

# 語尾・人称表現の自動抽出に基づく 対話システムにおけるキャラクター性の構築

長島 大和<sup>†</sup> 阿倍 博信<sup>†</sup>

<sup>†</sup> 東京電機大学 〒120-8551 東京都足立区千住旭町5番

E-mail: <sup>†</sup>20jkm20@ms.dendai.ac.jp, <sup>††</sup>hironobu.abe@mail.dendai.ac.jp

**あらまし** 対話システムに一貫したキャラクター性を構築することにより、その機械と対話する際の人間らしさの向上と、対話の満足感の向上を狙うことができる。現在、対話システムに一貫したキャラクター性を付与するためには、応答をすべて手作業で入力する、もしくは徹底したチューニングを行う必要があるが、その作業を手動で行うには多くの手間がかかるという問題があった。このような問題に関して、本研究では、対話システムにおいてキャラクタの発言からキャラクター性を構築する方式として、重要と思われる語尾・人称の表現を自動的に抽出する方式について提案する。次に、抽出結果を修正するアプリケーションを構築するとともに、抽出結果の対話システムへの適用を行う。最後に、評価を行った結果、提案方式の有効性について確認することができた。

**キーワード** 自然言語処理, 対話システム, 語尾抽出, キャラクター性

## 1 はじめに

機械に自然言語を入力すると、それに対応した応答を返す、現代社会においてはそうした対話システムは、ショッピングやIoTとの連携などにより我々の生活の中に浸透し始めており、将来的にはタスク指向型対話システムによる作業の自動化や、非タスク指向型対話システムによる機械との雑談対話などにおいての活躍が期待される。しかしながら、雑談対話を行う非タスク型対話システムにおいて、話し方の一貫性が保たれていない場合に、対話システムが「情緒が不安定である・説得力がない・人間関係の距離感がうまく取れていない」といった感覚をユーザに抱かせてしまう[1]ことが分かっている。こうした問題は対話システムの性能に直接的に関わる部分ではないが、システムとユーザの間に親和的関係を築こうとする場合において違和感を抱かせてしまう原因となる。

これらの解決に関して、対話システムに一貫したキャラクター性を持たせることが有効であると考えられる。キャラクター性とは、そのキャラクタを特徴付ける重要な要素のことである。例えば「おれは今腹が減っているのだ」「あたしは今お腹が空いているの」という2つの文章の場合、一般的に前者は男性、後者は女性の発言であるというように認識できる。このように、発話者に付与される性別などの属性と結びついた要素のことをキャラクター性と呼ぶ。こうしたキャラクター性を対話システムに付与することにより、その機械と対話する際の人間らしさの向上と、対話の満足感の向上を狙うことができると予測される。

現在、対話システムに一貫したキャラクター性を付与するためにはルールベースの対話システムを構築することが多い。ルールベース型対話システムとはユーザが入力した特定のキーワードに対応した応答を行う技術であり、AIML[2]などの技術を用いて実装される。しかし、こうしたルールベース型対話シ

ステムは人間が設定した以上の性能を発揮できず、現在の対話システムにおいて主流となる機械学習による生成ベース型対話システムによる柔軟かつ自然な応答生成という部分において劣ることが予想される。また、生成ベース型対話システムを利用する場合においてはキャラクター性を確保することがより難しいという問題がある。機械学習においては大量のデータを元に学習する必要性が存在するため、一般的にはTwitterなどのSNSからデータを収集し対話ペアを取得することが求められるが、SNSなどで無作為に取得した複数人の人物が混ざった学習データは、それぞれの会話においてキャラクター性が大きく異なる場合があることがキャラクター性構築という観点から見れば問題となり、最終的に完成した対話システムが一貫したキャラクター性を保持することができなくなってしまう。故に、それぞれの対話の応答自体が適切であったとしてもキャラクター性を保てないことにより性能に影響を及ぼす。そして、それを回避するために1人のキャラクタの発言を元に学習することも考えられるが、その場合一貫したキャラクター性を持った対話文章のデータ量を確保することが大きな問題となり、難易度が高い。

このような問題に関して、本研究では、対話システムにおいてキャラクタの発言からキャラクター性を構築する方式として、先行研究である自動語尾抽出方式[3]に、人称の表現を自動的に抽出する機能を追加した方式について提案する。次に、抽出結果を修正するアプリケーションを構築するとともに、抽出結果の対話システムへの適用を行う。最後に、評価を行った結果、提案方式の有効性についての報告を記載する。

## 2 関連研究

関連研究として、文中のキャラクター性表現がどこに現れるのかを調査した研究や、言語処理用のライブラリ、創作物データセットなどが存在する。

## 2.1 キャラクター性を持った対話システムの研究

まず、キャラクター性を保持した対話システムの例として、Microsoft により提供された人工知能キャラクターであるりんな [4] があげられる。LINE や Twitter などのサービスにおいて活動するチャットボットの一つであり、日本の女子高生をベースとして作成され、応答がそのキャラクター性に統一されていることが大きな特徴である。

2015 年の登場以来から改良を重ね、キャラクタービジネスにおいて成功を収めており、利用ユーザー数は 700 万人以上にのぼる。これらの点から対話モデルにおけるキャラクター性の重要性を示している。しかしながらこれらの技術は生成ベース型対話システム自体に多くの工夫を盛り込んでいるため、同様の対話システムの構築には多くのコストがかかることが懸念される。

次に、対話モデルの学習を工夫することにより、小説の登場人物のような喋り方をする対話システムの構築を目指した DeepEVE [5] が存在する。

当該研究においては発話者情報を考慮した A Persona-Based Neural Conversation Model モデル [6] を、自動構築した発話応答のペアと登場人物名によって学習し人物を指定することで登場人物らしい応答を行うものであり、一定のキャラクター性の維持が確認できる。

しかし、学習データの絶対量の不足により人間らしい自然な応答の量自体は少なかったことが指摘されている。そのため、小説の登場人物のように限定されたデータ量のキャラクター性を保ちつつ、満足度の高い応答を得るためには、純粋な機械学習によるモデル構築以外の方法を模索する必要性がある。

関連して、機械学習によるアプローチの一つである転移学習 (Transfer Learning) は、教師となる対話文のデータが少ない場合に他の十分な量のデータで学習した異なるモデルを引き継ぎ、さらにそこから学習を行う方法である。これらはデータを十分に確保できないタスクにおけるの精度向上を目的として行われるものであり、対話学習におけるキャラクターの語尾スタイルの制御という形での研究が赤間らによって行われている [7]。

ただし学習方式のフォーマットにおいて、キャラクターのセリフは対話形式の必要があるが、創作物におけるキャラクターの発言には対話以外にも独り言やモノログ、長文による解説などといった、明確な対話ではなく一方的に発話が繰り返されるケースがある。これらの場合においては転移学習用のデータを揃えることが困難となる。

また、事前学習とのデータの性質の差異により転移学習後の対話モデルの正確さに影響が出ることが確認されており、これらの問題の解決には他の種類のアプローチが必要であると考えられる。

## 2.2 文中のキャラクター性表現に関する研究

上述したキャラクター性について特に焦点を当てた研究の一つに刀山らの研究が存在する [8]。

当該研究は年齢・親密さなどのキャラクター性を中心に、内容表現、機能表現の書き換えによってどのような変化が発生するのかを調査したものとなり、結果として文章のほぼ全域におい

てキャラクター性を構築する自由度があり、中でも特にキャラクター性が強く結びつく要素として「終助詞を含む文末表現と、特定の機能を表す機能表現の 2 つが有力な候補である」と考察されている。

さらに、キャラクター性を再現するための変換方式として、宮崎らの研究が存在する [9]。当該研究はキャラクター性によって付与される役割語などの言語的特徴を抽出し、発話テキストの自動生成を行うことを目標とした研究であり、文末表現に注目した自動抽出などが行われており、性別・年齢・親密度より人物のキャラクターらしさを推定し、それらを自動書き換えすることによりキャラクター性を付与させられることを確認している。

その一方で、創作物内部で時折出現する特徴的な語尾などにおいての自動抽出の結果は記載されておらず、それらの特徴的な語尾を抽出するためにはより特化したアプローチが必要であると考えられる。

## 3 提案方式

### 3.1 方針

関連研究を受け、これまで対話システムにキャラクター性を付与するための方法においては、創作物などのキャラクター性に着目した研究が少なく、また機械学習においてはキャラクター性を持ったデータの量が不足しているということが強く課題として存在していることが分かった。

そのため、本研究ではキャラクターの対話を入力することで、そのキャラクター性が含まれた要素を出力し、キャラクター性の分析や変換などを行うツールの作成を目標としている。

作成するアプリケーションは、「抽出・修正・変換」の三つの機能を保持する。図 1 にアプリケーションの全体図を示す。

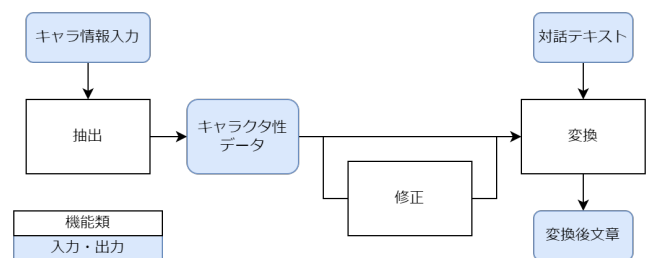


図 1 アプリ全体図

### 3.2 実装

全体のプログラムの作成にあたり、Python 3.9.7 [10] を言語として選択、文章保存などに pandas 1.1.3 [11]、修正機能の一部要素として wordcloud 1.8.1 [12]、Web アプリケーションの構築には Streamlit 0.89.0 [13] を利用した。

また、先行研究 [3] にて実装した自動語尾抽出機能では、形態素解析器として Janome 0.3.10 [14] を利用したが、本研究においては GiNZA 5.0.0 [15] およびその内部の形態素解析器 SudachiPy 0.5.4 [16] を総合的な自然言語処理ライブラリとして採用し、抽出機能の部分を作成しなおした。

### 3.3 抽出機能

抽出機能は、キャラクターの発言を入力することで、重要なキャラクター性にあたる語尾・人称表現を抽出し、データ化する機能である。

#### 3.3.1 背景

刀山らの研究 [8] より、キャラクター性の多くは文末表現である語尾部分や機能表現とされる部分にあらわれていることから、キャラクターの発言に含まれている語尾を個別に検証し分析することが必要であると考えられる。

それと同様にいわゆる一人称や三人称などといった人称表現に関しても、対話中に含まれるキャラクター性として強力であると考えられ、実際に先行研究においても抽出・書き換えが行われていた。そのため、これらの抽出も必要であるとも考えられた。

また、機械学習を用いた関連研究 [5] [7] より、キャラクター性の含まれるデータ量は多くの場合は機械学習などにおける必要なデータ数に達さないことが予想されるため、キャラクターの発言をより深く解析する処理が必要となる。

上記の理由から、キャラクターの発言のみを比較的少量かつ対話ペアという形で収集することなく動作し、キャラクター性の重要な要素である語尾・人称表現を抽出する方式が必要であることが推測できる。以下、表 1 に抽出を目指す表現の説明を示す。

表 1 抽出する表現の説明

表現	説明
語尾表現	「～です」のような文末の助動詞・終助詞表現
人称表現	「僕」「あなた」のような人称代名詞

#### 3.3.2 従来方式との差異

従来方式においては語尾を自動的に抽出する方針においてはいわゆるパターンマッチの方式を用いた変換器 [1] などが実装されることが多く、これらは例えば「です。」という特定の語尾を、例えば大阪弁のような「やん。」という表現に句点を含めたパターンにより変換するようなものである。しかしながらこうした方式は、例えば「ですね。」というように類似しているがパターン上記載されていない単語が出現した際に、抽出・変換することができず精度が低くなるという問題を抱える。

また、自動抽出などが用いられた上記の研究 [9] においても、「です。」「ます。」などの一般的な語尾の自動抽出に基づいたデータによるものが多いことが予想でき、創作物の登場キャラクターの発言や方言などといったものを抽出・変換することが難しいことが予想される。

先行研究 [3] において、創作物や方言などのキャラクターの語尾を抽出するための工夫を行うことで、対話の中に含まれている語尾を特殊な形態のものを含め自動的に抽出する方式を提案した。図 2 に本研究の想定するイメージについて示す。

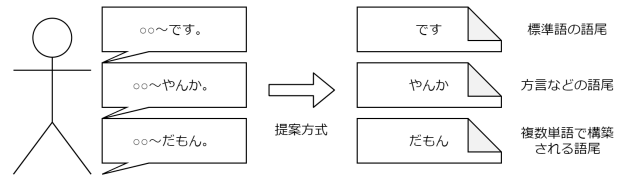


図 2 提案方式の動作イメージ図

この方式の利点として、キャラクターの発言のみを集めることで動作し対話ペアの作成を省略可能であり、一人のキャラクターに対しての特徴を明示的に抽出し、それを元にキャラクターの分析に役立てることが想定できた。このため、本研究の抽出機能においてもこれらの方針に沿った実装を行った。

また、人称表現に関してはそもそも明示的に抽出するアルゴリズムが少なかったため、キャラクターの発言のうち代名詞、及び人名の固有名詞と分類された単語を抽出することとした。ただし、今回の研究では特徴的なキャラクター性の抽出を主としているため、あまりにも一般的に利用されうると考え、抽出結果などを見比べて除外するべきだと判断した代名詞表現である「それ、そこ、そっち、これ、ここ、こっち、どれ、どこ、どっち、あれ、あそこ、あっち、何、何々、何何、なに、なん、いつ、誰」などの表現を除外している。

#### 3.3.3 処理

提案方式においては、キャラクターの名前、年齢と性別、性格分類法であるエゴグラムを参考にした性格(選択 2 つ、自由入力 1 つ)、発言を Web サービス上の入力フォームに入力し、それぞれに処理を行う。

図 3 に本研究で構築した提案方式のフローチャートを示す。

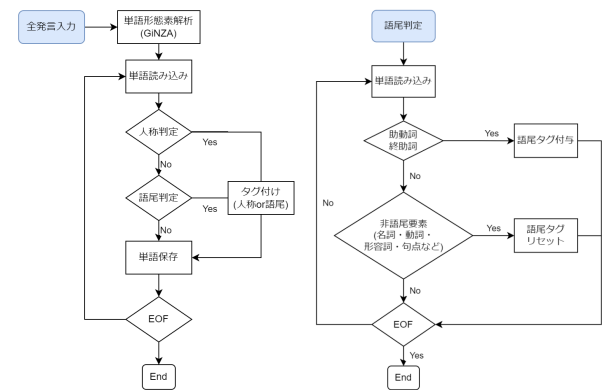


図 3 提案方式のアルゴリズムと語尾抽出の詳細

入力されたキャラクターの発言一覧を読み込み、日本語言語処理ライブラリである GiNZA [15] を用いて解析を行う。文中表現などを一度整形し、また、相槌などに代表されるような短い発言を除外するために一定文字数以下の応答を除外し、また改行、文節ごとに単語を分割した後、後述するキャラ語尾の辞書を読み込み表現抽出を行う。

この時点で、多くの場合においてキャラクター性を持った語尾は助動詞、終助詞などに分類される文末表現が中心となると関





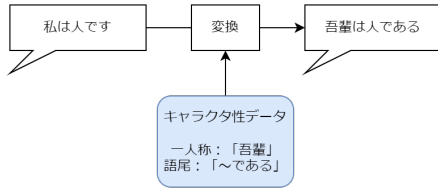


図 7 変換イメージ

変換に利用する抽出機能で作成されたキャラクタの TSV ファイルから、さらにどのようにすればキャラクタらしい変換ができるのかを検証するため、Random モデルと TFIDF モデルの 2 つのモデルを用意した。前者は変換時に扱う語尾・人称表現を全てランダムにシャッフルし利用するモデルである。後者は単語の重要度が高い順にソートを行い、それらを利用して変換するモデルである。

比較した場合、文章に頻出する表現が優先的に利用される後者の方がそのキャラクタの話し方に頻出する表現が選ばれる確率が高くなる傾向にあると考えられた。以下、2 つのモデルのイメージ図を図 8 に記載する。

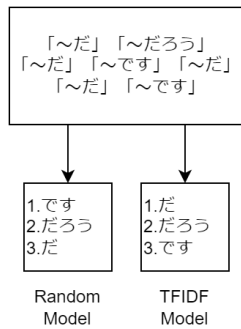


図 8 Random / TFIDF Model

## 4 評価

作成したアプリケーションの抽出・変換の二種類の機能に関して評価を行った。抽出部分は定量評価を、変換部分はアンケートによる定性評価を実施した。

### 4.1 抽出機能

#### 4.1.1 評価方針

語尾抽出の有効性を評価するために、それぞれのキャラクタの発言データに加え、その発言に対する語尾部分のみをまとめた Ground Truth(以下 GT) を作成した。

抽出結果を元にそれぞれを TP(True Positive), FP(False Positive), FN(False Negative), TN(True Negative) と分類し、これらの評価値から式 1,2,3 よりそれぞれ Recall(再現率), Precision(精度), F(F 値) を算出することで提案方式の評価とした。それぞれの評価値についての説明を表 2 に示し、FP, FN の評価値についての詳細を表 3 に示す。

$$Rec = \frac{TP}{TP + FN} \quad (1)$$

$$Prec = \frac{TP}{TP + FP} \quad (2)$$

$$F = \frac{2 \cdot Rec \cdot Prec}{Rec + Prec} \quad (3)$$

表 2 評価指標の一覧

評価値	説明
True Positive	GT と完全に一致する抽出
False Positive	GT に含まれない抽出 (過検出)
False Negative	GT の一部のみの抽出 (未検出)
True Negative	GT の空値と一致する抽出
Recall	抽出した語尾が未検出なく抽出できた割合
Precision	抽出した語尾が過検出せず抽出できた割合
F	総合的な精度

表 3 FP, FN の詳細

評価値	説明	想定された GT	抽出失敗例
FP	語尾ではないものを 真と予測 (過検出)	「～けど」	「～知れんけど」
FN	語尾の一部以外を 偽と予測 (未検出)	「～やろうな」	「～な」

#### 4.1.2 評価データ

評価データとして Manga109 [17] [18] データセットを用いた。Manga109 は東京大学相澤・山崎・松井研究室により取りまとめられた、109 の漫画で構成されたデータセットである。これらのアノテーションデータを利用することで、創作物におけるキャラクタの発言表現の抽出のため利用を試み、本研究においては『ベルモンド © 石岡ショウエイ』[19] 及び、『燃える!お兄さん © 佐藤正』[20] の二冊を選択し、キャラクタの発言を抽出する方法を取った。

以下、抽出したキャラクタの評価データの詳細を表 4 に、データセットごとの発言のサンプルを表 5 に示す。

表 4 評価用データの概要

キャラ名	説明	データ数
ベルモンド	青年・男性:冷淡・閉鎖的・主人公	200
ロラン	青年・男性:優しさ・社交的・協調的	100
セリーヌ	青年・女性:感情的・協調的・我儘	129
エスタフェ	青年・男性:感情的・怠惰・悪役	58
バレアー	青年・男性:厳格・論理的・悪役	81
国宝憲一	少年・男性:感情的・マイペース・兄	163
国宝雪絵	少年・女性:優しさ・妹	35
清田光太郎	青年・男性:厳格・優しさ・教師	156
合計		922

表5 評価用データの発言例

キャラクタ名	発言のサンプル
ベルモンド	のんびりはしてられんようだ
ロラン	僕はホッとします
セリーヌ	私はネズミが一番嫌いな
エスタフェ	あっしは無実ですってェ
バレアー	なるべく静かにやるべきだ
国宝憲一	みんなで外に行って遊ぶのだーっ
国宝雪絵	ほらおにーちゃん授業が始まるよ
清田光太郎	ちょっと国宝手伝ってくれーっ

表8 評価3 - 代名詞抽出

キャラ名	TP	FP	FN	TN	Rec	Prec	F
ベルモンド	30	46	18	18	0.625	0.395	0.484
ロラン	14	18	17	17	0.452	0.438	0.444
セリーヌ	18	27	23	23	0.439	0.4	0.419
エスタフェ	10	11	5	5	0.667	0.476	0.556
バレアー	5	18	5	5	0.5	0.217	0.303
国宝憲一	0	17	29	29	0	0	0
国宝雪絵	0	3	12	12	0	0	0
清田光太郎	1	32	20	20	0.048	0.03	0.037
平均					0.341375	0.2445	0.28

### 4.1.3 評価結果

キャラ語尾辞書を非適用の語尾抽出の結果を表6 キャラ語尾辞書を適用した語尾抽出の結果を表7, 人称表現の抽出結果を表8に示す。

また、実際に抽出されタグを付けられた単語類に関して、語尾・人称表現に関しての一部の例を表9に示す。

表6 評価1 - 語尾抽出・キャラ語尾辞書非適用

キャラ名	TP	FP	FN	TN	Rec	Prec	F
ベルモンド	69	12	97	97	0.416	0.852	0.559
ロラン	30	17	39	39	0.435	0.638	0.517
セリーヌ	51	7	49	49	0.51	0.879	0.646
エスタフェ	15	2	32	32	0.319	0.882	0.469
バレアー	24	3	37	37	0.393	0.889	0.545
国宝憲一	7	2	126	126	0.053	0.778	0.099
国宝雪絵	8	2	17	17	0.32	0.8	0.457
清田光太郎	45	6	68	68	0.398	0.882	0.549
平均					0.3555	0.825	0.48

表7 評価2 - 語尾抽出 - キャラ語尾辞書適用

キャラ名	TP	FP	FN	TN	Rec	Prec	F
ベルモンド	72	16	90	90	0.444	0.818	0.576
ロラン	31	18	37	37	0.456	0.633	0.53
セリーヌ	53	7	47	47	0.53	0.883	0.662
エスタフェ	14	2	33	33	0.298	0.875	0.444
バレアー	26	3	35	35	0.426	0.897	0.578
国宝憲一	103	4	27	27	0.792	0.963	0.869
国宝雪絵	11	3	13	13	0.458	0.786	0.579
清田光太郎	45	8	66	66	0.405	0.849	0.549
平均					0.476125	0.838	0.598

表9 語尾・人称表現の抽出結果例

キャラ名	語尾サンプル	人称サンプル
ベルモンド	たぞ、だ、するか、しれんしな	お前、私、貴様
ロラン	です、ます、なあ、ました	僕、あいつ、ベルモンド
セリーヌ	でしょ、わよ、でしょうね、たわ	私、あなた、ベルモンド
エスタフェ	やがる、だ、やがったな	あっし、アンタ、こいつ
バレアー	てる、べきだ、だ、かああ	俺、ベルモンド、
国宝憲一	のだーっ、なのだ、のだ、か	(検出不可)
国宝雪絵	た、ですか、ました、ますーっ	(検出不可)
清田光太郎	ですか、だよ、だぞーっ、だな	ボク

## 4.2 変換機能

### 4.2.1 評価方針

上記アプリケーションにて抽出されたキャラクタ性を用いて、実際に変換された文章がキャラクタらしいかどうかに関して、変換によりどれだけ元の文章の自然さが損なわれたかを検証するため、10名に協力してもらい、アンケートによる主観評価を実施した。評価の流れのイメージを図9に示す。

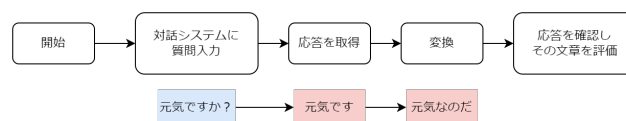


図9 評価の流れ

評価方法として、まずキャラクタの選択を行い、変換先のキャラクタ性を指定する。選択はこちら側で行い、その順番は「ベルモンド/Randomモデル」「ベルモンド/TFIDF」「国宝憲一/Randomモデル」「国宝憲一/TFIDFモデル」で固定である。次に、評価者に固定の質問を6個、自由に入力可能な質問を4個入力してもらい、実際にそれらの文章を対話システムに送信する。対話システム側から返却された応答文を、変換機能を用いてキャラクタらしく変換し、文章を表示する。これらの表示結果を表示したのち、Googleフォーム上にてその文章をそれぞれキャラクタらしさ、文章の自然さに関して評価を行

う。表 10,11 にそれぞれの評価項目の詳細を示す。

表 10 キャラクターらしさかどうかの指標

評価値	説明
1	キャラクター性が存在していない
2	わずかにキャラクターらしさが存在していると感じられる
3	比較的そのキャラクターらしさがあると感じられる
4	文面上、そのキャラクターであると強く感じられる

表 11 文章の自然さの指標

評価値	説明
1	文章が完全に破綻しており、読み解けない
2	ニュアンスは伝わるが、文法的な破綻が存在している
3	自然な文章である

なお、今回対話システムの実装には Web 上の会話 AI 構築プラットフォームである mebo [21] を利用した。mebo に存在する機能として、対話の応答に関してルールベースで決定することができたため、一部の頻出の質問に関しては設定を行い、AI の自動生成により評価が左右されることを抑えた。

#### 4.2.2 評価データ

評価に用いたデータとして、前述の抽出機能により抽出されたキャラクターの中から『ベルモンド』及び『国宝憲一』の二人を選択し、それらのデータを用いて変換を行った。

実験対象者には二名の簡単なキャラクターとしての性格や特性、サンプルのセリフなどを提示し、事前にキャラクターの理解を得た状態での実験とした。以下表 12,13 にそれぞれのキャラクターの詳細を示す。

表 12 ベルモンドのキャラクター説明

評価値	説明
名前	ベルモンド
属性	青年の男性、修道院内の拷問官という職業
性格	冷淡・論理的・ダーク シリアスな場面とおちゃらけた場面があり、 掴みどころがない
語尾	「だ」「だろう」「だな」 時折「だ♪」などの音符がついた表現
人称	一人称「私」、三人称「お前」など
発言	「私はここで生まれここで死ぬ」 「我ながら美しくできたな」

表 13 国宝憲一のキャラクター説明

評価値	説明
名前	国宝憲一 (こくほうけんいち)
属性	野生児から高校生?
性格	感情的・マイペース・ギャグ クセが強い性格
語尾	「のだ」「なのだ」を中心としたキャラ語尾
人称	一人称「おにーさん」、三人称「お前」など
発言	「あうーっ やったなーっ」 「お絵かきは一応得意分野なのだ」

#### 4.2.3 評価結果

キャラクター・モデルの組み合わせごとによるキャラクター性の平均値を図 10、自然さの平均値を図 11 に示す。

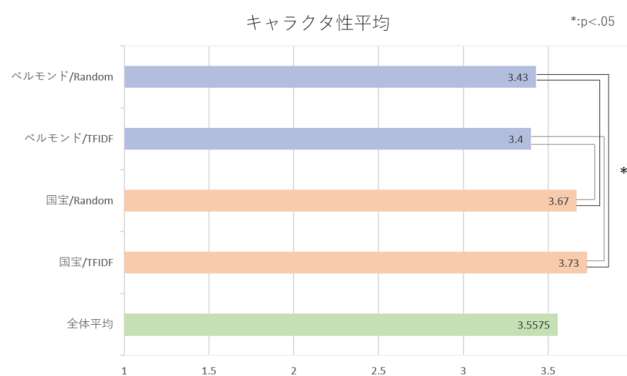


図 10 変換評価 - キャラクター性

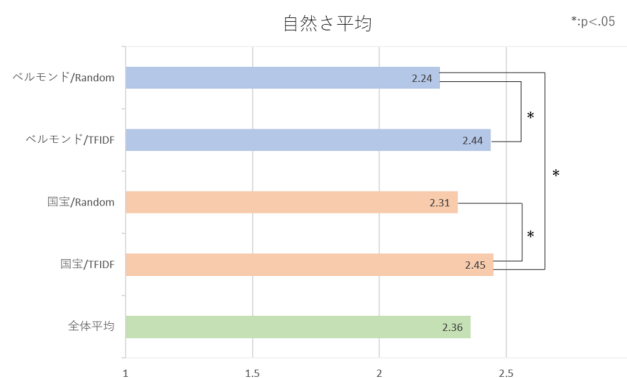


図 11 変換評価 - 自然さ

## 5 考察

### 6 抽出機能

まず、表 6 よりキャラ語尾辞書を適用しなかった場合の F 値が 0.480、表 7 よりキャラ語尾辞書を適用した場合の F 値が 0.598 であり、キャラ語尾辞書の適用により語尾抽出精度の改善

が確認できた。特に、『国宝憲一』のキャラクタに関しては「～なのだ」「～のだ」という語尾が中心となっており、これらを中心とした抽出がキャラ語尾辞書を利用しなかった場合の0.099という数値に対し、利用した場合に非常に高い0.859というスコアを出すことができ、有効性が確認できたと思われる。

ただし、一部分においては不適切な語尾抽出結果なども存在しており、助動詞・終助詞に限らず動詞などの活用形などを考慮することでより性能が改善できると考えられる。

人称表現に関しては一部キャラクタにおいてはまったく抽出できないなど、大きく差が出る結果となり、また同時に代名詞表現の中でもどれを人称表現と定義するのが難しく、抽出方式や定義といった段階から課題を感じられた。しかし、「お前」や「私」などの表現と、「僕」「私」などの表現が抽出できていること、また、一部人物は「ベルモンド」という人物名を正しく抽出していることから、現段階でもより正しく分類できればキャラクタ性の分析などに活用が可能だと感じられた。

## 6.1 変換機能

キャラクタ性に関しては、キャラクタ間での有意差が確認された。このことから、大本のキャラクタの影響が強くなる傾向、また、キャラ語尾などの特徴的な要素が存在する場合に変換時に強くキャラクタ性を認識させることができると考察できた。

また、文章の自然さは変換モデル間において有意差が現れていることが確認された。そのため、Random モデルと TFIDF モデルにおいては、文章頻度に影響を受ける TFIDF モデルの方が変換時の効果が高いと言える。ただし、Random モデル、TFIDF モデルに置いても、国宝の Random モデルとベルモンド TFIDF モデルには有意差が確認されなかった。このことから元々のキャラクタ性に影響を受ける部分もあることが推測される。

また、対話システム側で体言止めや倒置法など、変換が難しい文章が送信された場合に不自然な変換となる事例が存在し、前の単語の活用形の変化などが含まれてしまう場合は、現時点でのルールでは対応しきれない場合が多い。これらは実運用において問題となりうるため、この部分に対応する必要があると感じられた。

ほか、キャラ語尾の接続の問題として、母音・子音などが一致しなくても変換可能である「～なのだ」のような語尾は変換の自由度が高く、置き換えルールを再度考慮する必要が感じられた。

最後に、今回の変換では表面上のキャラクタ性の付与を行っているため、アンケートにおいて「この性格のキャラクタであればこのような発言はしないと思う」などのような備考が存在した。この部分に関しては提案方式をより拡張する必要があると考えられた。

## 7 おわりに

当研究では、キャラクタ性の抽出・修正・変換機能のシステム化や、変換処理などの開発・評価などを行った。

その結果として、語尾表現の抽出における全体的な抽出が可能という結果と、一部キャラクタに関しては高スコアを記録することができ、また、変換機能において原文から変換した文章においてキャラクタ性の付与に成功したが、人称表現が困難であったことや、変換時の文章の自然さなどに関しては課題が残る形となった。そのため、上述の人称表現の抽出の改善や、変換時の文章の自然さを増すための変換ルールなどの再定義が必要と感じられた。

## 文 献

- [1] 東中竜一郎, 稲葉通将, 水上雅博: Python で作る対話システム, オーム社, 2020.
- [2] W.S.Richard: The Anatomy of A.L.I.C.E., A.L.I.C.E. Artificial Intelligence Foundation, Inc. 2004.
- [3] 長島 大和, 阿倍 博信: 対話システムにおけるキャラクタ性構築のための自動語尾抽出方式の検討, 第 13 回データ工学と情報マネジメントに関するフォーラム, H25-5, 2021.
- [4] X.Wu, K.Ito, K.Iida, K.Tsuboi and M.Klyen: りんな: 女子高生人工知能, 言語処理学会 第 22 回年次大会 発表論文集 (2016 年 3 月), pp.306-309, 2016.
- [5] 小倉 拓人, 谷津 元樹, 原田 実: 人物ベースの Seq2Seq モデルを用いた対話システム DeepEVE における小説中の登場人物らしい応答文の作成, 情報処理学会第 80 回全国大会講演論文集, 2018(1), pp.423-424, 2018.
- [6] J.Li, G.Michel, B.Chris, S.Georgios, J.Gao and D.Bill: A Persona-Based Neural Conversation Model, Association for Computational Linguistics, pp.994-1003, 2016.
- [7] 赤間 怜奈, 稲田 和明, 小林 颯介, 佐藤 祥多, 乾 健太郎: 転移学習を用いた対話応答のスタイル制御, 言語処理学会 第 23 回年次大会 発表論文集, pp.338-341, 2017.
- [8] 刀山 将大, 佐藤 理史, 松崎 拓也, 宮崎 千明, 平野 徹, 松尾 義博: 文のどこにキャラクタ性を埋め込む自由度があるか, 言語処理学会第 22 回年次大会 発表論文集, pp.721-724, 2016.
- [9] 宮崎 千明, 平野 徹, 東中 竜一郎, 牧野 俊朗, 松尾 義博, 佐藤 理史: 文節機能部の確率的書き換えによる言語表現のキャラクタ性変換, 人工知能学会論文誌 vol.31 No.1 sp2-E, pp.1-9, 2016.
- [10] Python Software Foundation: Python, <https://www.python.org> (アクセス日:2021/12/21)
- [11] pandas-dev: pandas, <https://pypi.org/project/pandas> (アクセス日:2021/12/21)
- [12] Andreas Mueller: WordCloud for Python, [http://amueller.github.io/word\\_cloud/](http://amueller.github.io/word_cloud/) (アクセス日:2021/12/29)
- [13] Streamlit Inc.: Streamlit <https://streamlit.io/> (アクセス日:2021/12/21)
- [14] T.Uchida: Janome, <https://mocabeta.github.io/janome> (アクセス日:2021/12/21)
- [15] Megagon Labs : GiNZA - Japanese NLP Library <https://megagonlabs.github.io/ginza/> (アクセス日:2021/12/21)
- [16] WAP Tokushima Laboratory of AI and NLP: SudachiPy, <https://github.com/WorksApplications/SudachiPy> (アクセス日:2021/12/21)
- [17] K.Aizawa, A.Fujimoto, A.Otsubo, T.Ogawa, Y.Matsui, K.Tsubota and H.Ikuta: Building a Manga Dataset "Manga109" with Annotations for Multimedia Applications, IEEE MultiMedia, 2020.
- [18] Y.Matsui, K.Ito, Y.Aramaki, A.Fujimoto, T.Ogawa, T.Yamasaki and K.Aizawa: Sketch-based Manga Retrieval using Manga109 Dataset, Multimedia Tools and Applications vol.76, pp.21811-21838, 2017.
- [19] 石岡ショウエイ: ベルモンド Le Visiteur R 1, 集英社. 2007.
- [20] 佐藤正: 燃える!お兄さん 第 19 巻, 集英社, 1991.
- [21] mebo(ミーボ) - 会話 AI 構築プラットフォーム <https://mebo.work/> (アクセス日:2021/12/21)