

話し言葉における言語情報の個人性変換手法の拡張と評価*

☆水上雅博, Graham NEUBIG, Sakriani SAKTI, 戸田 智基, 中村 哲 (奈良先端大)

1 まえがき

単純な言語情報のみでなく、誰が話しているかといった個人性や、それに伴う非言語情報も伝達することは、円滑な意思疎通を行う上で有効である。例を挙げれば、対話相手が子供であれば柔らかい話し方を、大人であれば硬い話し方を用いることで、より円滑な対話を実現できると考えられる。そのため、対話エージェントの上でも豊かな個人性と雰囲気表現することは重要であると考えられ、対話システムの応答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。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 [1] 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、言語情報における話者ごとの特徴（以下、言語情報の個人性と呼ぶ）までも制御する研究はほとんどない。類似の研究として性格や礼儀正しさなどの心理学的要因を考慮し、話者性を考慮したルールベースの文生成を行う研究がある [2]。

そこで我々は、翻訳辞書と言語モデルを用い、話し言葉の書き起こし文章に対して個人性を変換する手法を提案している [3]。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を制御可能であり、すなわち「こんにちは、何かご用ですか」のような既存のシステムの出力文章を「ハロー、なんか用かな」のようなユーザの望む個人性を持った文章へ変換できる。また、翻訳モデルの学習に対訳コーパスを必要とせず、置き換え可能な語彙を集めた翻訳辞書と目標話者の言語モデルを目標話者の発話から学習するのみで利用できる。

しかしながら、翻訳モデルの代用としている翻訳辞書は変換の対象が限定的であった。そこで本稿では話し言葉における言語情報の個人性を自由に変換する技術の実現をめざし、この問題への対処として個人性変換手法の拡張と評価を行う。提案法は、従来法の翻訳辞書と言語モデルを用いた個人性変換手法に加え、シソーラスやN-gramを用いた翻訳辞書の自動構築について提案し、それぞれ実験を通じた評価結果から、有効性を示す。また、評価についても、人手による主観評価と容易に利用可能な自動評価につい

てそれぞれ提案する。

2 言語モデルと翻訳辞書を用いた個人性変換手法

まず、本研究の先行研究である話し言葉の書き起こし文章の話者性の変換 [3]（以下、個人性変換と呼ぶ）について説明する。先行研究では、変換の対象となる話し言葉のテキスト V と変換先の目標となる話者性を持つテキスト W の間の翻訳問題として扱い、統計的機械翻訳で翻訳を行う [4]。統計的機械翻訳では利用可能なコーパスの分量を考慮し、大量に確保することが難しい対訳コーパスを用いる翻訳モデル確率 $P(V|W)$ と大量に確保可能な出力側コーパスを用いる言語モデル確率 $P(W)$ の二つに分解し、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、個人性変換は一般的な統計的機械翻訳と異なり、対訳コーパスから翻訳モデルを学習することは容易ではない。これは、ある話者の個人性を持ったテキストと同じ意味で別の話者の個人性を持ったテキストという特殊な設定の対訳コーパスを収集することが非常に困難であり、翻訳モデルを構築できるほどの分量を用意できないためである。これを解決するために、先行研究 [3] では翻訳モデルを対訳コーパスから学習せず、文章の意味を変化させずに交換可能な機能語であるフィラーと感動詞を抽出した翻訳辞書を構築し、これに一律な翻訳モデル確率を割り当てることで翻訳モデルを構築している。

3 翻訳辞書の拡張

先行研究において、翻訳辞書はフィラーや感動詞を対象として構築した。その理由としては、フィラーや感動詞に対して交換を行っても文章の意味を変化させず、ほぼ無条件で置き換えを行うことができるためである。しかしながら、フィラーや感動詞のみの置き

*Expand and Evaluation of a Method for Transforming Individuality of Spoken Language. by MIZUKAMI Masahiro, NEUBIG Graham, SAKTI Sakriani, TODA Tomoki, NAKAMURA Satoshi (NAIST)

換えでは柔軟な個人性変換を行うことができず、より多様な変換が可能な翻訳辞書を構築する必要がある。そこで、本研究ではこの翻訳辞書の構築を自動で行う手法について提案する。このような言い換え可能な語彙を検討する様々な先行研究 [5] が提案されている。代表的な手法としてシソーラスを用いた言い換えや単語 N-gram を用いた類似語クラスタリングなどがあげられる。今回は、シソーラスを用いた名詞、動詞、形容詞表現などの言い換えを目的とした翻訳辞書の自動構築と、3-gram を利用した類似の助詞、助動詞などを対象とした翻訳辞書の自動構築について述べる。

3.1 シソーラスを用いた翻訳辞書の自動構築

同義語、類義語を取得可能な語彙資源としてシソーラスがあげられる。日本語の代表的なものでは日本語 WordNet[6] や EDR 日本語単語辞書などが存在し、非常に細かい意味分類に基づく単語間の同義関係が与えられている。今回は、WordNet を用いて翻訳辞書を構築した。

翻訳辞書の構築は以下の手順で行われる。

1. 変換対象の文章を形態素解析する。
2. 形態素解析して得られた形態素を取り出し、それをクエリとして WordNet のデータベースを検索する。
3. 語彙が見つかった場合はその語の持つ同義語集合 ID を利用して、WordNet のデータベースから同義語を取得し、辞書に格納する。

3.2 N-gram を用いた翻訳辞書の自動構築

同義語や類義語を得る手法の一つとして単語 N-gram を用いた類似語クラスタリングがあげられる [7]。今回は、この類似語クラスタリングを参考として単語 3-gram を利用した類似の助詞、助動詞を抽出し、翻訳辞書を構築する。助詞および助動詞を対象を絞った理由は、シソーラスが助詞、助動詞の言い換えに対応していない点と、終助詞および助動詞が日本語において話者の個人性をあらわす重要な要素として取り上げられている [8] ためである。

翻訳辞書の構築は以下の手順で行われる。

1. 様々な話者の発話を集めたコーパスから 3-gram 言語モデルを学習する。
2. 3-gram において、2 番目の要素に助詞および助動詞を含むデータを抽出する。
3. 以下のように、前後の単語が一致している助詞、助動詞を集めて辞書に格納する。

例) カメラ は 高い、カメラ も 高い

4 言語モデルの適応

先行研究では目標話者の発話を集めて言語モデルを構築していた。しかしながら、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、また内容も限定的である。そのため、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個人性変換を行うことは困難である。特に、本研究で提案する翻訳辞書の自動構築においては、辞書中の単語の翻訳に対して翻訳確率を与えず、単純に言語モデルの生成確率のみを用いて文章の変換を行うため、精度のよい言語モデルの生成が必要不可欠となる。

これを解決するために、少量の目標話者データから学習された言語モデルと大量の全話者混合データから学習された言語モデルを線形補間して利用する。これにより、少量の目標話者データから、個人性変換に十分利用可能なデータ量の言語モデルを構築することが可能となる。このような大規模なコーパスから学習した言語モデルと、取り組むタスクやドメインに注目した少量のコーパスから学習した言語モデルを混合することにより、効率的に分野適応を行う手法は音声認識などで用いられている [9]。

5 評価指標の提案

一般的な統計的機械翻訳においては BLEU[10] に代表される多くの評価指標が提案されている。これらは対訳コーパスが与えられたとき、システムによる翻訳結果と人間の翻訳者による翻訳文との距離を測ることで翻訳システムの精度を評価する。しかしながら、本研究では対訳コーパスを得ることが困難であり、このような自動評価指標は利用できない。そのため、システムの評価に必要な評価指標をいくつかに分けて定義し、それらの評価から総合的にシステムの評価を行う。今回提案する評価指標は「目標話者らしさ」、「変換の幅広さ」と「日本語として正しいか」の三つについて評価する。

5.1 自動評価

先述の通り、自動評価では以下の二つの要素についてそれぞれ評価する。

エントロピー 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。¹

カバレッジ 変換対象の文章のうち、翻訳辞書に変換候補が存在する単語の割合を示す。カバレッジが高いほど、翻訳辞書が様々な語彙を変換する

¹今回は利用できる目標話者データが少ないため、複数の言語モデルを用意することができない。そのため、エントロピー評価と個人性変換に用いる言語モデルは同じものを利用した。

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られる。

5.2 人手による主観性評価

上記の評価指標は自動で評価できるが、細かいニュアンスや、同義語の微細な意味の差などは評価することができず、目標話者らしさや日本語としての正しさを評価するには人手による評価も必要となる。以下の項目に対して、人手による主観評価を行った。

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) 変換後の文章に誤りが含まれている割合を評価し、変換結果が日本語として正しいかを示す。被験者に与えられた文章の文法および語彙の用法の間違いを指摘してもらうことにより得られる。

6 評価の実験

6.1 実験条件

実験の題材として3名の店員と19名の客による一対一のカメラ販売に関する対話をまとめたカメラ販売対話コーパス [11] を利用した。店員3名のコーパスをそれぞれ話者ごとに分け、さらに学習データと評価データに分けて利用した。

言語モデルの分野適応を行うために、個人性変換の対象であるカメラ販売対話コーパスの各話者から学習した言語モデルと、BTEC コーパス [12] および英辞郎の辞書例文コーパス [13] から学習した一般言語モデルを用意した。この両者の言語モデルを、変換の対象とするカメラ販売対話コーパスのトピックに合致するような線形結合パラメータをそれぞれ話者ごとに計算した。

辞書拡張の評価実験では、まず変換目標の話者の学習データを被験者に読んでもらう。次に、変換前の文章とそれを各辞書を使用して個人性変換を行った結果4文を読んでもらう。各辞書の組み合わせパターンを Table 1 に示す。それぞれの変換結果に対して WER (変換されたせいで意味が破壊された単語のカウント) とどれだけ目標話者らしいかを5段階で評価してもらった。この実験は被験者3名に対して行い、各被験者には3話者、1話者あたり10個の計30の変換結果に対して評価を行ってもらった。また、各評価指標の信頼区間を有意水準 $p < 0.05$ の Bootstrap Resampling を用いて求める [14]。

Table 1 各辞書の組み合わせと表記

パターン (表記)	提案手法 (フィルラー)	N-gram 類似 (助詞など)	シソーラス (名詞など)
PD	○	×	×
ND	○	○	×
TD	○	×	○
ND+TD	○	○	○

Table 2 各辞書の組み合わせと変換結果の例

辞書	変換結果
PD	あー、やっぱり携帯と比べて撮れる画は全然違いますね
ND	あー、やっぱり携帯に比べて撮れる画も全然違いますね
TD	あー、やっぱり携帯と比べて撮れる写真は全く違いますね
ND+TD	あー、やっぱり携帯に比べて撮れる写真も全く違いますね

6.2 実験結果

翻訳辞書の拡張の評価を行う。個人性変換の際に各辞書を利用し、その変換結果のエントロピーを Fig. 1 に、カバレッジを Fig. 2 に示す。また、変換結果の一例を Table 2 に示す。

この結果から、利用する辞書を多くするほどエントロピーおよびカバレッジの評価は向上している。今回提案した N-gram 類義語およびシソーラスによる辞書の両者を同時に用いることで、エントロピーは平均 5.95bit/word、カバレッジは平均 49%まで改善された。

次に、人手による評価の結果として、WER を Fig. 3 に、主観評価の結果を Fig. 4 に示す。全体として目標話者らしさは3.8を超えており、高い値となっているが、その中でも N-gram 類義語を用いた場合の主観評価値が最良で4.4を超え、フィルラー・感動詞のみを対象にした辞書を用いた場合の評価より改善した。しかしながら、シソーラスまたは N-gram 類義語とシソーラスの辞書の両者を利用した場合は先行研究に比べ目標話者らしさは低下した。これはシソーラスによる置き換えが間違っており、「候補は3万くらいのカメラですか」を「選挙は数字10と1000の積である基数くらいのカメラで」など、でたらめな意味の文章に変換したため、正しく個人性を評価できなくなってしまったためである。

7 まとめ

本稿では、先行研究である言語モデルと翻訳辞書を用いた個人性変換に対して、シソーラスおよび N-gram 類義語による翻訳辞書の自動構築と、個人性変換の評価指標を提案し、実験的評価を通して有効性を示した。今後の課題として、シソーラスを用いた翻

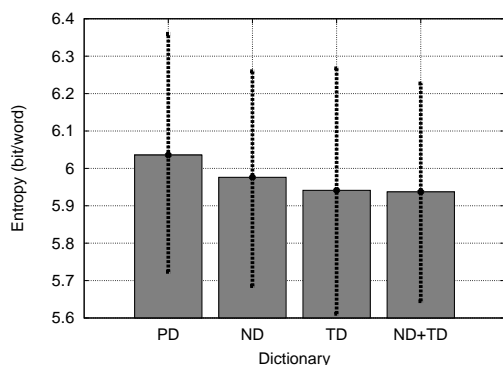


Fig. 1 各辞書使用時のエントロピー

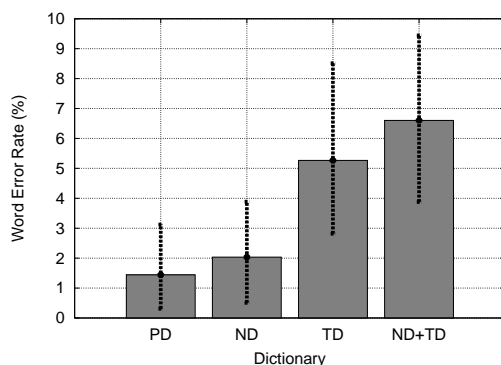


Fig. 3 各辞書使用時の WER

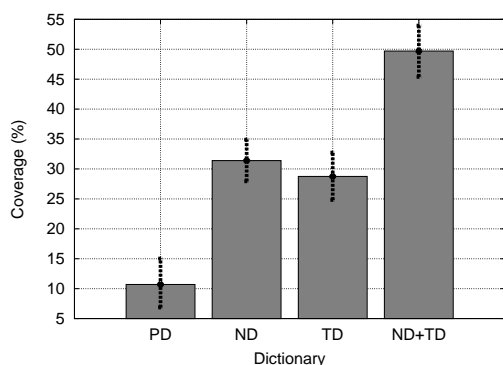


Fig. 2 各辞書使用時のカバレッジ

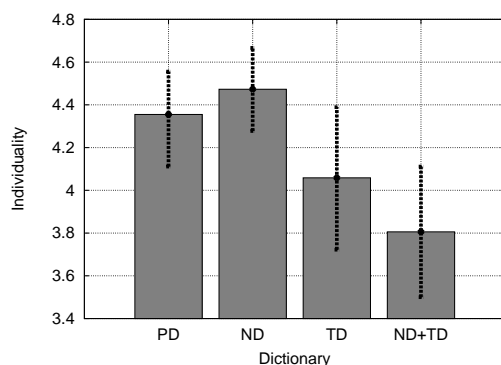


Fig. 4 各辞書使用時の話者らしさ

訳辞書の精度向上を進める。

参考文献

- [1] Junichi Yamagishi et al. Thousands of voices for HMM-based speech synthesis-analysis and application of TTS systems built on various ASR corpora. *IEEE Transactions on, Audio, Speech, and Language Processing*, Vol. 18, No. 5, pp. 984-1004, 2010.
- [2] François Mairesse, Marilyn A Walker. Controlling user perceptions of linguistic style: Trainable generation of personality traits. *Computational Linguistics*, Vol. 37, No. 3, pp. 455-488, 2011.
- [3] 水上 雅博ほか. 話し言葉の書き起こし文章の話者性の変換. 2013 年度人工知能学会全国大会 (JSAI2013), 富山, 6 2013.
- [4] Graham Neubig et al. A monotonic statistical machine translation approach to speaking style transformation. *Computer Speech & Language*, 2012.
- [5] 乾 健太郎, 藤田 篤. 言い換え技術に関する研究動向. *自然言語処理*, Vol. 11, No. 5, pp. 151-198, 2004.
- [6] Bond Francis et al. Enhancing the Japanese wordnet. In *Proceedings of the 7th Workshop on Asian Language Resources*, pp. 1-8. Association for Computational Linguistics, 2009.
- [7] Regina Barzilay, Lillian Lee. Learning to paraphrase: An unsupervised approach using multiple-sequence alignment. In *HLT-NAACL 2003: Main Proceedings*, pp. 16-23, 2003.

- [8] Mihoko Teshigawara, Satoshi Kinsui. Modern Japanese 'Role Language' (Yakuwarigo): fictionalised orality in Japanese literature and popular culture. In *Sociolinguistic Studies Vol 5-1*. Sheffield: Equinox Publishing, 2012.
- [9] 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, dec 1996.
- [10] Papineni Kishore et al. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pp. 311-318. Association for Computational Linguistics, 2002.
- [11] 平岡 拓也ほか. 説得対話システム構築のための対話コーパス分析. *日本音響学会 2013 年春季研究発表会 (ASJ)*, 東京, 3 2013.
- [12] Toshiyuki Takezawa et al. Toward a broad-coverage bilingual corpus for speech translation of travel conversations in the real world. In *LREC*, pp. 147-152, 2002.
- [13] アルク. 英辞郎. <http://shop.alc.co.jp/cnt/eijiro/>. Accessed: 2013-07-26.
- [14] Philipp Koehn. Statistical significance tests for machine translation evaluation. In *EMNLP*, pp. 388-395, 2004.