

印象の合成に基づくキャラクタ性表現モデルの検討

宮崎 千明¹ 平野 徹¹ 東中 竜一郎¹ 松尾 義博¹ 佐藤 理史²

¹ 日本電信電話株式会社 NTT メディアインテリジェンス研究所

² 名古屋大学大学院工学研究科

{miyazaki.chiaki, hirano.tohru, higashinaka.ryuichiro, matsuo.yoshihiro}@lab.ntt.co.jp, ssato@nuee.nagoya-u.ac.jp

1 はじめに

ユーザの意図に応じて適切な応答を返すことのできる対話エージェントは、検索等のための便利なツールとしてだけでなく、コミュニケーションのツールとしても期待され始めている。会話相手としての価値や娯楽性を高めるには、エージェントに対して何らかのキャラクタ設定(人物像)を与え、より人間らしく親しみやすい存在にすることが重要だと考えられる。

従来の対話エージェントサービスでは、様々なキャラクタに対応した発話データを人手で作成してきたが、全ての発話を人手で作成するのはコストが大変高い。そこで我々は、特定のキャラクタ性(人物像らしさ)を持った発話の自動生成に取り組み、なるべく人手をかけずに、大量に、目的のキャラクタに適した発話データを作成できる仕組みの実現を目指している。

日本語の発話テキストに対して特定のキャラクタ性を自動的に付与することを目的とした先行研究[5, 6, 7]では、キャラクタごとの言語表現の使用パターンを学習するために、個々のキャラクタに応じたテキストコーパスを用意する必要があり、キャラクタの種類を増やすのが容易ではなかった。

我々はこの課題を解決するために、キャラクタごとのテキストコーパスを用意せずに、簡便に多様なキャラクタ性を感じさせる発話を生成する手法を検討している。本研究においては、キャラクタと言語表現との間に、両者を一段抽象化した「印象」のレイヤを設け、印象の合成によってキャラクタ性を表現し分けるモデルを提案する(例えば、くだけた + 柔らかい + 明るい + 軽い = 若者キャラ など)。さらに、評価実験を通して、提案するモデルの下地となる仮説の妥当性を検証するとともに、言語表現に表れるキャラクタ性のうち、提案する観点では捉えきれない部分があるか、あるとしたらどのような部分であるかなど、このモデルの改善点について考察する。

なお、本研究において「キャラクタ」とは、話し手・書き手の「人物像」を指す。例えば、「年齢、性別、職業、階層、時代、容姿・風貌、性格等」[1]の要素で構成されるものを想定している。1つの構成要素のみが設定された「女性」のようなものもあれば、複数の構成要素が組み合わさった「女性・20代・大学生・陽気」のようなものもある。

2 関連研究

2.1 キャラクタに応じた発話生成

特定のキャラクタにふさわしい言語表現の生成を目的とした研究には、人物属性らしさ(年齢・性別・居住地)[5, 6]や、特定の話者らしさ(話者の個人性)[7]、話者の性格[3]を対象とした研究がある。

人物属性らしさを対象とした研究[5, 6]や、特定の話者らしさを対象とした研究[7]は、目的の人物像に合った言語表現の使い方を学習するために、キャラクタごとのテキストコーパスを必要とするものであり、すなわち本研究で解決したい課題をはらんでいる。

話者の性格を対象とした研究[3]では、外向性/内向性に関わる言語的特徴(発話内容の評価極性、統語構造の複雑さなど)に関するパラメータの設定によって外向性/内向性が感じられる発話を生成する言語生成器(対象言語は英語)の設計を行っている。本研究では言語的特徴をさらに一段抽象化した「印象」を調整可能なパラメータとして利用することを考える。

2.2 キャラクタと印象、印象と言語表現

特定のキャラクタにふさわしい発話の生成を目的とした研究には、話者の印象を対象とした研究[8]もある。この研究では、まじめさ、かわいさ/かっこ良さ、感情豊かさという観点で、印象形成に有効な言語パターンを音声転記テキストから人手で抽出し、その言語パターンを用いて作られたテキストから目的の印象を感じ取ることができるかを主観評価している。この研究

表 1: 本研究で使用する印象体系．小磯ら (2011) の評定尺度を基に，1 軸につき 5 段階の評価語を作成．印象軸の名称は筆者（宮崎）による．

印象軸	印象 (評価語)	
くだけ度	1: 改まった	~ 5: くだけた
柔らかさ	1: 硬い	~ 5: 柔らかい
軽さ	1: 重い	~ 5: 軽い
個性	1: 型にはまった	~ 5: 個性的な
話しことば度	1: 書きことば的な	~ 5: 話しことば的な
興奮度	1: 冷静な	~ 5: 興奮した
明るさ	1: 暗い	~ 5: 明るい
冗長さ	1: 簡潔な	~ 5: 冗長な
雑然度	1: 整然とした	~ 5: 雑然とした
わざとらしさ	1: 自然な	~ 5: わざとらしい
単調さ	1: めりはりのある	~ 5: 単調な
テンポ	1: テンポの良い	~ 5: テンポの悪い

では，印象（評価語）を用いて「キャラクタ像」を捉えており（かわいさ = 女性など），印象がキャラクタ性を捉える上で有用な要素であることが暗示されている．

言語表現の与える印象を扱う研究としては，小磯ら (2011) [2] のものがある．小磯らは『現代日本語書き言葉均衡コーパス』に収録されている新聞・雑誌・ブログ等の様々なテキストから人間が感じ取る印象（評価語）を収集し，その中から，テキストの多様性を捉えるための分類指標として有用な 12 軸の評定尺度（評価語の対）を求めている．これら 12 軸は書き言葉を対象として求められたものであるため，対話エージェントの発話テキストを扱う上で適切なものであるかは自明ではないが，本研究では，印象の合成によってキャラクタ性を表現し分けるモデルを考える上での出発点として，これら 12 軸 24 評価語を基にして作成した印象体系（12 軸 60 評価語から成る）を利用する．本研究で使用する印象体系を表 1 に示す．

3 言語表現が与える印象の合成に基づくキャラクタ性の表現モデル

3.1 提案モデルの概要

本節では，本研究で提案するキャラクタ性表現モデルについて，関連研究 [5, 6, 7] で用いられている従来のモデルと比較しながら説明する．図 1 に従来のモデルと本研究で提案するモデルの直観的なイメージを示す．

従来のモデルでは，目的のキャラクタごとに，そのキャラクタの言語表現の使い方を把握するための発話テキストコーパスを用意する必要があり，発話生成の対象キャラクタを新たに増やす際には，そのキャラクタの発話テキストを収集するコストがかかってしまう．

これに対し，提案モデルでは，キャラクタを印象という構成要素に分解し，それらの印象に応じた言語表現を用いた発話を生成することで，生成された発話に

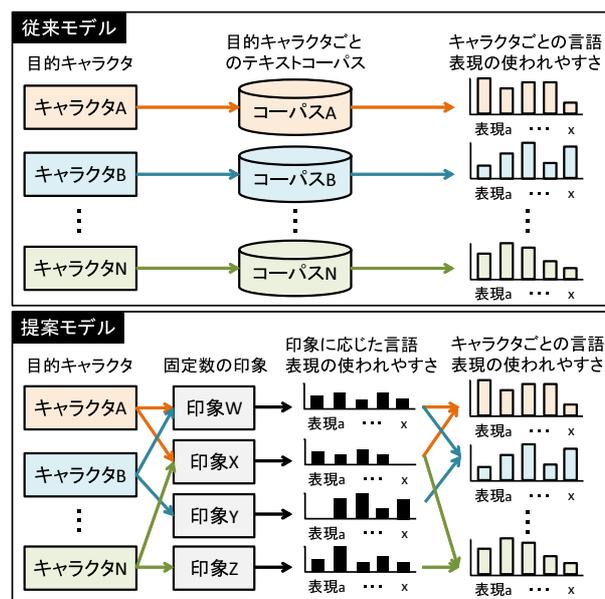


図 1: 従来モデルと提案モデルの比較

おけるキャラクタ性を表現することを考える．このとき，キャラクタの特徴を分解するために用いられる印象体系は，表 1 に示されるような固定数の印象（本研究では 12 軸 60 評価語）で構成されることを想定している．このモデルでは，予め定められた個数の印象（評価語）ごとに，その印象を与えるテキストにおいてどのような表現が使われやすいかを把握しておきさえすれば，発話生成の対象キャラクタを増やす際には，新たな発話テキストコーパスを収集する必要はない．なお，キャラクタから印象への分解は，対話エージェントの運営・作成者が人手で行うことを想定している．

3.2 提案モデルの前提となる仮説

我々は，この提案モデルがうまく機能するには，下記の 2 つの仮説が成り立つことが前提となると考える．

1. 聞き手・読み手（日本語母語話者）は，日本語の言語表現の違いから，印象の違いをある程度共通して感じ取ることができる
2. 印象の違いによって，キャラクタ性の違いを区別することができる

1 つ目の仮説は，言語表現の与える印象を意図的に操作することができるかどうかにかかわる仮説である．そもそも聞き手・読み手が日本語の言語表現から共通した印象を感じ取ることができないのであれば，言語生成器が発話生成で用いる言語表現やその使用頻度をいくら調整しても，生成された発話の聞き手・読み手に対して，対話エージェントの運営・作成者が意図した通りの特定の印象を感じ取らせることはできない．

2つ目の仮説は、キャラクターの特徴を分解する構成要素として、現段階で想定している印象体系(表1)で十分であるかどうかにかかわる仮説である。予め用意された印象(評価語)の組み合わせの違いに基づいて、言語表現に表れるキャラクター性を区別できないのであれば、本研究ならびに提案モデルの目的は達成されない。この場合、キャラクターを何らかの構成要素に分解し、その構成要素の合成によってキャラクター性を表現するというモデルを成立させるためには、キャラクター性の区別に寄与する印象軸をさらに追加するか、印象以外の構成要素の追加を検討する必要があるだろう。

4 仮説検証実験

3.2節の仮説を検証するため、2種類の実験を行った。

4.1 実験1: 印象判定の評価者間一致度の測定

聞き手・読み手が日本語の言語表現から共通した印象を感じ取ることができるかを検証するために、発話テキストに対する印象判定の評価者間での一致度を測定した。評価対象は、平均約22文字で構成される1000発話である。1000発話は、意味内容が同一で言語表現が異なる10発話(同一の発話を、指定された10種のキャラクターらしさを強調するように人手で書き換えて作成したもの[4])を1組として評価者に提示し、発話同士を比較しながら、各発話に対して言葉遣いから感じ取った印象を付与させた。付与させた印象は、表1に示す12軸60評価語である。評価者は筆者らを含まない2名の日本語母語話者(評価者A, B)である。

評価指標としては、評価者2名による評価の間での κ 係数、およびポリコリック相関係数を用いた。 κ 係数は、2名の評価の偶然によらない一致度を測るための指標で、今回は、各印象軸の両極の評価語に該当する評価1と5のみを対象として κ 係数を算出した。ポリコリック相関係数は、順序尺度(5段階評価の値)同士の相関を測る目的で用いた指標である。これによって評価者2名の評価の間に関係があるかを確認した。

評価の結果を表2に示す。評価1と5のみを対象として算出した κ 係数は、「わざとらしさ」と「テンポ」以外の10軸については中程度の一致以上の値(0.41以上)を示しており、各印象軸の両極の印象(評価語)は、評価者2名の間で一致して判定されていることが分かった。5段階全ての評価語を対象として算出したポリコリック相関係数は、「テンポ」を除く11軸については弱い相関(絶対値が0.2~0.4)から中程度の相関(絶対値が0.4~0.7)を示しており、評価者2名の評価の間には関係があることが確認できた。

以上の結果から、我々は、3.2節で示した1つ目の

表 2: 印象判定の評価者間一致度

印象軸	κ 係数 (評価1対5)	ポリコリック 相関係数
くだけ度	0.77	0.64
柔らかさ	0.81	0.53
軽さ	0.85	0.62
個性	0.56	0.51
話しことば度	0.67	0.51
興奮度	0.78	0.33
明るさ	0.77	0.45
冗長さ	0.97	0.52
雑然度	0.97	0.35
わざとらしさ	0.03	0.22
単調さ	0.45	0.32
テンポ	-0.38	-0.16

仮説「聞き手・読み手(日本語母語話者)は、日本語の言語表現の違いから、印象の違いをある程度共通して感じ取ることができる」は成り立つと考える。印象軸「わざとらしさ」の一致度が低かったことについては、評価対象の発話が、キャラクターらしさを人為的に強調したものであったため、そもそも大半が「自然なものではなかったことが原因として考えられる。「テンポ」の一致度が低かったことについては、評価対象とした発話が平均約22文字で構成される短いものであったため、「テンポ」の良し悪しを判断することが困難だったことが原因ではないかと考える。これら2軸の印象については、今後も引き続き分析を行いたい。

4.2 実験2: 印象によるキャラクター分類

印象の違いによってキャラクター性の違いを区別することができるかどうかを検証するために、実験1で付与された印象12軸の値(評価語)が一致する発話をグループ化し(以降これを「印象グループ」と呼ぶ)、1つの印象グループにいくつのキャラクター性を持った発話が属しているかを調べた。

評価対象は、実験1で用いた1000発話のうち360発話である。この360発話の各々に対し、実験1と同一の評価者2名に、年齢(男性/女性)、性別(10歳/25歳/50歳/75歳)、会話相手との親密度(親密/非親密)の3属性で構成される10種のキャラクター(女性・10歳・親密など)[4]のうち最もふさわしいと考えるものを付与させた。これにより、人間には違いを感じとることができるのに、提案する印象体系では区別できないキャラクターがないかを確認する。

実験結果を表3に示す。評価者Aの場合は275、評価者Bの場合は344の印象グループがあり、そのうち、グループに属する発話に対して全て同一(1種類)のキャラクターが付与されたものは、評価者Aで255グループ(約93%)、評価者Bでは336グループ(約98%)であった。大半の印象グループについては特定

表 3: 各印象グループに属する発話に付与されたキャラクターの異なり数, およびキャラクターの異なり数別の印象グループ数

キャラクターの異なり数	印象グループ数	
	評価者 A	評価者 B
1	255 (92.6%)	336 (97.7%)
2	15 (5.5%)	7 (2.0%)
3	4 (1.5%)	1 (0.3%)
4	1 (0.4%)	0 (0.0%)
計	275 (100.0%)	344 (100.0%)

表 4: 発話に付与された印象とキャラクターの例 (発話中の*は固有名)

印象	評価者	発話	キャラクター
くれた 柔らかい 軽い 個性的 話しことば的 興奮した 明るい やや冗長な やや整然とした わざとらしい ややめりはり のある ややテンポの 良い	A	はいはいする姿は愛らしいわよね。	F50p
		はいはいしているのって、可愛いよね。	M10p
		経堂にある*というお店のティラミスは最高よ!	F75p
	B	私の友人に、すごい*ちゃんファンがいるわよ。	F50p
		*水族館のジンベエザメはすごい迫力だったよ。	M25p
		最近はライトアップもしてるらしいよ。	F25p
		地域おこしにいるんな B 級グルメが開発されてるよね。	F25p

のキャラクター性との結び付きが確認できたが、評価者 A の場合は約 7%, 評価者 B の場合は 2% のグループにおいて、異なるキャラクターだと判断された発話が混在しており、現状の印象体系では、完全にはキャラクター性を区別しきれていない。

提案モデルおよび印象体系の改善に向け、評価対象の発話と付与された印象およびキャラクターを観察した。表 4 の発話は全て同じ印象 (評価語) が付与されたものであるが、それぞれの発話に付与されたキャラクターは異なっている。この例は評価対象とした発話の一部でしかないが、この例からは少なくとも、「わよ」のような女性に特有の表現や「愛らしい」のような古風な表現を捉えることが、性別や年齢に関係する特徴を分解するうえで必要だと分かる。

参考情報として、実験 2 で用いた 360 発話について、評価者 A によって付与された印象 (5 段階) の誤差の分布を図 2 に示す。この誤差は、キャラクター (実験 2 で付与) ごとの発話の印象 (実験 1 で付与) の平均値と、各発話に付与された印象の値との差である。図 2 より、どの発話からも、キャラクター内で一貫した印象が感じ取られた軸もあれば (「雑然度」「くだけ度」は誤差 ±1 点以下の発話が 80% 以上)、発話ごとに印象がばらつく軸もある (「興奮度」「軽さ」などは 40% 以

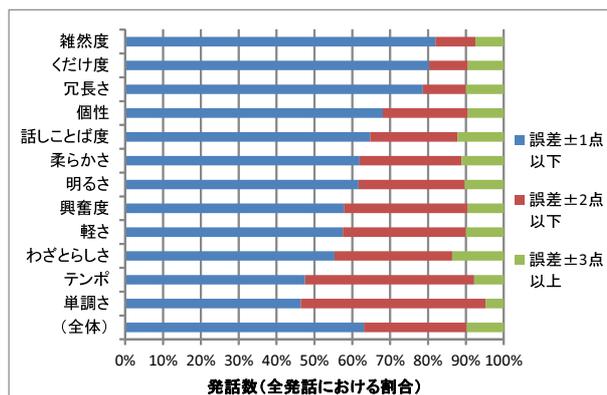


図 2: 評価者 A によって各発話に付与された印象 (5 段階) のキャラクター別平均値からの誤差の分布

上の発話が誤差 ±1 点を超える) ことが分かる。提案モデルに基づく言語生成器を設計する際は、こうした印象軸ごとの性質の違いを考慮する必要がある。

5 まとめ

本研究では、キャラクターと言語表現との間に、両者を一段抽象化した「印象」のレイヤを設け、「印象」の合成によってキャラクター性を表現し分けるモデルを提案し、キャラクター性の分解に必要な印象 (および構成要素) を検討した。今後は、モデルの改善およびこのモデルに基づく言語生成器の設計を行う。

参考文献

- [1] 金水敏. ヴァーチャル日本語役割語の謎. 岩波書店, 2003.
- [2] 小磯花絵, 田中弥生, 小木曾智信, 近藤明日子. テキストの多様性をとらえる分類指標の体系化の試み. 言語処理学会第 17 回年次大会発表論文集, pp. 683-686, 2011.
- [3] François Mairesse and Marilyn Walker. PERSONAGE: Personality generation for dialogue. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pp. 496-503, 2007.
- [4] 宮崎千明, 平野徹, 東中竜一郎, 牧野俊朗, 松尾義博, 佐藤理史. 話者のキャラクター性に寄与する言語表現の基礎的分析. 言語処理学会第 20 回年次大会発表論文集, pp. 232-235, 2014.
- [5] 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ー性変換. *人工知能学会論文誌*, Vol. 31, No. 1, 2016.
- [6] Chiaki Miyazaki, Toru Hirano, Ryuichiro Higashinaka, Toshiro Makino, and Yoshihiro Matsuo. Automatic conversion of sentence-end expressions for utterance characterization of dialogue systems. In *29th Pacific Asia Conference on Language, Information and Computation*, pp. 307-314, 2015.
- [7] 水上雅博, Graham Neubig, Sakriani Sakti, 戸田智基, 中村哲. 特徴的話を対象とした言語的個人性変換. 情報処理学会研究報告. 自然言語処理研究会報告, No. 20, pp. 1-7, 2014.
- [8] 沈睿, 菊池英明, 太田克己, 三田村健. 音声生成を前提としたテキストレベルでのキャラクター付与. *情報処理学会論文誌*, Vol. 53, No. 4, pp. 1269-1276, 2012.