

評価用データセットを用いた口調エンコーダの性能評価と改良

川北雄大^{1,2} 小川 浩平¹ 佐藤 理史¹¹ 名古屋大学大学院工学研究科² kawakita.yudai.w4@s.mail.nagoya-u.ac.jp

概要

口調エンコーダとは、セリフを口調の特徴を埋め込んだベクトル(口調ベクトル)に変換する変換器である。口調ベクトルは、小説のセリフの話者同定タスクにおいてその有効性が示されているが、その能力は十分には解明されていない。本研究では、口調ベクトルの口調弁別性能を評価するためのデータセットを作成し、既存の口調エンコーダの性能を評価した。その結果に基づき、学習データの増強・再編成を行い、性能向上を確認した。

1 はじめに

日本語の話し言葉には多様な口調(話し方のスタイル)が存在する。口調は主にエンターテインメント小説などで、登場人物の性格等を際立たせたり、セリフの話者を暗黙的に伝えたりすることを可能とする[1]。本研究では、セリフの「口調」を、セリフの表記に現れるスタイル的特徴を包括する概念と捉える。

石川ら[2]は、小説のセリフを、そのセリフの口調の特徴を埋め込んだベクトル(口調ベクトル)に変換する方法(口調エンコーダ)を提案するとともに、口調ベクトルを使用して、セリフの話者同定を実現する方法を示した。2種類の口調タイプを弁別できるかどうかを調査した後続研究[3]では、口調エンコーダの学習に使用していない口調タイプであっても、もう一方の口調タイプと特徴が大きく異なれば、多くの場合、弁別が可能であることが示された。

本研究では、口調エンコーダの口調弁別性能の向上を目的として、次のことを行った。

- 20種類の典型的な口調タイプを設定し、これらの口調タイプの弁別性能を評価するためのデータセットを作成して、既存の口調エンコーダを評価した。このデータセットは、600セリフ(20タイプ×30セリフ)から構成されている。

- 典型的な口調タイプに基づいて、口調エンコーダの学習データを増強・再編成することで、既存の口調エンコーダより性能が高い口調エンコーダを実現した。

2 関連研究

小説等のセリフの口調に大きな影響を与えると考えられる要素のひとつに「役割語」がある。役割語とは、話者の特定の人物像を想起させる特定の言葉遣いを指す[4]。金水編の『〈役割語〉小辞典』[5]では、役割語を「性差、年齢・世代、職業・階層、地域、時代、人間以外」の6種類に分けて整理し、総計42種類の役割語ラベルを設定している。役割語は、主にセリフの語尾に現れる特定の語彙に着目している。

メイナード[6]は「あるキャラクターに限って一貫して使われるカラフルな話のスタイル」を「キャラ語」と名付けた。そして、「キャラ語にはコミュニケーションのありとあらゆるレベルの操作が含まれる」と説明する。キャラ語は役割語を包含する概念であり、本論文の「口調」は、概念としては、このキャラ語にほぼ対応する。

口調のベクトル化を志向した研究として夏目らの研究[7]がある。この研究ではベクトルを手で付与している。また、セリフを口調ベクトルに自動変換する研究として、銭本らの研究[8]がある。この研究では、対照学習を用いて口調の特徴を埋め込んだベクトルを学習し、そのベクトルから口調の類似性を評価している。しかし、その口調弁別性能は十分には示されていない。

3 口調エンコーダ

口調エンコーダは、セリフを口調の特徴を埋め込んだベクトル(口調ベクトル)に変換する変換器である[2]。口調エンコーダの構成を図1に示す。

口調エンコーダは、次の2つのモジュールから構

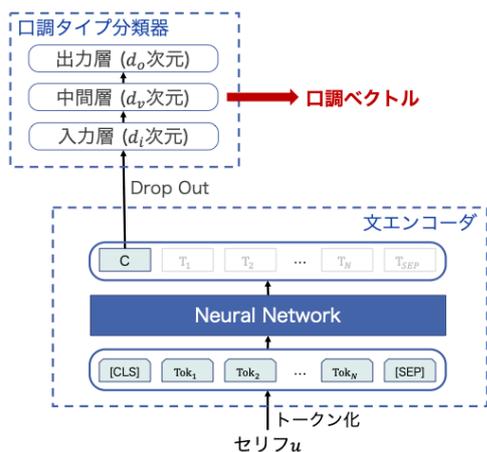


図1 口調エンコーダの構成

成されている。

1. 文エンコーダ：セリフを文埋め込みに変換する
2. 口調タイプ分類器：文埋め込みを入力とし、セリフの口調タイプを分類する

口調エンコーダの学習には、セリフと口調タイプの組からなる学習データを用い、口調タイプ分類器と文エンコーダのパラメータを同時に学習する。口調ベクトルの作成では、口調エンコーダにセリフを入力して動作させ、口調タイプ分類器の中間層(出力層の1層前)の値を口調ベクトルとして採用する。

口調エンコーダの実装には多くのバリエーションが考えられるが、以前の調査 [3] で口調弁別性能が最も高かったのは、次のモデルである。

- 文エンコーダ：日本語 wikipedia で事前学習済の BERT[9] に、「なろうセリフデータ」¹⁾を用いて継続事前学習を行ったモデル
- 口調タイプ分類器： $d_i=768$ 次元， $d_v=64$ 次元， $d_o=10$ 次元
- 学習データ
 - 口調タイプ数： $d_o=10$ 種類
 - 総セリフ数：16,705 セリフ
 - セリフを抽出した作品数：7 作品
 - セリフを抽出したキャラクター数：50 名

本研究では、この口調エンコーダの性能を、次節で述べる評価用データセットを用いて評価する。以降、この口調エンコーダを E_B と表記する。

1) 小説投稿サイトである「小説家になろう」から収集した 5,297,896 セリフ

4 評価用データセットの作成

ライトノベル等の小説において多くの口調が観察されるが、口調のタイプ分類は自明ではなく、確立されたタイプ分類も存在しない。

本研究では、「ピクシブ百科事典」²⁾等のインターネット上の各種情報を参考に、以前の 19 種類の口調タイプ [3] を再編成し、第 1 著者の主観に基づき 20 種類の典型的な口調タイプを設定した。ここで典型的な口調とは、多くのエンターテインメント作品で観察される口調のことである。20 タイプのうち 9 タイプは、口調エンコーダの学習 [2] で用いた 10 タイプから「女性-てよだわ-乱暴」を除外した 9 タイプに対応する。残りの 11 タイプのうち 8 タイプは、以前の 19 タイプで導入した口調タイプに対応するタイプが存在する³⁾。実質的に新たに設定した口調タイプは「男性的-関西弁」、「女性的-乱暴」、「寡黙」の 3 種類である⁴⁾。

まず、それぞれの口調タイプに対して、小説等のキャラクターを一人定めた。ここで設定した 20 名のキャラクターは、口調エンコーダの学習データで利用したキャラクター 50 名とは完全に異なる⁵⁾。

次に、20 名のキャラクターそれぞれに対して、出典作品から 30 セリフを収集した。セリフの収集では、長さが 10 文字以上 100 文字未満⁶⁾で、口調タイプの特徴が比較的よく現れていると思われるセリフを第 1 著者が選んだ。

以上の手順で作成した口調弁別評価データセットの概要を表 1 に示す。タイプ名をゴシック体で表記したものは、「口調エンコーダ E_B の学習データに含まれない口調タイプ」である。各口調タイプのセリフ例を付録の表 6 に示す。

5 口調エンコーダ E_B の評価

本節では、前述の口調エンコーダ E_B の口調弁別性能を調査する。具体的には、評価用データセットに含まれる 2 つの口調タイプ i, j のペアに対して、以下の手順で口調弁別性能を調査する。

1. それぞれの口調タイプの 30 セリフを 15 セリフずつに分割し、一方の 15 セリフを用いて、代表

- 2) <https://dic.pixiv.net>
- 3) 以前のタイプと名前が違うものも存在する。例：「執事口調」→「慇懃」
- 4) 以前の 19 タイプのうち、「こども」「猫言葉」「京ことば」「中華風」は採用しない。
- 5) 作品は一部重複する。
- 6) 文字数は、前後の引用記号を含む。

表1 口調弁別評価用データセットの20種類の口調タイプ

ID	口調タイプ	キャラクタ	出典作品	セリフ長(文字数)			[2]の口調タイプとの対応
				中央値	最短	最長	
1	男性的-敬語	フレデリク	『ファイアーエムブレム 覚醒』(ゲーム)	34.5	14	61	男性-敬語
2	男性的-優しい	福部里志	『氷菓』(小説)	32.5	14	69	男性的-優しい
3	男性的-普通	スバル	『かがみの孤城』(小説)	29.5	12	93	男性-普通
4	男性的-関西弁	塚本秀一	『響け!ユーフォニアム』(小説)	29	13	73	
5	男性的-乱暴	ナツキ-スバル	『Re: ゼロから始める異世界生活』(小説)	31.5	13	79	男性-乱暴
6	男性的-厳格	提督	『鶴翼の絆』(小説)	32.5	12	71	
7	女性的-敬語	千反田える	『氷菓』(小説)	24.5	12	61	女性-敬語
8	ボーイッシュ	東郷あい	『アイドルマスター シンデレラガールズ スターライトステージ』(ゲーム)	29.5	22	79	
9	女性的-普通	アキ	『かがみの孤城』(小説)	30	12	94	女性-普通
10	女性的-関西弁	難波笑美	『アイドルマスター シンデレラガールズ スターライトステージ』(ゲーム)	32	24	44	女性-関西弁
11	女性的-乱暴	フェルト	『Re: ゼロから始める異世界生活』(小説)	33.5	14	71	
12	女性的-厳格	ダクネス	『この素晴らしい世界に祝福を!』(小説)	47	13	96	女性-男性っぽい
13	女性的-てよだわ	クラミー・ツェル	『ノーゲーム・ノーライフ』(小説)	29.5	11	94	女性-てよだわ言葉
14	お嬢様	ステファニー・ドーラ	『ノーゲーム・ノーライフ』(小説)	30.5	12	76	
15	～なの	ピエリ	『ファイアーエムブレム if』(ゲーム)	36.5	13	78	
16	老人口調	ロム爺	『ノーゲーム・ノーライフ』(小説)	36	13	77	
17	～っス	荒木比奈	『アイドルマスター シンデレラガールズ スターライトステージ』(ゲーム)	31	20	48	
18	武士口調	緋村剣心	『るろうに剣心』(マンガ)	36.5	18	80	
19	慥懃	影山	『謎解きはディナーのあとで』(小説)	29	13	69	
20	寡黙	和泉紗霧	『エロマンガ先生』(小説)	27.5	13	58	
				32	11	96	

口調ベクトル(口調ベクトルの平均)を求め、他方の15セリフをテストデータとする

- テストデータ30(=15×2タイプ)セリフに対して、口調タイプ判定(2クラス分類)を行う。具体的には、セリフの口調ベクトルと2つの代表口調ベクトルの距離を計算し、距離が小さい方を推定結果とする
- 以上の手続きを代表口調ベクトルの算出に使用する15セリフとテストデータの15セリフを交換して2回実施し、計60(=30×2)セリフに対して口調タイプ判定を行う

20の口調タイプから2つの口調タイプを選ぶ組み合わせは190(=20×19/2)通り存在し、その全てに対して上記を実行する。これらをすべて合わせた総計11,400(=60×190)件に対する精度を全体精度と呼び、この値を口調エンコーダの性能を表す指標として用いる。なお、このタスクは2クラス分類問題なので、チャンスレベルは0.5である。

評価結果を表2に示す。この表の平均は、シード値を変えて学習した5つの口調エンコーダの全体精度の平均値である。この表に示すように、口調エン

表2 E_B の評価

シード	42	43	44	45	46	平均
E_B	0.905	0.903	0.899	0.917	0.909	0.906

コーダは平均値で9割のセリフの口調を正しく判定できていることがわかる。

しかしながら、弁別精度が低い口調タイプペアも存在する。以下に例を示す。

- ・「女性的-乱暴(11)」と「老人口調(16)」:0.550
- ・「女性的-厳格(12)」と「武士口調(18)」:0.583
- ・「男性的-乱暴(3)」と「老人口調(16)」:0.767

これらの口調ペアは、我々日本人であれば比較的容易に弁別可能である。口調エンコーダ E_B がこれらの口調ペアをうまく弁別できないのは学習データに口調タイプが不足しているためと考えられる。

6 学習データの再編成

前述の評価結果に基づき、評価用データセットの20口調タイプに対してそれぞれ学習データを準備し、口調エンコーダを再学習する。変更するのは学習データと口調タイプ分類器の出力層の次元数 d_o のみである。

表3 新たに作成した学習データ

タイプ	数	判定基準
女性的-乱暴	907	一人称、語彙が女性的であり、「～だぜ」など乱暴な表現を含む
男性的-厳格	1000	男性的な語彙を用い、「～だな」「～だぞ」など固い表現を多く用いる
ボーイッシュ	1001	文末形式が「～さ」「～だね」など中性的な話し方をする
男性的-関西弁	1003	男性的な語彙を用い、「～やねん」などいわゆる関西弁を用いる
武士口調	1161	「ござる」の活用形を文末形式として含む
老人口調	1000	「～じゃ」「～わい」などいわゆる老人語を用いる
お嬢様	1000	「ですわ」「ですの」などの表現を用いる
慇懃	1053	「ごぞいます」「お嬢様」「であります」など特定の表現を含む
～っス	1013	文末形式として「っす」「っス」「っす」「っス」のいずれかを含む
～なの	1000	「の」「なの」という文末形式を含み、幼い表現が多い
寡黙	1001	「…」や体言止めを用いる

20の口調タイプのうち9タイプの学習データは、以前のもの[2]をそのまま使用する。残りの11タイプの学習データは以下の手順で作成した。

1. 大量のセリフデータを用意する⁷⁾
2. 口調エンコーダ E_B を用いて、全てのセリフに口調ベクトルを付与する
3. それぞれの口調タイプに対して、以下の手順で学習データを作成する
 - (a) 評価用データセットの30セリフを用いて、代表口調ベクトルを求める
 - (b) 2.のセリフデータを、代表口調ベクトルとの距離が近い順にソートする
 - (c) ソートされたセリフリストの上位から、口調タイプの定義に当てはまるものを1000件程度選定する。この選定は、第1著者が行った

新たに作成した学習データの概要を表3に示す。

7 口調エンコーダの再学習

再編成した20口調タイプからなる学習データを用い、学習に使用するタイプ数を段階的に増やし、口調エンコーダの性能がどのように変化するか調べた。その結果を表4に示す。表4の順序は、第1著者が一般的な口調と考える順序である。

この表に示すように、学習データに含まれる口調

7) 前述の「なろうセリフデータ」と「BCCWJ小説会話文」[10]の一部。セリフ数は5,443,992。

表4 段階的な口調タイプの追加

タイプ数	追加した口調タイプ	全体精度
4	男性的-普通(3), 女性的-普通(9), 男性的-敬語(1), 女性的-敬語(7)	82.17
5	女性的-てよだわ(13)	82.47
7	男性的-乱暴(5), 女性的-乱暴(11)	83.11
9	男性的-厳格(6), 女性的-厳格(12)	86.27
11	男性的-優しい(2), ボーイッシュ(8)	88.38
13	男性的-関西弁(4), 女性的-関西弁(10)	89.49
15	武士口調(18), 老人口調(16)	91.66
17	お嬢様(14), 慇懃(19)	92.96
19	～っス(17), ~なの(15)	93.00
20	寡黙(20)	93.40

表5 E_R の評価 (太字は中央値)

シード	42	43	44	45	46	平均
E_B	0.905	0.903	0.899	0.917	0.909	0.907
E_R	0.934	0.928	0.927	0.922	0.929	0.928

タイプ数を増やすと全体精度が向上し、20タイプ全てを用いた時に全体精度が最も高くなった。20タイプの学習データで学習した口調エンコーダ(記号 E_R で表す)の精度を E_B の精度と対比する形で表5に示す。

この表に示すように、口調エンコーダ E_R は E_B と比較して平均精度が上昇していることがわかる。全体精度が中央値となる E_B (シード42)と E_R (シード43)の実行結果に対して、2つのエンコーダの正解と不正解のペアを利用してマクネマー検定を行ったところ、有意水準0.05で E_R の方が性能が高いという結果が得られた。

5節で示した、 E_B の弁別精度が低かった口調タイプペアに対する E_R の弁別精度は、以下のように大きく改善した。

- 「女性的-乱暴(11)」と「老人口調(16)」: 0.900
- 「女性的-厳格(12)」と「武士口調(18)」: 0.850
- 「男性的-乱暴(3)」と「老人口調(16)」: 0.917

このことから、口調タイプの弁別においては、弁別したい口調タイプを学習データに含めることが有効であることが示唆される。

参考文献

- [1] 大沢在昌. 小説講座 売れる作家の全技術. 角川書店, 2012.
- [2] 石川和樹, 佐藤理史, 小川浩平. 口調エンコーダを用いた小説発話の話者推定. 自然言語処理, 第 31 巻, pp. 894–934, 2024.
- [3] 川北雄大, 石川和樹, 夏目和子, 佐藤理史, 小川浩平. 口調弁別評価データセットの作成と口調エンコーダの評価. 情報処理学会研究報告, pp. 1–10, 2024.
- [4] 金水敏. ヴァーチャル日本語 役割語の謎. 岩波書店, 2003.
- [5] 金水敏 (編). 〈役割語〉小辞典. 研究社, 2014.
- [6] メイナード泉子・K. ライトノベル表現論. 明治書院, 2012.
- [7] 夏目和子, 佐藤理史. 話し言葉生成のための発話文表現文型辞書. 言語処理学会第 28 回年次大会発表論文集, pp. 1612–1617, 2022.
- [8] 錢本友樹, 古俣槇山, 長谷川遼, 宇津呂武仁. 日本語の口調類似度評価データセットの作成および口調埋め込みモデルの構築・評価. 特集論文「人工知能と物語応用」, 第 40 巻, pp. 1–9, 2025.
- [9] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In **Proc. of NAACL-2019**, pp. 4171–4186, 2019.
- [10] 山崎誠, 宮寄由美, 柏野和佳子. 小説会話文への話者情報付与. Technical report, 国立国語研究所, 2022. <https://www2.ninjal.ac.jp/conversation/report/report05.pdf>.

A 口調弁別評価用データセットのセリフ例

口調弁別評価用データセットのセリフ例を表6に示す。

ID	口調タイプ	セリフ例
1	男性的-敬語	「はい、構いません。どこからでもどうぞ」 「ええ。私にできることであれば何なりと」
2	男性的-優しい	「そうだね。ホータローはその手のは全然向いてないんだよ」 「ホータロー、その顔はなんか良かったね」
3	男性的-普通	「僕も、マサムネから聞いて驚いたんだ」 「兄ちゃんにやられちゃってさ」
4	男性的-関西弁	「またとか言うなや、またとか！そういうお前はどこ入んねん」 「そりゃあまあ、低音王国はあれやしな。田中先輩の領地やし」
5	男性的-乱暴	「てめえ、ふざけんなよー！！」 「ってことはこの子、俺がいなけりゃこんだけ早く辿り着けたってことか」
6	男性的-厳格	「君たちに沖ノ島沖の敵泊地への出撃を命じる」 「気を落とすな。まだ、チャンスはある」
7	女性的-敬語	「折木さん、なにか心当たりがあるんですか？」 「折木さんのちょっと前です。三分前ぐらいだったと思います」
8	ボーイッシュ	「コーヒープレイクして、また気合いを入れ直そうじゃないか」 「格好いいでも、可愛いでもお好きなように。称賛には変わらないさ」
9	女性的-普通	「私たちも、みんな、今来たばかりなの」 「私、今日、部屋にいるね。フウカも来てるみたいだから、誘ってあげたら」
10	女性的-関西弁	「今更、関西弁直せとか言わへんよな？ウチ、この喋りしかでけへんで」 「ひゃっ！ツ、ツッコミはもっとビシっと頼むわ！優しすぎるで！」
11	女性的-乱暴	「納得いかねーのはアタシも一緒だよ。兄ちゃん、アンタ、なんなんだよ？」 「おい！どーいうことだよ！」
12	女性的-厳格	「えっ？……なんだ？私は何か、おかしな事を言ったか？」 「……ふむ、なんとも無いのだが」
13	女性的-てよだわ	「そう、チェスよ。でもこれはーただのチェスじゃないわ」 「あら、いいの？そんな悠長なことしてて」
14	お嬢様	「ふ、ふふ……もう……いいですわよ、どうにでもしてくればいいですわ」 「ジャンケンー？それは……まあ、知ってますわよ」
15	～なの	「よかったの。さっそく食べてほしいの！」 「ふええええ…体中びちゃびちななの…」
16	老人口調	「なんじゃ小僧ー盗品に興味があるのか？」 「話にならんじゃろうが！」
17	～っス	「事務所の近くに、いい画材屋さんないっスカねー」 「最近デジタルが主流だから、高スペックPC買うため、がんばるっス」
18	武士口調	「これで人を斬れると思うでござるか？」 「拙者の剣は、後世に残す気は無いでござるよ」
19	懇勤	「さようございます。そろそろご帰還されるころかと思ひまして」 「では、率直に思うところを述べさせていただきます」
20	寡黙	「……………じゃあ、わ、私に何を要求するつもり？」 「え？……どうしたって……なにが？」