

語りの傾聴における補完応答の生成のための話し手の発話の予測

海野 博揮¹ 大野 誠寛² 伊藤 滉一郎³ 松原 茂樹^{3,4}

¹ 名古屋大学 情報学部 ² 東京電機大学 未来科学部

³ 名古屋大学 大学院情報学研究科 ⁴ 名古屋大学 情報基盤センター

unno.hiroki.t9@s.mail.nagoya-u.ac.jp ohno@mail.dendai.ac.jp

{ito.koichiro.v1@s,matsubara.shigeki.z8@f}.mail.nagoya-u.ac.jp

概要

円滑に会話を進めるには、話し手の語りに対して聞き手が相槌などの応答をすることが重要であり、対話システムが語りを聞く際にも、傾聴態度を示す応答（傾聴応答）の生成が望まれる。傾聴応答の1つに話し手の発話を補完する応答、すなわち、補完応答がある。補完応答を適切に生成できれば、語りを理解していることを効果的に伝えることができる。本論文では、補完応答の生成に向けた、話し手の発話の予測について述べる。まず、補完応答において話し手のどのような発話が補完対象であるかを分析する。次に、分析結果に基づき、本研究では、話し手の既発話文節と係り受け関係にある未発話文節を補完対象として予測する問題に取り組む。

1 はじめに

傾聴応答は、話し手の語りに対して傾聴の態度を示す応答であり、人間同士の会話を円滑に進めるうえで重要な役割を果たしている。そのため、対話システムが語りを聞く際にも、傾聴応答の生成が望まれる。

傾聴応答は、その機能の観点から、相槌、感心、評価、同意、補完など、いくつかの種類に分類することができる（例えば16種類[1]）。このうち、代表例と考えられる相槌については、その生成方法がいくつか提案されている[2, 3, 4]。しかし、「うん」や「はい」のような相槌は、話し手の語りの内容を必ずしも傾聴していなくても表出できるため[5]、傾聴の態度を示す効果が必ずしも高いとはいえない。

一方、話し手の語りの内容を傾聴しないとできない傾聴応答の種類の一つとして補完応答がある。補完応答は、話し手の発話を補う応答であり、聞き手

が話し手のこれまでの語りを理解し、その先に発話される内容を予測したうえで発話される。そのため、補完応答を適切に生成することができれば、話し手の語りを理解し傾聴している態度をより効果的に示すことができると考えられる。

本論文では、補完応答の生成に向けた、話し手の発話の予測について述べる。まず、補完応答において話し手のどのような発話が補完対象であるかを分析する。次に、分析結果に基づき、本研究では、話し手の既発話文節と係り受け関係にある未発話文節を補完対象として予測する問題に取り組むこととし、その予測手法を提案する。最後に、評価実験を行い、その予測可能性について検証する。

2 補完応答とその分析

2.1 補完応答

語りの傾聴における補完応答とは、話し手がこれから発話する内容を予測し、話し手に先んじて話し手の発話を補う応答である。図1に補完応答の例を2つ示す。いずれも、「甘いものが好きだった」という事柄の後に逆説の接続詞「けど」が続いているという内容をもとに、これから語られる内容（「辛いもの」や「変わった」）が予測され、補完応答が行われていると考えられる。

このような補完応答の出現割合は、伊藤ら[1]によると、傾聴応答の中で1.29%であると報告されており、その割合は必ずしも高くない。しかし、補完応答は、話し手の語りの内容を理解し話し手の発話を予測しないと表出できないため、傾聴態度を示すうえで、その生成を行う効果は大きいといえる。

一方、聞き手への入力となる発話、すなわち、話し手がそれまでに発話した語りに対して、聞き手が

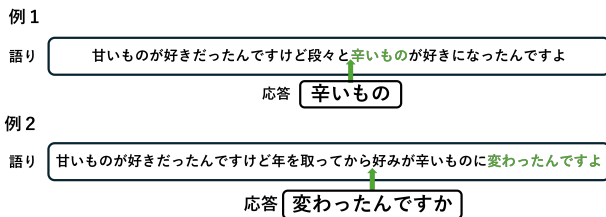


図 1 補完応答の例

予測し補完する対象がどのような要素となっているのかは明らかにされていない。図 1 の例 1 では、それまでに発話された語り「～ですけど段々と」の直後に来る発話「辛いもの」を予測し補完している。これに対して、例 2 は、「～けど年を取ってから」の直後の発話である「好みが」ではなく、その先に出現する「変わったんですよ」を予測し補完している。また、例 1 の補完対象「辛いもの」は、それまでに発話された語りのいずれの文節とも係り受け関係にないのに対して、例 2 の補完対象「変わったんですよ」は、それまでに発話された語りの文節「好きだったんですけど」や「取ってから」の係り先となっている。これらの関係の一端を明らかにすることができれば、補完応答の生成に向けた知見が得られると考えられる。

2.2 補完応答の分析

本節では、聞き手が補完応答を行ったタイミングの前後で話し手の語りを 2 つに分割し、それらと補完応答の内容との関係を分析する。

分析には伊藤らが作成した傾聴応答データ [1] を使用する。このデータには、聞き手役の 11 人のアナウンサーが語りの音声聞きながら発声した傾聴応答 148,962 個が収録されている。語りの音声データには、高齢者のナラティブコーパスである JELiCo[6] が用いられている。収集された傾聴応答は、その機能の観点から、補完応答を含む 16 種類に分類されており、614 個の補完応答が収録されている。

このうち、ランダムにサンプリングした 300 個について、補完応答の内容がその生成タイミング以降に話し手の語りに現れるか否かを判定し、現れたものから 95 個を本分析の対象とした。以降では、語りに現れた補完応答の内容を補完対象と呼ぶ。なお、傾聴応答データは語りの録音データに対してアナウンサーが応答をしているため、応答内容がその後の語りの展開に影響を与えることがない。そのため、応答で補完したことが原因で、補完された内容

表 1 補完応答の内容の発話中での出現位置

	応答直後	応答直後よりも後	合計
応答数	25 (26.3%)	70 (73.7%)	95

表 2 補完対象が既発話文節の係り先であるか否か（応答直後よりも後に出現したもののみ）

	係り先である	係り先でない	合計
応答数	56 (80.0%)	14 (20.0%)	70

の発話を控えるという現象は起きない。

まず、各補完応答に対して、補完対象がその応答タイミングの直後に位置するかそれよりも後かで分類し、頻度を集計した。その結果を表 1 に示す。補完対象が補完応答タイミングの直後よりも後に出現するものは 70 個あり、全体の 73.7% に相当する。補完応答の生成の実現のためには、単に応答タイミング直後に出現すると思われる話し手の発話を予測できるだけでは十分でなく、応答タイミング直後よりも後の話し手の発話も含めた予測が必要であるといえる。

次に、補完対象が補完応答タイミングの直後よりも後に出現する 70 個の補完応答に対して、補完応答タイミング以前に発話された語りの中のいずれかの文節が補完対象に係っている（すなわち、補完対象が既発話文節の係り先である）か否かに分類し、頻度を集計した。その結果を表 2 に示す。補完対象が既発話文節の係り先であるものは全体の 80.0% (56/70) を占め、その多くはそれまでの語りと補完対象との間に係り受け関係があることがわかる。係り受け関係にないものも存在したが、それらの多くは、語りが独話であるために長文となる傾向にあり、従属節が繋がった先に補完対象が偶然現れたものであった。以上より、応答タイミング直後よりも後の話し手の発話を予測するための戦略の 1 つとして、既発話文節の係り先を予測することが考えられる。

3 係り先文節の予測

本研究では、2.2 節の分析に基づき、語りの傾聴における補完応答生成に向けて、話し手の既発話文節のうち、係り先が未入力である文節の係り先を予測する問題に取り組む。具体的には、発話文が文節列 $b_1 \cdots b_n$ で表される場合、文節 $b_t (1 \leq t \leq n)$ が発話されるたびに、既発話文節列 $B_t = b_1 \cdots b_t$ と、その係り受け構造 D_t をもとに、 B_t の中の係り先が未入力である文節の係り先を予測する。

本研究では、既発話文節列 B_t だけでなく、その係

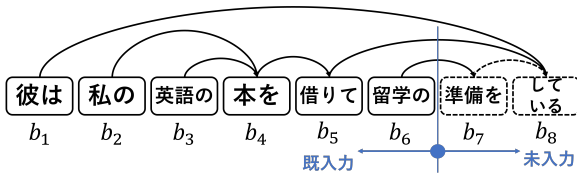


図2 従来手法 [7, 8, 9] の係り受け構造

り受け構造 D_t を必要とするが、文節が入力されるごとに、それまでの既入力文節列に対する係り受け構造を漸進的に同定し出力する係り受け解析手法はいくつか存在している [7, 8, 9, 10, 11, 12]. 特に、従来手法 [7, 8, 9] では、文節が入力されるごとに、図2に示すような係り受け構造を出力する。図2は、「彼は私の英語の本を借りて留学の準備をしている」という文のうち、「留学の」までが入力された時点で出力する係り受け構造を表している。係り先が未入力の文節が複数存在するときに、それらの係り先が同一か否かの情報を含んだ係り受け構造となっている。本研究では、 D_t がこの係り受け構造となっていることを前提として、係り先の予測可能性を検証する。ただし、本論文では、自動解析結果を用いるのではなく、正解の係り受け構造を利用することとする。

本手法では、 D_t に基づいて、 B_t の中で同じ未入力文節に係る文節（複数あれば複数）を取り出し、それらと既発話文節列全体 B_t を MLM (Masked Language Model) に入力し、その係り先文節の主辞¹⁾を予測する。図3に本手法の概要を示す。図3は、図2において b_6 まで入力された時点での MLM への入力と出力を示している。まず、未入力文節 b_8 に係る文節として、 b_1 と b_5 が取り出された後、これら係り元文節の係り先の主辞が予測されている。既発話文節列 (B_6)、係り元の各文節 (b_1 と b_5)、係り先の主辞を表す [MASK] の順に連結し、先頭に [CLS] を、各々の後に [SEP] を付与したうえで、サブワード分割を施したものを MLM に入力し、[MASK] の予測結果を出力している。さらに、未入力文節 b_7 に係る文節として、 b_6 が取り出され、同様の予測が行われている。各時点で、異なる未入力文節に係る文節があるだけ、このような予測が繰り返される。なお、モデルには、Encoder モデルの BERT[14], RoBERTa[15], DeBERTa[16] をそれぞれ使用する。

1) 予測対象としての主辞になりうる品詞は、内元らの定義 [13] から、助動詞を除いたものとする。

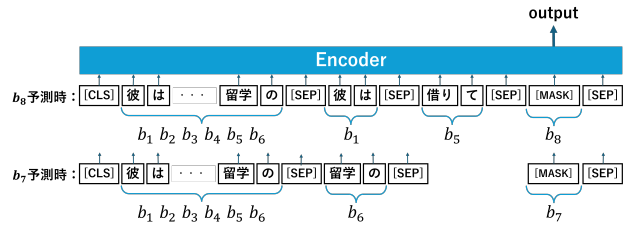


図3 本手法の概要

4 実験

本手法の予測性能を評価するために実験を行った。

4.1 実験の概要

実験データには、同時通訳データベース (SIDB) の日本語講演の書き起こしデータ [17] を使用した。講演の書き起こしデータは文節で区切られており、文節番号と係り先の文節番号が付与されている。講演データは 16 講演分を使用し、そのうちの 14 講演を訓練データ、1 講演を検証データ、1 講演をテストデータとした。本手法の実装には PyTorch²⁾ と Hugging Face³⁾, Juman++[18]⁴⁾ を用いた。モデルは BERT⁵⁾, RoBERTa⁶⁾, DeBERTa⁷⁾ の事前学習済みの公開モデルを使用した。fine tuning 時のハイパーパラメータは、学習率を $5e-7$ 、バッチサイズを 16、エポック数を 50、Early Stopping を 10 とした。また、fine tuning を行わない DeBERTa のモデルを作成し、この性能を baseline とした。評価方法は top- k accuracy を使用し、 k を 1 から 20 の範囲で変化させて測定した。

4.2 実験結果

図4に実験結果を示す。本手法が3つのモデルすべてで baseline を大きく上回った。baseline はいずれの k においても最も性能が低く、高くても 10% 程度であり、既存の事前学習済みモデルでは係り先を予測することが困難であった。また、3つのモデルのうち DeBERTa がいずれの k においても最も高い性能を示し、 $k = 7$ の時点で 50% を超える性能を示

- 2) <https://pytorch.org/>
- 3) <https://huggingface.co/>
- 4) <https://github.com/ku-nlp/jumanpp>
- 5) <https://huggingface.co/cl-tohoku/bert-base-japanese-whole-word-masking>
- 6) <https://huggingface.co/nlp-waseda/roberta-base-japanese>
- 7) <https://huggingface.co/ku-nlp/deberta-v2-base-japanese>

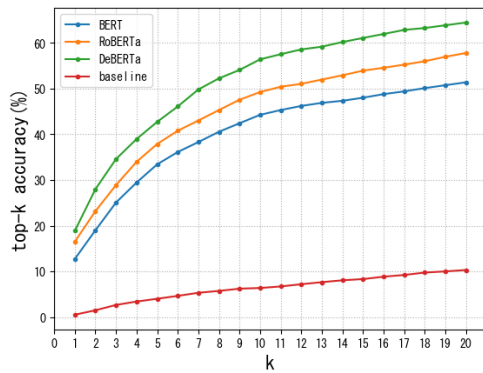


図4 実験結果

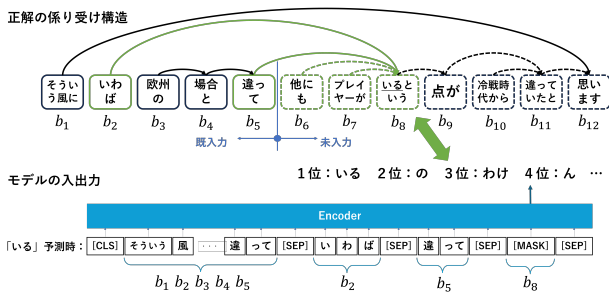


図5 本手法の正解例

した。

図5に本手法 (DeBERTa) が正解した具体例を示す。図5は「違って」までの文節が入力されたときの「いるという」という文節の主辞を予測した結果である。モデルの出力 top-1 が「いる」であり、正解の「いるという」の主辞を正しく予測できている。

以上より、本手法によって、一定の水準で係り先の主辞を予測できることを確認した。

4.3 考察

本節では、予測対象の係り先文節の品詞や、それに係る文節数によって、予測しやすさがどのように変わるかを考察する。以下では、最も高い性能を示した本手法 (DeBERTa) の実験結果を用いる。

4.3.1 品詞による正解率の比較

予測対象の係り先文節の品詞による正解率の変化を図6に示す。動詞、名詞、その他、形容詞の順に正解率が高いことがわかる。形容詞とその他については、その頻度が少なく、テストデータにおいて、形容詞が31回、その他が14回であった。そのため、十分に学習できなかった可能性がある。一方、動詞や名詞は多く出現するが、名詞に比べて動詞は講演時に使用される種類が少ないため、動詞の方が予測しやすかったと考えられる。

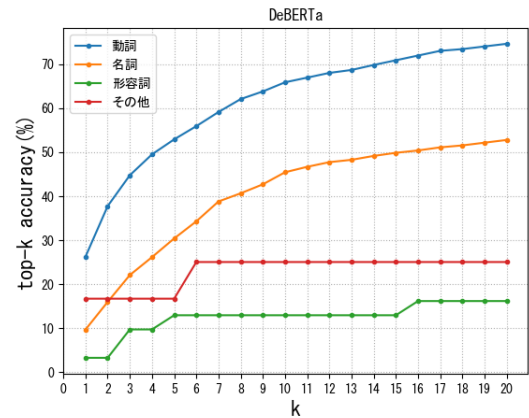


図6 品詞による正解率の変化 (DeBERTa)

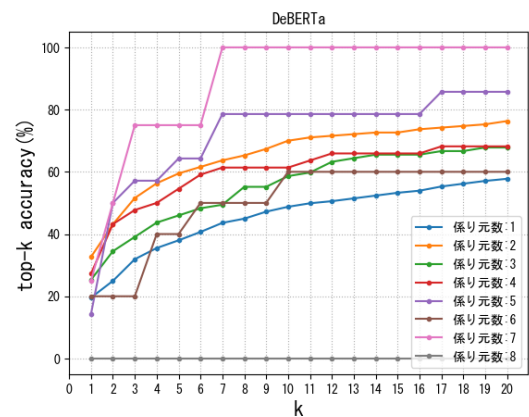


図7 係り元文節数による正解率の変化 (DeBERTa)

4.3.2 係り元文節数による正解率の比較

予測対象の係り先文節に係る文節数 (係り元文節数) による正解率の変化を図7に示す。この図は、係り元文節数により予測対象を分類し、その数が1から8となるものの正解率をそれぞれ示している。既入力文節列に含まれる係り元文節の数が少ないと正解率は低下し、係り元文節の数が多いと正解率は比較的高い値を示すことを確認できる。係り先文節の主辞を予測する際に、係り受け関係にある文節の情報は有益であると考えられ、その数が多いほど、予測に有利に働くといえる。

5 おわりに

本論文では、補完応答の生成に向けて、話し手の既発話文節と係り受け関係にある未発話文節を補完対象として予測する手法を提案した。実験の結果、本手法により一定の水準で補完対象を予測可能であることを確認した。今後は、人間が本手法と同じ作業を行ったときに、どの程度の正解率を示すのか調査し、比較したい。

謝辞

高齢者のナラティブコーパスは、奈良先端科学技術大学院大学ソーシャル・コンピューティング研究室から提供いただいた。本研究は、一部、科学研究費補助金基盤研究(C) No. 19K12127により実施した。

参考文献

- [1] Koichiro Ito, Masaki Murata, Tomohiro Ohno, and Shigeki Matsubara. Construction of responsive utterance corpus for attentive listening response production. In **Proceedings of the 13th Language Resources and Evaluation Conference**, p. 7244–7252, 2022.
- [2] Nigel G. Ward and Wataru Tsukahara. Prosodic features which cue back-channel responses in English and Japanese. **Journal of Pragmatics**, Vol. 32, pp. 1177–1207, 2000.
- [3] Jin Yea Jang, San Kim, Minyoung Jung, Saim Shin, and Gahgene Gweon. BPM_MT: Enhanced backchannel prediction model using multi-task learning. In **Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing**, pp. 3447–3452, 2021.
- [4] Wencke Liermann, Yo-Han Park, Yong-Seok Choi, and Kong Lee. Dialogue act-aided backchannel prediction using multi-task learning. In **Findings of the Association for Computational Linguistics: EMNLP 2023**, pp. 15073–15079, 2023.
- [5] 心が感じられる音声対話システム “heartalk™”. <http://www.y2lab.com/project/heartalk/>, 2016.
- [6] Eiji Aramaki. Japanese elder’s language index corpus v2. https://figshare.com/articles/dataset/Japanese_Elder_s_Language_Index_Corpus_v2/2082706, 2016.
- [7] 村田匡輝, 大野誠寛, 松原茂樹. 未入力文節との構文的関係を考慮する漸進的な係り受け解析. 言語処理学会第20回年次大会発表論文集, pp. 193–196, 2014.
- [8] 相津徹也, 大野誠寛, 松原茂樹. 漸進的係り受け解析における未入力文節との構文的関係の同定. 情報処理学会第82回全国大会講演論文集, pp. 441–442, 2020.
- [9] 橋本優希, 大野誠寛, 松原茂樹. 漸進的係り受け解析における BERT を用いた未入力文節との構文的関係の同定. 情報処理学会第85回全国大会講演論文集, No. 2, pp. 803–804, 2023.
- [10] 加藤芳秀, 松原茂樹, 外山勝彦, 稲垣康善. 確率文脈自由文法に基づく漸進的構文解析. 電気学会論文誌 C (電子・情報・システム部門誌), Vol. 122, No. 12, pp. 2109–2119, 2002.
- [11] 大野誠寛, 松原茂樹, 柏岡秀紀, 加藤直人, 稲垣康善. 節境界単位での漸進的な独話係り受け解析. 情報処理学会研究報告, 音声言語情報処理 (SLP), No. 12, pp. 53–58, 2005.
- [12] 大野誠寛, 松原茂樹. 文節間の依存・非依存を同定する漸進的係り受け解析. 電子情報通信学会論文誌, Vol. J98-D, No. 4, pp. 709–718, 2015.
- [13] 内元清貴, 関根聡, 井佐原均. 最大エントロピー法に基づくモデルを用いた日本語係り受け解析. 情報処理学会論文誌, Vol. 40, No. 9, pp. 3397–3407, 1999.
- [14] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**, pp. 4171–4186, 2019.
- [15] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. RoBERTa: A robustly optimized BERT pretraining approach, 2019.
- [16] Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. DeBERTa: decoding-enhanced BERT with disentangled attention. In **Proceedings of the 9th International Conference Learning Representations**, pp. 1–21, 2021.
- [17] Shigeki Matsubara, Akira Takagi, Nobuo Kawaguchi, and Yasuyoshi Inagaki. Bilingual spoken monologue corpus for simultaneous machine interpretation research. In **Proceedings of the 3rd International Conference on Language Resources and Evaluation**, pp. 153–159, 2002.
- [18] Arseny Tolmachev, Daisuke Kawahara, and Sadao Kurohashi. Juman++: A morphological analysis toolkit for scriptio continua. In **Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations**, pp. 54–59, 2018.