一度話した内容についてもう一度質問してしまう可能性があり、それによって文脈に合わなくなることがあるが、自己開示については、自分の情報を公開するため、そのようなことが起こりにくい。~~そのため、(要約+質問) の提案システムの評価は従来システムより有意に低くなったのに対し、(要約+自己開示) の提案システムは従来システムとの間に有意差がなかったと考えられる。~~

次に、理解の表出の項目 (Q3) について考察する。文脈整合性が低くなってい

るために、提案システムにおける評価が低くなっていると考えられる。実際に、文脈整合性に関して従来システムと有意差がなかった(要約+自己開示)の提案システムについては、理解の表出に関しても従来システムとの間に有意差がないが、それ以外の提案システムに関しては従来システムより有意に低い評価となっている。

次に、関心の表出の項目(Q4)について考察する。理解を示さずに関心を示すことは難しいと考えられる。しかし、簡単応答については、2文目が自己開示の場合も質問の場合も、理解の表出に関しては従来システムより有意に低い評価であるにも関わらず、この項目については従来システムとの間に有意差がない。このことから、話題を用いた応答生成が関心の表出に効果的であると考えられる。また、要約応答の方が簡単応答よりも評価が高く、(要約+質問)の提案システムに関しては、従来システムより有意に評価が高くなっていることから、焦点語を用いた要約応答の生成が関心の表出に非常に有効であると考えられる。

次に、共感の表出も項目(Q5)について考察する。要約応答、簡単応答ともに2文目が質問の場合の評価は人間よりも有意に低くなっているが、2文目が自己開示のものに関しては有意差が見られなかった。このことから、質問よりも自己開示の方が共感を示すには適切だと言える。

次に、印象の項目(Q6)について考察する。要約応答、簡単応答ともに2文目が質問の場合よりも自己開示の場合の方が、従来システムや人間と比較した場合の結果が良くなっている。

また、提案システムの生成する応答文の自然さが、自然さ以外の評価にも影響している可能性がある。そこで、要約応答と簡単応答の評価用データうち、項目Q1の評価が4.2を超えていたもののみを取り出し、それらに対応する従来システムの評価用データ、人間の評価用データとの間でTukey検定を行った。その結果を表6と表7に示す。要約応答においては、文脈整合性、理解の表出、共感の表出、印象のそれぞれの項目について従来システムとの間に有意差がなかった。そして、関心の表出の項目については有意水準1%で従来システムより高い評価結果となった。簡単応答においては、文脈整合性の項目については有意水準5%で従来システムより低い評価結果となったが、理解の表出、共感の表出、印象のそれぞれの項目について従来システムとの間に有意差がなかった。また、関心の表出の項目については有意水準1%で従来システムより高い評価結果となった。これらのことから、日本語としての自然さを改善することにより、提案システムは

従来システムより有意に低い評価となる項目を高々1つに抑え、関心の項目については従来システムより有意に高い評価を得られることが分かった。

7 おわりに

本研究では、傾聴対話において、GPT-3 を用いたプロンプティングにより、関心と共感の表出のために、話題や焦点語、感情語を用いた応答を生成するシステムを構築した。応答は2文セットであり、1文目は簡単応答または要約応答である。簡単応答は、「~~おにぎりがおいしかったんですね~~」などの話題と感情語から生成される応答である。要約応答は、「~~旅行先で食べたおにぎりが手作りでとてもおいしかったんですね~~」などの焦点語を用いた要約を伴う応答である。2文目は自己開示または質問である。自己開示は、「~~私もおにぎりが好きなんです~~」などのシステムの情報をユーザに開示する応答である。質問は、「~~どんなおにぎりが好きですか？~~」などのユーザへの質問である。提案システムの応答を、1文目が簡単応答である場合と要約応答である場合の2つのパターンを用意し、人間のオペレータによる応答、そして従来の傾聴対話システムによる応答と比較するための評価実験を実施した。~~多重分析を行ったところ~~、日本語としての自然さ、文脈整合性、理解の表出において、提案システムは従来システムよりも有意に低い評価結果が示されが、関心の表出においては、従来システムより有意に高い評価結果が示された。また、~~提案システムの生成する応答文の自然さが、自然さ以外の評価にも影響している可能性があるため~~、提案システムの応答のうち、日本語としての自然さに関する評価が高かったものに着目して分析を行った。~~その結果~~、日本語として自然なものについては他の項目で従来システムと同程度の評価となっており、関心の表出の項目において従来システムより有意に高い評価を得ていることが分かった。今後の課題点としては、日本語としての自然さの改善が挙げられる。また、今回は1文目が簡単応答であるものと要約応答であるものを分けて生成したが、効果的にどちらかを選択する基準を調べることは今後の課題である。

謝辞

本研究、論文執筆にあたり多くのご指導、ご指摘を賜りました河原達也教授に深く感謝の意を表します。また、研究や論文執筆に関して多くの助言を頂いた井

上昂治助教, 研究員の Divesh Lala 氏, 博士後期課程の山本賢太氏, 対話研究チームをはじめとする音声メディア研究室の皆様に深く感謝の意を表します.

参考文献

- [1] 山本賢太, 井上昂治, Lala, D., 中村静, 高梨克也, 河原達也: 自律型アンドロイド ERICA による傾聴対話, 人工知能学会研究会資料, SLUD-B802-13, pp. 58–61 (2018).
- [2] 河原達也: 音声対話システムの進化と淘汰 ー歴史と最近の技術動向ー, 人工知能学会誌, Vol. 28 巻 1 号 (2013).
- [3] F., G. D., Minaot, T., Ishi, C. T., Kawahara, T. and Ishiguro, H.: The ERATO intelligent conversational android, pp. 22–29 (2016).
- [4] Inoue, K., Milhorat, P., Lala, D., Zhao, T. and Kawahara, T.: Talking with ERICA, an autonomous android, pp. 212–215 (2016).
- [5] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I. and Amodei, D.: Language Models are Few-Shot Learners, *Advances in Neural Information Processing Systems*, Vol. 34 (2020).
- [6] 丸田要: 自然な対話継続のための推移する話題推定, 言語処理学会 第 24 回年次大会 発表論文集, pp. 528–531 (2018).
- [7] Sidner, C.L.: *Focusing in the Comprehension of Definite Anaphora*, MIT Press, pp. 267–330 (1983).
- [8] 金子綾, 吉村枝里子, 土屋誠司, 渡部広一: 話題を考慮した自然な会話システムの構築 (2016).
- [9] 星野綾子, 細見格: 句構造解析とクラスタリングを用いた会話履歴の要約, 人工知能学会全国大会論文集, Vol. 32 (2018).
- [10] 松尾豊, 大澤幸生, 石塚満: 電子掲示板における会話からのトピックの発見と要約, 人工知能学会全国大会論文集, pp. 206–206 (2002).
- [11] 井上昂治, Lala, D., 山本賢太, 中村静, 高梨克也, 河原達也: アンドロイド

- ERICA の傾聴対話システム-人間による傾聴との比較評価-, 人工知能学会論文誌, Vol. 36 巻 5 号, pp. H-L51_1-12 (2021).
- [12] Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, *arXiv:1810.04805* (2018).
- [13] 松本裕治, 黒橋禎夫, 山地治, 妙木裕, 長尾真: 日本語形態素解析システム JUMAN 使用説明書, 京都大学工学部長尾研究室. (1997).
- [14] 春日悠生, 井上昂治, 山本賢太, 高梨克也, 河原達也: ヒューマンロボットインタラクションコーパスへの焦点アノテーションの基準と予備的分析, 人工知能学会研究会資料, SLUD-B901-03, pp. 7-12 (2019).
- [15] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一: 意見抽出のための評価表現の収集, 自然言語処理, Vol. 12 巻 3 号, pp. 203-222 (2005).
- [16] 東山昌彦, 乾健太郎, 松本裕治: 述語の選択選好性に着目した名詞評価極性の獲得, 言語処理学会第 14 回年次大会論文集, pp. 584-587 (2008).

付録

A.1 簡単応答の生成文

- 鯉のお造りは好きでしたね。鯉のお造りは美味しいですよ。私も好きです
- 鯉のお造りがおいしかったんですね。私も鯉のお造りは美味しいですよ。私も好きです
- 生菓子はおいしいですよ。生菓子はやっぱりおいしいです
- 春の餅はおいしいですよ。春の餅が好きなんです。私も好きです
- お茶はめでたいですね。お茶が好きですね。私も好きです
- デザートはおいしいですよ。私もデザートが好きです
- コーヒーがおいしいですね。私もコーヒーが好きです
- サーロインステーキがおいしかったんですね。サーロインステーキ以外にはおいしいものはありましたか？
- 淡路島の玉ねぎはおいしかったんですね。淡路島の玉ねぎが好きです
- 色んな種類のお塩が好きになりました。ほかに好きなお塩はありますか？
- ご馳走が楽しみです。ご馳走以外に楽しみで美味しいものはありましたか？
- 中国風テーブルでおいしいお酒を飲みながらわいわい言いながら食べるのは楽しいですね。美味しいお酒を飲みながらわいわい言いながら食べるのも楽しいですね
- 中国風テーブルはおいしいですよ。中国風テーブル以外にはおいしいと思うものはありましたか？
- ちょっとしたお肉はおいしいですよ。私もちょっとしたお肉が好きです
- 白菜が好きなんです。白菜は私も好きです
- バターが好きですね。バター以外には好きなものはありますか？
- あら汁はおいしいですよ。あら汁以外にはおいしい飲み物はありましたか？
- 味噌汁が大好きですね。私も味噌汁が大好きです
- いいお出汁はおいしいですよ。私もいいお出汁が好きです
- 鯛そうめんがおいしいんですね。鯛そうめん以外にはおいしいものはありましたか？
- おからがおいしいんですね。おからって美味しいですよ。私も好きです
- おからはおいしいですよ。おから以外にもおいしいものはありましたか？

- 柚子味噌はおいしいですよ。柚子味噌もおいしいですよ。私も好きです
- 柚子味噌はおいしいですよ。柚子味噌以外にはおいしいものはありましたか？
- 柚子味噌はおいしいですよ。柚子味噌以外にはおいしいものはありましたか？
- レモンの汁を入れるとおいしいですよ。レモン汁もおいしいですよ。私も好きです
- 豪華なお料理はおいしかったですね。豪華なお料理以外にもおいしいお料理はありましたか？
- オードブルはおいしいですよ。オードブル以外には美味しいお料理はありますか？
- フランスワインはおいしいですね。フランスワイン好きですよ。私も好きです
- 塩を沸騰させた状態にして、一定の時間経つまで、甲羅の方を上、脚も上手く鍋の中に入
- れて。塩以外にはこう炊いておいしいものはありましたか？
- エノキのご飯はとってもおいしいですね。エノキ以外にご飯がとってもおいしいものはありましたか？
- エノキのご飯はおいしいですね。いろいろなご飯に出会えていいですか？
- バイキングがおいしかったんですね。私もバイキングが好きです
- バイキングは楽しかったんですね。食べるのが好きなんですか？
- バイキングは楽しくておいしかったんですね。私もバイキングが好きです
- 湯快リゾートで楽しんだんですね。リゾートを楽しむのは好きなんですか？
- 外国で楽しんだんですね。今度一緒に外国に行きましょう
- ハーレーダビッドソンはよかったんですね。他に面白い場所はありましたか？
- お花畑で感動したんですね。お花畑が好きなんですか？
- お花畑に日本の総理大臣がいたんですね。私もお花が好きなんで、お花畑に行ったことがあります
- ビクトリア島は楽しい場所ですね。他にも楽しい場所はありましたか？
- 山温泉で元気になったんですね。弱った体が元気になったんですね？
- エベレストは面白かったんですね。エベレストが好きなんですか？
- スイスアルプスで面白くなかったんですね。スイスは美しいですよ

- スイスアルプスでおいしいものを食べたんですね。他の場所でおいしいものを食べると何が違うんですか？
- スイスでおいしい料理を食べたんですね。おいしい料理は好きなんですか？
- 竹富島で元気だったんですね。元気が好きなんですか？
- ゴールドコーストで大切にしたんですね。人を大切にすることは大事だと思いますか
- 国民宿舎で楽しく過ごしたんですね。私も国民宿舎に行きたいです
- 八角形の天守閣で感銘を受けたんですね。八角形が好きなんです

A.2 要約応答の生成文

- 鯉のお造りが好きだったんですね。どうして鯉のお造りが好きなんですか？
- お造りも一夜漬けも新鮮でおいしいんですね。お造り以外にも新鮮で美味しいものはありましたか？
- 季節のお菓子は本当においしいんですね。季節のお菓子以外にはおいしいお菓子はありましたか？
- 紅葉襲の春のお餅はいろんな味ですね。お餅が好きなのでいろんな味があるのは楽しいです
- めでたいですね。他にもめでたいことはありましたか？
- デザートは満足でおいしいんですね。デザート以外にも満足でおいしいものは食べたんですか？
- コーヒーは温度が濃くておいしいんですね。コーヒー以外にも温度が濃く美味しいものはありましたか？
- サーロインステーキはおいしかったんですね。サーロインステーキ以外のものはおいしかったですか？
- 上等の鉄板で淡路島玉ねぎはおいしかったんですね。淡路島玉ねぎ以外にはおいしいものはありましたか？
- 色んな種類のお塩がおいしかったんですね。色んな種類のお塩以外にはおいしいものはありましたか？
- ご馳走がおいしかったんですね。私もおいしいご馳走を食べてみたいです
- 中国風の円卓のテーブルと友達と一緒にわいわい言い非常においしいですね。私の母が中国風の円卓のテーブルでわいわい言いなんです

- 中国風の円卓のテーブルで一緒にお酒も飲みながらおいしいんですね。お酒以外には一緒に季節の手料理を食べながら楽しいことはありましたか？
- ちょっとしたお肉はおいしそうだったんですね。ちょっとしたお肉は食べてみたいです
- 白菜もブロッコリーも甘くなるのが好きです。白菜以外にも食べると甘くて美味しいものはありますか？
- 意外とバターはおいしいですよ。バター以外には意外に美味しいものはありましたか？
- あら汁はおいしいですね。私もあら汁が好きなんです
- 海の幸が大好きで味噌汁も良かったです。味噌汁以外には和食中心で好きなものはありましたか？
- いいお出汁が漁港ならではの料理だったんでおいしいですね。村沢牛も新鮮でおいしそうです
- 鯛そうめんはおいしいですよ。鯛そうめん以外にはおいしいものはありますか？
- おからはクタクタと炊かなくてもおいしそうだったんですね。おから以外にはクタクタと炊かなくてもおいしいものがありましたか？
- おからをそのまま食べるとおいしいんですよ。おから以外にもそのまま食べると美味しいものはありましたか？
- 柚子味噌はおいしいからパンに付けたりと、おナスを焼いたりと楽しいんですね。柚子味噌以外のナスでおいしいものはありますか？ 柚
- 子味噌はその方がおいしいんですね。柚子味噌以外にもその方がおいしいものはありましたか？
- その方がおいしいんですね。その方がおいしいのは何ですか？
- レモンの汁を入れても全部おいしいんですね。レモンの汁以外にも入れても美味しいものはありますか？
- 豪華なお料理がおいしかったんですね。ほかに豪華なお料理がありましたか？
- オードブルが一番おいしいと提供してますよね。一番おいしいところというのはどちらですか？
- ドイツワインの白とアンジュロゼはおいしそうだったんですよ。ドイツワインの白とアンジュロゼ以外にはおいしそうだったものはありましたか？
- 塩が沸騰させた状態でおいしかったんですね。塩以外は沸騰させた状態で

おいしいものはありましたか？

- エノキの混ざったご飯もとってもおいしいんですね。エノキ以外にもとってもおいしいご飯というのはあったでしょうか？
- エノキの混ざったご飯はとってもおいしいんですね。エノキ以外にはとっても美味しいものはありましたか？