

# 言語的個人性変換における言語モデルの適応と分析

## Language Model Adaptation and Analysis for Individuality Transforming

水上 雅博<sup>1\*</sup> Graham Neubig<sup>1</sup> Sakriani Sakti<sup>1</sup> 戸田 智基<sup>1</sup> 中村 哲<sup>1</sup>  
Masahiro MIZUKAMI<sup>1</sup> Graham NEUBIG<sup>1</sup> Sakriani SAKTI<sup>1</sup> Tomoki TODA<sup>1</sup> Satoshi NAKAMURA<sup>1</sup>

<sup>1</sup> 奈良先端科学技術大学院大学  
<sup>1</sup> Nara Institute of Science and Technology

**Abstract:** In text and speech, there are various features that express the individuality of the writer or speaker. We proposed a method for transforming individuality using a technique inspired by statistical machine translation (SMT), and showed the effectiveness. In previous work, we proposed a method for paraphrasing for characteristic words using n-gram clustering. However, the method can be improved, because it considers only short context. In this paper, we propose a model of transforming individuality that considers longer contexts. To achieve this, we suggest adaptation of the language models and expansion of paraphrasing for characteristic words.

## 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、内容に直接影響しない要素も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。

対話システムの話者性に関して、個人性を考慮した音声合成の研究等が進んでおり、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、言語的個人性と称す）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 [1] 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。

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な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

[2, 3]. 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、言語モデルおよび言い換えモデルを学習するための少量の目標話者コーパスのみで利用できる。しかしながら、現状では、ごく少量の目標話者コーパスから言語モデルを構築するため、高い精度を保証できない問題があった。また、言い換えモデルにおいても、話者依存性のない一般的な言い換えデータか、事前に選ばれたわずかな特徴語に対して、前後の  $n$ -gram 分布から置き換え可能な単語を探すのみであった。

本稿では、より高精度な言語的個人性変換を行うために、言語モデルの適応と、より広い文脈を考慮できるように特徴語の言い換え手法の拡張を提案する。言語モデルの適応では、目標話者のコーパスから学習された言語モデルと、目標話者に類似するドメインのコーパスで学習された言語モデル、そして、一般的な日本語コーパスから学習された言語モデルをそれぞれ適応し、目標話者に最も近い言語モデルを構築する。言い換えモデルの拡張では、今まで前後 1 単語の  $n$ -gram 分布のみを考慮していた言い換え抽出を、より広い範囲の  $n$ -gram 分布を考慮する手法を提案する。

なお、本稿では複数の言語モデル、言い換えモデルを組み合わせる個人性変換モデルを構築し、評価を行う。評価を効率よく行うため、人手によってつくられた特定の入力に対する理想的な変換結果（以下、参照文）を用意し、参照文ベースの自動評価を行う。また、自動評価を用いて、個人性変換モデルの変換パラメータをチューニングし、最も自動評価において高い評価を得られた変換モデルを用いた個人性変換結果を対象

\*連絡先：奈良先端科学技術大学院大学  
奈良県生駒市高山町 8916-5  
E-mail: masahiro-mi@is.naist.jp

に、被験者による主観評価実験を行う。

(3)

## 2 言語的個人性変換

言語的個人性変換 [2, 3] 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 [4] でも利用されている。

一般的な統計的機械翻訳においては、変換する要素を対としたコーパス（以下、対訳コーパス）を利用し、学習を行う。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「同じ意味を持ち、異なる個人性を持つ」特殊な対訳コーパスを必要とする。個人性を再現したい各個人に対して、この対訳コーパスを収集するのは労力がかかるため、我々は今まで、 $n$ -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 [2, 3, 5]。特に、文献 [5] で提案された特徴語の言い換えモデルは、固有のキャラクターや特殊な話し方をする話者の個人性を変換するために非常に効率的であることが示されている。

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、言い換え抽出によって得られる言い換えモデル確率  $P(V|W)$  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、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。

$$\hat{W} = \operatorname{argmax}_W \log P(V|W) + \lambda_{LM} * \log P(W) + \lambda WP * N$$

なお、変換モデルに存在する  $\lambda_{LM}$  は言語モデルに対する重み、 $\lambda_{WP}$  は文章長に対する重みを示すパラメータである。先行研究においては、これらのパラメータには経験的に一意な値を与えていたが、本稿では6.4節で説明する参照文を用いた自動評価指標を利用し、尤もらしい変換が行われるパラメータを推定する。

## 3 言語モデルの適応

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、先行研究では、少量の目標話者コーパスから学習された言語モデル  $P_t(w)$  と、大量の全話者混合コーパスから学習された言語モデル  $P_g(w)$  を、与えられた目標話者のテキストに最も一致するように線形結合を行っていた。また、目標話者コーパスが少量であることから、非常に短い文脈のみを考慮する  $n$ -gram モデルを利用していた。

本稿では、先行研究で利用した目標話者の言語モデル  $P_t(w)$  と大規模な全話者混合の言語モデル  $P_g(w)$  に加えて、目標話者に類似したドメインの言語モデル  $P_d(w)$  を結合することで、少量の目標話者データから高精度で長い文脈を考慮できる言語モデルを構築する。なお、言語モデルを線形結合する際には、与えられた文章に対して最適な結合係数  $\lambda_t, \lambda_d, \lambda_g$  をそれぞれ求める。

$$P(w) = \lambda_t P_t(w) + \lambda_d P_d(w) + \lambda_g P_g(w) \quad (4)$$
$$\lambda_t + \lambda_d + \lambda_g = 1$$

## 4 言い換えによる翻訳モデル

先行研究 [2, 3] では、言語的個人性変換に利用可能な対訳コーパスを収集する労力を考慮し、対訳コーパスを用いずに学習可能な一般的な言い換えとその言い換え確率を用いて擬似的に翻訳モデルを構築している。しかしながら、一般的な言い換えのみでは、特徴的な話者や固有のキャラクターの個人性を変換できないという問題があった。

これに対して文献 [5] では、ごく少量の目標話者のコーパスと大規模な全話者混合のコーパスを比較することで、単言語のコーパスから話者特有の語彙（以下、特徴語と呼ぶ）の言い換えを抽出する手法を提案し、特徴的な個人性を持つ目標話者への言語的個人性変換を可能にしている。

表 1: 特徴語の言い換えの一例 (文脈長  $m = 3$ )

$\mathbf{v}$	$\mathbf{w}$	$P(\mathbf{w} \mathbf{v})$
だ。	だ モン !	0.15
	だ モン。	0.08
	だ モン ☆	0.06
	か モン ?	0.06
	モン !	0.05
	計 13 語	

この手法では、目標話者の言語モデルと全話者混合の言語モデルの  $n$ -gram 確率から、言語モデル間の差異に寄与する量を示す  $\chi^2$  値を計算し、 $\chi^2$  値が大きい語を対象として言い換えの抽出を行った。言い換えの抽出、言い換え確率の計算では、単語間の  $n$ -gram 分布の類似性 [6] を用いて、類似の文脈で登場する単語対は言い換えられるという仮定から、言い換え可能な語を抽出している。しかしながら、この手法は前後 1 単語の  $n$ -gram 分布のみを考慮するため、大域的な文脈は考慮できていなかった。

本稿では、この手法を拡張し、任意の前後  $m$  語の  $n$ -gram 分布の類似性を考慮した言い換え確率を計算する手法を提案する。また、文献 [5] においては、類似性を JS ダイバージェンスを負の乗数とした指数関数として計算したが、本稿ではコサイン類似度を利用する。この理由として、計算量の問題が挙げられる。JS ダイバージェンスは言い換えの候補となる任意のフレーズ  $\mathbf{v}$  およびフレーズ  $\mathbf{w}$  の持つ全ての  $n$ -gram を考慮して計算するが、これは考慮する文脈  $m$  を大きくすると、組合せ爆発が起こり、実時間内の計算は不可能となる。これに対してコサイン類似度は、各要素の積として表現されるため、どちらか一方でも 0 となる要素は計算しなくてよい。すなわち、任意のフレーズ  $\mathbf{v}$  およびフレーズ  $\mathbf{w}$  の持つ全ての  $n$ -gram のうち少ない方を要素数のみで計算が終了する。

以下に特徴語言い換えの抽出手法を示す。

1. 一般的コーパスから出現頻度の高いフレーズの集合  $\mathcal{V}$  を、目標話者コーパスから出現頻度の高いフレーズの集合  $\mathcal{W}$  を抽出する。
2. あるフレーズ  $\mathbf{v} \in \mathcal{V}$ ,  $\mathbf{w} \in \mathcal{W}$  の前後  $m$  単語における  $n$ -gram 分布仮定を両コーパスから計算する。
3. あるフレーズ  $\mathbf{v} \in \mathcal{V}$ ,  $\mathbf{w} \in \mathcal{W}$  間の  $n$ -gram 分布仮定のコサイン類似度  $\text{Cos}(\mathbf{w}, \mathbf{v})$  を計算する。
4. あるフレーズ  $\mathbf{v} \in \mathcal{V}$  に対して、言い換え候補  $\mathbf{w} \in \mathcal{W}$  について以下の式を用いて周辺化を行い、言い換え確率  $P(\mathbf{w}|\mathbf{v})$  を求める。  

$$P(\mathbf{w}|\mathbf{v}) = \frac{\text{Cos}(\mathbf{w}, \mathbf{v})}{\sum_{\mathbf{w}' \in \mathcal{W}} \text{Cos}(\mathbf{w}', \mathbf{v})}$$
5. 言い換え確率  $P(\mathbf{w}|\mathbf{v})$  が閾値<sup>1</sup> 以上のフレーズ  $\mathbf{v}$ ,  $\mathbf{w}$  を言い換えとして抽出する。

本手法によって得られた言い換えの例を表 1 に示す。

<sup>1</sup>本稿では、経験的に閾値を 0.01 とした。

表 2: 参照文の一例

入力文	見てるだけです。
参照文	見てるだけだモン☆ 見てるだけだモン。 見てるだけだモン。 見てるだけだモン!
入力文	ありがとう。
参照文	ありがとうだモン! サンクマー☆ ありがとうだモン! サンくまー☆

## 5 参照文の収集

2 節で述べたとおり、言語的個人性変換における対訳コーパスを変換モデルの学習が可能な量を収集することは労力がかかる。そのため、文献 [2, 3, 5] では言い換え資源を用いて擬似的な翻訳モデルを構築している。

本稿では、人手で作られた理想的な入力文と、入力文に目標話者の言語的個人性を与えた対訳コーパス（以下、参照文と呼ぶ）をモデルの学習ではなく、評価とチューニングのみに用いることで、少量の参照文から高精度な言語的個人性変換のモデルを構築する手法を提案する。言語的個人性変換のモデルのチューニングは、6.4 節にて説明する。参照文の収集は以下の手順で行われる。

1. 作業者に目標話者の発話を集めたコーパスを熟読してもらい、目標話者の言語的個人性を学習する。
2. 作業者に全話者混合の一般的な話し言葉コーパスから、ランダムに抽出した 20 文章を与える。
3. 作業者は、与えられた文章を先に学習した目標話者の言語的個人性を持つように書き換える。
4. 作業者によって書き換えられた、目標話者の言語的個人性を持つ文章を入力文と対して参照文として収集する。

今回は、4 名の作業者に対して参照文の収集を行った。表 2 に与えた入力文と被験者から得られた参照文を例として示す。

## 6 評価的実験

### 6.1 実験条件

実験条件として、先行研究 [5] と同様に、特定のキャラクターを目標話者として設定した。また、参照文収集には、目標話者の言語的個人性を十分に理解した作業者が複数人必要であり、複数のキャラクターの参照文を収集することは難しい。そのため、本稿では目標話者を 1 つのキャラクターに限定した。

評価実験に用いる言語的個人性変換の言語モデルの学習と適応、言い換えモデル、チューニングに用いる

表 3: 変換用言語モデルの緒元

コーパス名	種類	文数	単語数
Character	目標話者	1304	16.0k
Twitter	ドメイン	390k	5178k
BTEC[7]	旅行会話	186k	1645k
REIJIRO	一般例文	169k	3560k
Dev	目標話者	326	4.2l

表 5: 評価用言語モデルの緒元

コーパス名	種類	文数	単語数
Character	目標話者	1304	16.3k
Twitter	ドメイン	7371k	83724k
BTEC[7]	旅行会話	186k	1648k
REIJIRO	一般例文	169k	3555k
Dev1	目標話者	326	4.1k
Dev2	一般例文, 旅行会話	88.9k	1029k

評価用の言語モデルの学習と適応の手順を次節より示す。

## 6.2 言語モデルの構築と適応

言語的個人性変換に用いる言語モデルを学習、適応する。なお、目標話者と類似のドメインのコーパスとして、Twitter から収集したコーパスを利用する<sup>2</sup>。学習に用いたコーパスの緒元を表 3 に示す。なお、結合の際は Dev コーパスを利用して適応する。適応の結果を 4 に示す。また、言い換えモデルも表 3 のコーパスを用いて、4 節の手法で構築する。

6.4 節において、言語的個人性変換の結果を自動評価するために、目標話者の言語モデルと、話者混合の一般的な言語モデルの二つの評価用の言語モデルが必要となる。評価用の言語モデルの学習に用いたコーパスの緒元を表 5 に示す。評価用の目標話者の言語モデルは表 5 の全てのコーパスから学習した言語モデルを Dev1 を用いて適応、一般言語モデルは、BTEC と REIJIRO から学習した言語モデルを Dev2 を用いて適応することで得られる。

## 6.3 言語モデル適応の評価

6.2 で構築した言語モデルの評価として、適応によって構築されたそれぞれの組合せの言語モデルを用いて、目標話者の学習に利用していないテストコーパス<sup>3</sup>に対するパープレキシティを計算した。言語モデルの組合せ、文脈長  $n$ 、パープレキシティの関係を図 1 に示す。

<sup>2</sup>ノイズを除去するために、Twitter から得られたコーパスは、リプライ、RT、ハッシュタグ、URL を除外した。

<sup>3</sup>326 文、合計単語数 4.1k。

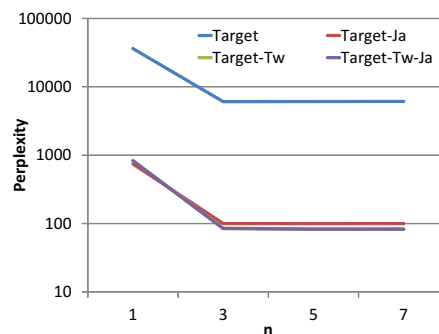


図 1: 各言語モデルとパープレキシティ

図 1 から、全ての言語モデルにおいて長い文脈を考慮することでパープレキシティは減少することが分かった。また、適応に用いる言語モデル数を増やすことでパープレキシティは減少することが明らかになった。全言語モデルのパープレキシティを比較したところ、最小のパープレキシティとなった言語モデルは Target-Tw-Ja 言語モデルの 5-gram であった。この結果を踏まえ、6.4 節で利用する評価用の言語モデルは Target-Tw-Ja 言語モデルの 5-gram とした。

## 6.4 自動評価とチューニング

パラメータのチューニングを行うために、先行研究 [8, 5] に基づいて自動評価手法を行う。自動評価の指標は以下の二つを選択した。

**言語モデル指標 (LMM)** Xu らによって提案された話し方に関する自動評価指標の一つで、目標話者らしさと高い相関を示す指標である [8]。変換前の言語モデル確率  $P_g(W)$  と目標話者の言語モデル確率  $P_t(W)$  を用いて以下のように表現される<sup>4</sup>。  

$$P(\text{style} = \text{target} | W) = \frac{P_t(W)}{P_t(W) + P_g(W)}$$

**BLEU** 先行研究 [8] において自然性と高い相関を示す指標であり、統計的機械翻訳の自動評価で最も頻繁に利用される自動評価指標の一つである。BLEU は参照文と出力文においてどれだけ  $n$ -gram が一致するかを計算する [9]<sup>5</sup>。

2 節で述べたとおり、先行研究においては、言語的個人性変換で利用されるモデルのパラメータは経験的に決定されていた。本稿では 5 節で得られた参照文と 6.2 節で得られた評価用の言語モデルを用いて、言語的個人性変換の最適なモデル構成とパラメータをチューニングする。

<sup>4</sup>この指標は目標話者の言語モデルと一般言語モデルを必要とするが、これは 6.2 節で示した評価用の言語モデルを利用する。

<sup>5</sup>一般的に、BLEU を計算する際の  $n$ -gram は  $n=4$  とされる。本稿でも、 $n=4$  として BLEU を計算した。

表 4: 言語モデル適応の組合せ

言語モデル	利用コーパス (結合係数 $\lambda$ )			
	Character(0.61)	Twitter(0.35)	BTEC(0.01)	REIJIRO(0.03)
Target+Tw+Ja	Character(0.61)	Twitter(0.35)	BTEC(0.01)	REIJIRO(0.03)
Target+Tw	Character(0.62)	Twitter(0.38)	-	-
Target+Ja	Character(0.76)	-	BTEC(0.06)	REIJIRO(0.18)
Target	Character(1.00)	-	-	-

チューニングの際は、先述の自動評価指標である LMM と BLEU、そして LMM と BLEU の幾何平均を最大化する。このような自動評価指標を用いたチューニング手法は MERT[10] と呼ばれ、統計的機械翻訳の分野では幅広く利用されている。

## 6.5 チューニングと分析

6.4 節で説明した自動評価を用いたチューニングを行い、その結果を表 6 に示す。

チューニングの結果、全ての指標において、言い換えモデルにおける文脈長を  $m$  を長く考慮した方が評価指標は向上した。BLEU では、変換なしに比べてわずかに向上したが、LMM 値も変換なしに比べてわずかに向上するのみであった。LMM 値の向上値が小さいことから、変換はあまり行われなかったと考えられる。LMM では、LMM 値が向上したが、BLEU 値は低下した。これらの値から、多くの単語が変換によって追加されたと考えられる。BLEU と LMM の幾何平均である BLEU+LMM では、変換なしに比べて BLEU 値はわずかに低下したが、LMM 値のは向上した。BLEU+LMM は BLEU を低下させないように、LMM を最大化するような変換結果が得られたと考えられる。

## 6.6 主観評価実験

先行研究 [5] と同様に、被験者による主観評価を通して変換結果の目標話者らしさと変換精度の評価を行った。主観評価の対象となる文は、BTEC および REIJIRO のコーパスから、言語モデルの学習に用いられなかった文からランダムに 10 文選択され、言語的個人性変換は 6.4 節で得られたモデル、パラメータを用いて行われる。なお、今後はそれぞれ言語的個人性変換に利用したモデル、パラメータをチューニング時に利用した評価指標名 (BLEU, LMM, BLEU+LMM) で表記する。

得られた変換結果、計 30 文に対して、被験者 5 名が目標話者らしさおよび変換で生じた単語誤り率を評価する。評価結果に対して信頼区間を有意水準  $p < 0.05$  の Bootstrap Resampling[11] を用いて求める。また、文献 [5] と同様の手法で個人性変換を行った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。

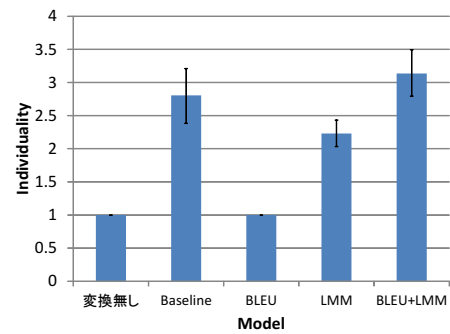


図 2: 変換モデルと目標話者らしさ

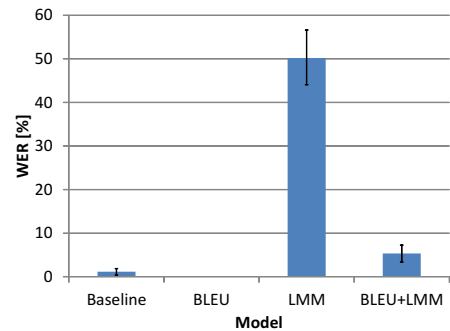


図 3: 変換モデルと WER

## 6.7 主観評価実験結果

図 2 に変換モデルと目標話者らしさの関係を、図 3 に変換モデルと WER の関係を示す。

変換前の目標話者らしさは 1.0 (目標話者らしくない) であり、ベースラインの目標話者らしさは 2.79 まで向上した。提案法による目標話者らしさは、LMM を用いた際に 2.22、BLEU+LMM を用いた際に 3.13 まで向上したが、BLEU を用いた際には変換が行われなかった。

また、WER はベースラインが最小で 1.2%、提案法中では変換が行われなかった BLEU を除いて、BLEU+LMM を用いた際に最小となり、5.3% となった。LMM では、言語モデル確率のみを考慮するため、しばしば過剰な変換が行われ、WER が大きくなる。Xu らの研究 [8] において、LMM 値は目標話者らしさと高い相関を示すことがすでに明らかになっているが、本手法においては、非常に単語誤りの多い文章が生じるため、LMM

表 6: チューニングとパラメータ

Model	LM $n$	TM $m$	$\lambda_{LM}$	$\lambda_{WP}$	BLEU	LMM	BLEU+LMM
変換無し	-	-	-	-	0.894	0.267	0.489
BLEU	7	3	0.1	0.2	0.903	0.305	0.525
LMM	7	3	0.1	5.3	0.071	0.995	0.266
BLEU+LMM	7	3	0.6	1.6	0.790	0.841	0.815

表 7: 変換結果の一例

モデル	変換結果
入力文	食べたいものはありますか。
ベースライン	食べたいものはありますモン。
提案法	食べたいものはありますかモン？
入力文	参加しています。
ベースライン	参加していたモン！
提案法	参加してきたモン。

値のみを最大化しても、目標話者らしさは最大化されなかった。これに対して、BLEU は  $n$ -gram 一致と単語数を考慮するため、個人性を最大化するために過剰に特徴語を言い換えたり、文章長を変化させることは少なくなり、結果として BLEU+LMM を用いた場合、LMM に比べて低い WER になり、BLEU のみを用いた場合は変換が行われなかったと考えられる。

BLEU をチューニングに用いた際に変換が行われなかった原因として、変換前の BLEU 値が 0.894 と高い値であり、変換を行って誤った文章を得るよりも、変換を行わないようなパラメータがチューニングによって得られたと考えられる。

主観評価実験の結果から、提案法 BLEU+LMM はベースラインに比べて目標話者らしさが向上したが、WER もベースラインに比べて増加した。

最後に、表 7 にベースラインと提案手法 (BLEU+LMM) の変換結果の一例を示す。

## 7 まとめ

本稿では、言語的個人性変換の高精度化を目標とし、言語モデルの適応とより大域的な文脈を考慮した特徴語の言い換え、参照文を用いたチューニングを行った。言語モデルを適応することにより、従来よりも高精度で長い文脈を考慮可能な目標話者言語モデルを獲得した。また、より大域的な文脈を考慮した特徴語の言い換えでは、先行研究に比べて長い文脈を考慮することにより、高精度な特徴語言い換えモデルを獲得した。これらを組み合わせた言語的個人性変換モデルとチューニングの有効性を主観評価実験を通して示した。

今後は、文脈のみでなく、語彙を形成する内部構造の類似性を用いた言い換え抽出や既知の言い換え規則を利用した特徴語の言い換え抽出手法を提案する。

## 参考文献

- [1] François Mairesse and Marilyn A Walker. Controlling user perceptions of linguistic style: Trainable generation of personality traits. *Computational Linguistics*, Vol. 37, No. 3, pp. 455–488, 2011.
- [2] 水上雅博, Graham Neubig, Sakriani Sakti, 戸田智基, 中村哲. 話し言葉における言語情報の個人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, 9 2013.
- [3] 水上雅博, Graham Neubig, Sakriani Sakti, 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人性変換への応用. 言語処理学会第 20 回年次大会 (NLP2014), pp. 773–776, 北海道, 3 2014.
- [4] Graham Neubig, Yuya Akita, Shinsuke Mori, and Tatsuya Kawahara. A monotonic statistical machine translation approach to speaking style transformation. *Computer Speech & Language*, Vol. 26, No. 5, pp. 349–370, 2012.
- [5] 水上雅博, Graham Neubig, Sakriani Sakti, 戸田智基, 中村哲. 特徴的語者を対象とした言語的個人性変換. 情報処理学会 第 216 回自然言語処理研究会 (SIG-NL), 東京, 5 2014.
- [6] Regina Barzilay and Lillian Lee. Learning to paraphrase: An unsupervised approach using multiple-sequence alignment. In *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Main Proceedings*, pp. 16–23, 2003.
- [7] Toshiyuki Takezawa, Eiichiro Sumita, Fumiaki Sugaya, Hirofumi Yamamoto, and Seiichi Yamamoto. Toward a broad-coverage bilingual corpus for speech translation of travel conversations in the real world. In *The International Conference on Language Resources and Evaluation*, pp. 147–152, 2002.
- [8] Wei Xu, Alan Ritter, Bill Dolan, Ralph Grishman, and Colin Cherry. Paraphrasing for style. In *Proceedings of COLING 2012*, pp. 2899–2914, Mumbai, India, December 2012. The COLING 2012 Organizing Committee.
- [9] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: A method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL '02, pp. 311–318, Stroudsburg, PA, USA, 2002. Association for Computational Linguistics.
- [10] Franz Josef Och. Minimum error rate training in statistical machine translation. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*, pp. 160–167. Association for Computational Linguistics, 2003.
- [11] Philipp Koehn. Statistical significance tests for machine translation evaluation. In *Conference on Empirical Methods on Natural Language Processing*, pp. 388–395, 2004.