

# 対話システムにおけるキャラクター性構築のための 自動語尾抽出方式の検討

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    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。そこで、本研究では、対話システムにおいてキャラクター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を自動的に抽出する方式について提案する。また、提案方式の有効性について確認した結果について報告する。  
キーワード    自然言語処理, 対話システム, キャラクター性, 語尾抽出

## 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文を返す、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な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、そのキャラクター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。

しかしながら、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。

また、それを回避するために1人のキャラクター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。

そこで、本研究では対話システムにおいてキャラクターの発言からキャラクター性を構築するために重要と思われる語尾を自動的に抽出する方式について提案する。また、提案方式の有効性について確認した結果について報告する。

## 2 関連研究

以下は、いずれもキャラクター性という部分に着目した関連研究である。

### 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[3]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 [4] 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 [5] 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の語尾スタイルの制御という形で研究が赤間らによって行われている [6]。

ただし学習方式のフォーマットにおいて、キャラクタ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 [7]。

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 [8]。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

機械学習を用いた関連研究 [4] [6] より、キャラクター性の含まれるデータ量は多くの場合においては機械学習などにおいて必要数に達さないことから、別のアプローチが必要である。

さらに、刀山らの研究 [7] より、キャラクター性の多くは文末表現である語尾部分にあらわれていることから、キャラクタ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。

上記の理由から、キャラクタ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表現を抽出する方式が必要であることが推測できる。そこで、語尾部分の抽出を行う方式に関して調査を行ったところ、いくつかの問題を確認した。図1に従来方式における語尾抽出の問題を示す。

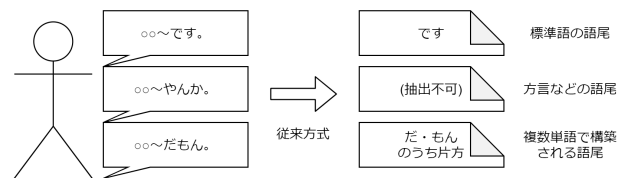


図1 従来方式の動作イメージ図

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い。

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 [8] においても、「です。」「ます。」などの一般的な語尾の手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。

そこで、本研究では創作物や方言などのキャラクタの語尾を

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する。図2に本研究の想定するイメージについて示す。

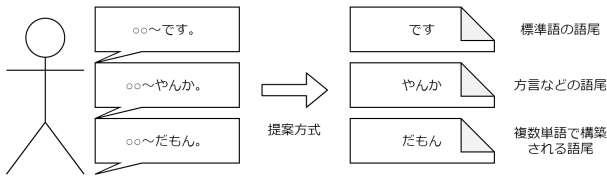


図2 提案方式の動作イメージ図

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ることがあげられる。このことより、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る。

### 3.2 自動語尾抽出方式の概要

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。

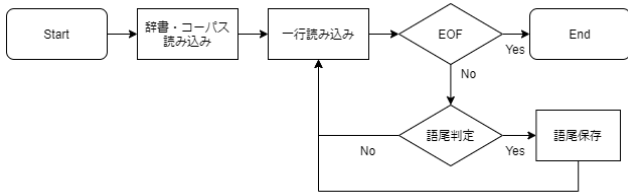


図3 提案方式のアルゴリズム

提案方式においては、キャラクターの発言をまとめたテキストファイルを読み込み、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する前処理を実行する。

次に、後述するキャラ語尾の辞書を読み込んだ後、条件付き確率場を用いた形態素解析器を利用し解析を行う。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連研究より判明している[7]ため、これら2属性を持った単語に限って抽出し語尾候補とする。

これらの処理をテキストファイルのEOFまで継続して行い、最終的に語尾が抽出されたファイルを出力する。

### 3.3 従来方式との差異

まず、創作物に存在するキャラクターの語尾抽出の際に問題になるものとして、キャラクターを際立たせるために付与されるキャラ語尾というものが存在する。例えば犬のキャラクターであれば「～ワン」、猫のキャラクターであれば「～にゃ」などといった語尾が付与されることがある。

こうしたキャラ語尾の表現の幅は非常に広く、基本的には助動詞などとしてではなく、未知語として認識される場合が多い。その上、一般的な雑談対話には出現することが少ないために創作物などのキャラクターから語尾を抽出する際に課題になること

が予想された。

そこで、本研究では事前にアニメ・漫画などの作品より196単語程度抽出し、事前にmecab系統のユーザ辞書(ipadic)形式[9]で辞書を作成することとした。構築したキャラ語尾辞書に関してのイメージを図4に示す。

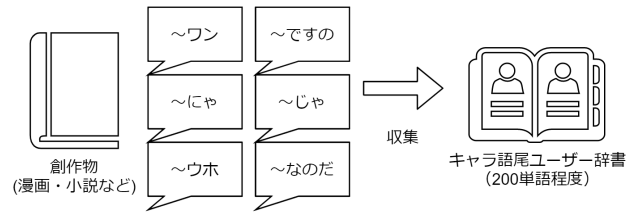


図4 キャラ語尾辞書のイメージ

キャラ語尾辞書に登録されたそれぞれの単語には助動詞の属性を付与されており、これらを語尾抽出の際に読み込むことでこうした特殊な表現への対策を試みた。

ほか、いくらかの文章においては、複数の助動詞・終助詞が組み合わさることで語尾として成立する語尾がいくらか存在することが分かっている。例えば「だもん」という語尾は、「だ」という一般的な助動詞と、「もん」という終助詞によって構成される単語である。そのため、単純な単語一つを抽出するのみでは、「だもん」が出現した際に、「もん」のみを抽出してしまうことが予想された。これらの表現をキャラ語尾辞書に記載することも考えられたが、作業コストの削減のため、事前の単語の属性が語尾になりうる場合にはそれらをまとめて抽出することで、パターンマッチ方式での弱点になりうる、表現の変化による精度低下の改善を目指した。

ほか、時折語尾が存在しない文章(体言止め、長い領きなど)が存在することが分かっており、これらを適切に処理するため上記の語尾になりうる属性を持つ単語が存在しなかった場合には抽出結果を空とした。

### 3.4 実装

全体のプログラムの作成にあたり、Python 3.8.3[10]、整形部分にbeautifulsoup4 4.9.3[11]、emoji 0.6.0[12]、文章保存などにpandas 1.1.3[13]、形態素解析器としてJanome 0.3.10[14]を使用した。

## 4 評価

### 4.1 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ー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関しての説明を表1に示し、FP, FNの評価値に関し

ての詳細を表 2 に示す。

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

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

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

表 1 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 2 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」

## 4.2 評価データ

評価データとして 3 種類のコーパスからそれぞれ 2 種類ずつのデータを選定した。

関西弁コーパス [15] はヘファナン・ケビンにより構成されたコーパスであり、大学生とその家族もしくは親密な知り合いなどとのインタビュー結果を収集したものである。年齢幅 10 代から 80 代と幅広く、キャラクター性として大阪弁、京都弁などの語尾表現の変化が確認できるものが付与されている点、インタビューの回答者が基本的に 1 名であったことから評価用データとして利用した。

名大会話コーパス [16] は国立国語研究所によって公開されているコーパスの一つであり、日本語話者同士の雑談対話が文字化されたものである。インタビュー形式よりもよりリラックスした雑談という条件と、また標準語での対話の語尾抽出などの評価のため利用した。

Manga109 [17] [18] 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、プロ漫画家によって描かれた 109 の漫画で構成されたデータセットである。創作物のキャラクターのテキスト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ーの発言の語尾、特にキャラ語尾の評価を行うため利用を試みた。ただし、アノテーションデータ形式の都合上、キャラクターの発言を固定して取り出すということに困難が生じたた

め、漫画全体としての語尾抽出器の有効性の調査、またジャンルごとにどのような差異が発生するのかの検討のため、漫画内にある全テキストを抽出しそれを評価用データとして用いることとした。

それぞれのデータセットは方言話者などといったキャラクターが含まれており、これらの発言の語尾部分を抽出可能かどうかの検討を行った。また、年代によっても語尾表現には差が発生することが関連研究 [7] から予想されたため、その点も考慮し選出した。

これらを、それぞれ関西弁コーパスを KSJ、KYT、名大会話コーパスを NUCC1、NUCC2、Manga109 を M109B、M109M として評価データとし、それぞれの評価データの説明に関しての詳細を表 3 に、データセットごとの発言のサンプルを表 4 に示す。

表 3 評価用データの概要

データ名	説明	データ数
KSJ	20 代女性、大阪在住 (大阪弁)	201
KYT	60 代女性、京都在住 (京都弁)	200
NUCC1	20 代男性、愛知在住 (関西弁・標準語混合)	204
NUCC2	60 代女性、東京在住 (標準語)	208
M109B	『ベルモンド』© 石岡ショウエイのテキスト	253
M109M	『燃える!お兄さん』© 佐藤正のテキスト	209

表 4 評価用データの発言例

データセット名	発言のサンプル
KSJ	「はいってないのとちゃうやろうな」 「結構普通ぐらいなんちゃうん」
KYT	「次から次から入ってきはんねん」 「ものすごく辛い思いをしたさかいにな」
NUCC1	「今日何しようかな」
NUCC2	「向こうも慣れてんじやない」
M109B	「のんびりはしてられんようだ」 「だまされるお前が悪いんだぜエ」
M109M	「みんなで外に行って遊ぶのーっ」 「先生がきたよーっ」

## 4.3 評価結果

前章で提案した自動語尾抽出方式を Python で実装し、評価データを用いて有効性について評価を行った。その結果について、キャラ語尾辞書を利用しない場合の抽出結果を評価 1、キャラ語尾辞書を適用した場合を評価 2 とし、それぞれ表 5、表 6 に示す。

また、今回の提案方式との比較のため、従来方式であるパターンマッチに基づく方式を関連研究 [1] を参考に構築した。これを同じくキャラ語尾辞書のありなしを比較し、同様に評価 3、評価 4 とし、表 7、表 8 に示す。こちらの方式はパターンによ

る抽出が中心となっており、語尾として選択されているのは標準語と大阪弁の2種類によるもので、「です、ます、だ、である、だよ、ですか、や、やねん、やろな、やろなあ」を抽出する。キャラ語尾辞書を利用する際には、前述の語尾パターンのリストに直接キャラ語尾 196 種類を追加した形となる。

これらの評価全体に対して、F 値のみを比較したものを図 5 に示す。

表 5 評価 1 - 提案方式・キャラ語尾辞書非適用

評価値	KSJ	KYT	NUCC1	NUCC2	M109B	M109M	平均
TP	62	70	88	75	93	45	72.2
FP	56	27	45	39	46	45	43
FN	50	68	44	59	59	86	61
TN	33	35	27	35	55	33	36.3
Rec	0.554	0.507	0.667	0.56	0.612	0.344	0.54
Prec	0.525	0.722	0.662	0.658	0.669	0.5	0.623
F	0.539	0.596	0.664	0.605	0.639	0.407	0.575

表 6 評価 2 - 提案方式 - キャラ語尾辞書適用

評価値	KSJ	KYT	NUCC1	NUCC2	M109B	M109M	平均
TP	68	70	85	80	96	73	78.6
FP	62	27	50	40	47	59	48.1
FN	40	68	42	54	55	47	51
TN	31	35	27	34	55	30	35.3
Rec	0.629	0.507	0.669	0.597	0.635	0.608	0.607
Prec	0.523	0.721	0.629	0.667	0.671	0.553	0.628
F	0.571	0.596	0.649	0.63	0.653	0.579	0.613

表 7 評価 3 - 比較方式 (パターンマッチ) - キャラ語尾辞書非適用

評価値	KSJ	KYT	NUCC1	NUCC2	M109B	M109M	平均
TP	5	5	5	3	21	8	7.8
FP	29	27	18	37	31	24	27.7
FN	128	134	149	131	140	138	136.7
TN	39	34	32	37	61	39	40.3
Rec	0.038	0.036	0.032	0.022	0.13	0.055	0.052
Prec	0.147	0.156	0.217	0.075	0.404	0.25	0.208
F	0.06	0.058	0.056	0.034	0.197	0.09	0.083

表 8 評価 4 - 比較方式 (パターンマッチ) - キャラ語尾辞書適用

評価値	KSJ	KYT	NUCC1	NUCC2	M109B	M109M	平均
TP	16	8	35	26	21	19	20.8
FP	102	98	74	85	146	89	99
FN	65	85	75	77	57	82	73.5
TN	18	9	20	20	29	19	19.2
Rec	0.198	0.086	0.318	0.252	0.269	0.188	0.219
Prec	0.136	0.075	0.321	0.234	0.126	0.176	0.178
F	0.161	0.08	0.32	0.243	0.171	0.182	0.193

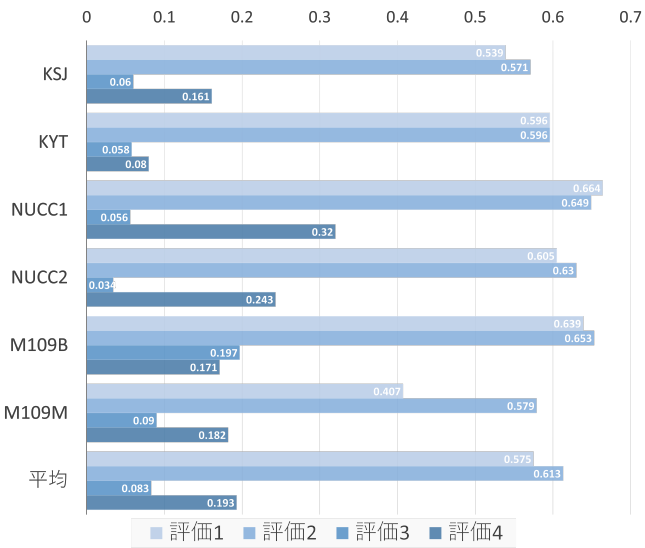


図 5 評価ごとの F 値比較

## 5 考察

### 5.1 提案方式の有効性

図 5 より、提案方式において、従来方式と比較して語尾の抽出率が高く、F 値も向上していることから、その有効性について確認することができた。

今回従来方式としたパターンマッチに基づく語尾抽出方式はいささか語尾パターン自体が少ないことも理由となるが、例えば「です」という語尾に関しても「でしょうか、でしょうかね、ですね、ですかね、でしょ、でしょう、でしょー」のように、多くの変化型があり、これらを全てパターンで抽出することは多大なコストが懸念されるため、提案方式により効率化されたと考えられる。

ただし、提案方式の場合でも、例えば「～ちゃうやろうな」という文末の文章の場合、「ちゃう」「やろ」「う」「な」と分割されるが、語尾として「～やろうな」を GT として設定していた場合においても、「やろ」が動詞、「う」が助動詞と辞書で定義されるため、今回の助動詞・終助詞に限った方式においては「うな」が語尾候補として抽出される。これは語尾としては一般的な認識として適当ではないと考えられる。

そのため、今後の課題として、助動詞・終助詞に限らず動詞などの活用形などを考慮すべきかが考えられる。

## 5.2 漫画ジャンルによる差異とキャラ語尾辞書の有効性

M109B と M109M においては、漫画としてのジャンルがシリアスなファンタジー漫画とギャグ漫画という形で分かれており、その中身のテキストの性質の差異によりスコアに影響が出ていることが分かった。

登場人物が一般的な標準語に近いキャラクタ性を持っている場合は、創作物というカテゴリにおいても抽出の難易度に差異はほとんど生じない。しかし、ギャグ漫画のように特殊な発言が使用される場合、例えば、「～だーっ」や「～ねーっ」などの特徴的な文末表現が行われるため、特に M109M のデータにおいてはキャラ語尾辞書が存在しないことによる影響が強く現れていることが確認できた。

その他、今回のキャラ語尾辞書の制作にあたっては当該の漫画「燃えるお兄さん」を参照せずに作成したが、別の創作物のキャラクタの語尾である「～なのだ」「～のだ」とほぼ同じ喋り方をしていた影響により、キャラ語尾辞書を利用した際の F 値の向上に繋がったことが確認できた。

ただし、キャラ語尾辞書の特性上、今回の方式では検知しきれない語尾も存在していることが予想される。そのため精度がより必要な場合には、未知の類似したキャラ語尾を自動的に辞書に追加する方式などを追加することで精度の向上を図るなどといった改善案が考えられる。

また、NUCC1 以外のデータにおいて F 値が向上する傾向にあったことから、事前にこのように特徴的なキャラ語尾を調査し、辞書として構築するのは語尾抽出方式全体として効果があるかと予想される。

## 5.3 キャラ語尾辞書適用・非適用による F 値の変化

KYT および、NUCC1 とともに、キャラ語尾辞書を適用したことによっての改善が行われていないことが読み取れる。

まず、KYT に関しては京都弁が含まれるデータセットとして作成したが、内部の確認を行った所、話し方などが基本的な MeCab 上の辞書で取得できる範囲での語尾が大部分となっており、キャラ語尾辞書によるスコア上の影響が少ない状態となっていた。また、その上でキャラ語尾辞書を確認したところ、京都弁に類する語尾が少なく、追加で獲得できた語尾がテストデータの範囲内で存在しない状態となっていた。そのため、今後はこうした方言に注力したキャラ語尾辞書の補充をすることでのスコアの向上や、もしくはより創作物に特化する形の研究方針への移行という形が考えられる。

次に、NUCC1 に関して内部の分析を行った結果として、一部 FP(過検出)として抽出されているものが、キャラクタ性が含まれる語尾としては正しい結果であることが判明した。そのため、実質的にはキャラ語尾辞書により精度は改善しているものと思われる。当該データセットの発言はため口や、ら抜き言葉などが多いカジュアルなものが大多数を占め、語尾抽出の正解データの設定の規定が難しかったことが要因となっている

と考えられるため、キャラクタ性の含まれる語尾の定義や正解データ作成の方針などを再考することとした。

## 6 おわりに

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性を構築する際に重要となる語尾を自動的に抽出する方式について提案した。

提案方式について評価を行った結果、F 値として 0.613 という値を得ることができ、従来方式と比較して有効性について確認することができた。

今後の予定であるが、提案方式によって抽出された語尾を用いたキャラクタごとの語尾のカテゴリ分け機能やその結果を用いてキャラクタ性を付与した対話システムの構築及び評価を行っていく予定である。

### 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] 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.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] T.Kudo: MeCab, <http://taku910.github.io/mecab/> (アクセス日:2020/12/16)
- [10] Python Software Foundation: Python, <https://www.python.org> (アクセス日:2020/12/12)
- [11] R.Leonard: beautifulsoup4, <https://pypi.org/project/beautifulsoup4/> (アクセス日:2020/12/12)
- [12] T.Kim, W.Kevin: emoji, <https://pypi.org/project/emoji/> (アクセス日:2020/12/12)
- [13] pandas-dev: pandas, <https://pypi.org/project/pandas> (アクセス日:2020/12/12)
- [14] T.Uchida: Janome, <https://mocabeta.github.io/janome> (アクセス日:2020/12/12)
- [15] ヘファナン・ケビン: 関西弁コーパスの紹介, 総合政策研究 No.41, pp.157-164. 2012.
- [16] 藤村逸子, 滝沢直宏編: 言語研究の技法: データの収集と分析, ひつじ書房. 2011.
- [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.