

# 人間と共同でキャッチコピーを作成する対話システムの試作

## A prototype dialogue system for co-writing taglines with users

周旭琳\* 市川拓茉 東中竜一郎  
Xulin Zhou Takuma Ichikawa Ryuichiro Higashinaka

名古屋大学大学院情報学研究科  
Graduate School of Informatics, Nagoya University

**Abstract:** In this study, we developed a prototype system for co-writing taglines with users. Specifically, the system combines utterance generation, tagline generation, and automatic tagline evaluation modules created by using a large-scale language model and template-based utterance generation. We report the collaborative work currently possible with this prototype system.

### 1 はじめに

近年、対話システムが普及している。しかし、従来の対話システムでは人間の仲間として協働し新たな価値を生み出すことはほとんど考えられておらず、人間と創造的な共同作業を行う対話システムの研究はいまだ少ない。人間同士が協力し合い新しい価値を創造するように、対話システムが人間と創造的な協働作業を行うことができれば、対話システムの可能性がより広がり、多くの領域で使われることにつながると考えられる。

共同作業を行う対話システムに関する研究には、システムがユーザの要求に応じて種々の情報の提供や表示を行い、交通手段の計画を策定する TRIPS [1]、ビデオオカセットレコーダやガスタービンシミュレータなどの操作においてユーザを補助する COLLAGEN [2]、画像に描画されている色や大きさの異なる点の配置から、同じ点を二人で対話を通して特定する OneCommon [3] に取り組む対話システム [4] などがある。創造的な共同作業を人間と行う対話システムに関する関連研究には、マイクラフトにおいて二者が話し合いながら新しい構造物を作るタスクを題材とした研究がある [5]。

本研究では、創造的な共同作業が可能な対話システムの実現に向けて、人と共同でキャッチコピーを作るシステムのプロトタイプを作成を行った。システムのプロトタイプは、キャッチコピー共同作成タスク（次節参照）を用いて収集したデータで学習されたモデルを、ルールに基づく選択やテンプレートベース発話生成と組み合わせることで構築した。そして、その動作を確認するとともに、ユーザ評価により定性的な評価を行った。

### 2 キャッチコピー共同作成タスク

我々は、創造的な共同作業としてキャッチコピー共同作成タスクを設定し、人間同士の共同作業のデータを収集した [6]。<sup>1</sup>キャッチコピー共同作成タスクでは、提示される両作業員共通の商品説明を参考に、テキストチャットを用いた対話を行いながら、作業相手とテキストボックスを共同編集しキャッチコピーを作成する。本研究では 782 対話のデータを使用する。

データ収集はクラウドソーシングを用い、オンラインで実施した。収集されるデータには、キャッチコピーの編集内容およびチャットの内容があり、編集・送信を行った作業員およびタイムスタンプも記録した。1 回の作業は 30 分とし、作業員ペアは 1 回の作業の間で少なくとも 3 つ以上のキャッチコピーを作成した。作業員にはユニークで印象に残るキャッチコピーを作成するように指示した。

図 1 はデータ収集時に使用したインタフェースの画面である。画面中段には、キャッチコピーの対象となる商品の商品名、商品説明、商品画像が表示されている。商品は 100 種使用した。<sup>2</sup>画面左下のキャッチコピー編集欄は、作業員が思いついたキャッチコピーに使えるような単語やフレーズ、キャッチコピーの案を書き込む場所であり、両作業員が共有・編集する。キャッチコピー編集欄にはテキストボックスが 8 つあり、それぞれに A~H のアルファベットが付与されている。キャッチコピー編集欄のテキストボックス内に文字を入力すると、その変更が作業相手の見ている画面にも即座に反映される。画面右下には、作業員が対話を行う際に使用するチャット欄がある。表 1 に 782 対話の統計情報を示す。これらの統計情報は作業開始時点から 30 分の制限

\*連絡先：名古屋大学大学院情報学研究科  
〒464-8601 名古屋市中種区不老町  
E-mail:zhou.xulin.j3@s.mail.nagoya-u.ac.jp

<sup>1</sup>文献 [6] は 300 対話収集した時点での報告である。

<sup>2</sup>商品の情報は楽天商品検索 API を用いて取得した。

ルームID: Baxo ユーザID: j2wPDI


**タスクの説明**

- テキストチャットで相手の方に挨拶してください。次に、商品説明を**すべて読んでください**。読み終わったらその旨を相手に伝えてください。
- タイマーをスタートした後**、相手の方と2人で話し合い、キャッチコピー編集欄を書き換えながら、ユニークで印象に残るキャッチコピーを3つ以上作成してください。制限時間が過ぎたら、きりのよいところで会話を終了してください。
- その後、**終了後アンケートに回答してください**。

**Already Finished**

---

**商品説明**



**商品名: サマンサタバサ バスケース**  
Samantha Thavasa Petit Choice (サマンサタバサプチチョイス) のフルール バスケースソフトなベロア素材とレザーを組み合わせ、切り返しのリボンベルトがポイント。優しいダスティバステルのカラー展開に。ロゴ金具にはさりげないハートをプラス。素材違いとさりげない部分の可愛らしさを楽しんでみてはいかがでしょうか?

**キャッチコピー編集欄**

A	THE女の子になりたい人、必携!	16
B	映えるためのバスケース	11
C	大人になっても、お嬢様になってみませんか?	21
D	初めてのデートに 可愛いワタシ、連れて行こう。	23
E	持てば、キラキラ プリンセス	14
F	初めてのバイト代で、「小さい頃の夢」を手に入れた。	25
G	甘いだけじゃない! 私スタイルのバスケース	20
H		00

[14:53:51] (あなた): 思いつかなければ、今までのを見直す感じでも  
[14:55:21] (お相手): そうですね。ちょっとEのたまちって言葉が可愛く感じられないんですよね。  
残り時間は5分です。制限時間が終了するまで、作業を続けてください。  
[14:55:29] (あなた): ふむふむ、語感いい感じですけどね...ちょっと他の言葉探してみましようか!  
[14:57:04] (お相手): キラキラ?  
[14:57:06] (あなた): うんうん、なんかシンデレラが魔法にかけられてドレス着るイメージが湧きました~  
[14:58:21] (お相手): こっちの方がよさそうですね。ありがとうございます!  
[14:58:29] (あなた): こちらこそです~  
[14:59:39] (お相手): なんか、いい感じですね。けっこう、まとまって良かったです。  
[14:59:34] (あなた): ですね!  
制限時間が経過しました。確認コードは「75GDGe」です。  
キャッチコピーが少なくとも3つ完成していない場合は、作業を続けてもかまいません。  
[15:00:48] (お相手): 私はこれで満足です!  
[15:00:31] (あなた): 同じくです! アンケートに移りましょうか

作業終了後、次のURLからアンケートにお答えください。 [アンケートはこちら](#)

図 1: データ収集に使用したインタフェース

表 1: 収集した共同作業データの統計情報

総作業数	105
総作業ペア数	398
総対話数	782
<b>チャット</b>	
対話の平均発話数	46.91
発話の平均文字数	20.31
<b>キャッチコピー編集欄</b>	
対話で記入されている欄の数の平均	7.60
対話の書き込み文字数の平均	235.76
対話の削除文字数の平均	98.30

表 2: キャッチコピーの第三者評価の結果 (7段階)

質問項目	平均	標準偏差
Q1. 興味を引く	4.35	1.30
Q2. 想像を膨らませる	4.20	1.41
Q3. ありふれている	4.20	1.54
Q4. 特徴をふまえている	4.58	1.46

(1) ~ 「とてもそう思う」(7) の7段階のリッカート尺度を用いた。各キャッチコピーに対する評価者は5名である。表2は評価結果の平均および標準偏差である。

時間が経過した時点までのデータを対象とした。

さらに、キャッチコピー共同作成作業で作成されたキャッチコピーそれぞれについて、第三者による評価を収集した。評価は、Q1「このキャッチコピーは目にした人の興味をを引くと思いますか」(興味を引く)、Q2「このキャッチコピーは見る人の想像をを膨らませることができると思いますか」(想像を膨らませる)、Q3「あなたはこのキャッチコピーをありふれていると思いますか」(ありふれている)の3項目に評価作業の質を担保するための質問としてQ4「このキャッチコピーは商品の特徴をふまえているをふまえていますか」(特徴をふまえている)という質問項目を加えた合計4項目について収集した。収集には「とてもそう思わない」

### 3 システムの試作

収集したデータを用いて大規模言語モデルをファインチューニングすることで、キャッチコピー生成モデル、発話生成モデル、キャッチコピー評価モデルを作成した。それらと収集したデータを参考に作成したテンプレートによる発話生成を組み合わせ、システムを試作した。データ収集時はいつでも編集や発話を行うことができたが、このシステムではターン制とした。

図2はシステムのプロトタイプのGUIである。共同作業開始時は左上のテキストボックスに商品名および商品説明を入力し確定ボタンを押す。左下には履歴の欄があり、各ターンの発話とキャッチコピー編集欄の状態が表示される。右側にはキャッチコピー編集欄と

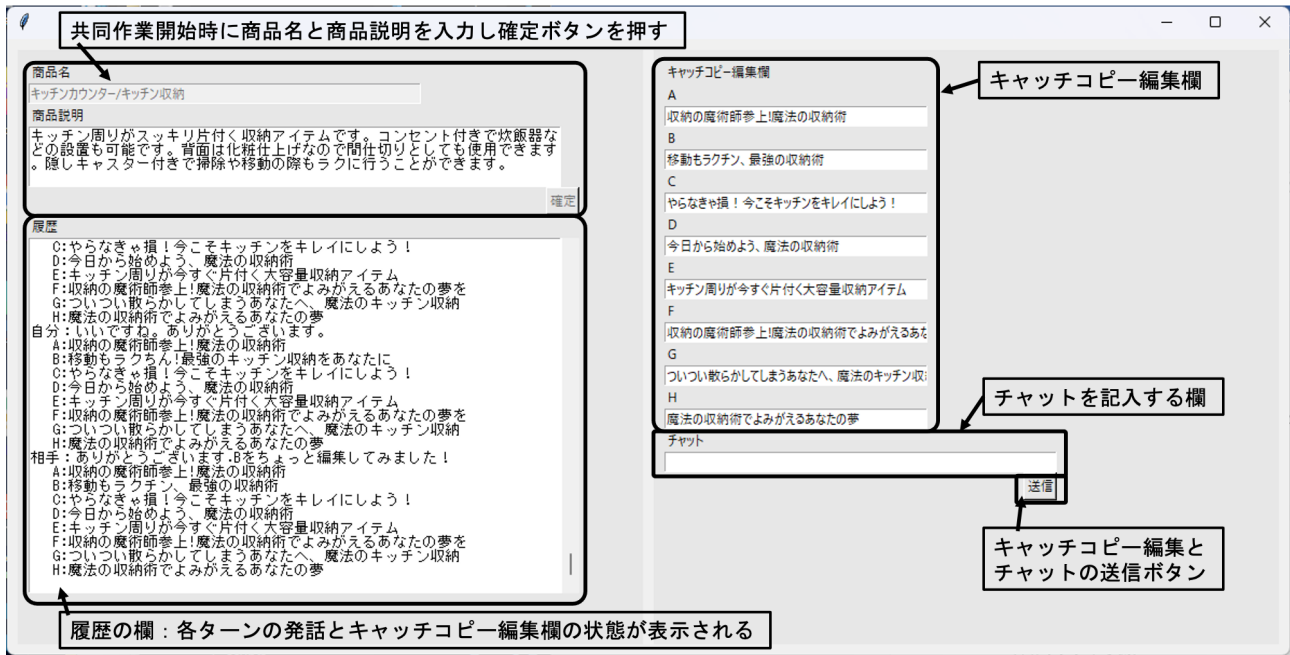


図 2: 試作システムの GUI

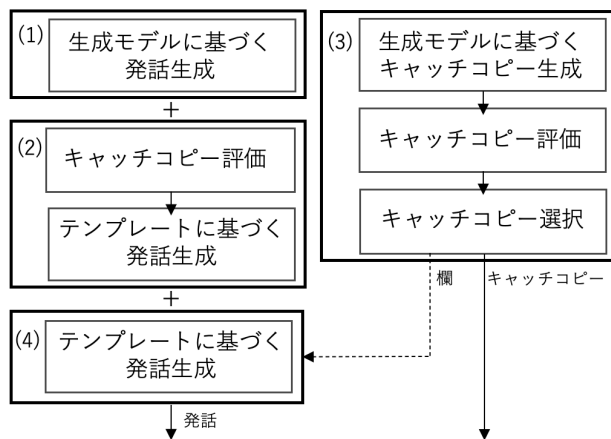


図 3: システムの処理の流れ

チャットを記入する欄があり、送信ボタンを押すことでそれらが送信される。GUIの作成にはTkinter<sup>3</sup>を利用した。

### 3.1 処理の流れ

図3はシステムの大まかな処理の流れを示している。ユーザがキャッチコピー編集欄の編集やチャットの入力を行い、送信ボタンを押すと、システムにターンが渡される。システムのターンでは、システムは基本的に以下の処理を経てユーザに編集後のキャッチコピーおよび発話を返す。

- (1) 商品情報、対話履歴、キャッチコピー編集欄の状態を入力とする生成モデルが発話を生成する。
- (2) ユーザがキャッチコピー編集欄を編集した場合、キャッチコピー評価モデルを使用し、編集されたキャッチコピーの評価を行う。評価値が一定以上であるならば、ユーザの編集への反応を返すテンプレートにより、発話を生成する。
- (3) 生成モデルでキャッチコピーの候補を複数生成し、それぞれの候補を(2)で使用したものと同一キャッチコピー評価モデルを使用して評価する。評価値やキャッチコピー編集欄の履歴をふまえ、生成した候補のうち1つを、実際に編集を行うキャッチコピーとして選択する。
- (4) 所定の確率で、どのキャッチコピー編集欄を編集したかユーザに伝える発話をテンプレートにより生成する。
- (5) (1), (2), (4) を連結した発話、(3) で選択されたキャッチコピーを表示画面に反映する。

### 3.2 作成したモデル

ここでは、システムの処理で使用する各モジュールおよびモデルの学習や評価について述べる。データを train 80%, validation 10%, test 10%に分割し、発話生成モデル、キャッチコピー生成モデル、キャッチコピー評価モデルの学習および評価に用いた。ここで、同じ

<sup>3</sup><https://docs.python.org/ja/3/library/tkinter.html>

商品に関するデータは, train, validation, test のいずれか一つにのみ割り当てられるよう分割した。

### 3.2.1 発話生成モデル

商品名, 商品説明文, ある時点のキャッチコピー編集欄の状態, ある時点の対話履歴を入力とし, 次発話を出力するようにモデルのファインチューニングを行った。GPT2<sup>4</sup>[7] を事前学習済みモデルとして利用した。発話のデータおよびキャッチコピー編集をタイムスタンプの順に並べ, 作業員, 発話かキャッチコピー編集かの種別, 編集対象のキャッチコピー編集欄のいずれかが変化している箇所でのデータの分割を行った。分割箇所より前の行動が入力に含まれる。分割箇所後の最初の行動が発話であるデータのみ用い, その発話を出力の正解データとした。BLEU (unigram) [8] による評価値は 0.107, perplexity は 13.459 であった。

### 3.2.2 キャッチコピー生成モデル

キャッチコピーの自動生成に関する研究は盛んに行われている [9, 10, 11]。本研究では, 発話生成モデルと同様の入力形式で, 次発話ではなく次のキャッチコピー編集を出力するようにファインチューニングを行った。分割箇所後の最初の行動がキャッチコピー編集であるデータのみ用い, その編集後のキャッチコピーを出力の正解データとした。BLEU (unigram) による評価値は 0.268, perplexity は 6.29 であった。

### 3.2.3 キャッチコピー評価モデル

キャッチコピー評価モデルは, BERT モデル<sup>5</sup>を使用し, キャッチコピーとその評価のペアのデータを利用して評価の2クラス分類を行うようファインチューニングすることで作成した。2クラス分類では1~4をラベル0, 5~7をラベル1とした。評価モデルはQ1 (興味を引く), Q2 (想像を膨らませる), Q3 (ありふれている) の評価項目に基づくものを別々に作成した。キャッチコピー評価モデルを test データで評価した正解率は, Q1 は 59.3%, Q2 は 58.6%, Q3 は 62.2% であった。テストデータにおいて多数派のラベルが占める割合は, Q1 は 58.4%, Q2 は 50.6%, Q3 は 51.5% である。キャッチコピーの評価としては, ラベルの予測値をソフトマックス関数に入力して得た確率を利用した。今回はありふれていないキャッチコピーを高く評価しつつある程度の質が保たれたキャッチコピーを作成させたいと考え, Q1, Q2 のモデルはラベル1になる確率を, Q3 はラベ

表 3: 作業員とプロトタイプシステムとの対話。U はユーザ, S はシステムを表す。キャッチコピー生成対象の商品はキッチンでの使用を想定した棚である。背景色が灰色の箇所は編集された欄のアルファベットとその内容である。

U1	何に注目してキャッチコピーを作りますか?
S1	収納、掃除、移動、掃除の5つのことを軸に考えてみます～ A: 収納の魔術師
U2	いいですね
S2	ですねー あとは、写真が白くて清潔感があるのと、おしゃれなのもポイントになってくるのかな。B書いてみました。 B: おうちの
U3	コンセント付きなので、それを強みとしてキャッチコピーを作りました。c書いてみました。 C: 多機能収納アイテム
S3	はい、ありがとうございます! B: おうちの魔法の収納
U4	キッチン収納なので、キャッチコピーにキッチンと入れてみませんか。
S4	はい!D書いてみましたが、どうでしょうか? D: 今日から、魔法の収納術
U5	Dを変更してみました。どうですか。 D: 今日から始めるキッチン整頓
S5	ありがとうございます!B どうでしょうか?ちょっと変更してみました B: おうちの魔法の収納術
U6	A からDのキャッチコピーを参考にして、新しくEを考えてみました。 E: キッチン周りが今すぐ片付く大容量収納アイテム
S6	了解です!D、変えてみました。 D: 今日から始めよう、多機能収納術
U7	Dはかなりいい感じになったと思います。
S7	そうなのですね!よかったです!ありがとうございます!そうですね,d,cの2つがいい感じだと思います.Fに新しく書いてみました F:i've will be
U8	Fはどういう意味ですか?
S8	私も気になります.F どうでしょうか?ちょっと変更してみました F:i've will be ~ あなたの夢を実現化させます ~
U9	Fは抽象的なので具体的なキャッチコピーに変更してみませんか。
S9	了解です!他の案もぜひぜひお願いします!F どうでしょうか?ちょっと変更してみました F:i've will be ~ あなたの夢を実現
...	...
U11	Bはシンプルになっていいと思います。今までの案をまとめてHに新しくキャッチコピーを書いてもらえますか?
S11	はい!大丈夫ですよ! F: 魔法の収納術でよみがえるあなたの夢を実現
...	...
U14	Cを大幅に変更しました。何か加えた方がいい要素はありますか? C: キッチンをキレイにしよう!
S14	そうですね、cはもう少しインパクトのある方がいいかなーと思います! B: 魔法の収納術でよみがえるあなたの夢を実現!
U15	なるほど。インパクトのあるように変更してみます。 C: やらなきゃ損!今こそキッチンをキレイにしよう!
S15	いい感じですよ!C、言葉の使い方が素敵です! H: 魔法の収納術でよみがえるあなたの夢
...	...

<sup>4</sup><https://huggingface.co/rinna/japanese-gpt2-medium>

<sup>5</sup><https://huggingface.co/cl-tohoku/bert-base-japanese-whole-word-masking>

ル0になる確率を評価値とし、それらの評価値を1:1:2の比率で足し合わせ0~1に正規化した値を用いた。

### 3.3 対話例および所感

表3に、著者ではない作業者が本システムを利用したときの対話例を示す。

作業者からは、「欄が埋まりはじめてからは、要求しないと新しい案を相手側があまり出してくれなくなった」、「相手が考える系の質問をすると、不自然な回答になるときがあった(S8, S8)。ちゃんと返ってくるときもあった(S1, S14)」、「急に「i've will be」のようなよくわからないものが出てきた(S7)」、「チャットはきちんとしているが記入欄は別の欄が更新されている(S11)」、「序盤で軸を提示してもらったのはやりやすかった(S1)」、「システムが出してくるキャッチコピーに「魔法」が多い」、「自分が入れたキャッチコピーと同じような文言を他のキャッチコピーに入れてくるような動作があった(U3, S6)」などの意見を得た。

指摘された点に加え、軸の列挙の際に「掃除」を複数回含め、列挙している軸の数を「5つ」と間違えている(S1)など発話生成の質や、商品情報として画像を使用していないのに写真に言及し誤った情報を含んだ発話を行っている(S2)ことが挙げられる。より性能の高い発話生成や編集が可能なモデルの作成が必要である。

## 4 おわりに

本研究では、人間同士のキャッチコピー共同作成作業データによって大規模言語モデルをファインチューニングし作成したモデルを用いて、人と共同でキャッチコピーを作るシステムを試作し、定性的な評価を行った。その結果、システムの発話やキャッチコピーに誤った情報が含まれる問題や、ユーザの質問にうまく応答できていない問題、発話と編集行動の矛盾が起こる問題があるとわかった。今後は、より性能の高い発話生成や編集が可能なモデルの作成に取り組みたい。

## 参考文献

- [1] George Ferguson, James F Allen, et al. TRIPS: An integrated intelligent problem-solving assistant. In *Proceedings of Association for the Advancement of Artificial Intelligence*, pp. 567–572, 1998.
- [2] Charles Rich and Candace L Sidner. COLLAGEN: A collaboration manager for software interface agents. *Computational models of mixed-initiative interaction*, pp. 149–184, 1999.
- [3] Takuma Udagawa and Akiko Aizawa. Maintaining common ground in dynamic environments. *Transactions of the Association for Computational Linguistics*, Vol. 9, pp. 995–1011, 2021.
- [4] Daniel Fried, Justin Chiu, and Dan Klein. Reference-centric models for grounded collaborative dialogue. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 2130–2147, 2021.
- [5] Takuma Ichikawa and Ryuichiro Higashinaka. Analysis of dialogue in human-human collaboration in minecraft. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pp. 4051–4059, 2022.
- [6] 周旭琳, 市川拓茉, 東中竜一郎. キャッチコピー共同作成タスクにおける対話の収集と分析. 人工知能学会第36回全国大会論文集, pp. 2A6GS603–2A6GS603, 2022.
- [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] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pp. 311–318, 2002.
- [9] 梅村奏子, 狩野芳伸. メタファーの自動生成による意味的な重ね合わせのあるキャッチコピー生成器の構築. 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. 91, pp. 24–30, 2021.
- [10] A. Ray, P. Agarwal, C. K. Maurya, and G. B. Dasgupta. Creative tagline generation framework for product advertisement. *IBM J. Res. Dev.*, Vol. 63, No. 1, pp. 6:1–6:10, 2019.
- [11] Gözde Özbal, Daniele Pighin, and Carlo Strapparava. BRAINSUP: Brainstorming support for creative sentence generation. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1446–1455, 2013.