

特徴的話し手を対象とした言語的個人性変換

水上 雅博^{1,a)} Graham Neubig^{1,b)} Sakriani Sakti^{1,c)} 戸田 智基^{1,d)} 中村 哲^{1,e)}

概要: ユーザや状況に合わせて、対話システムが個人性を発揮することは円滑なコミュニケーションを行う上で重要である。我々は、話し言葉を対象として、機械翻訳の技術を用いて文章における個人性を変換する手法について提案し、その有効性を示してきた。しかしながら、今まで提案してきた手法では、文脈に意味を与えないフィラー等の機能語や、一般的な言い換えのみを変換の対象としていた。そのため、変換の目標話者に依らず常に一意な翻訳モデルを利用して、目標話者のみが用いる独特な言語表現は変換できなかった。そこで本稿では、目標話者のテキストに対して大規模で一般的な話し言葉の言語モデルを用いた特徴的分析を行うことで、目標話者のみが用いる独特な言語表現を抽出し、さらに、先行研究で有効性の示されている n -gram 類似度を用いた翻訳モデル確率の付与を行うことで、目標話者により適した翻訳モデルの構築を提案する。また、提案法について先行研究同様に主観評価実験を通して有効性を示す。

1. はじめに

円滑な意思疎通を行う上で、伝えたい内容のみならず、その内容を伝えるために用いる語彙や、発話の際に用いる韻律や声質など、内容に直接しない要素も非常に重要な役割を果たす。これらの要素は人間とコンピュータのコミュニケーションにおいても重要であると考えられる。例を挙げれば、対話システムにユーザの望む性格や性別、または特定の有名人、キャラクターの話し方を再現することで、より良い印象を与えられる。そのため、対話システムにおいて、話し方を適切に制御する技術の構築が望まれる。

対話システムの応答生成に関連する技術の一つである音声合成においては、非言語情報に着目した研究が盛んに行われている [1]。特に、話者性に関しては、個人性を考慮した音声合成の研究が進んでおり、その発展として、話者の声質を考慮した音声翻訳システムなども実現されている。その一方で、話し言葉特有の言語表現や、文章における話者ごとの特徴（以下、言語的個人性と呼ぶ）までも制御する研究は少なく、社交性や知性など、個人性のごく一部の要素を考慮し、ルールベースの文章生成でその特徴を再現する研究 [2] などに限定される。個人性を扱うに当たって、このような心理学的要素を個人性の表現モデルとして汎化

した場合、目的の個人性を柔軟に制御できるものの、特定の個人を再現することや架空のキャラクターなどのステレオタイプな個人性を再現できない問題がある。

そこで我々は、統計的機械翻訳の技術を用い、話し言葉の書き起こし文章に対して言語的個人性を変換する手法（以下、言語的個人性変換と呼ぶ）を提案している [3], [4]。言語的個人性変換は、文章生成ではなく文章変換を行うことで既存の対話システムの内部に変更を加えることなく個人性が制御可能となる。また、対訳コーパスを必要とせず、言い換え表現を集めた翻訳モデルと目標話者の言語モデルを学習する発話コーパスのみで利用できる。しかしながら、今までの言語的個人性変換において、翻訳モデルは文脈の意味に影響を与えないフィラー等の機能語や、一般的な言い換えに限定されており、目標話者のみが用いる独特の言語表現（以下、特徴語と呼ぶ）については変換できない。

本稿では、先行研究に比べて特徴的な話し手を対象とした変換を実現するために、目標話者の話し方に関する特徴の分析を行い、その結果から特徴語を言い換える方法について提案する。目標話者の話し方に関する特徴の抽出では、大規模で一般的な話し言葉の言語モデルと目標話者の発話コーパスから学習された言語モデルを比較することで、特徴語を抽出する。抽出された特徴語を対象に、先行研究で有効性が示された n -gram 類似度を用いた類義語クラスタリングを用いて言い換えを抽出し、目標話者の話し方に適合した翻訳モデルを構築する。また、構築された翻訳モデルを利用して言語的個人性変換を行い、その結果から提案法の有効性を示す。

¹ 奈良先端科学技術大学院大学
Nara Institute of Science and Technology.

a) masahiro-mi@is.naist.jp

b) neubig@is.naist.jp

c) ssakti@is.naist.jp

d) tomoki@is.naist.jp

e) s-nakamura@is.naist.jp

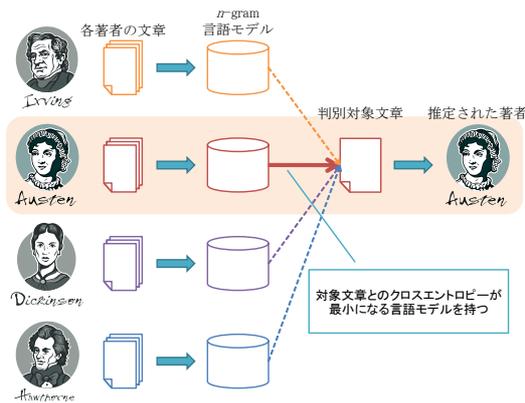


図 1 言語モデルを用いた文章の著者推定の概要 [7]

Fig. 1 The research of author identification using language models.

2. 言語的個人性に関する研究

言語的個人性に関する研究は言語学等の分野で行われており、それらの研究の結果として機能語が個人性を認識するための重要な要素であることが明らかとなっている [5], [6].

言語的特徴と個人性の関連性をより明らかにする関連研究として、言語モデルを用いた文章の著者推定 [7] があげられる。文献 [7] では言語特徴を「言葉の指紋」と表現し、言語特徴がそれぞれの著者において唯一無二のものであると提言している。著者推定の具体的な手法として、与えられたテキストと各著者の言語モデルを用いてエントロピーを計算し、テキストに対して最も著者らしい言語モデルを推定する。この手法の概要を図 1 に示す。

また、対話エージェントや対話システムに限定した個人性の制御を行う関連研究では、Isard らによる対話生成における個人性とアライメント [8] がある。Isard らの手法では、対話エージェントの持つ複数の出力候補に対して、外交性、神経質、情緒安定性などの性格を示す 5 大要素 [9], [10] のそれぞれを表す言語モデルを構築し、目的の性格となるような出力を言語モデルの尤度に基づいて選択する。こちらも個人性を分析するための手法として n -gram を用いており、 n -gram が個人性を推定するモデルとして有効であることが分かる。

さらに、言語全般における個人性の制御を行う関連研究では、個人性を考慮した文章生成の研究が行われている [2]. この手法では、Isard らの用いた性格を示す 5 大要素に基づいて、複数の話者から集めたコーパスに対してそれぞれの性格をタグ付し、それぞれの性格タグを与えられたコーパスから性格と関係する言語的特徴を分析する。また、言語生成の際に言語的特徴をルールベースで考慮し、任意の性格を持つ話者の文章を生成する。

しかしながら、個人性を考慮した文章生成の問題として、個人性を性格の要素の組み合わせで表現した場合、特定の

有名人や架空のキャラクター等の特定個人の個人性、および日本語を喋る中国人、老人、武士などの非現実的なステレオタイプ性を持つ個人性を再現することはできない。本研究では、このような特定の話者の個人性を捉える手法に着目する。

3. 言語的個人性変換

言語的個人性変換 [3], [4] では、与えられたテキスト V を目的の話者の個人性を持つテキスト W へと変換する。基本となる変換のモデルは、統計的機械翻訳において代表的なモデルである雑音のある通信路モデルを用いる。雑音のある通信路モデルは、同言語間における翻訳の関連研究である話し言葉の整形 [11] でも利用されている。

統計的機械翻訳では翻訳モデルの学習の際に変換する要素を対としたコーパス（以下、対訳コーパス）を要する。一般的には、同じ意味の文章を英語と日本語で対とした日英対訳コーパス等を利用する。しかしながら、言語的個人性変換において翻訳モデルの学習を行うためには「同じ意味でありながら別の話者の個人性を持つ」特殊な対訳コーパスを必要とする。そのような対訳コーパスを収集するのは容易ではなく、対訳コーパスから翻訳モデルを学習するのは困難である。この問題に対して、先行研究 [3], [4] では、 n -gram やシソーラス、日英対訳コーパス等から学習された言い換え資源を用いて構築した擬似的な翻訳モデルを構築している。

言語的個人性変換のモデル化では、十分な量が確保可能な目標話者の個人性を持つ言語モデル確率 $P(W)$ と、類義語とその言い換え確率によって構築された擬似的な翻訳モデル確率 $P(V|W)$ の二つを用いて、事後確率 $P(W|V)$ を以下のようにモデル化する。

$$P(W|V) = \frac{P(V|W)P(W)}{P(V)} \quad (1)$$

与えられた V に対して $P(W|V)$ が最大となる \hat{W} を探索する。 $P(V)$ は W の選択によらず変動しないため、以下のように表せる。

$$\hat{W} = \operatorname{argmax}_W P(V|W)P(W) \quad (2)$$

しかしながら、翻訳モデルにおいて W と V が異なる文章長を持つような変換が行われる場合、言語モデル確率を最大化するために文章を短くするような変換候補ばかりが選択され、不当に短い変換結果が出力されることがしばしばある。これを防ぐために、機械翻訳では出力単語数に比例してスコアに加算されるワードペナルティを導入し、文章の単語数を制御している。本研究でも同様に、変換モデルにおける候補 \hat{W} のスコアを、対数スケールで表現される事後確率 $P(W|V)$ とワードペナルティ WP と \hat{W} の単語数 N の積を用いて、以下のように定義する。

$$\hat{W} = \underset{W}{\operatorname{argmax}} \log P(V|W) + \log P(W) + WP * N \quad (3)$$

この \hat{W} の探索を行う具体的な方法は、図 2 に示すような探索グラフを作成し、各単語または単語列の言い換えをに対して $P(W|V)$ が最大となる経路 \hat{W} をビタビアルゴリズムを用いて探索する。

また、ワードペナルティ WP のチューニングを行う。具体的には、チューニングコーパス $\mathcal{V} = \{V_1, \dots, V_k\}$ に対して、様々な WP を用いた際の出力 $\hat{W} = \{W_1, \dots, W_k\}$ を計算し、評価関数 ω が最も大きくなるものを選択する。

$$\begin{aligned} \hat{W}_k &= \underset{W}{\operatorname{argmax}} P(W|V_k; WP) \\ \hat{WP} &= \underset{WP}{\operatorname{argmax}} \omega(\hat{W}; WP) \end{aligned} \quad (4)$$

ここで利用する評価尺度は、5.1.2 節で述べるエントロピーである。

3.1 個人性評価のための目的言語モデル

目的言語モデルは言語的個人性変換において、変換候補の文章がどの程度目標話者らしいかを評価するために用いられる。言語モデルによる個人性の評価が有効であることは、関連研究 [7] で示されている。従って、今回の研究においても n -gram 言語モデルを用いた個人性の評価は可能であると考えられる。しかしながら、目標話者の発話を集めて目的言語モデルを構築する上で問題となるのは、一人の話者から集めることのできる発話は限られており、内容も限定的である点である。このような少量の目標話者の発話のみから構築された言語モデルを用いて正確に言語モデル確率 $P(W)$ を推定することは困難である。

これらの問題を解決するために、大規模なコーパスから学習した言語モデルと、対象とするタスクやドメインに応じた少量のコーパスから学習した言語モデルを線形補間することにより、効率的に分野適応を行う手法 [12] の利用を提案する。つまり、本研究においては、少量の目標話者データから学習された言語モデル $P_t(W)$ と、大量の全話者混合データから学習された言語モデル $P_g(W)$ を線形補間することで、目的言語モデルを構築する。線形補間によって生成される言語モデル $P(W)$ は、学習に利用していないテスト用の目標話者データに対して最大の言語モデル確率が得られるように λ を計算することで得られる。

$$\begin{aligned} P(W) &= \lambda P_t(W) + (1 - \lambda) P_g(W) \\ 0 \leq \lambda \leq 1 \end{aligned} \quad (5)$$

3.2 言い換えの抽出による翻訳モデル

翻訳モデルを学習する際に、機械翻訳では異なる言語の

対訳コーパスを用いて、話し言葉の整形では話し言葉と書き言葉の対訳コーパスを用いている。しかしながら、言語的個人性変換においては、対訳コーパスを得ることは困難である。すなわち、翻訳モデルを先行研究と同様の手法で構築することは難しい。

そこで言語的個人性変換では、翻訳モデルの代用として、言い換えとその言い換えが起きる確率（以下、言い換え確率と呼ぶ）を用いる。言い換えを用いれば、文章の意味を変えない別の言語表現を探すことが可能となり、変換の際にはこれらの言い換え候補の中から目的の個人性に近いものを選択する。これにより、対訳コーパスを用いずに翻訳モデルと等価の変換が可能となる。言い換えの抽出手法については、次節で詳しく説明する。

4. 話者特徴の分析と翻訳モデルの構築

先行研究 [3], [4] では、変換の目標話者の言語的個人性が「全話者において共通の語彙上で再現される」と仮定し、目標話者に関わらず一意な翻訳モデルを構築していた。しかしながら、ステレオタイプなキャラクター性を持つ場合（例えば、「ござる」「候」等の武家言葉や「ちゃん」「だ〜」等のギャル語など）、目標話者における特徴的な語彙が全話者において共通の語彙上に存在しない場合、先行研究で用いた全話者一様の翻訳モデルでは目標話者の個人性を再現できない。そこで、本研究では目標話者の言語モデルと標準的な話者の言語モデルを比較し、目標話者における特徴的な語彙（以下、特徴語と呼ぶ）を分析し、特徴語に焦点を当てることで効率的に目標話者に適応した翻訳モデルの構築を行う。

4.1 χ^2 値による特徴語の抽出

本稿では、目標話者言語モデルと一般言語モデル間の n -gram 確率から χ^2 値を計算し、言語モデル間の差異に寄与する量が多い語を特徴語として抽出する手法を提案する。 χ^2 値とはピアソンの χ^2 値検定で用いられる統計量であり、自由度に依る χ^2 分布と比較することで「観測値の相対的頻度が理論値に従う」という帰無仮説を検証する。特徴語の抽出では、帰無仮説とは逆に「観測値（目標話者言語モデル P_t における、ある n -gram 確率 $P_t(\mathbf{w})$) が理論値（一般的な話者言語モデル P_g における、ある n -gram 確率 $P_g(\mathbf{w})$) に従わない」こと、すなわち目標話者の話し方が一般的な話者の話し方と異なることを検証する。さらに χ^2 値への個々の n -gram がどの程度寄与しているかを調べることで、一般的な話者と異なる話し方となっている要因、すなわち特徴語を明らかにする。具体的には、ある語 w における χ^2 値である χ_w^2 を計算し、 χ_w^2 の上位 100 語を特徴語とする。

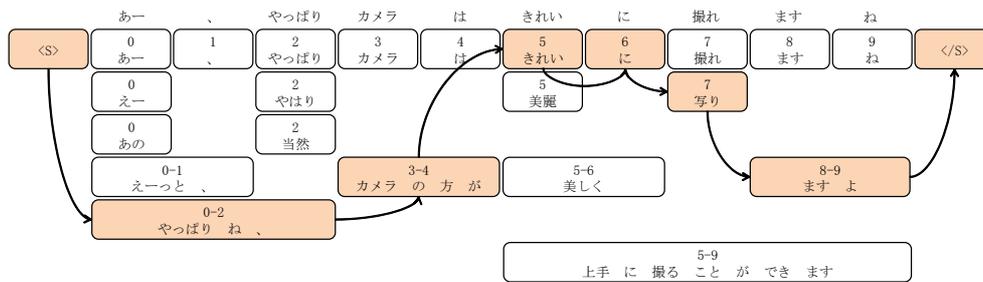


図 2 グラフ探索による変換の例

Fig. 2 Sample of transformation using a search graph.

$$\chi_w^2 = \frac{(P_t(\mathbf{w}) - P_g(\mathbf{w}))^2}{P_g(\mathbf{w})} \quad (6)$$

4.2 特徴語に関する言い換え抽出

抽出された特徴語の集合と、目標話者コーパス、一般的な話し言葉コーパスから特徴語に関する言い換えを抽出し、翻訳モデルを構築する。言い換えを用いた翻訳モデルの構築は、先行研究でその有効性が示されている [3], [4].

提案法では、 n -gram による類似語クラスタリング [13] を用いて、前後の文脈が類似しているフレーズを言い換え候補 \mathbf{w}_{pp} , \mathbf{v}_{pp} はそれぞれ 30 フレーズとする。特徴語に関する言い換え抽出の手順を以下に示す。

- (1) 一般的な話し言葉コーパスから文あたりの出現頻度の高いフレーズの集合 \mathbf{v}_{pp} を抽出する。
- (2) 目標話者コーパスから文あたりの出現頻度が高く、特徴語を含むフレーズの集合 \mathbf{w}_{pp} を抽出する。
- (3) あるフレーズ $\mathbf{v} (\in \mathbf{v}_{pp})$ と、特徴語を含むフレーズ $\mathbf{w} (\in \mathbf{w}_{pp})$ 間の n -gram 分布の距離 $D_{JS}(\mathbf{w}||\mathbf{v})$ を計算する。
- (4) 各フレーズ間の距離に対して周辺化を行い、言い換え確率 $P(\mathbf{w}|\mathbf{v})$ を求める。
- (5) 言い換え確率 $P(\mathbf{w}|\mathbf{v})$ が閾値以上のフレーズ \mathbf{v} , \mathbf{w} を言い換えとして抽出する。

なお、抽出する言い換えの候補は、できる限り多くの文章で言い換えが利用可能なように、文あたりの出現頻度が高いフレーズを選択する。

具体的な n -gram 分布の距離 $D_{JS}(\mathbf{w}||\mathbf{v})$ と、距離に基づく言い換え確率 $P(\mathbf{w}|\mathbf{v})$ の計算方法を示す。まず、一般的な話し言葉コーパスにおいて、あるフレーズ \mathbf{v} の出現数 $C_g(\mathbf{v})$ と、あるフレーズ \mathbf{v} の前後に単語 x, y が出現した数 $C_g(x, \mathbf{v}, y)$ を数え上げる。同様に、目標話者コーパスにおいても、あるフレーズ \mathbf{w} の出現数 $C_t(\mathbf{w})$ と、あるフレーズ \mathbf{w} の前後に単語 x, y が出現した数 $C_t(x, \mathbf{w}, y)$ を数え上げる。これらを利用して、一般的な話し言葉のコーパスにおいてあるフレーズ \mathbf{v} が生じたときに、その前後にある単語 x, y が生じる条件付き言語モデル確率 $P_g(x, y|\mathbf{v})$ と、目

標話者コーパスにおいてあるフレーズ \mathbf{w} が生じたときに、その前後にある単語 x, y が生じる条件付き言語モデル確率 $P_t(x, y|\mathbf{w})$ を求める。しかしながら、目標話者コーパスは一般的な話し言葉コーパスに比べ少量である点、特徴語を含むフレーズ \mathbf{w} が一般的な話し言葉コーパスには登場しにくい語である点を考慮し、ラプラススムージングを用いてそれぞれの条件付き言語モデル確率を計算する。なお、 N_g および N_t はそれぞれのコーパスにおける n -gram の異なり単語数である。

$$P_g(x, y|\mathbf{v}) = \frac{C_g(x, y|\mathbf{v}) + 1}{C_g(\mathbf{v}) + N_g} \quad (7)$$

$$P_t(x, y|\mathbf{w}) = \frac{C_t(x, y|\mathbf{w}) + 1}{C_t(\mathbf{w}) + N_t} \quad (8)$$

計算された条件付き言語モデル確率 $P_g(x, y|\mathbf{v})$ および条件付き言語モデル確率 $P_t(x, y|\mathbf{w})$ は、あるフレーズ \mathbf{v} , \mathbf{w} に対して、前後の文脈 x, y を考慮した確率となっている。すなわち、この両者の条件付き言語モデル確率が全ての前後の文脈 x, y において一致していれば、両者は類似の文脈において登場するフレーズであると言える。単語列 \mathbf{w}, \mathbf{v} の類似性を測るために、Jensen-Shannon ダイバージェンスを用いて距離 $D_{JS}(\mathbf{w}||\mathbf{v})$ を計測する。なお、Jensen-Shannon ダイバージェンスの係数 λ は 0.5 とする。このような確率分布間の距離尺度では他に Kullback-Leibler ダイバージェンスやピアソン距離などがあげられる。しかしながら、言い換え関係において両確率分布間に非対称な距離関係を与えることは妥当ではないため、対称性を持つ Jensen-Shannon ダイバージェンスを利用した。

$$\begin{aligned} D_{JS}(\mathbf{w}||\mathbf{v}) &= \frac{1}{2}D_{KL}(\mathbf{w}||\mathbf{v}) + \frac{1}{2}D_{KL}(\mathbf{v}||\mathbf{w}) \\ &= \sum_{x, y \in \mathcal{X}, \mathcal{Y}} \left(P_t(x, y|\mathbf{w}) - P_g(x, y|\mathbf{v}) \right) \log \frac{P_t(x, y|\mathbf{w})}{P_g(x, y|\mathbf{v})} \end{aligned} \quad (9)$$

前後の文脈が共通している場合、 $D_{JS}(\mathbf{w}||\mathbf{v})$ は小さくなり、フレーズ \mathbf{w}, \mathbf{v} は類似していると言える。

距離 $D_{JS}(\mathbf{w}||\mathbf{v})$ から言い換え確率を計算するためには、フレーズ \mathbf{v} を固定して、負の距離 $-D_{JS}(\mathbf{w}||\mathbf{v})$ のソフト

表 1 特徴語の言い換えの一例

Table 1 Sample of paraphrasing with Characteristic words.

v	w	$P(\mathbf{w} \mathbf{v})$
。	モン。	0.108
。	モン！	0.106
。	モン☆	0.076
。	だモン！	0.029
。	たモン！	0.027
計 30 語		

頻出語列			
入力	コーヒー	が 欲しい	です 。
出力	コーヒー	が 欲しい	モン ！
		特徴語	+フレーズ化

図 3 文章における言い換え関係の一例

Fig. 3 Sample of paraphrasing relation in text.

マックス関数をすべての言い換え候補 \mathbf{w}_{pp} について周辺化することで得られる.

$$P(\mathbf{w}|\mathbf{v}) = \frac{-\exp(D_{JS}(\mathbf{w}|\mathbf{v}))}{\sum_{\mathbf{w}' \in \mathbf{w}_{pp}} \exp(-D_{JS}(\mathbf{w}'|\mathbf{v}))} \quad (10)$$

全ての言い換え候補間における言い換え確率の計算が終わった時点で、言い換え確率 $P(\mathbf{w}|\mathbf{v}) > 0.01$ を条件として、言い換え候補の中から言い換えとして採用するフレーズ \mathbf{w}, \mathbf{v} の組を決定する. これにより、一般的な話し言葉と目標話者の特徴語との言い換えと、その言い換え確率を得ることができる.

得られた言い換えを表 1 に、文章における言い換え関係を図 3 に示す.

5. 評価の実験

5.1 実験条件

5.1.1 実験データ

実験の題材として、Twitter 上で活動する 3 名のキャラクターを目標話者に選択する. この題材選択の理由として、キャラクターがステレオタイプな個性を持つことを望める点、コーパスから目標話者の発話のみを抽出しやすい点、目標話者本人が書き起こし行為を行っており、発話内容と書き起こしに不一致がない点があげられる.

目標話者のコーパスは、それぞれのキャラクターのツイートを収集することで得られる. また、正確に目標話者の文章のみを抽出するために、以下の条件を設ける.

- 目標話者以外の文章を含むツイートの除外
- URL の除外
- ユーザ ID の除外
- ハッシュタグの除外

収集したコーパスを話者ごとに、変換時に利用する目的言語モデルの学習データ Train, ワードペナルティやエント

表 2 キャラクターコーパスの緒元

Table 2 Number of utterances and words in the character corpus.

用途	キャラクター	文章数	単語数
Train.	A	880	17.4k
	B	276	5.5k
	C	288	3.2k
Eval.	A	880	15.9k
	B	276	5.5k
	C	288	2.8k
Dev.	A	220	3.2k
	B	69	1.1k
	C	72	0.8k

表 3 一般コーパスの緒元

Table 3 Number of sentences and words in BTEC, and REI-JIRO.

用途	文章数	単語数
Train.	390k	5262k
Eval.	390k	5178k
Dev.	97k	1281k
Test1.	10	129
Test2.	97k	1322k

ロピー計算に利用する評価言語モデルの学習データ Eval, 言語モデル適応のチューニングデータに利用する Dev の 3 つに分割する. コーパスの緒元を表 2 に示す.

また、先行研究と同様に言語モデルの適応を行う. 適応に用いる一般言語モデルは BTEC コーパス [14], 英辞郎辞書例文コーパス [15] を混合して学習する. 一般コーパスの緒元を表 3 に示す. なお、一般コーパスの Dev データはワードペナルティのチューニングに利用し、Test1 および Test2 は実験の変換対象として利用する.

言語モデルの適応では、各キャラクターのコーパスから学習された言語モデルと一般言語モデルをキャラクターのチューニングデータ Dev に合致するように線形結合パラメータを設定する. これをそれぞれ学習データ Train, 評価データ Eval に対して行い、学習言語モデル、評価言語モデルを得る.

5.1.2 評価

先行研究と同様に、以下の 4 種の指標に基づいて評価を行う.

エントロピー $H(W) = \frac{1}{|W|} \sum_{w \in W} -\log P(w)$

評価言語モデルにおける負の底 2 の対数尤度を単語数で正規化して得られる. 目標話者の発話としての尤もらしさを評価する. エントロピーは低いほどよい評価となる. また、変換に用いる言語モデルと評価に用いる言語モデルが異なるため、クロスエントロピーの一種であるともいえる.

カバレッジ 変換対象の文章のうち、翻訳モデルに変換候

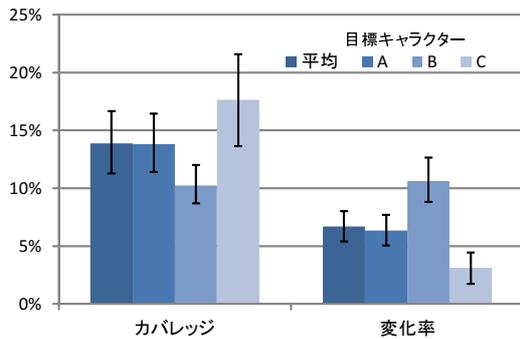


図 4 各目標話者におけるカバレッジおよび変化率
 Fig. 4 Coverage and Change rate in each character.

補が存在する単語の割合を示す。カバレッジが高いほど、翻訳モデルが様々な語彙を変換することが可能であるため、変換の幅広さを評価するのに用いる。

目標話者らしさ 被験者に目標話者の話し方を学習データから学習してもらい、与えられた文章がどの程度その話し方に一致しているかを「1:一致しない」～「5:一致する」の5段階で評価してもらう。

単語誤り率 (Word Error Rate; WER) 変換によって文章に間違った意味が与えられた割合を示す。被験者に与えられた変換前と変換後の文章対において、意味を変質させてしまった変換を指摘してもらうことにより得られる。

また、文章に対して実際に変化が起きた割合を示すために、変化率の統計も追加する。

変化率 変換対象の文章のうち、翻訳モデルによって変換によって変化した単語の割合を示す。カバレッジと異なり、変化が起きた単語のみを数えるため、実際の文面がどの程度変化したかを示す。

被験者は3名として、それぞれ3キャラクター、1キャラクターあたり一般コーパスのTest1セットに対して変換を行った結果10文の計30文に対して評価を行う。また、エントロピー、カバレッジ、変化率は被験者を必要とせず評価できるため、一般コーパスのTest2データを用いて評価を行う。各評価指標は信頼区間を有意水準 $p < 0.05$ のBootstrap Resampling[16]を用いて求める。また、変換前の文章に対しても同様の評価も行い、ベースラインとしてそれを示す。

5.2 実験結果

図4にカバレッジと変化率を、図5に変換前後におけるエントロピーを示す。提案法によるカバレッジは平均で14%、変化率は平均で7%であった。カバレッジは高くないものの、変化率から確実に変換が行われていることが分かる。エントロピーの平均値は、変換を通して減少していることが分かる。

主観評価の結果として、目標話者らしさの主観評価スコ

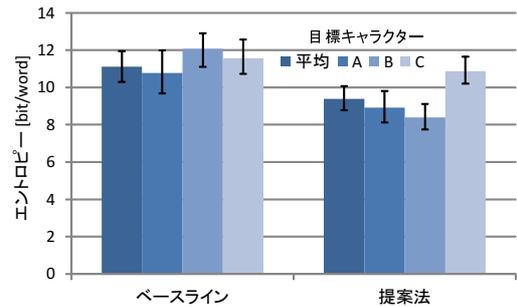


図 5 各目標話者におけるエントロピー
 Fig. 5 Entropy in each character.

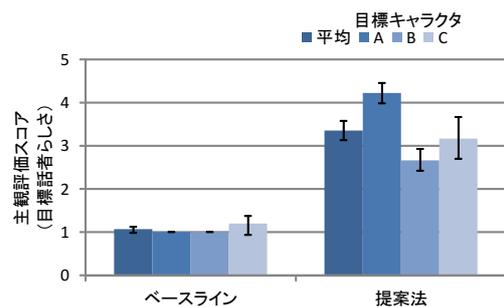


図 6 各目標話者における主観評価スコア
 Fig. 6 Individuality in each character.

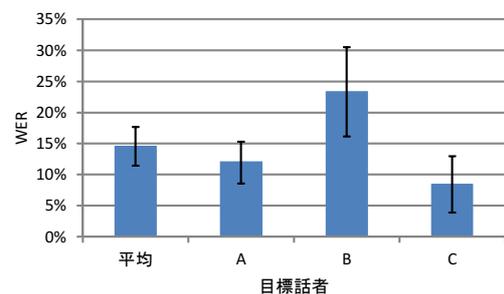


図 7 各目標話者における WER の評価
 Fig. 7 WER in each character.

アの平均と、各キャラクターにおけるスコアを図6に、各話者の変換による WER を図7に示す。変換によって、全てのキャラクターにおける目標話者らしさがベースラインに比べて向上している。しかしながら、各キャラクターごとのスコア間にも差が存在しており、提案手法はキャラクターによっても効果に差があることが分かる。

提案法の平均 WER は15%であった。提案法で用いている n -gram による類似語クラスタリングを用いた言い換えは、フレーズ前後の単語のみを考慮して言い換えを抽出するため、隣接する単語との不一致は少ないものの、それより広い範囲に対して文脈への一致を期待することはできない。そのため、言語モデルは広く文脈を評価可能な(最低でも3-gramの)言語モデルを要する。しかしながら、今回は提案法の言語モデルを先行研究と同様に言語モデルを

表 4 各目標話者の変換結果の一例
Table 4 Sample of transforming result in each character.

キャラクター	目標話者らしさ	入力	変換結果
A	5	おいしそう。	おいしそうだモン！
B	3	おいしそう。	おいしそう°▽°)ノ
C	4.3	多くても二十人だった。	多くても二十人だったッ！

2-gram で構築しているため、誤った変換が多く起きたと考えられる。また、類似語クラスタリングによる言い換えは、文脈に一致するフレーズを言い換えるため、「(は) できます。((/s))」と「(は) できません。((/s))」のような意味が異なるが前後の文脈が類似するフレーズを言い換えとして取得してしまう問題がある。これらの問題が WER を高くする原因であると考えられる。

最後に、表 4 に各目標話者の変換結果と主観評価スコアの一例を示す。変換結果から、提案法による変換では終助詞を中心に言い換えが行われ、これにより目標話者の言語的個性が得られていることが分かる。終助詞をはじめとした機能語が個性に対して影響を与えることが言語学においても知られており [5], [6], 言語的個性変換においても機能語の言い換えが重要であることが先行研究で明らかとなっている [3], [4]。本稿の実験結果は、先行研究から得られる傾向および言語学的知見と一致していると言える。

6. まとめ

本稿では、より特徴的な個性を持つ話者を対象とした特徴語の分析と言い換え抽出を行った。得られた特徴語の言い換えを用いて言語的個性変換を行い、評価実験を通してその有効性を示した。特徴語の言い換えでは、従来の機能語に焦点を当てた言い換えでは変換できなかった目標話者独特の言葉使いを変換可能で、特徴的個性話者の言語的個性を変換する上で大きく貢献した。

今後は、特徴語のみではなく、より広い範囲に対する文脈の整合性を取れる言い換えができるように言い換え抽出手法を改良し、特徴的個性話者の言語的個性をより高い精度で実現することを目標とする。

参考文献

[1] 橋本 誠, 北川 敏, 樋口宜男: 音声の個性知覚に影響を及ぼす音響的特徴の定量的分析, 日本音響学会誌, Vol. 54, No. 3, pp. 169–178 (オンライン), 入手先 (<http://ci.nii.ac.jp/naid/110003136431/>) (1998).

[2] Mairesse, F. and Walker, M. A.: Controlling user perceptions of linguistic style: Trainable generation of personality traits, *Computational Linguistics*, Vol. 37, No. 3, pp. 455–488 (2011).

[3] 水上雅博, Neubig, G., Sakti, S., 戸田智基, 中村 哲: 話し言葉における言語情報の個性変換における変換辞書拡張, 電子情報通信学会音声研究会 (SP), 千葉 (2013).

[4] 水上雅博, Neubig, G., Sakti, S., 戸田智基, 中村 哲: 日本語言い換えデータベースの構築と言語的個性変換

への応用, 言語処理学会第 20 回年次大会 (NLP2014), 北海道, pp. 773–776 (2014).

[5] Teshigawara, M. and Kinsui, S.: Modern Japanese ‘Role Language’ (Yakuwarigo): fictionalised orality in Japanese literature and popular culture, *Sociolinguistic Studies Vol 5-1*, Sheffield: Equinox Publishing (2012).

[6] Chung, C. and Pennebaker, J.: The psychological functions of function words (2007).

[7] Juola, P. and Baayen, R. H.: A controlled-corpus experiment in authorship identification by cross-entropy, *Literary and Linguistic Computing*, Vol. 20, No. Suppl, pp. 59–67 (2005).

[8] Isard, A., Brockmann, C. and Oberlander, J.: Individuality and Alignment in Generated Dialogues, *Proceedings of the Fourth International Natural Language Generation Conference*, INLG '06, Stroudsburg, PA, USA, Association for Computational Linguistics, pp. 25–32 (online), available from (<http://dl.acm.org/citation.cfm?id=1706269.1706277>) (2006).

[9] Gosling, S. D., Rentfrow, P. J. and Swann Jr, W. B.: A very brief measure of the Big-Five personality domains, *Journal of Research in personality*, Vol. 37, No. 6, pp. 504–528 (2003).

[10] Matthews, G., Deary, I. J. and Whiteman, M. C.: *Personality traits*, Cambridge University Press (2003).

[11] Neubig, G., Akita, Y., Mori, S. and Kawahara, T.: A monotonic statistical machine translation approach to speaking style transformation, *Computer Speech & Language*, Vol. 26, No. 5, pp. 349–370 (2012).

[12] 伊藤彰則, 好田正紀: 対話音声認識のための事前タスク適応の検討, 情報処理学会研究報告. SLP, 音声言語情報処理, Vol. 96, No. 123, pp. 91–98 (オンライン), 入手先 (<http://ci.nii.ac.jp/naid/110002916939/>) (1996).

[13] Barzilay, R. and Lee, L.: Learning to Paraphrase: An Unsupervised Approach Using Multiple-Sequence Alignment, *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies : Main Proceedings*, pp. 16–23 (2003).

[14] Takezawa, T., Sumita, E., Sugaya, F., Yamamoto, H. and Yamamoto, S.: Toward a Broad-coverage Bilingual Corpus for Speech Translation of Travel Conversations in the Real World., *The International Conference on Language Resources and Evaluation*, pp. 147–152 (2002).

[15] アルク: 英辞郎, <http://shop.alc.co.jp/cnt/eijiro/>.

[16] Koehn, P.: Statistical Significance Tests for Machine Translation Evaluation., *Conference on Empirical Methods on Natural Language Processing*, pp. 388–395 (2004).